Single Technology Appraisal

Tirzepatide for treating type 2 diabetes [ID3938]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Tirzepatide for treating type 2 diabetes [ID3938]

Contents:

The following documents are made available to stakeholders:

The **final scope** and **final stakeholder list** are available on the NICE website.

1. **Company submission** from Eli Lilly & Company
   a. **Company summary of information for patients (SIP)**

2. **Clarification questions and company responses**

3. **Patient group, professional group and NHS organisation submissions** from:
   a. Diabetes UK

4. **Expert personal perspectives** from:
   a. Prof. Michael Cummings, Consultant, Diabetes and Endocrinology – clinical expert, nominated by Eli Lilly & Company
   b. Prof. Stephen C Bain, Professor of Medicine (Diabetes) – clinical expert, nominated by the Association of British Clinical Diabetologists and NovoNordisk

5. **External Assessment Report** prepared by KSR Ltd

6. **External Assessment Report – factual accuracy check**

7. **Additional information from company**
   a. Company response to EAG report
   b. Additional sensitivity analyses requested by EAG

8. **External Assessment Report critique of company additional information** prepared by KSR Ltd

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Tirzepatide for the treatment of patients with type 2 diabetes

[ID3938]

Document B

Company evidence submission

9th August 2022

File name	Version	Contains confidential information	Date
ID3938_Eli Lilly_Tirzepatide for T2D_Document B [ACIC]_v3_170323	3	Yes	17/03/23

Instructions for companies

This is the template for submission of evidence to the National Institute for Health and Care Excellence (NICE) as part of the single technology appraisal (STA) process. Please note that the information requirements for submissions are summarised in this template; full details of the requirements for pharmaceuticals and devices are in the user guide.

This submission must not be longer than 150 pages, excluding appendices and the pages covered by this template. If it is too long it will not be accepted.

Companies making evidence submissions to NICE should also refer to the NICE health technology evaluation guidance development manual.

In this template any information that should be provided in an appendix is listed in a box.

Highlighting in the template (excluding the contents list)

Square brackets and grey highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [grey highlighting] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Grey highlighted text in the footer does not work as an automatic form field, but serves the same purpose – as prompt text to show where you need to fill in relevant details. Replace the text highlighted in [grey] in the header and footer with appropriate text. (To change the header and footer, double click over the header or footer text. Double click back in the main body text when you have finished.)

Contents

Tables and figures

Abbreviations

Abbreviation	Definition
ADA	American Diabetes Association
ADDQoL	Audit of Diabetes-Dependent Quality of Life
AE	Adverse Event
AER	Albumin Excretion Rate
ALT	Alanine Transaminase
ANCOVA	Analysis of Covariance
APPADL	Ability to Perform Physical Activities of Daily Living
BG	Blood Glucose
BID	Twice Daily
BMI	Body Mass Index
BRAVO	Building, Relating, Assessing, and Validating Outcomes
CDK-EPI	Chronic Kidney Disease-Epidemiology
CEC	Clinical Endpoint Committee
CEM	Cost-Effectiveness Model
CfB	Change from Baseline
CHMP	Committee for Medicinal Products for Human Use
CIs	Confidence Intervals
CKD	Chronic Kidney Disease
COVID-19	Coronavirus Disease 2019
CRD	Centre for Review and Dissemination
CrI	Credible Interval
CSR	Clinical Study Reports
CT	Clinical Trials
CV	Cardiovascular
CVD	Cardiovascular Disease
CVOT	Cardiovascular Outcomes Trial
DBP	Diastolic Blood Pressure
DDD	Defined Daily Dose
DIC	Deviance Information Criterion
DPP-4	Dipeptidyl-Peptidase 4
DSU	Decision Support Unit
DTSQc	Diabetes Treatment Satisfaction Questionnaire-change
DTSQs	Diabetes Treatment Satisfaction Questionnaire-status
DULA	Dulaglutide
EASD	European Association for the Study of Diabetes
ECDRP	EC Decision Reliance Procedure
ECG	Electrocardiogram
eGFR	estimated Glomerular Filtration Rate
EMA	European Medicines Agenda
EPAR	European Public Assessment Report

EQ-5D-5L	5-Level European Quality of Life 5 Dimension score
ESRD	End Stage Renal Disease
ET	Early Termination
ETD	Estimated Treatment Difference
FAS	Full Analysis Set
FBG	Fasting Blood Glucose
FDA	Food and Drugs Administration
FSG	Fasting Serum Glucose
GBP	Great British Pounds
GE	General Electric
GI	Gastrointestinal
GIP	Glucose-Dependent Insulinotropic Polypeptide
GLP-1	Glucagon-Like Peptide-1
Hb	Haemoglobin
HbA1c	Glycated Haemoglobin
HDL	High-Density Lipoprotein
HE	Health Economic
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment
ICER	Incremental Cost-Effectiveness Ratio
IDF	International Diabetes Federation
IHD	Ischaemic Heart Disease
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITT	Intent-to-Treat
IU	Insulin Units
IV	Intravenous
IWQOL-LITE-CT	Impact of Weight on Quality of Life-Lite Clinical Trials Version
IWRS	Interactive Web Response System
IW-SP	Impact of Weight on Self-Perception
JAGS	Just Another Gibbs Sampler
KDIGO	Kidney Disease Improving Global Outcomes
LDL	Low-Density Lipoprotein
LIRA	Liraglutide
LOCF	Last Observation Carried Forward
LS	Least Squares
LSM	Least Squares Mean
MA	Meta-Analysis
MACE	Major Adverse Cardiovascular Events
MCMC	Markov Chain Monte Carlo
MHRA	Medicines and Healthcare products Regulatory Agency
MI	Myocardial Infarction
mITT	modified Intent-to-Treat

MMRM	Mixed Model for Repeated Measures
MTC	Medullary Thyroid Cancer
NAFLD	Non-Alcoholic Fatty Liver Disease
NCPE	National Centre for Pharmacoeconomics
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network Meta-Analysis
NPH	Neutral Protamine Hagedorn
OAD	Oral Antidiabetic Drug
OM	Outcomes Model
ORs	Odds Ratios
OUS	Outside the USA
PBO	Placebo
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic Sensitivity Analysis
PSSRU	Personal Social Services Research Unit
QALY	Quality-Adjusted Life Year
QD	Once Daily
QoL	Quality of Life
QW	Once Weekly
RA	Receptor Agonist
RCT	Randomised Controlled Trials
REML	Restricted Maximum Likelihood
RNG	Random Number Generator
SAE	Serious Adverse Event
SBP	Systolic Blood Pressure
SD	Standard Deviation
SEMA	Semaglutide
SF-36	Short-Form 36-item survey
SGLT2i	Sodium-Glucose Co-Transporter-2 Inhibitor
SLR	Systematic Literature Review
SMBG	Self-Monitored Blood Glucose
SmPC	Summary of Product Characteristics
SOC	System Organ Class
SPSL	Severe Pressure Sensation Loss
SS	Safety Population
STA	Single Technology Appraisal
SU	Sulfonylurea
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TEAEs	Treatment-Emergent Adverse Events
TG	Triglycerides
TIA	Transient Ischaemic Attack

TSD	Technical Support Documents
TTT	Treat-to-Target
TZD	Thiazolidinediones
TZP	Tirzepatide
UK	United Kingdom
UKPDS	United Kingdom Prospective Diabetes Study
ULN	Upper Limit of Normal
USA	United States of America
VLDL	Very Low-Density Lipoprotein

Decision problem, description of the technology and clinical care pathway

B.1.1 *Decision problem*

The decision problem addressed within this submission is broadly consistent with the National Institute of Health and Care Excellence (NICE) final scope for this appraisal. The population defined in the final scope is consistent with anticipated marketing authorisation of tirzepatide for the treatment of type 2 diabetes mellitus (T2D). The decision problem is summarised in Table 1.

The full anticipated marketing authorisation for tirzepatide (Mounjaro®) is:

- for the treatment of adults with insufficiently controlled type 2 diabetes mellitus as an adjunct to diet and exercise
    - as monotherapy when metformin is considered inappropriate due to intolerance or contraindications
    - in addition to other medicinal products for the treatment of diabetes

The expected eligible population for tirzepatide in NHS clinical practice, and the focus of this submission, is a narrower population than the anticipated marketing authorisation: it is expected that clinicians would use tirzepatide in patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. This anticipated position aligns with current NHS clinical practice in England and reflects the highest unmet need for a more effective treatment option for patients for whom the alternative is a GLP-1 RA, which may not sufficiently control their HbA1c level and/or provide sufficient weight loss.

Table 1: The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	Tirzepatide monotherapy: • Adults with type 2 diabetes that is inadequately controlled with diet and exercise alone and in whom the use of metformin is considered inappropriate Tirzepatide with other antidiabetic agents: • Adults with type 2 diabetes that is inadequately controlled with one or more antidiabetic agents	Tirzepatide with other antidiabetic agents: • Adults with T2D that is inadequately controlled with **three or more** antidiabetic agents	This submission positions tirzepatide for use in patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. This is the anticipated positioning of tirzepatide in UK clinical practice.
Intervention	Tirzepatide alone or with other antidiabetic agents	Tirzepatide with other antidiabetic agents	As above
Comparator(s)	The following interventions as monotherapy or in combination regimens, in accordance with NICE guidance: • sulfonylureas • DPP-4 inhibitors • pioglitazone • GLP-1 mimetics • SGLT-2 inhibitors • insulin	The following interventions in combination regimens: GLP-1 RAs: • Dulaglutide • Exenatide (standard and modified-release formulations) • Liraglutide • Lixisenatide • Semaglutide (oral and injectable formulations)	GLP-1 RAs are considered the only relevant comparators for tirzepatide in this submission, as this aligns with the anticipated position for tirzepatide in the UK clinical pathway of care (see above).
Outcomes	The outcome measures to be considered include: • HbA1c/glycaemic control • complications of diabetes, including cardiovascular, renal and eye • mortality	The outcome measures to be included are: • Glycaemic control (HbA1c) • Change in body weight • Body Mass Index • Frequency and severity of	Aligned with the final NICE scope. A CV safety meta-analysis confirming CV safety is described in Section B.2.9. Further data on cardiovascular outcomes are not yet available; they are expected to become available upon completion of the SURPASS-CVOT trial

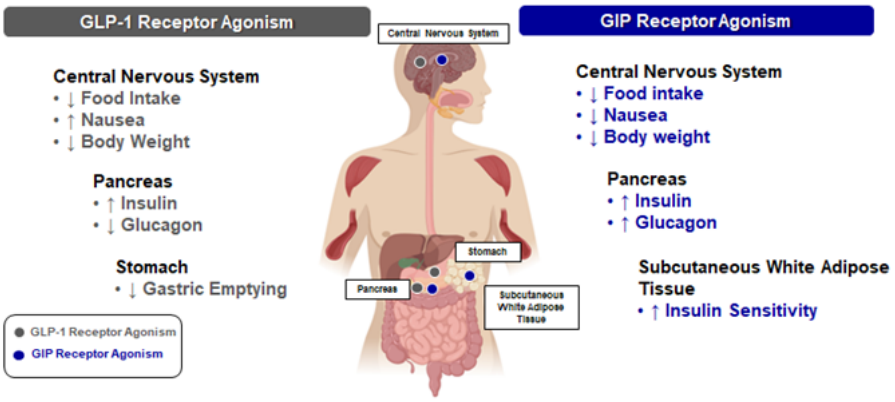
	• body mass index • frequency and severity of hypoglycaemia • changes in cardiovascular risk factors • adverse effects of treatment • health-related quality of life	hypoglycaemia • Adverse effects of treatment • Health-related quality of life (APPADL and IWQOL-LITE-CT)	in 2025.[1] A dedicated addendum study to SURPASS-CVOT is ongoing to further investigate the impact of tirzepatide treatment on diabetic retinopathy progression.

Abbreviations: APPADL: Ability to Perform Physical Activities of Daily Living; CVOT: cardiovascular outcomes trial; DTSQs: Diabetes Treatment Satisfaction Questionnaire-Status; DTSQc: Diabetes Treatment Satisfaction Questionnaire-Change; EQ-5D-5L: 5-level European quality of life 5 dimension score; GLP-1: glucagon-like peptide-1; HbA1c: glycated haemoglobin; IW-SP: Impact of Weight on Self-Perception; IWQOL-LITE-CT: Impact of Weight on Quality of Life-Lite Clinical Trials Version; T2D: type 2 diabetes.

B.1.2 *Description of the technology being evaluated*

A summary of the mechanism of action, marketing authorisation status, costs and administration requirements associated with tirzepatide for adults with T2D is provided in Table 2. The draft Summary of Product Characteristics (SmPC) is presented in Appendix C.

Table 2: Technology being appraised

UK approved name and brand name	Tirzepatide (Mounjaro®)
Mechanism of action	Tirzepatide is a first-in-class, long-acting single molecule designed to activate both glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 RA. It is a 39 amino acid peptide with a C20 fatty diacid moiety which is highly selective to both GIP and GLP-1 receptors. Its unique structure and receptor pharmacology distinguish the pharmacological profile of tirzepatide from that of the selective GLP-1 RA class. GIP and GLP-1 are incretin hormones which have multiple glucoregulatory actions, including a key role in enhancement of glucose-stimulated insulin secretion in pancreatic beta cells and control of glucagon secretion from pancreatic alpha cells (Figure 1). In people with T2D, the effect of incretins is diminished. In vitro studies demonstrated that the affinity of tirzepatide for both receptor types is high, with affinity to the GIP receptor comparable to native GIP and affinity for the GLP-1 receptor weaker than native GLP-1.[2] Preclinical and phase 2 clinical data indicated that co-stimulation of GIP and GLP-1 receptors may enhance insulin secretion, improve insulin sensitivity and reduce body weight beyond the effects produced by selective GLP-1 receptor stimulation. Based on preclinical studies, the improvement in insulin sensitivity is both weight-independent and -dependent. A clamp study in patients with T2D showed improved beta-cell insulin secretion and increased whole-body insulin sensitivity with tirzepatide vs selective GLP-1 RAs.[3] **Figure 1.Complementary actions of GLP-1 and GIP**  Actions of GIP reported are proposed and based on clinical and preclinical studies, and not all have been confirmed in humans. Data presented in the figure come from human and animal studies. **Abbreviations:** GIP: glucose-dependent insulinotropic polypeptide; GLP-1: glucagon-like peptide-1. **Source:** Samms et al. (2020).[4]
Marketing authorisation/CE mark status	An application for marketing authorisation was submitted to the Medicines and Healthcare products Regulatory Agency (MHRA) through the EC Decision Reliance Procedure (ECDRP) on 26th July 2022.

<table>
<tr><td></td><td>A marketing authorisation application for tirzepatide for the treatment of T2D was submitted to the European Medicines Agenda (EMA) in October 2021. A positive EMA Committee for Medicinal Products for Human Use (CHMP) opinion was adopted on 21st July 2022 and EC marketing authorisation decision is expected in [redacted]. The MHRA marketing authorisation through ECDRP is anticipated in [redacted].</td></tr>
<tr><td>Indications and any restriction(s) as described in the summary of product characteristics (SmPC)</td><td>Tirzepatide is anticipated to be licenced for the treatment of adults with insufficiently controlled type 2 diabetes mellitus as an adjunct to diet and exercise
as monotherapy when metformin is considered inappropriate due to intolerances or contraindications
in addition to other medicinal products for the treatment of diabetes</td></tr>
<tr><td>Method of administration and dosage</td><td>Tirzepatide is administered via injection once weekly (QW), using a single-dose pre-filled autoinjector pen device. The dose should be injected in the abdomen, thigh or upper arm, rotating the injection site with each dose. The dose can be administered at any time of day, with or without meals.[5]

Tirzepatide is initiated at 2.5 mg QW. After 4 weeks, increase to 5 mg QW. If needed, the dose can be increased in 2.5 mg increments every 4 weeks up to 15 mg. The recommended maintenance doses are 5 mg, 10 mg and 15 mg.</td></tr>
<tr><td>Additional tests or investigations</td><td>No additional tests are needed.</td></tr>
<tr><td>List price and average cost of a course of treatment</td><td>The list price for tirzepatide has not been agreed yet and remains under consideration. The following has been proposed and is confidential subject to approval. The prices are for each pack of 4 pre-filled single-dose autoinjector pen devices. Packs are available for the recommended maintenance doses (5 mg, 10 mg and 15 mg), and for the intermediate titration doses required when following the dose escalation recommendations.

<table>
<tr><th>Tirzepatide dose</th><th>List price</th></tr>
<tr><td>2.5 mg</td><td>[redacted]</td></tr>
<tr><td>5 mg</td><td>[redacted]</td></tr>
<tr><td>7.5 mg</td><td>[redacted]</td></tr>
<tr><td>10 mg</td><td>[redacted]</td></tr>
<tr><td>12.5 mg</td><td>[redacted]</td></tr>
<tr><td>15 mg</td><td>[redacted]</td></tr>
</table></td></tr>
<tr><td>Patient access scheme (if applicable)</td><td>The fixed discounted prices for tirzepatide 5 mg, 10 mg and 15 mg are £[redacted], £[redacted] and £[redacted] per pack of 4 pre-filled single-dose autoinjector pen devices, respectively.

The annual cost of tirzepatide 5 mg, 10 mg and 15 mg at the fixed discounted prices are £[redacted], £[redacted] and £[redacted] per patient, respectively.</td></tr>
</table>

Abbreviations: CHMP: Committee for Medicinal Products for Human Use; ECDRP: EC Decision Reliance Procedure; EMA: European Medicines Agency; EPAR: European public assessment report; GIP: glucose-dependent insulinotropic polypeptide; GLP-1: glucagon-like peptide-1; MHRA: Medicines and Healthcare products Regulatory Agency; SmPC: summary of product characteristics; T2D: type 2 diabetes; QW: every week.

B.1.3 *Health condition and position of the technology in the treatment pathway*

Type 2 diabetes

- T2D is a progressive metabolic condition characterised by impaired glycaemic control and caused by increased insulin resistance, progressive pancreatic beta cell failure, and inadequate insulin secretion[6]
- It is estimated that one in ten adults over 40 years of age in the UK have a T2D diagnosis[7]
- Poor management of T2D can increase the risk of a range of chronic, potentially life-threatening complications including cardiovascular disease (CVD), retinopathy, neuropathy and peripheral vascular disease.[8-12] People with T2D have a poorer quality of life (QoL) compared to the general population[13]
- Obesity (BMI≥30 kg/m^2) is the strongest risk factor for T2D,[14-16] and is a leading contributor to insulin resistance in patients with T2D.[17-21]
- There is an association between increasing BMI or obesity and poorer health related quality of life in people with T2D.[22-25]

Unmet need

- Improving T2D care has been recognised as a priority in the NHS Long-Term Plan[26]
- Despite the availability of multiple classes of T2D treatment, substantial numbers of patients with T2D do not reach their goals for glycaemic control, weight loss, blood pressure control or lipid control[21, 27-29]
- In 2019–20, only 65.6% of patients with T2D in England achieved a target level of HbA1c ≤7.5% (<58 mmol/mol).[10] More recent data from the National Diabetes Audit Quarterly Reports reported that this proportion was 63.7% in 2020–21. Even fewer patients achieved more stringent targets of ≤6.5% (≤48 mmol/mol; when on a drug not associated with hypoglycaemia) or <7.0% (<53 mmol/mol) when on a drug associated with hypoglycaemia) as recommended by NICE: 32.0% and 50.4%, respectively
- 90% of adults with T2D aged 16–54 years are overweight or obese[30]
- Higher BMI is associated with a higher proportion of patients with uncontrolled HbA1c[23, 31-33]
- For adults with T2D, modest and sustained weight loss is associated with improvement in glucose control, blood pressure, lipids and overall health[34, 35]
- There is therefore a clear unmet need for new treatments which help patients achieve a body weight reduction alongside further improvements in glycaemic control beyond those currently available in the NHS

Clinical pathway of care

- Metformin is used as a first-line treatment for T2D, unless contraindicated or not tolerated. Metformin may be prescribed in combination with a sodium-glucose co-transporter-2 inhibitor (SGLT2i) in patients with a high risk of CVD as a first-line treatment
- HbA1c levels are monitored to assess the efficacy of treatment. If HbA1c levels are not controlled by the patient's treatment, switching treatments or adding a second and third oral drug can be considered
- If dual therapy is not adequately controlling HbA1c levels, either triple therapy by adding another oral drug, or starting insulin-based treatment (±other drugs), is considered
- Triple therapy including a GLP-1 RA is considered for patients for whom triple therapy with metformin and two other oral drugs is not effective or tolerated, and who:
    - have a BMI ≥35 kg/m^2 (adjusted accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity **or**
    - who have a BMI <35 kg/m^2 but for whom insulin therapy would have significant occupational implications or when weight loss would benefit other significant obesity related complications[36]

Tirzepatide positioning

- The anticipated positioning of tirzepatide is for patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered

B.1.3.1 Disease overview

T2D is a progressive metabolic condition characterised by impaired glycaemic control and caused by increased insulin resistance, progressive pancreatic beta cell failure, and inadequate insulin secretion.

Risk factors for developing T2D include obesity, physical inactivity and high carbohydrate consumption.[30, 37] Around 90% of people with T2D aged 16–54 years are overweight or obese, and a higher BMI is associated with worse glycaemic control.[30, 31]

Epidemiology

It is estimated that one in ten adults over 40 years of age in the UK have a T2D diagnosis, with that number expected to increase in the future.[7] As of 2019, over 3.9 million people in the UK were living with a diabetes diagnosis, and 90% of those cases were T2D. In addition, it was estimated that almost a million people had T2D but were undiagnosed in the UK in 2019.[7]

The population of patients with T2D is shifting over time. Since 1998, the number of people living with T2D has more than doubled, an increase largely believed to be driven by the rising global prevalence of obesity.[38, 39 40] While there is an association between increasing age and greater diabetes prevalence,[30] the proportion of children and young people diagnosed with T2D is growing.[41, 42] This is a sub-population of particular concern as age at T2D diagnosis and T2D duration are independently associated with macrovascular events; diabetes duration is also independently associated with microvascular events, an effect which is greater in the youngest patients.[43]

Morbidity and mortality

T2D is associated with significant morbidity and mortality.[44] Poor management of T2D can increase the risk of a range of chronic, potentially life-changing or even life-threatening complications including CVD, retinopathy, nephropathy, neuropathy and peripheral vascular disease.[8-12]Overall life expectancy is reduced, on average, by up to 10 years in patients with T2D.[45]

Glycaemic control

T2D is characterised by impaired glycaemic control, and studies have shown that achieving glycaemic control reduces the risk of T2D-associated complications.[46-49] HbA1c levels controlled to <7.0% (<53 mmol/mol) and ≤6.5% (≤48 mmol/mol) were associated with lower risks of macrovascular and microvascular events, respectively,[49] whilst higher glycaemic variability was associated with significantly higher risk of nephropathy, macrovascular events and mortality.[46]

Further evidence suggests that achieving early glycaemic control may generate a legacy effect of reduced risk of microvascular complications, myocardial infarction and mortality for up to 10 years.[50, 51]

Whilst published guidance recommends general HbA1c targets, individualised HbA1c targets are recommended for use in routine clinical practice. For example, for patients who manage their T2D using a single drug not associated with hypoglycaemia, NICE guidelines recommend to support patients to a target HbA1c level of ≤6.5% (48 mmol/mol); for patients with T2D on dual or triple therapy, or a single drug associated with hypoglycaemia, a target of <7.0% (53 mmol/mol) is recommended. If patients with T2D reach a lower HbA1c level than their target and are not experiencing hypoglycaemia, they are encouraged to maintain that level.[36]

T2D and obesity

90% of patients with T2D aged 16–54 years are overweight or obese.[30] Obesity (BMI≥30 kg/m^2) is the strongest risk factor for T2D,[14-16] and is the leading contributor to insulin resistance in patients with T2D.[17-21] Higher BMI is associated with a higher proportion of patients with uncontrolled HbA1c levels.[31, 40, 52] Further, obesity contributes to the mortality and morbidity experienced by patients with T2D, with patients with T2D and higher BMI scores having a greater risk of all-cause mortality.[53]

According to the American Diabetes Association (ADA) 2022 guidelines, weight loss improves glycaemic control, blood pressure, and lipids in people with T2D who are overweight or obese. The benefits of weight loss are progressive: clinical benefits typically begin upon achieving 3–5% weight loss, and more intensive weight-loss goals (>5%, >7%, >15%, etc.) may be pursued to achieve further health improvements. A post hoc secondary analysis of intensive lifestyle intervention study demonstrated a mitigation in cardiovascular risk in those who achieved and maintained >10% weight loss.[35]

Health-related quality of life

Patients with T2D have a poorer health-related quality of life (HRQoL) compared to the general population and this has been shown to worsen as complications develop.[13, 22, 54-57] HRQoL was reduced in patients with T2D compared to a control population in all 8 dimensions of a French Short-Form 36-item (SF-36) survey.[58]

Poor glycaemic control independently contributes to the reduced HRQoL experienced by patients with T2D.[59-61] A study of 510 patients across a 4-year period measured an association between higher median HbA1c, a lower median Audit of Diabetes-Dependent Quality of Life (ADDQoL) scores and mean physical health scores. After adjustment for other factors that may affect HRQoL, it was estimated that for every 1% increase in HbA1c there was a 38% increase in reporting a negative impact of diabetes on HRQoL, as measured by the ADDQoL scale.[59]

Multiple studies in patients with T2D have demonstrated that overweight and obesity is associated with poorer HRQoL.[22-25] T2D-associated complications also contribute to reduced HRQoL; patients with T2D and complications experienced a greater impact on health and a reduced ability to complete daily activities, compared with those without complications.[62] As well as physical impairment, the impact of T2D on HRQoL includes emotional distress and depression; this is linked to self-perception, long-term health and ability to perform daily activities.[63] Higher emotional distress scores are also associated with higher HbA1c and BMI. [63] Weight loss and decreased BMI are associated with significant improvements in HRQoL, with the same trend emerging across several measures of HRQoL.[64-66]

B.1.3.2 Unmet need in the treatment of T2D

Despite the availability of multiple classes of T2D treatment , substantial numbers of patients with T2D do not reach their goals for glycaemic control, weight loss, blood pressure control or lipid control.[21, 27-29] Improving T2D care through continued investment has been recognised as a priority in the NHS Long-Term Plan with the aim to enable more people to achieve the recommended diabetes treatment targets.[26] In 2019–20, only 65.6% of patients with T2D in England achieved an HbA1c ≤7.5%, and that dropped to 63.7% in 2020-21. Even fewer achieved the NICE recommended targets in 2020–21: 50.4% reached HbA1c <7.0% (53 mmol/mol; targeted when on a drug associated with hypoglycaemia), and 32.0% reached HbA1c ≤6.5% (≤48 mmol/mol; targeted when on a drug not associated with hypoglycaemia).[67]

A high proportion of the people with T2D who do not succeed in meeting glycaemic control targets are overweight or obese, and many patients do not achieve adequate weight loss on current T2D treatments.[31, 68, 69] The reduced HRQoL experienced by patients with T2D is influenced by poor glycaemic control, obesity and complications, such as CVD, nephropathy, proteinuria and end-stage renal disease.

There is a clear unmet need for more efficacious treatment options to help more patients with T2D achieve a body weight reduction alongside improvements in glycaemic control, beyond what can be achieved with currently available therapies. Treatment options which offer the added benefit of weight loss may provide additional benefits for patients with T2D, including greater glycaemic control which is an important risk factor in the development of complications.

B.1.3.3 Economic burden

T2D exerts economic strain on both the healthcare system and on individuals. In the UK, diabetes accounts for approximately 10% of the total health resource expenditure, with the NHS spending over £6 billion on T2D and its complications in 2018.[70] Around 80% of the total cost of diabetes to the NHS is spent on complications;[71, 72] CVD and renal complications may increase direct costs by up to 6.3-fold.[73] A study in 2012 estimated that by 2035/36, T2D will cost the UK £15.1bn in direct costs and £20.5bn in indirect costs.[72]

People with both T2D and obesity incur higher costs of all-cause and diabetes-related drugs, outpatient care, and diabetes-related acute care.[74-77] One contributor to these increased costs is that patients with both T2D and obesity have a higher incidence of diabetes-related complications and cardiovascular events.[74]

Improving T2D care will help reduce the economic burden of T2D.[72] Good glycaemic control when compared with poor glycaemic control reduces all-cause total healthcare costs, diabetes-related healthcare costs, and diabetes-related hospitalisation costs.[78, 79] Additionally, weight loss in patients with T2D decreases healthcare costs, demonstrating the further benefits of treatments which help patients achieve a body weight reduction alongside improvements in glycaemic control.[68, 74, 80-82]

B.1.3.4 Clinical pathway of care

Clinical guidelines in NHS England

Treatment decisions in NHS England clinical practice are largely guided by the NICE T2D guideline NG28.[36]

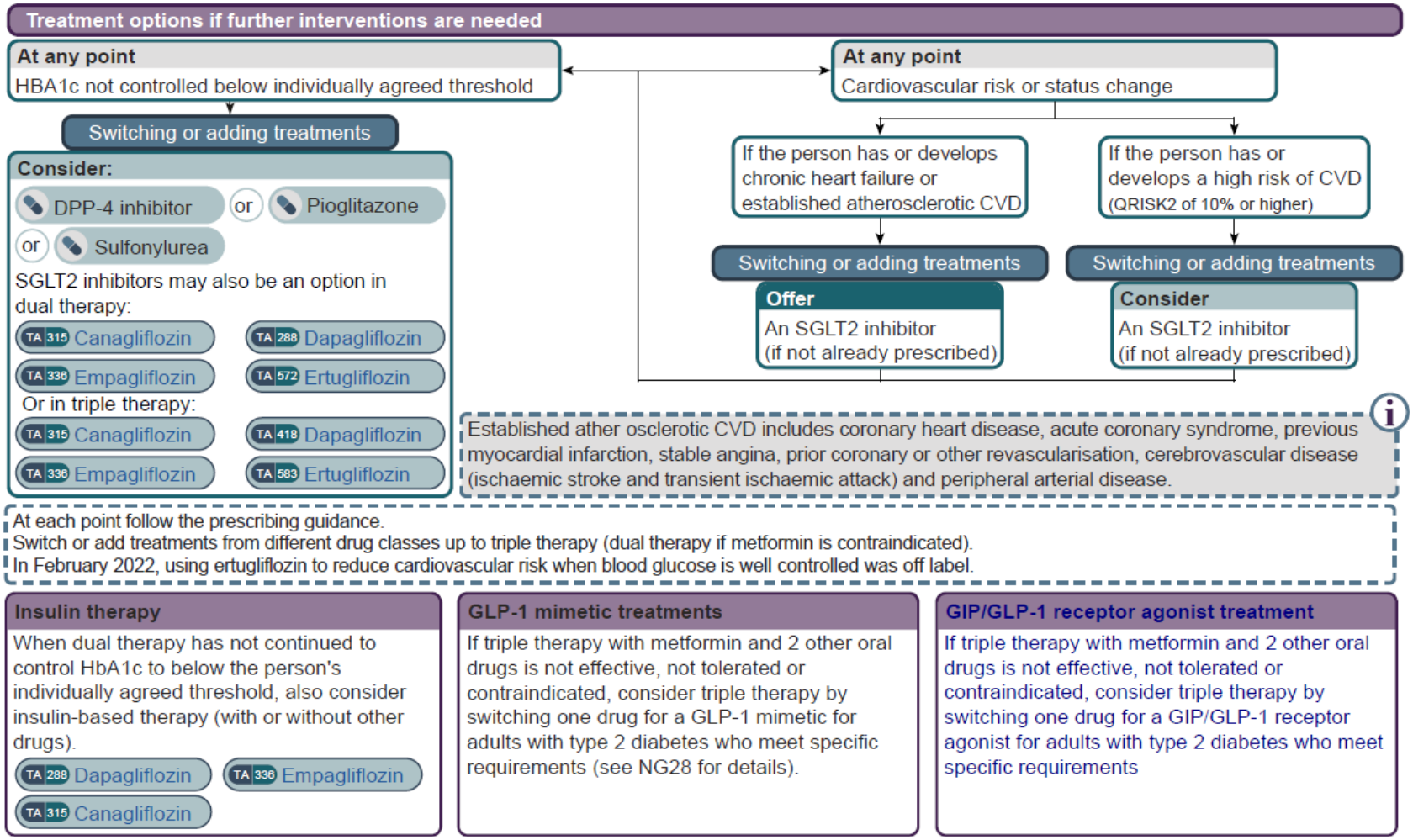
Initial management of T2D includes structured education, dietary and lifestyle advice. Where pharmacological management is preferred or required, metformin is used as a first-line treatment, unless contraindicated or not tolerated. Metformin may be prescribed in combination with an SGLT2i in patients at a high risk of developing CVD or who have established CVD.

HbA1c levels are monitored to assess the efficacy of treatment. If HbA1c levels are not controlled by the patient's current treatment, switching treatments or adding a second and third oral drug can be considered. If dual therapy is not adequately controlling HbA1c levels, either triple therapy by adding another oral drug, or starting insulin-based treatment (with or without other drugs), is considered. Triple therapy including a GLP-1 RA is considered for patients for whom triple therapy with metformin and two other oral drugs is not effective or tolerated and who:

- have a BMI ≥35 kg/m^2 (adjust accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity, or
- who have a BMI <35 kg/m^2 **and**:
    - for whom insulin therapy would have significant occupational implications **or**
    - when weight loss would benefit other significant obesity related comorbidities.[36]

The clinical pathway of care for T2D following insufficient control of HbA1c levels on first-line therapy, as recommended by NICE, and the proposed positioning of tirzepatide within this pathway are summarised in Figure 2.

Figure 2: Anticipated positioning of tirzepatide alongside NICE T2D NG28 clinical guidelines for patients following insufficient control of HbA1c levels on first-line therapy



Abbreviations: BMI: body mass index; CVD: cardiovascular disease; DPP-4: dipeptidyl peptidase-4; GIP: glucose-dependent insulinotropic polypeptide; GLP-1: glucagon-like peptide-1; HbA1c: glycated haemoglobin; OAD: oral antidiabetic drug; SGLT2i: sodium glucose cotransporter-2 inhibitor.
Source: NICE guidelines on management type 2 diabetes.[36]

International guidelines

International guidance from the European Association for the Study of Diabetes (EASD) and ADA prioritise patient-centred care and self-management education for T2D, including advising overweight and obese patients of the health benefits of weight loss.[83]

Where pharmacological management is preferred or required, metformin continues to be the first-line recommended therapy for patients with T2D, which is aligned with NICE NG28.[36] When required due to inadequately controlled HbA1c levels, add-on medications are selected based on patient preference and clinical characteristics, such as CVD, kidney disease, the need to optimise weight loss or minimise weight gain and risk of adverse medication events (e.g. hypoglycaemia). GLP-1 RAs are currently recommended as the first injectable medication for the treatment of T2D.[83]

This guidance is currently under revision and an updated ADA-EASD Consensus Report is due to be presented at EASD in September 2022. Updates under consideration include further clarity on the importance of weight control in T2D, as well as new therapeutic options that have launched since the last version, including tirzepatide.

B.1.3.5 Anticipated use of tirzepatide in NHS England clinical practice

As presented above, substantial numbers of patients with T2D do not achieve their treatment targets for glycaemic control, weight loss, blood pressure control or lipid control, and therefore experience increased morbidity, an increased risk of mortality, and poorer HRQoL.[21, 27-29] Higher BMI is also associated with a higher proportion of people with uncontrolled T2D (HbA1c ≥7% / 53mmol/mol). .[31, 52]

Tirzepatide is the first member of the new GIP/GLP-1 RA class. Treatment with tirzepatide has demonstrated significant improvements in both glycaemic control and body weight reduction compared with placebo, insulin degludec, insulin glargine, and most notably semaglutide 1.0mg (an established GLP-1 RA) in the SURPASS trials (Section B.2.6).[84-88] Tirzepatide represents an important new treatment option to help more patients achieve greater glycaemic control and body weight reduction than if they receive a GLP-1 RA. Therefore, the anticipated positioning of tirzepatide is for patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever a GLP-1 RA would otherwise be considered.

B.1.4 *Equality considerations*

No equality issues related specifically to the use of tirzepatide are foreseen.

Clinical effectiveness

SURPASS trial programme

- The efficacy and safety of tirzepatide for the treatment of T2D were evaluated in five global randomised, controlled, phase 3 studies (SURPASS-1–5) including 6,263 treated patients (4,199 treated with tirzepatide)
  - The primary efficacy endpoint in all trials was glycaemic control
  - Secondary endpoints included body weight, fasting serum glucose (FSG) and the proportion of patients reaching HbA1c targets
  - All five phase 3 studies assessed tirzepatide 5 mg, 10 mg and 15 mg. All patients treated with tirzepatide followed a dosing algorithm beginning with 2.5 mg for 4 weeks, then increasing the dose of tirzepatide by 2.5 mg every 4 weeks until the assigned dose was reached
- SURPASS-2–5 form the main clinical evidence within this appraisal and data from SURPASS-1 are presented in the appendices as supporting evidence
- The SURPASS trials, including comparators and background medications, are summarised below

Only Tirzepatide	Tirzepatide + Other T2D Medicines			Tirzepatide + Insulin
SURPASS-1 vs placebo **478 patients** 40 weeks	**SURPASS-2** vs subcutaneous semaglutide In patients also taking metformin **1,878 patients** 40 weeks	**SURPASS-3** vs insulin degludec In patients also taking metformin with or without SGLT2i **1,437 patients** 52 weeks	**SURPASS-4** vs insulin glargine In patients also taking ≥1 and ≤3 OADs (metformin, SGLT2i, or sulfonylurea) **1,995 patients** 52 weeks (sub-set of patients followed up to 104 weeks)	**SURPASS-5** vs placebo In patients also taking insulin glargine with or without metformin **475 patients** 40 weeks

Abbreviations: OAD: oral antidiabetic drug; SGLT2i: sodium glucose co-transporter 2 inhibitor; T2D: type 2 diabetes.

Summary of efficacy

- Across all studies, treatment with tirzepatide at all doses demonstrated statistically significant reductions in HbA1c from baseline to the primary endpoint (Week 40 or 52) compared with either placebo or the active comparator (injectable semaglutide 1 mg, insulin degludec and insulin glargine) for up to 1 year. In SURPASS-4, effects were sustained in a subset of the population for up to 2 years

HbA1c change from baseline, % (mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	−2.1% (−22.8)	2.4% (−25.9)	−2.5% (−26.9)	−1.9% (−20.3)
SURPASS-3 (vs insulin degludec)	−1.9% (−21.1)	−2.5% (−24.0)	−2.4% (−26.0)	−1.3% (−14.6)
SURPASS-4 (vs insulin glargine)	−2.2% (−24.5)	−2.4% (26.6)	−2.6% (−28.2)	−1.4% (−15.7)
SURPASS-5 (vs placebo)	−2.2% [redacted]	−2.6% [redacted]	−2.6% [redacted]	−0.9% [redacted]

Source: Frías *et al*, 2021;[86] Ludvik *et al,* 2021;[87] Del Prato *et al,* 2021;[85] SURPASS-5 CSR.[89]

- Significantly higher proportions of patients achieved an HbA1c target of <7.0% (<53 mmol/mol) on all three doses of tirzepatide at the primary endpoint (Week 40 or 52) compared with either placebo or the active comparator in all four of the SURPASS trials presented. Similarly, significantly higher proportions of patients on all tirzepatide doses achieved the more stringent HbA1c targets of ≤6.5% (≤48 mmol/mol) and <5.7% (<39

mmol/mol)

Patients achieving HbA1c <7.0% (<53 mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	85.5%	88.9%	92.2%	81.1%
SURPASS-3 (vs insulin degludec)	82.4%	89.7%	92.6%	61.3%
SURPASS-4 (vs insulin glargine)	81.0%	88.2%	90.7%	50.7%
SURPASS-5 (vs placebo)	93.0%	97.4%	94.0%	33.9%

Source: Frias *et al*, 2021;[86] Ludvik *et al,* 2021;[87] Del Prato *et al,* 2021;[85] Dahl *et al*, 2021.[84]

- Since substantial numbers of patients do not reach their glycaemic goals on current standard of care, these results demonstrate that tirzepatide is well-placed to meet this considerable unmet need
- Across the studies, treatment with tirzepatide at all doses was associated with significant reductions in body weight from baseline to the primary endpoint (Week 40 or 52) compared to placebo and all active comparators

Body weight change from baseline (kg)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	[redacted] (−7.8)	[redacted] (−10.3)	[redacted] (−12.4)	[redacted] (−6.2)
SURPASS-3 (vs insulin degludec)	[redacted] (−7.5)	[redacted] (−10.7)	[redacted] (−12.9)	[redacted] (2.3)
SURPASS-4 (vs insulin glargine)	[redacted] (−7.1)	[redacted] (−9.5)	[redacted] (−11.7)	[redacted] (1.9)
SURPASS-5 (vs placebo)	[redacted] (−6.2)	[redacted] (−8.2)	[redacted] (−10.9)	[redacted] (1.7)

Source: SURPASS-2 CSR;[90] Frias *et al*, 2021;[86] SURPASS-3 CSR;[91] Ludvik *et al,* 2021;[87] SURPASS-4 CSR;[92] Del Prato *et al,* 2021;[85] SURPASS-5 CSR;[89] Dahl *et al*, 2021.[84]

- Significantly higher proportions of patients achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% from baseline to the primary endpoint (Week 40 or 52) compared with either placebo or the active comparator in all four of the SURPASS trials presented
- Tirzepatide therefore represents a treatment option that helps patients achieve body weight reductions alongside improvements in glycaemic control; this is particularly important given the relationship between weight and HbA1c
- Across the studies, from baseline to the primary endpoint (Week 40 or 52), treatment with tirzepatide demonstrated significant improvements in the ability to engage in activities of normal daily living and a significant reduction in the impact of weight on function and daily activities. Tirzepatide 15 mg also demonstrated statistically significant improvements in these patient-reported outcomes compared with semaglutide 1 mg. These results demonstrate that tirzepatide treatment improves patient HRQoL as well as helping them meet treatment targets
- Across the studies, treatment with all doses of tirzepatide demonstrated reductions in triglycerides, total cholesterol, low density lipoprotein cholesterol (LDL-C) and very low density lipoprotein cholesterol (VLDL-C) and increases in high density lipoprotein cholesterol (HDL-C) from baseline, demonstrating favourable changes in lipid markers for patients with T2D

Summary of safety

- Like the well-established safety profile of the GLP-1 RA class, the most commonly reported

treatment-emergent adverse events (TEAEs) in patients treated with tirzepatide included gastrointestinal disorders. Reports of nausea, vomiting and diarrhoea were mostly mild to moderate in severity and most frequently reported in the dose-escalation periods.

- In the placebo-controlled analysis set, [redacted]% of tirzepatide-treated patients and [redacted]% of placebo-treated patients reported ≥1 TEAE
- In the placebo-controlled analysis set, the percentage of patients reporting SAEs was similar across tirzepatide doses and placebo
- Overall, [redacted] of tirzepatide-treated patients and [redacted] of placebo-treated patients discontinued from the study due to an AE
- The risk of severe hypoglycaemia with tirzepatide treatment is low; [redacted] patients reported [redacted] episodes of severe hypoglycaemia across the phase 3 global studies, [redacted] of which occurred prior to reaching the maintenance dose. Of the [redacted] patients, [redacted] patients were on a background of insulin glargine or SU. Overall, the risk of severe hypoglycaemia was:
    - Higher when tirzepatide was used in combination with insulin glargine or SU, compared with other background glucose-lowering therapies studied, which has also been observed with other GLP-1 RAs
    - Similar between tirzepatide and GLP-1 RAs (semaglutide 1 mg and dulaglutide 0.75 mg)
    - Lower in tirzepatide-treated patients compared with basal insulin-treated patients
- This safety profile of tirzepatide will be familiar to the healthcare community and is readily managed by following the guidance in the SmPC and monitored via routine pharmacovigilance
- Full efficacy and safety results from SURPASS 2–5 are presented in Section B.2.6

Summary of results from the network meta-analysis

- As it is not feasible to conduct randomised controlled trials (RCTs) versus all relevant comparators in all clinical settings, a network meta-analysis (NMA) has been conducted to assess the relative efficacy and safety of tirzepatide versus GLP-1 RAs available in NHS practice; the results of this NMA inform clinical inputs within the cost-effectiveness model
- As GLP-1 RAs and tirzepatide exhibit a dose-response relationship in terms of efficacy and gastrointestinal side-effects, when interpreting the NMA, comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps; for example, tirzepatide 5 mg is compared to the lowest recommended maintenance dose of each comparator
- For HbA1c change from baseline, all three doses of tirzepatide demonstrated a statistically significantly greater reduction in HbA1c from baseline compared to all GLP1-RAs within the same recommended maintenance dose step
- For body weight change from baseline, all three doses of tirzepatide demonstrated a significantly greater reduction in body weight from baseline compared to all GLP-1 RAs within the same recommended maintenance dose step
- All doses of tirzepatide demonstrated significantly greater reductions in BMI compared to all other GLP-1 RAs at the same recommended maintenance dose step; although, BMI data for studies in the main analyses were limited.
- Sensitivity analyses were conducted to assess the robustness of the findings of the main analyses and were largely consistent with the main analyses
- The NMA provides robust results that are generalisable to UK clinical practice. Baseline characteristics were largely consistent across the included treatment arms and as such, the results are likely to be robust with minimal impact from prognostic variables. In addition, numerous sensitivity analyses were conducted to assess the robustness of the findings of the main analyses; results of the sensitivity analyses demonstrate the robustness of the results of the main analyses. Limited concerns with regards to inconsistency and heterogeneity were identified
- Overall, for glycaemic control and weight loss, tirzepatide demonstrated statistically significant improvements when compared to all GLP-1 RAs at the same recommended maintenance dose step

Summary

- The SURPASS trial programme and NMA have demonstrated the clinical efficacy and

safety of tirzepatide for the treatment of T2D throughout the clinical pathway of care used in the UK

- Given the high proportion of patients who do not meet glycaemic and weight loss targets on currently available treatments in the UK and the superior glycaemic control and weight loss results seen with tirzepatide treatment, tirzepatide represents an important treatment option that will help address the considerable unmet need of T2D in the UK

B.2.1 *Identification and selection of relevant studies*

The clinical evidence base for tirzepatide as a treatment for T2D is based on the phase 3 randomised, controlled SURPASS trials.

A clinical systematic literature review (SLR) was conducted in September 2021 to identify further relevant clinical evidence on the efficacy and safety of treatment of T2D, including tirzepatide, in patients with T2D who match the patient population of interest for this appraisal. The SLR was subsequently updated in October 2021 to ensure recently published evidence was included. The SLR was performed in alignment with review conduct guidelines including Cochrane, the Centre for Review and Dissemination (CRD) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

A total of 246 publications reporting on 205 unique studies were identified in the SLR. Of those, seven phase 3 studies with tirzepatide as the primary intervention were identified: SURPASS 1–5, SURPASS-J-Mono and SURPASS-J-Combo. CSRs were available for all tirzepatide studies while full text publications were available for SURPASS 1–5.

Full details of the SLR search strategy, study selection process and results can be found in Appendix D.

B.2.2 *List of relevant clinical effectiveness evidence*

The phase 3 randomised, controlled SURPASS trials provide evidence for the clinical effectiveness and safety of tirzepatide as a treatment for T2D. SURPASS 1–5 are complete whilst SURPASS-6 and SURPASS-CVOT are ongoing. SURPASS-AP-Combo has recently completed, but data are not yet available. Data from this study are not relevant to this submission because the study was conducted in an Asian population. A summary of the completed SURPASS trials is presented in Table 3.

SURPASS-2–5 are most relevant to this submission and are presented below as the main clinical effectiveness evidence for this submission. SURPASS-1 compares the efficacy and safety of tirzepatide monotherapy for the treatment of T2D with placebo in patients who were naïve to antihyperglycaemic injectable therapy; data from SURPASS-1 are therefore of limited relevance to this submission due to the disparity in treatment type and patient population to the anticipated use of tirzepatide in UK clinical practice, but are presented in Appendix M for completeness. SURPASS-J-Mono and SURPASS-J-Combo were conducted in a Japanese population and are therefore not considered generalisable to the UK population; they are not presented as part of the clinical evidence in this appraisal. Data from SURPASS-J-Mono and SURPASS-J-Combo are included in the safety analysis in Section B.2.9.

The trials are presented in order based on the relevance of the comparators and patient population to the anticipated positioning of tirzepatide in the clinical treatment pathway, for patients with T2D that is inadequately controlled with three or more antidiabetic agents, when the GLP-1 RA class would otherwise be considered:

- SURPASS-2 is the most relevant to this decision problem because injectable semaglutide (a GLP-1 RA) is the comparator in this trial. SURPASS-2 allows the efficacy of tirzepatide to be assessed in comparison to a well-established and efficacious standard of care on a background of metformin (Section B.2.9). The clinical effectiveness evidence for SURPASS-2 is summarised in Table 4
- SURPASS-3 is relevant to this decision problem with a patient population that is close to the anticipated positioning of tirzepatide in the UK clinical treatment pathway: those who have received 1–2 prior therapies (metformin with or without an SGLT2i). However, the comparator for this trial is insulin degludec, which is not a relevant comparator for this submission due to the anticipated positioning of tirzepatide. The clinical effectiveness evidence for SURPASS-3 is summarised in Table 5
- SURPASS-4 is also relevant to this decision problem, as the population is the closest to the anticipated positioning of tirzepatide with some patients having previously received triple oral therapy (metformin, an SU and an SGLT2i). Whilst the population is narrower than the anticipated positioning due to all patients having high CV risk, this study provides important CV safety data in patients treated with tirzepatide. However, the comparator for this trial is insulin glargine, which is not a relevant comparator for this submission due to the anticipated positioning of tirzepatide. The clinical effectiveness evidence for SURPASS-4 is summarised in Table 6
- SURPASS-5 also provides relevant supportive data, comparing the efficacy and safety of tirzepatide for the treatment of T2D with placebo in patients with background therapy of insulin glargine, with or without metformin. The clinical effectiveness evidence summaries for SURPASS-5 are presented in Table 7

Table 3: Summary of the SURPASS trials

Study	Background Therapy	Comparator	Comparator or Background Therapy Titration Regimen*	Time to Primary Endpoint
SURPASS-2 (GPGL)	Metformin	Injectable semaglutide 1 mg	**Comparator**: Starting dose of semaglutide was 0.25 mg once weekly, the dose was doubled every 4 weeks until the 1 mg dose was reached.	40 weeks
SURPASS-3 (GPGH)	Metformin ± SGLT2i	Insulin degludec	**Comparator**: Starting dose of insulin was 10 units once daily. Patients adjusted their insulin degludec doses once weekly to a target fasting blood glucose of <90 mg/dL (5.0 mmol/L) based on the median value of the last 3 self-monitored blood glucose (SMBG) values according to a treat-to-target algorithm	52 weeks
SURPASS-4** (GPGM)	Metformin ± SU ± SGLT2i	Insulin glargine	**Comparator**: Starting dose of insulin was 10 units once daily. Patients adjusted their insulin glargine doses once weekly to a target fasting blood glucose of <100 mg/dL (5.6 mmol/L) based on the median value of the last 3 SMBG values according to a treat-to-target algorithm	52 weeks***
SURPASS-5 (GPGI)	Insulin glargine ± metformin	PBO	**Background therapy**: insulin glargine was titrated by patients using a protocol defined treat-to-target algorithm to reach a target fasting blood glucose of <100 mg/dL (5.6 mmol/L).	40 weeks

*All patients treated with tirzepatide started with 2.5 mg for 4 weeks, the dose of tirzepatide was then increased by 2.5 mg every 4 weeks until they reached their assigned dose. **High CVD risk population; ***A subset of patients were followed up to 104 weeks.
Abbreviations: PBO: placebo; SGLT2i: sodium glucose cotransporter 2 inhibitor; SMBG: self-monitored blood glucose; SU: sulfonylurea; TZP: tirzepatide.
Sources: SURPASS-2 CSR,[90] SURPASS-3 CSR,[91] SURPASS-4 CSR,[92] SURPASS-5 CSR.[89]

Table 4: Clinical effectiveness evidence for SURPASS-2

Study	SURPASS-2 (NCT03987919)		
Study design	Randomised, open-label, dose-blind, active-controlled, international, multicentre phase 3 trial assessing the efficacy and safety of tirzepatide for the treatment of T2D, compared to semaglutide		
Population	Patients with T2D, who had inadequate glycaemic control with metformin monotherapy (≥1500 mg/day) and had not been treated with any other OADs during the 3 months prior to the start of the study		
Intervention(s)	Tirzepatide 5 mg, 10 mg, 15 mg administered once weekly via single-dose pen. The dose of tirzepatide received was double-blinded Tirzepatide dosing algorithms started at 2.5 mg accompanied by dose escalation of 2.5 mg-increments every four weeks until the treatment dose (5 mg, 10 mg or 15 mg) was reached		
Comparator(s)	Semaglutide 1 mg administered once weekly via single-dose pen Semaglutide dosing started at 0.25 mg once weekly and the dose was doubled every four weeks until the 1 mg dose was reached.		
Indicate if trial supports application for marketing authorisation	Yes	**Indicate if trial used in the economic model**	Yes
Rationale for use/non-use in the model	The analysis population used in the NMA and therefore in the model included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. See Section B.2.9.5.1 for further information.		
Reported outcomes specified in the decision problem (outcomes in bold are incorporated into the model base-case)	• **HbA1c CfB** • Proportion of patients achieving HbA1c targets of <7.0% (<53 mmol/mol), ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol) • BMI CfB • Safety (including **SBP CfB**) ○ Adverse events (**nausea**, hypoglycaemia)		

All other reported outcomes
(outcomes in bold are incorporated into the model base-case)
	• **Body weight CfB**
	• Proportion of patients achieving weight loss targets of ≥5%, ≥10% and ≥15%
	• **Lipids CfB** (triglycerides, total cholesterol, **HDL-C, LDL-C** and VLDL-C)
	• Health-related quality of life
	  ○ APPADL scores CfB
	  ○ IWQOL-Lite-CT scores CfB		

Abbreviations: APPADL: ability to perform physical activities of daily living; BMI: body mass index; CfB: change from baseline; HbA1c: glycated haemoglobin; HDL-C: high-density lipoprotein cholesterol; IWQOL-Lite-CT: Impact of Weight on Quality of Life-Lite Clinical Trials Version; LDL-C: low-density lipoprotein cholesterol; OAD: oral antidiabetic drug; NMA: network meta-analysis; SBP: systolic blood pressure; T2D: type 2 diabetes; VLDL-C: very-low-density lipoprotein cholesterol.
Source: Frias et al. (2021);[86] SURPASS-2 CSR.[90]

Table 5: Clinical effectiveness evidence for SURPASS-3

Study	SURPASS-3 (NCT03882970)		
Study design	Randomised, open-label, international, multicentre phase 3 trial assessing the efficacy and safety of tirzepatide for the treatment of T2D, compared to insulin degludec		
Population	Patients with T2D, who had inadequate glycaemic control on stable doses of metformin with or without an SGLT2i		
Intervention(s)	Tirzepatide 5 mg, 10 mg, 15 mg administered once weekly via single-dose pen Tirzepatide dosing algorithms started at 2.5 mg accompanied by dose escalation of 2.5 mg-increments every four weeks until the treatment dose (5 mg, 10 mg or 15 mg) was reached		
Comparator(s)	Titrated insulin degludec (titrated to a fasting blood glucose of <90 mg/dL [5.0 mmol/L])		
Indicate if trial supports application for marketing authorisation	Yes	**Indicate if trial used in the economic model**	Yes
Rationale for use/non-use in the model	The analysis population used in the NMA and therefore in the model included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. See Section B.2.9.5.1 for further information.		
Reported outcomes specified in the decision problem (outcomes in bold are incorporated into the model base-case)	• **HbA1c CfB** • Proportion of patients achieving HbA1c targets of <7.0% (<53 mmol/mol), ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol) • BMI CfB • Safety (including **SBP CfB**) ○ Adverse events (**nausea**, hypoglycaemia)		
All other reported outcomes (outcomes in bold are incorporated into the model base-case)	• **Body weight CfB** • Proportion of patients achieving weight loss targets of ≥5%, ≥10% and ≥15% • **Lipids CfB** (triglycerides, total cholesterol, **HDL-C, LDL-C** and VLDL-C) • APPADL scores CfB		

Abbreviations: APPADL: ability to perform physical activities of daily living; BMI: body mass index; CfB: change from baseline; HbA1c: glycated haemoglobin; HDL-C: high density lipoprotein cholesterol; HRQoL: health-related quality of life; LDL-C: low density lipoprotein cholesterol; NMA: network meta-analysis; OAD: oral antidiabetic drug; SBP: systolic blood pressure; SGLT2i: sodium glucose cotransporter 2 inhibitor; T2D: type 2 diabetes; VLDL-C: very low density lipoprotein cholesterol.
Source: Ludvik et al. (2021);[87] SURPASS-3 CSR. [91]

Table 6: Clinical effectiveness evidence for SURPASS-4

Study	SURPASS-4 (NCT03730662)		
Study design	Randomised, open-label, international, multicentre phase 3 trial assessing the efficacy and safety of tirzepatide for the treatment of T2D, compared to insulin glargine		
Population	Patients with T2D with high CVD risk, who had inadequate glycaemic control on stable doses of at least 1 and no more than 3 oral antidiabetic drugs (OADs), including metformin, an SGLT2i and/or an SU		
Intervention(s)	Tirzepatide 5 mg, 10 mg, 15 mg administered once weekly via single-dose pen Tirzepatide dosing algorithms started at 2.5 mg accompanied by dose escalation of 2.5 mg-increments every four weeks until the treatment dose (5 mg, 10 mg or 15 mg) was reached.		
Comparator(s)	Titrated insulin glargine (titrated to a fasting blood glucose of <100 mg/dL [5.6 mmol/L])		
Indicate if trial supports application for marketing authorisation	Yes	**Indicate if trial used in the economic model**	No
Rationale for use/non-use in the model	This trial was not included in the NMA. The analysis population used in the NMA and therefore in the model included studies conducted with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. SURPASS-4 was not included because it was only in a high CVD risk population of patients. See Section B.2.9.5.1 for further information.		

Reported outcomes specified in the decision problem
(outcomes in bold are incorporated into the model base-case)
	• HbA1c CfB
	• Proportion of patients achieving HbA1c targets of <7.0% (<53 mmol/mol), ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol)
	• BMI CfB
	• Safety (including SBP CfB)
	  ○ Adverse events (nausea, hypoglycaemia)
	  ○ Cardiovascular outcomes		

All other reported outcomes (outcomes in bold are incorporated into the model base-case)	• Body weight CfB • Proportion of patients achieving weight loss targets of ≥5%, ≥10% and ≥15% • Lipids CfB (triglycerides, total cholesterol, HDL-C, LDL-C and VLDL-C) • APPADL scores CfB		

Abbreviations: APPADL: ability to perform physical activities of daily living; BMI: body mass index; CfB: change from baseline; CVD: cardiovascular disease; HbA1c: glycated haemoglobin; HDL-C: high density lipoprotein cholesterol; HRQoL: health-related quality of life; LDL-C: low density lipoprotein cholesterol; NMA: network meta-analysis; OAD: oral antidiabetic drug; SBP: systolic blood pressure; SGLT2i: sodium glucose cotransporter 2 inhibitor; SU: sulfonylurea; T2D: type 2 diabetes; VLDL-C: very low density lipoprotein cholesterol.
Source: Del Prato et al. (2021);[85] SURPASS-4 CSR.[92]

Table 7: Clinical effectiveness evidence for SURPASS-5

Study	SURPASS-5 (NCT04039503)		
Study design	Randomised, double-blind, international, multicentre phase 3 trial assessing the efficacy and safety of tirzepatide for the treatment of T2D, compared to placebo		
Population	Patients with T2D, with background therapy of insulin glargine with or without metformin		
Intervention(s)	Tirzepatide 5 mg, 10 mg, 15 mg administered once weekly via single-dose pen Tirzepatide dosing algorithms started at 2.5 mg accompanied by dose escalation of 2.5 mg-increments every four weeks until the treatment dose (5 mg, 10 mg or 15 mg) was reached.		
Comparator(s)	Placebo		
Indicate if trial supports application for marketing authorisation	Yes	**Indicate if trial used in the economic model**	No
Rationale for use/non-use in the model	This trial was not included in the NMA. The analysis population used in the NMA and therefore in the model included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. See Section B.2.9.5.1 for further information.		
Reported outcomes specified in the decision problem (outcomes in bold are incorporated into the model base-case)	• HbA1c CfB • Proportion of patients achieving HbA1c targets of <7.0% (<53 mmol/mol), ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol) • BMI CfB • Safety (including SBP CfB) ○ Adverse events (nausea, hypoglycaemia)		
All other reported outcomes (outcomes in bold are incorporated into the model base-case)	• Body weight CfB • Proportion of patients achieving weight loss targets of ≥5%, ≥10% and ≥15% • Lipids CfB (triglycerides, total cholesterol, HDL-C, LDL-C and VLDL-C) • APPADL scores CfB		

Abbreviations: APPADL: ability to perform physical activities of daily living; BMI: body mass index; CfB: change from baseline; HbA1c: glycated haemoglobin; HDL-C: high density lipoprotein cholesterol; HRQoL: health-related quality of life; LDL-C: low density lipoprotein cholesterol; NMA: network meta-analysis; OAD: oral antidiabetic drug; SBP: systolic blood pressure; T2D: type 2 diabetes; VLDL-C: very low density lipoprotein cholesterol.
Source: Dahl et al. (2022);[84] SURPASS-5 CSR.[89]

B.2.3 *Summary of methodology of the relevant clinical effectiveness evidence*

All data and trial information to be presented primarily from publications where available and supplemented with data from the clinical study reports (CSRs). The endpoints most relevant to this appraisal have been presented in Section B.2.6; details on other endpoints recorded in the trials are available in the CSRs supplied alongside the submission.

B.2.3.1 Tirzepatide dosing in the SURPASS trials

All of the SURPASS trials followed the same treatment algorithm for tirzepatide dosing.

Patients were randomised to either 5 mg, 10 mg or 15 mg once weekly. Tirzepatide dosing algorithms started at 2.5 mg accompanied by dose escalation in 2.5 mg increments every four weeks until the treatment dose was reached. This dose escalation permitted time for development of tolerance to gastrointestinal (GI) effects. The tirzepatide dosing algorithm is summarised in Table 8 below. Following this dosing algorithm, it takes four weeks to reach a target dose of 5 mg, 12 weeks to reach a target dose of 10 mg and 20 weeks to reach a target dose of 15 mg.

Table 8: Tirzepatide dosing algorithm in the SURPASS trials

Week	Tirzepatide dose
Week 1–4	2.5 mg
Week 5–8	5 mg
Week 9–12	7.5 mg
Week 13–16	10 mg
Week 17–20	12.5 mg
Week 21 onwards	15 mg

Source: SURPASS-2 CSR; SURPASS-3 CSR; SURPASS-4 CSR; SURPASS-5 CSR.

All tirzepatide doses were administered once weekly via injection in the abdomen or thigh if self-administered, or upper arm if administered by a caregiver. There were no restrictions on the time of day each weekly dose of tirzepatide was administered. Patients were advised to administer the injections on the same day and same time each week and were asked to record the actual date and time of all dose administrations.

B.2.3.2 SURPASS-2

Trial design

SURPASS-2 is a phase 3, international, multicentre, randomised, open-label, parallel group, 40-week, active-controlled study designed to assess the efficacy and safety of three once-weekly doses of tirzepatide (5 mg, 10 mg and 15 mg) compared with once-weekly, injectable semaglutide (1 mg) in patients with T2D who have inadequate glycaemic control with metformin monotherapy (≥1500 mg/day) and had not been treated with any other OADs during the 3 months prior to the start of the study.

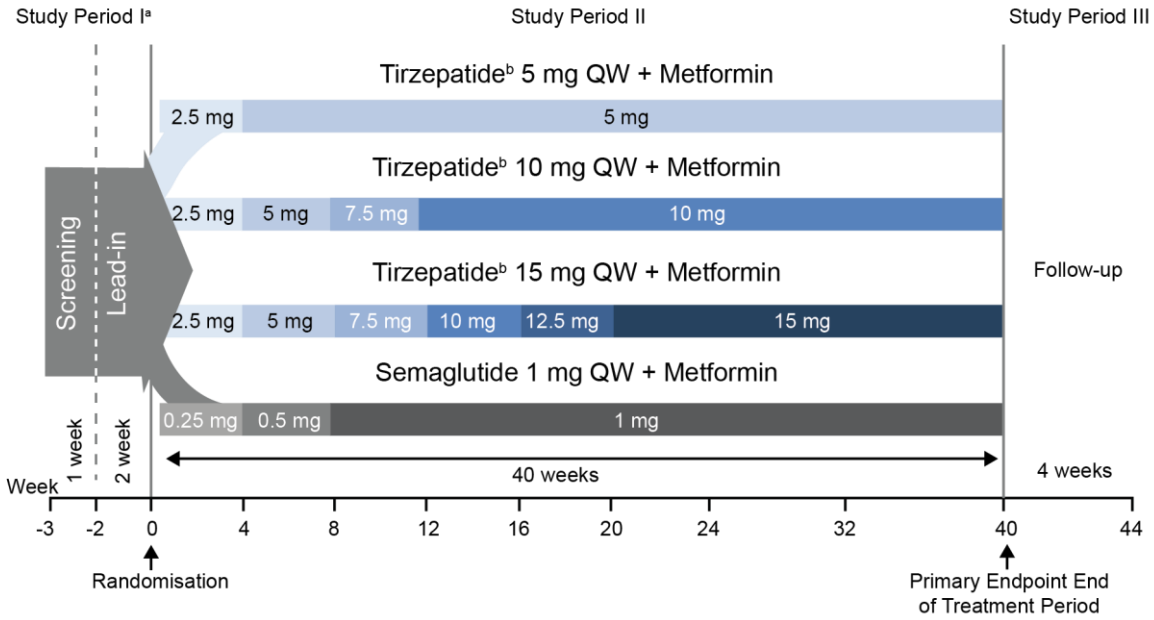
The trial had 3 study periods:

- Period I: screening and lead-in lasting 3 weeks
- Period II: treatment period lasting 40 weeks
- Period III: safety follow-up period lasting 4 weeks

The primary efficacy endpoint was the mean change in HbA1c from baseline to 40 weeks for tirzepatide 10 mg and 15 mg. Key secondary endpoints were mean change in HbA1c from baseline to 40 weeks for tirzepatide 5 mg, and body weight change from baseline to 40 weeks, the proportion of patients achieving the HbA1c target of <7.0%, (<53 mmol/mol) and <5.7% (<39 mmol/mol) at 40 weeks for all tirzepatide doses.

A summary of the trial design of SURPASS-2 is presented in Figure 3.

Figure 3: Study design of SURPASS-2



[a]Stable dose of metformin ≥1500 mg/day for at least 3 months prior to Visit 1 and during the screening/lead-in period. [b]All tirzepatide doses were double-blinded.
Abbreviations: QW: once weekly.
Source: SURPASS-2 CSR.[90]

Trial methodology

A summary of the methodology of SURPASS-2 is presented in Table 9. A summary of the pre-planned subgroups is presented in Section B.2.7.

Table 9: Summary of the methodology of SURPASS-2

Trial name	SURPASS-2 (NCT03987919)
Location	128 centres across 8 countries (Argentina, Australia, Brazil, Canada, Israel, Mexico, the United Kingdom, and the USA). Sites in the UK included both primary and secondary care facilities
Trial design	Phase 3, international, multicentre, randomised, open-label, 40-week study to assess the efficacy and safety of tirzepatide, compared with injectable semaglutide, for the treatment of patients with T2D as an add-on to metformin
Eligibility criteria for participants	**Eligibility criteria** • ≥18 years of age • HbA1c of ≥7.0% (≥53 mmol/mol) to ≤10.5% (≤91 mmol/mol) at Visit 1 • On stable diabetes treatment with metformin ≥1500 mg/day during the 3 months prior to visit 1 and between visits 1 and 3 • A stable weight for 3 months prior to Visit 1 and agreed not to initiate an organized diet or exercise program during the study with the intent of reducing body weight other than the lifestyle and dietary measures for diabetes treatment • BMI of at least 25 kg/m^2 at Visit 1

	Exclusion criteria
	• Type 1 diabetes mellitus
	• History of
	  ○ Proliferative diabetic retinopathy,
	  ○ Diabetic maculopathy, or
	  ○ Non-proliferative diabetic retinopathy that requires acute treatment
	• Used any antihyperglycemic medication (other than metformin) within the 3 months prior to Visit 1 (lead in, [Week-3]). An exception to this was the use of insulin to treat
	  ○ gestational diabetes
	  ○ acute conditions such as acute illness, hospitalization, or elective surgery (≤14 days)
	• Treated with prescription drugs that promote weight loss within 3 months prior to Visit 1 and/or between study entry (Visit 1) and randomization (Visit 3)

Intervention	After confirmation of the eligibility criteria, patients were randomised 1:1:1:1 to open-label once-weekly injectable tirzepatide 5 mg, 10 mg, 15 mg, or semaglutide 1 mg. Assignment to treatment group was determined by a computer-generated random sequence using an Interactive Web Response System (IWRS). This was an open-label study with respect to assignment to semaglutide versus tirzepatide due to differences in devices. Within the tirzepatide arms, the dose of tirzepatide was blinded to patients, investigators, and the sponsor.
Method of study drug administration	**Tirzepatide** • Tirzepatide doses: 5 mg, 10 mg, or 15 mg once weekly • The dosing algorithm for tirzepatide is summarised in Section B.2.3.1 **Semaglutide** • Semaglutide 1 mg • Semaglutide dosing began at 0.25 mg once weekly for 4 weeks. The dose was then increased to 0.5 mg once weekly for 4 weeks, then increased to and maintained at 1 mg once weekly for the duration of the study

Permitted and disallowed concomitant medication	The following concomitant medications were permitted during the study:
	• Metformin to treat T2D
	• After randomisation, discontinuation of metformin or change in dosage and formulation was only allowed in specific circumstances:
	  ○ The event of a hypoglycaemic episode(s)
	  ○ Certain clinical situations that required short-term discontinuation in line with the product(s) labelling for each country; for example, severe dehydration, elective surgery, or need for radiologic exam involving IV iodinated contrast dye
	  ○ Patient developed contraindications to metformin or an SGLT2i such that the use of

	the drug was contraindicated according to the country-specific label ○ Patient met the criteria for severe, persistent hyperglycaemia or discontinued study drug, then metformin dose could be increased according to country-specific label as long as that was not the sole intervention • Post-randomisation, patients were permitted to use concomitant medications that they required during the study, except certain medications that may interfere with the assessment of efficacy and safety characteristics of the study treatments • Antihyperglycaemic medications other than study drug were allowed only in specific circumstances, including severe persistent hyperglycaemic (rescue therapy) or early discontinuation of study treatment The following medications were prohibited during the study: • GLP-1 RAs • DPP-4 inhibitors • Pramlintide
Primary outcomes	Mean change in HbA1c values from baseline to 40 weeks for tirzepatide 10 mg and 15 mg.
Secondary and exploratory outcomes	**Key secondary efficacy endpoints (controlled for type 1 error)** • Mean CfB in HbA1c for tirzepatide 5 mg • Mean CfB in body weight for all tirzepatide doses • Proportion of patients achieving a target HbA1c <7% (53 mmol/mol) for all tirzepatide doses • Proportion of patients achieving HbA1c <5.7% (39 mmol/mol) for tirzepatide 10 mg and 15 mg **Additional secondary efficacy endpoints (not controlled for type 1 error; for all tirzepatide doses unless otherwise specified)** • Proportion of patients achieving a target HbA1c of ≤6.5% (48 mmol/mol) • Proportion of patients achieving HbA1c <5.7% (39 mmol/mol) for tirzepatide 5 mg • Mean CfB in FSG • Mean CfB in 7-point SMBG profiles • Proportion of patients who achieved weight loss ≥5%, ≥10% and ≥15% • Mean CfB in patient-reported outcomes, including DTSQs/DTSQc, IW-SP, and APPADL **Tertiary or exploratory efficacy endpoints (for all tirzepatide doses)** • Mean change in fasting glucose, C-peptide and insulin levels • Mean CfB in lipids (total cholesterol, HDL, VLDL, and triglycerides) • Mean CfB in BMI and waist circumference

	• Mean CfB in biomarkers • Mean CfB in patient-reported outcomes, including EQ-5D-5L scores and IWQOL-Lite-CT **Safety assessments** • AEs • Patient diaries • Concomitant medications • Dilated fundoscopic examinations were performed at baseline for all patients; follow-up dilated fundoscopic examinations were performed as deemed appropriate by the investigator • Vital signs • ECGs • Laboratory tests, including hepatic safety monitoring
Duration of study and follow-up	The study was initiated on 30th July 2019 and completed on 15th February 2021. The approximately 4-week safety follow-up period occurred after the last treatment visit for patients who either: • Completed the entire treatment period • Discontinued early and performed an early termination (ET) visit During the safety follow-up, patients did not receive study treatment and were treated with another glucose-lowering intervention decided upon by the investigator.

Abbreviations: AE: adverse event; ALT: alanine transaminase; APPADL: ability to perform physical activities of daily living; BMI: body mass index; CDK-EPI: chronic Kidney Disease-Epidemiology; CfB: change from baseline; DPP-4: dipeptidyl-peptidase 4; DTSQ(c/s): diabetes treatment satisfaction questionnaire (change/status); ECG: electrocardiogram; eGFR: estimated glomerular filtration rate; EQ-5D-5L: EuroQol-5 dimension-5 level descriptive system; ET: early termination; FSG: fasting serum glucose; GI: gastrointestinal; GLP-1 RA: glucagon-like peptide-1 receptor agonist; HbA1c: glycated haemoglobin; HDL: high density lipoprotein; IW-SP: impact of weight on self-perception; IWQOL-Lite-CT: impact of weight on quality of life lite clinical trials version; IWRS: Interactive Web Response System; MTC: medullary thyroid cancer; NAFLD: non-alcoholic fatty liver disease; OUS: outside the USA; SGLT2i: sodium glucose cotransporter 2 inhibitor; SMBG: self-monitored blood glucose; T2D: type 2 diabetes; ULN: upper limit of normal; USA: United States of America; VLDL: very low density lipoprotein.
Source: Frias *et al*, 2021;[86] SURPASS-2 CSR.[90]

Baseline characteristics

Baseline demographics and disease characteristics of the modified intent-to-treat (mITT) population of patients with T2D included in the final analysis of SURPASS-2 are presented in Table 10.

Table 10: Baseline characteristics of patients included in the SURPASS-2 trial

Characteristics	TZP 5 mg (N=470)	TZP 10 mg (N=469)	TZP 15 mg (N=470)	SEMA 1 mg (N=469)	Overall population (N=1878)
Demographics					
Age (years), mean ± SD	56.3 ± 10.0	57.2 ± 10.5	55.9 ± 10.4	56.9 ± 10.8	56.6 ± 10.4
Female, n (%)	265 (56.4)	231 (49.3)	256 (54.5)	244 (52.0)	996 (53.0)
Race, n (%)					
White	382 (81.3)	376 (80.2)	392 (83.4)	401 (85.5)	1551 (82.6)

Characteristics	TZP 5 mg (N=470)	TZP 10 mg (N=469)	TZP 15 mg (N=470)	SEMA 1 mg (N=469)	Overall population (N=1878)
American Indian or Alaska native	53 (11.3)	53 (11.3)	57 (12.1)	45 (9.6)	208 (11.1)
Asian	6 (1.3)	11 (2.3)	5 (1.1)	3 (0.6)	25 (1.3)
Black or African American	28 (6.0)	21 (4.5)	15 (3.2)	15 (3.2)	79 (4.2)
Multiple	1 (0.2)	8 (1.7)	0	3 (0.6)	12 (0.6)
Native Hawaiian or other Pacific Islander	0	0	1 (0.2)	2 (0.4)	3 (0.2)
Weight (kg), mean ± SD	92.5 ± 21.8	94.8 ± 22.7	93.8 ± 21.8	93.7 ± 21.1	93.7 ± 21.9
BMI (kg/m^2), mean ± SD	33.8 ± 6.9	34.3 ± 6.6	34.5 ± 7.1	34.2 ± 7.2	34.2 ± 6.9
BMI category, n (%)					
<30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30 to <35	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥35	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease Characteristics					
Duration of diabetes (years), mean ± SD	9.1 ± 7.2	8.4 ± 5.9	8.7 ± 6.9	8.3 ± 5.8	8.6 ± 6.5
HbA1c (%), mean ± SD	8.32 ± 1.08	8.30 ± 1.02	8.26 ± 1.00	8.25 ± 1.01	8.28 ± 1.03
HbA1c (mmol/mol), mean ± SD	67.46 ± 1.84	67.20 ± 11.20	66.78 ± 10.97	66.69 ± 10.99	67.03 ± 11.25
HbA1c category, n (%)					
≤8.5% (69 mmol/mol)	293 (62.3)	294 (62.7)	303 (64.5)	302 (64.4)	1192 (63.5)
>8.5% (69 mmol/mol)	177 (37.7)	175 (37.3)	167 (35.5)	167 (35.6)	686 (36.5)
FSG (mg/dL), mean ± SD	173.8 ± 51.9	174.2 ± 49.8	172.4 ± 54.4	171.4 ± 49.8	172.9 ± 51.5
FSG (mmol/L), mean ± SD	9.7 ± 2.9	9.7 ± 2.8	9.6 ± 3.0	9.5 ± 2.8	9.6 ± 2.9
Systolic blood pressure (mm Hg), mean ± SD	130.5 ± 14.1	131.5 ± 13.8	130.5 ± 14.3	130.0 ± 13.0	130.6 ± 13.8
Diastolic blood pressure (mm Hg), mean ± SD	78.6 ± 8.9	80.0 ± 9.6	79.0± 9.0	79.3 ± 8.6	79.2 ± 9.0
History of CV disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
eGFR (CKD-EPI, mL/min per 1.73 m2), mean ± SD	96.6 ± 17.5	95.5 ± 16.6	96.3 ± 16.9	95.6 ± 17.3	96.0 ± 17.1
eGFR category, n (%)					
<60 mL/min per 1.73 m^2	19 (4.0)	15 (3.2)	11 (2.3)	19 (4.1)	64 (3.4)

Characteristics	TZP 5 mg (N=470)	TZP 10 mg (N=469)	TZP 15 mg (N=470)	SEMA 1 mg (N=469)	Overall population (N=1878)
≥60 mL/min per 1.73 m^2	451 (96.0)	454 (96.8)	459 (97.7)	450 (95.9)	1814 (96.6)

Abbreviations: BMI: body mass index; bpm: beats per minute; CKD-EPI: chronic Kidney Disease-Epidemiology; CV: cardiovascular; eGFR: estimated glomerular filtration rate; FSG: fasting serum glucose; HbA1c: glycated haemoglobin; SD: standard deviation; SEMA: semaglutide; TZP: tirzepatide.
Source: Frias *et al,* 2021;[86] SURPASS-2 CSR.[90]

B.2.3.3 SURPASS-3

Trial design

SURPASS-3 is a phase 3, multicentre, randomised, open-label, dose-blind, parallel-group study that investigated the effects of treatment with tirzepatide 5, 10 and 15 mg every week, compared with insulin degludec (titrated to a fasting blood glucose of <90 mg/dL [5.0 mmol/L]) in patients with T2D naïve to insulin treatment who had inadequate glycaemic control on stable doses of metformin with or without an SGLT2i.

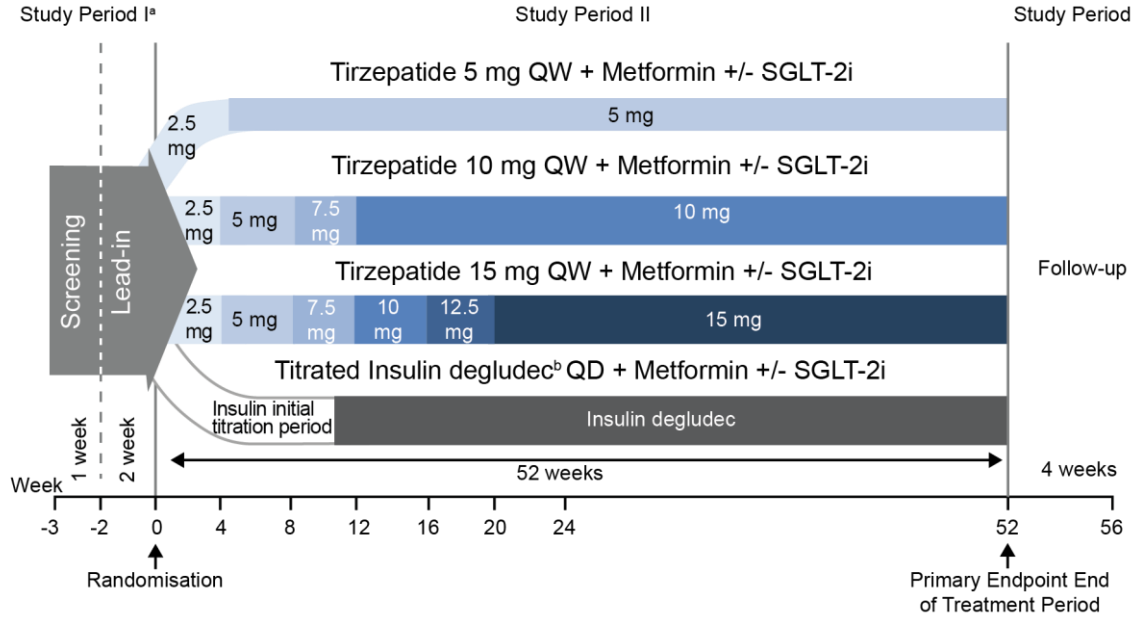
The trial had three study periods:

- Period I: screening and lead-in period lasted three weeks
- Period II: treatment period lasted 52 weeks
- Period III: safety follow-up period lasted four weeks

The primary efficacy endpoint was change in HbA1c from baseline to 52 weeks for tirzepatide 10 mg and 15 mg. Key secondary endpoints were change in HbA1c from baseline to 52 weeks for tirzepatide 5 mg, and change in body weight from baseline to 52 weeks and achievement of the HbA1c target of <7.0% (<53 mmol/mol) for all tirzepatide doses.

A summary of the trial design of SURPASS-3 is presented in Figure 4.

Figure 4: Study design of SURPASS-3



[a]Stable doses of metformin (≥1500 mg/day) ± SGLT-2i for ≥ 3 months prior to Visit 1 and during the screening/lead-in period. [b]The starting dose of insulin degludec was 10 IU/day ideally at bedtime, titrated to an FBG <90 mg/dL (5.0 mmol/L), following a TTT algorithm.
Abbreviations: FBG: fasting blood glucose; QD: once daily; QW: once weekly; SGLT-2i: sodium glucose co transporter 2 inhibitor; TTT: treat-to-target.
Source: SURPASS-3 CSR.[91]

Trial methodology

A summary of the methodology of SURPASS-3 is presented in Table 11. A summary of the pre-planned subgroups is presented in Section B.2.7.

Table 11: Summary of the methodology of SURPASS-3

Trial name	SURPASS-3 (NCT03882970)
Location	121 sites across 12 countries (Argentina, Austria, Greece, Hungary, Italy, Poland, Romania, South Korea, Spain, Taiwan, Ukraine, and the USA)
Trial design	A phase 3, international, multicentre randomised, open-label 52-week trial assessing the efficacy and safety of tirzepatide, compared with titrated insulin degludec, for the treatment of patients with T2D as an add-on to metformin with or without an SGLT2i

Eligibility criteria for participants	**Eligibility criteria**
	• ≥18 years of age
	• Naïve to insulin unless insulin was used to treat:
	  ○ Gestational diabetes
	  ○ Acute conditions such as acute illness, hospitalisation, or elective surgery (≤14 days)
	• Used metformin or metformin plus an SGLT2i for at least 3 months prior to Visit 1
	• Had HbA1c of ≥7.0% to ≤10.5% at Visit 1
	• Had a stable weight of 3 months prior to Visit 1
	• Had a BMI of at least 25 kg/m^2 at Visit 1
	• Agreed not to initiate an organised diet or exercise program during the study with the intent of reducing body weight other than the lifestyle and dietary measures for diabetes treatment
	
	Exclusion criteria
	• Type 1 diabetes mellitus
	• History of
	  ○ Proliferative diabetic retinopathy,
	  ○ Diabetic maculopathy, or
	  ○ Non-proliferative diabetic retinopathy that requires acute treatment
	• Used any glucose-lowering agent(s) other than metformin or metformin plus an SGLT-2i during the 3 months preceding Visit 1 and between Visit 1 (screening [Week −3]) and Visit 3 (randomization [Week 0])
	• Treated with prescription drugs that promote weight loss within 3 months prior to Visit 1 and/or between study entry (Visit 1) and randomization (Visit 3)

Intervention	After confirmation of the eligibility criteria, patients were randomised 1:1:1:1 to once-weekly injectable tirzepatide 5 mg, 10 mg, 15 mg, or titrated insulin degludec. Assignment to treatment group was determined by a computer-generated random sequence using an IWRS.

<table>
<tr>
<td>Method of study drug administration</td>
<td>
Tirzepatide

Tirzepatide doses: 5 mg, 10 mg, or 15 mg once weekly
The dosing algorithm for tirzepatide is summarised in Section B.2.3.1

Insulin degludec

A treat-to-target algorithm of dosing was used

Insulin degludec doses started at 10 units once daily
Patients adjusted their insulin degludec dose once weekly to a target fasting blood glucose (FBG) of <90 mg/dL (<5.0 mmol/L) based on the median value of the last 3 SMBG values, as summarized below

Titration of insulin degludec
<table>
<tr><th colspan="2">Median fasting blood glucose</th><th rowspan="2">Adjustment of insulin degludec dose</th></tr>
<tr><th>mg/dL</th><th>mmol/L</th></tr>
<tr><td>≤70</td><td>≤3.9</td><td>Decrease by 2 to 4 units</td></tr>
<tr><td>71 to 90</td><td>4.0 to 5.0</td><td>No adjustment</td></tr>
<tr><td>91 to 126</td><td>5.1 to 7.0</td><td>Increase by 2 units</td></tr>
<tr><td>127 to 144</td><td>7.1 to 8.0</td><td>Increase by 4 units</td></tr>
<tr><td>145 to 162</td><td>8.1 to 9.0</td><td>Increase by 6 units</td></tr>
<tr><td>≥162</td><td>≥9.0</td><td>Increase by 8 units</td></tr>
</table>
Source: SURPASS-3 CSR.[91]

Patients were advised to administer their daily doses at the time of day agreed upon between the patients and the investigator, ideally at bedtime
Doses were decreased by 2 to 4 units if:

Multiple episodes of non-severe hypoglycaemia were recorded during the assessment period at any time during the day
At least 1 episode that met the criteria for severe hypoglycaemia (events requiring assistance to administer therapy) or was associated with SMBG value <54 mg/dL (<3.0 mmol/L) was recorded during the assessment period

</td>
</tr>
<tr>
<td>Permitted and disallowed concomitant medication</td>
<td>
The following concomitant medications were permitted during the study:

Metformin and/or an SGLT2i to treat T2D
After randomisation, discontinuation of metformin or an SGLT2i or change in dosage was only allowed in specific circumstances:

The event of a hypoglycaemic episode(s)
Certain clinical situations that required short-term discontinuation in line with the product(s) labelling for each country; for example, severe dehydration, elective surgery, or need for radiologic exam involving IV iodinated contrast dye
Patient developed contraindications to metformin or an SGLT2i such that the use of the drug was contraindicated according to the country-specific label

Post-randomisation, patients were permitted to use concomitant medications that they required during the study, except certain medications that may interfere with the assessment of efficacy and safety characteristics of the study treatments

</td>
</tr>
</table>

	• Antihyperglycaemic medications other than study drug were allowed only in specific circumstances, including severe persistent hyperglycaemic (rescue therapy) or early discontinuation of study treatment The following medications were prohibited during the study: • GLP-1 RAs • DPP-4 inhibitors • Pramlintide • Use of other basal insulins was not allowed in the insulin degludec group
Primary outcomes	Mean change in HbA1c values from baseline to 52 weeks for tirzepatide 10 mg and 15 mg.
Secondary and exploratory outcomes	**Key secondary efficacy endpoints (controlled for type 1 error)** • Mean CfB in HbA1c for tirzepatide 5 mg • Mean CfB in body weight for all tirzepatide doses • Proportion of patients achieving a target of HbA1c <7% (53 mmol/mol) for all tirzepatide doses **Additional secondary efficacy endpoints (not controlled for type 1 error; for all tirzepatide doses)** • Proportion of patients achieving target HbA1c ≤6.5% (48 mmol/mol) and <5.7% (39 mmol/mol) • Mean CfB in FSG, measured in the central laboratory • Mean CfB in 7-point SMBG profiles • Proportion of patients who achieved weight loss ≥5%, ≥10% and ≥15% • Mean CfB in patient-reported outcomes, including DTSQs/DTSQc, IW-SP, and APPADL **Tertiary or exploratory efficacy endpoints (for all tirzepatide doses)** • Mean CfB in lipids (total cholesterol, HDL, LDL, VLDL, and triglycerides) • Mean CfB in BMI • Mean CfB in waist circumference • Mean CfB in biomarkers • Mean CfB in EQ-5D-5L scores **Safety assessments** • AEs • Patient diaries • Concomitant medications • Dilated fundoscopic examinations were performed at baseline for all patients; follow-up dilated fundoscopic examinations were performed as deemed appropriate by the investigator • Vital signs • ECGs • Laboratory tests, including hepatic safety monitoring
Duration of study and follow-up	The study was initiated on 1st April 2019 and completed on 4th January 2021. The approximately 4-week safety follow-up period occurred after the last treatment visit for patients who either:

	• Completed the entire treatment period • Discontinued early and performed an early termination (ET) visit During the safety follow-up, patients did not receive study treatment and were treated with another glucose-lowering intervention decided upon by the investigator.

Abbreviations: AE: adverse event; ALT: alanine transaminase; APPADL: ability to perform physical activities of daily living; BMI: body mass index; CDK-EPI: chronic Kidney Disease-Epidemiology; CfB: change from baseline; DPP-4: dipeptidyl-peptidase 4; DTSQ(c/s): diabetes treatment satisfaction questionnaire (change/status); ECG: electrocardiogram; eGFR: estimated glomerular filtration rate; EQ-5D-5L: EuroQol-5 dimension-5 level descriptive system; ET: early termination; FSG: fasting serum glucose; GI: gastrointestinal; GLP-1 RA: glucagon-like peptide-1 receptor agonist; HbA1c: glycated haemoglobin; HDL: high density lipoprotein; IW-SP: impact of weight on self-perception; IWQOL-Lite-CT: impact of weight on quality of life lite clinical trials version; IWRS: Interactive Web Response System; LDL: low density lipoprotein; MTC: medullary thyroid cancer; NAFLD: non-alcoholic fatty liver disease; OAD: oral antidiabetic medication; OUS: outside the USA; SGLT2i: sodium glucose cotransporter 2 inhibitor; SMBG: self-monitored blood glucose; T2D: type 2 diabetes; ULN: upper limit of normal; USA: United States of America; VLDL: very low density lipoprotein.
Source: SURPASS-3 CSR. [91]

Baseline characteristics

Baseline demographics, disease characteristics and a summary of prior therapies of the mITT population of patients with T2D included in the final analysis of SURPASS-3 are presented in Table 12.

Table 12: Baseline characteristics of patients included in the SURPASS-3 trial

Characteristics	TZP 5 mg (N=358)	TZP 10 mg (N=360)	TZP 15 mg (N=359)	Insulin degludec (N=360)	Overall population (N=1437)
Demographics					
Age (years), mean ± SD	57.2 ± 10.1	57.4 ± 9.7	57.5 ± 10.2	57.5 ± 10.1	57.4 ± 10.0
Female, n (%)	158 (44.1)	165 (45.8)	165 (46.0)	147 (40.8)	635 (44.2)
Race, n (%)					
White	323 (90.2)	328 (91.1)	327 (91.1)	329 (91.4)	1307 (91.0)
American Indian or Alaska native	0	1 (0.3)	1 (0.3)	2 (0.6)	4 (0.3)
Asian	20 (5.6)	19 (5.3)	20 (5.6)	17 (4.7)	76 (5.3)
Black or African American	13 (3.6)	12 (3.3)	8 (2.2)	11 (3.1)	44 (3.1)
Multiple	1 (0.3)	0	1 (0.3)	0	2 (0.1)
Native Hawaiian or other Pacific Islander	1 (0.3)	0	2 (0.6)	1 (0.3)	4 (0.3)
Weight (kg), mean ± SD	94.43 ± 18.86	93.80 ± 19.81	94.90 ± 20.98	93.98 ± 20.59	94.28 ± 20.06
BMI (kg/m^2), mean ± SD	33.58 ± 5.87	33.41 ± 6.21	33.68 ± 6.11	33.42 ± 6.06	33.52 ± 6.06
BMI category, n (%)					
<30 kg/m^2	104 (29.1)	116 (32.2)	109 (30.4)	117 (32.5)	446 (31.0)

Characteristics	TZP 5 mg (N=358)	TZP 10 mg (N=360)	TZP 15 mg (N=359)	Insulin degludec (N=360)	Overall population (N=1437)
30 to <35 kg/m^2	136 (38.0)	119 (33.1)	121 (33.7)	120 (33.3)	496 (34.5)
≥35 kg/m^2	118 (33.0)	125 (34.7)	129 (35.9)	123 (34.2)	495 (34.4)
Disease Characteristics					
Duration of diabetes (years), mean ± SD	8.47 ± 5.83	8.43 ± 6.59	8.52 ± 6.47	8.12 ± 6.04	8.38 ± 6.24
HbA1c (%), mean ± SD	8.17 ± 0.89	8.18 ± 0.89	8.21 ± 0.94	8.12 ± 0.94	8.17 ± 0.91
HbA1c (mmol/mol), mean ± SD	65.81 ± 9.69	65.91 ± 9.76	66.18 ± 10.24	65.20 ± 10.28	65.78 ± 9.99
HbA1c category, n (%)					
≤8.5% (69 mmol/mol)	248 (69.3)	249 (69.2)	252 (70.2)	256 (71.1)	1005 (69.9)
>8.5% (69 mmol/mol)	110 (30.7)	111 (30.8)	107 (29.8)	104 (28.9)	432 (30.1)
FSG (mg/dL), mean ± SD	171.73 ± 47.86	170.42 ± 47.64	168.42 ± 45.95	166.73 ± 41.90	169.33 ± 45.89
FSG (mmol/L), mean ± SD	9.53 ± 2.66	9.46 ± 2.64	9.35 ± 2.55	9.26 ± 2.33	9.40 ± 2.55
Systolic blood pressure (mm Hg), mean ± SD	130.73 ± 13.59	131.10 ± 13.12	131.85 ± 12.85	132.45 ± 13.63	131.53 ± 13.30
Diastolic blood pressure (mm Hg), mean ± SD	78.59 ± 8.52	79.22 ± 8.69	79.25 ± 9.16	79.57 ± 9.18	79.16 ± 8.89
History of CV disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
eGFR (CKD-EPI, mL/min per 1.73 m^2), mean ± SD	95.14 ± 17.22	93.65 ± 16.90	93.09 ± 17.25	94.63 ± 16.78	94.13 ± 17.04
eGFR category (CDK-EPI), n (%)					
<60 mL/min per 1.73 m^2	16 (4.5)	13 (3.6)	12 (3.3)	15 (4.2)	56 (3.9)
≥60 mL/min per 1.73 m^2	342 (95.5)	347 (96.4)	347 (96.7)	345 (95.8)	1381 (96.1)
Baseline antihyperglycaemic medications					
Metformin alone, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SGLT2i, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	458 (31.9)
Metformin dose (mg/day), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BMI: body mass index; bpm: beats per minute; CKD-EPI: chronic Kidney Disease-Epidemiology; CV: cardiovascular; eGFR: estimated glomerular filtration rate; FSG: fasting serum glucose; HbA1c: glycated haemoglobin; SD: standard deviation; SGLT2i: sodium glucose cotransporter 2 inhibitor; TZP: tirzepatide.

Source: Ludvik *et al,* 2021;[87] SURPASS-3 CSR.[91]

B.2.3.4 SURPASS-4

Trial design

SURPASS-4 is a phase 3, international, multicentre, randomised, open-label, parallel-group study that investigated the effects of treatment with tirzepatide 5, 10, and 15 mg, compared to insulin glargine in patients with T2D with increased CV risk who had inadequate glycaemic control on stable dose of at least one and no more than three oral antihyperglycaemic drugs, including metformin, SGLT-2i, and/or sulphonylurea. The study was open-label due to the differences between once-per-week tirzepatide and once-per-day insulin glargine in dosing schedule, dose escalation, and devices used.

The trial had four study periods:

- Period I: screening and lead-in period lasting two weeks
- Period II: treatment period lasting 52 weeks
- Period III: variable treatment period starting at 52 weeks and continuing up to 104 weeks
- Period IV: safety follow-up lasting four weeks

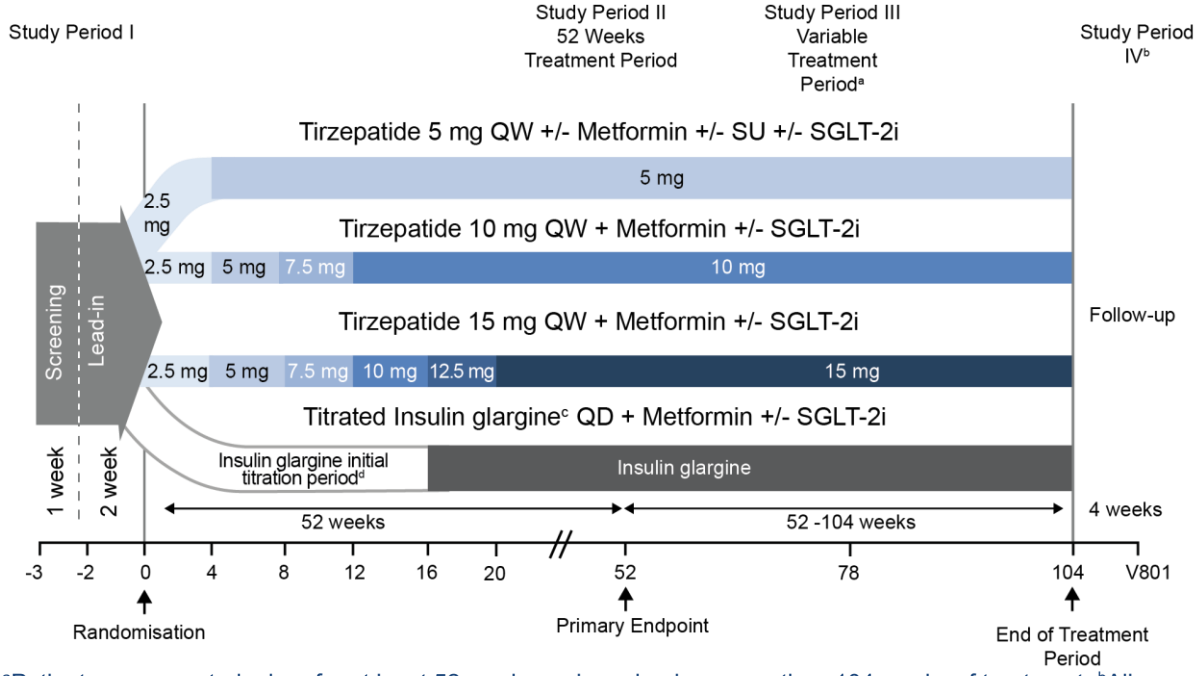
The primary endpoint was change in HbA1c from baseline to 52 weeks for tirzepatide 10 mg and 15 mg. Key secondary endpoints were change in HbA1c from baseline to 52 weeks for tirzepatide 5 mg, and change in body weight from baseline to 52 weeks and achievement of the HbA1c target of <7.0% (<53 mmol/mol) for all tirzepatide doses.

Patients were to remain in the study until 104 weeks or until all criteria for closing the study were met. Therefore, patients randomised earlier had a longer exposure than those randomised later. The study was planned to continue until all of the following criteria were fulfilled:

- At least 52 weeks from the time of the last patient randomised
- At least 300 patients assigned to the combined tirzepatide groups received at least 78 weeks of treatment
- Approximately 110 patients in this study experienced at least one component event of the composite CV endpoint of CV death, MI, stroke, or hospitalisation for unstable angina

A summary of the trial design of SURPASS-4 is presented in Figure 5.

Figure 5: Study design of SURPASS-4



[a]Patients were on study drug for at least 52 weeks and received no more than 104 weeks of treatment. [b]All patients completed a safety follow-up visit (Visit 801) four weeks after their last treatment visit. [c]The starting dose of insulin glargine was 10 IU/day at bedtime, titrated to an FBG <100 mg/dL, following a TTT algorithm. Patients titrated insulin glargine dose in a weekly manner and made the dose decision with the investigator for the first 8 weeks (phone or clinic visit). From Weeks 8 to 16, patients continued the titration by a phone consultation or clinic visit every other week, with 3 weeks between Visits 13 (Week 12) and 14 (Week 15). Note: // indicates the x-axis is not shown to scale from 20 to 52 weeks.
Abbreviations: FBG: fasting blood glucose; QD: once daily; QW: once weekly; SGLT-2i: sodium-glucose cotransporter-2 inhibitor; SU: sulfonylurea; TTT: treat-to-target.
Source: SURPASS-4 CSR. [92]

Trial methodology

A summary of the methodology of SURPASS-4 is presented in Table 13. A summary of the pre-planned subgroups is presented in Section B.2.7.

Table 13: Summary of the methodology of SURPASS-4

Trial name	SURPASS-4 (NCT03730662)
Location	187 sites in 14 countries (Argentina, Australia, Brazil, Canada, Greece, Israel, Mexico, Poland, Romania, Russia, Slovakia, Spain, Taiwan and the USA)
Trial design	International, multicentre, open-label, phase 3 study assessing the safety and efficacy of tirzepatide, compared with titrated insulin glargine, for the treatment of patients with T2D with high CV disease risk as an add-on to between 1 and 3 of metformin, SGLT2i, and SU. A long-term safety period allowed patients to continue treatment up to 104 weeks or until all criteria for closing the study were met; patients received treatment for 52 to 104 weeks, with patients randomised earlier having a longer exposure than patients randomised later.
Eligibility criteria for participants	**Eligibility criteria** • Adults (≥18 years) with T2D • Inadequately controlled HbA1c levels (7.5–10.5% [58–91 mmol/mol]) • BMI of 25 kg/m^2 or greater • Stable weight (≤5% fluctuation in either direction) during the

<table>
<tr><td></td><td>previous 3 months.
• Used ≥1 and ≤3 oral antihyperglycaemic drugs, which could only include metformin, an SGLT2i, and/or an SU, for ≥3 months prior to screening
• Increased risk of CV events (based on predefined list; see the Clinical Study Report for full details)
• Agreed not to initiate an organised diet or exercise program during the study with the intent of reducing body weight other than the lifestyle and dietary measures for diabetes treatment

Exclusion criteria
• Type 1 diabetes mellitus
• History of
○ Proliferative diabetic retinopathy,
○ Diabetic maculopathy,
○ Non-proliferative diabetic retinopathy that requires acute treatment,
○ Acute or chronic hepatitis,
○ Signs and symptoms of any other liver disease other than non-alcoholic fatty liver disease (if ALT level was ≤3.0 times the ULN of the reference range), or
○ ALT level >3.0 times the ULN of the reference range, as determined by the central laboratory at Visit 1
• Prior use of insulin therapy except for the use of insulin for treatment of gestational diabetes or acute, temporary use of insulin (≤14 days)
• Used any glucose-lowering agent(s) other than metformin, an SU and/or an SGLT2i during the 3 months preceding Visit 1
• Treated with prescription drugs that promote weight loss within 3 months prior to Visit 1</td></tr>
<tr><td>Intervention</td><td>After confirmation of the eligibility criteria, patients were randomised (1:1:1:3) to one of four cohorts: tirzepatide 5 mg, tirzepatide 10 mg, tirzepatide 15 mg, or titrated insulin glargine. Assignment to treatment group was determined by a computer-generated random sequence using an IWRS.</td></tr>
<tr><td>Method of study drug administration</td><td>Tirzepatide
• Tirzepatide doses: 5 mg, 10 mg, or 15 mg once weekly
• The dosing algorithm for tirzepatide is summarised in Section B.2.3.1

Insulin glargine
• A treat-to-target algorithm of dosing was used
○ Dosing started at 10 units daily
○ Patients adjusted their insulin glargine doses once weekly to a target fasting blood glucose (FBG) of <100 mg/dL (<5.6 mmol/L) based on the median value of the last 3 SMBG values, as summarized below

Titration of insulin glargine
<table>
<tr><th colspan="2">Median fasting blood glucose</th><th rowspan="2">Adjustment of insulin glargine dose</th></tr>
<tr><th>mg/dL</th><th>mmol/L</th></tr>
</table></td></tr>
</table>

	<table><tr><td>≤70</td><td>≤3.9</td><td>Decrease by 2 to 4 units</td></tr><tr><td>71 to 99</td><td>4.0 to 5.5</td><td>No adjustment</td></tr><tr><td>100 to 119</td><td>5.6 to 6.6</td><td>Increase by 2 units</td></tr><tr><td>120 to 139</td><td>6.7 to 7.7</td><td>Increase by 4 units</td></tr><tr><td>140 to 179</td><td>7.8 to 9.9</td><td>Increase by 6 units</td></tr><tr><td>≥180</td><td>≥10.0</td><td>Increase by 8 units</td></tr></table> **Source:** SURPASS-4 CSR.[92] • Patients were advised to administer their daily doses at the time of day agreed upon between the patient and the investigator, ideally at bedtime • Doses were decreased by 2 to 4 units if: ○ Multiple episodes of non-severe hypoglycaemia were recorded during the assessment period at any time during the day ○ At least 1 episode that met the criteria for severe hypoglycaemia (events requiring assistance to administer therapy) or was associated with SMBG value <54 mg/dL (<3.0 mmol/L) was recorded during the assessment period
Permitted and disallowed concomitant medication	The following concomitant medications were permitted during the study: • Metformin, an SU and/or an SGLT2i to treat T2D • Temporary discontinuation of concomitant antihyperglycemic medications was allowed for certain clinical situations; for example, severe dehydration, elective surgery, or need for radiologic exam involving IV iodinated contrast dye • Patients were permitted to use concomitant medications that they required during the study (e.g. blood pressure lowering, lipid lowering, anti-platelet), except certain medications that may interfere with the assessment of efficacy and safety characteristics of the study treatments • Antihyperglycaemic medications other than the study drugs were allowed only in specific circumstances, including severe persistent hyperglycaemia or early discontinuation of study treatment • Short-term treatment with insulin for ≤14 days was permitted in certain clinical situations (e.g. elective surgery, during hospitalisation, hyperosmolar states) The following medications were prohibited during the study: • GLP-1 RAs • DPP-4 inhibitors • Amylin analogues
Primary outcomes	Mean change in HbA1c values from baseline to 52 weeks for tirzepatide 10 mg and 15 mg.
Secondary and exploratory outcomes	**Key secondary efficacy endpoints (controlled for type 1 error)** • Mean CfB in HbA1c for tirzepatide 5 mg • Mean CfB in body weight for all tirzepatide doses • Proportion of patients achieving a target of HbA1c <7% (53 mmol/mol) for all tirzepatide doses

	Additional secondary efficacy endpoints (not controlled for type 1 error; for all tirzepatide doses) • Proportion of patients achieving target HbA1c ≤6.5% (48 mmol/mol) and <5.7% (39 mmol/mol) • Mean CfB in FSG, measured in the central laboratory • Mean CfB in 7-point SMBG profiles • Proportion of patients who achieved weight loss ≥5%, ≥10% and ≥15% • Mean CfB in patient-reported outcomes, including DTSQs/DTSQc, IW-SP, and APPADL **Tertiary or exploratory efficacy endpoints (for all tirzepatide doses)** • Mean CfB in lipids (total cholesterol, HDL, LDL, VLDL, and triglycerides) • Mean CfB in BMI • Mean CfB in waist circumference • Mean CfB in patient-reported outcomes, including APPADL, IW-SP, DTSQs/DTSQc and EQ-5D-5L scores **Safety assessments** • AEs • CV events (time to first occurrence of MACE-4) • Patient diaries • Concomitant medications • Dilated fundoscopic examinations were performed at baseline for all patients; follow-up dilated fundoscopic examinations were performed as deemed appropriate by the investigator • Vital signs • ECGs • Laboratory tests, including hepatic safety monitoring
Duration of study and follow-up	The study was initiated on 20th November 2018 and completed on 22nd April 2021. The approximately 4-week safety follow-up period occurred after the last treatment visit for patients who either: • Completed the entire treatment period • Discontinued early and performed an early termination (ET) visit During the safety follow-up, patients did not receive study treatment and were treated with another glucose-lowering intervention decided upon by the investigator.

Abbreviations: AE: adverse event; ALT: alanine transaminase; APPADL: ability to perform physical activities of daily living; BMI: body mass index; CDK-EPI: chronic Kidney Disease-Epidemiology; CfB: change from baseline; CV: cardiovascular; DPP-4: dipeptidyl-peptidase 4; DTSQ(c/s): diabetes treatment satisfaction questionnaire (change/status); ECG: electrocardiogram; eGFR: estimated glomerular filtration rate; EQ-5D-5L: EuroQol-5 dimension-5 level descriptive system; ET: early termination; FSG: fasting serum glucose; GI: gastrointestinal; GLP-1 RA: glucagon-like peptide-1 receptor agonist; HbA1c: glycated haemoglobin; HDL: high density lipoprotein; IW-SP: impact of weight on self-perception; IWQOL-Lite-CT: impact of weight on quality of life lite clinical trials version; IWRS: Interactive Web Response System; LDL: low density lipoprotein; MACE: major adverse cardiovascular events; MTC: medullary thyroid cancer; NAFLD: non-alcoholic fatty liver disease; OUS: outside the USA; SGLT2i: sodium glucose cotransporter 2 inhibitor; SMBG: self-monitored blood glucose; SU: sulfonylurea; T2D: type 2 diabetes; TIA: transient ischaemic attack; ULN: upper limit of normal; USA: United States of America; VLDL: very low density lipoprotein.
Source: Del Prato *et al,* 2021.[85]; SURPASS-4 CSR.[92]

Baseline characteristics

Baseline demographics, disease characteristics, history of cardiovascular disease (CVD), and a summary of prior therapies of the mITT population of patients with T2D included in the final analysis of SURPASS-4 are presented in Table 14.

Table 14: Baseline characteristics of patients included in the SURPASS-4 trial

Characteristics	TZP 5 mg (N=329)	TZP 10 mg (N=328)	TZP 15 mg (N=338)	Insulin glargine (N=1000)	Overall population (N=1995)
Demographics					
Age (years), mean ± SD	62.9 ± 8.6	63.7 ± 8.7	63.7 ± 8.6	63.8 ± 8.5	63.6 ± 8.6
Female, n (%)	131 (39.8)	119 (36.3)	135 (39.9)	364 (36.4)	749 (37.5)
Race, n (%)					
White	260 (79.3)	259 (79.0)	285 (84.6)	825 (82.7)	1629 (81.8)
American Indian or Alaska native	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Asian	15 (4.6)	16 (4.9)	8 (2.4)	31 (3.1)	70 (3.5)
Black or African American	13 (4.0)	17 (5.2)	11 (3.3)	32 (3.2)	73 (3.7)
Multiple	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Native Hawaiian or other Pacific Islander	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weight (kg), mean ± SD	90.3 ± 20.3	90.6 ± 18.2	90.0 ± 16.3	90.2 ± 19.0	90.3 ± 18.7
BMI (kg/m^2), mean ± SD	32.6 ± 6.1	32.8 ± 5.5	32.5 ± 5.0	32.5 ± 5.5	32.6 ± 5.5
BMI category, n (%)					
<30 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30 to <35 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥35 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease Characteristics					
Duration of diabetes (years), mean ± SD	11.14 ± 7.08	11.96 ± 7.45	11.48 ± 7.54	12.03 ± 7.66	11.78 ± 7.51
HbA1c (%), mean ± SD	8.52 ± 0.84	8.59 ± 0.91	8.52 ± 0.98	8.50 ± 0.85	8.52 ± 0.88
HbA1c (mmol/mol), mean ± SD	69.59 ± 9.21	70.43 ± 9.95	69.63 ± 10.68	69.41 ± 9.32	69.65 ± 9.65
HbA1c category, n (%)					
≤8.5% (69 mmol/mol)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>8.5% (69 mmol/mol)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
FSG (mg/dL), mean ± SD	172.27 ± 49.11	175.47 ± 51.93	174.14 ± 53.84	168.40 ± 49.72	171.17 ± 50.75

Characteristics	TZP 5 mg (N=329)	TZP 10 mg (N=328)	TZP 15 mg (N=338)	Insulin glargine (N=1000)	Overall population (N=1995)
FSG (mmol/L), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Systolic blood pressure (mm Hg), mean ± SD	133.28 ± 14.18	135.08 ± 16.11	134.34 ± 15.02	134.57 ± 15.67	134.40 ± 15.40
Diastolic blood pressure (mm Hg), mean ± SD	78.39 ± 8.75	78.60 ± 9.50	78.24 ± 9.16	78.41 ± 9.62	78.41 ± 9.38
Non-proliferative diabetes retinopathy	68 (20.7)	63 (19.2)	89 (26.3)	187 (18.7)	407 (20.4)
History of CV disease, n (%)	275 (83.6)	296 (89.7)	293 (86.7)	874 (87.0)	1738 (86.8)
Documented coronary artery disease	133 (40.4)	146 (44.2)	146 (43.2)	455 (45.3)	880 (44.0)
Myocardial infarction	109 (33.1)	87 (26.4)	106 (31.4)	344 (34.2)	646 (32.3)
Coronary revascularisation procedure	109 (33.1)	104 (31.5)	102 (30.2)	329 (32.7)	644 (32.2)
Hospitalisation for unstable angina	21 (6.4)	30 (9.1)	22 (6.5)	91 (9.1)	164 (8.2)
Hospitalisation for heart failure	22 (6.7)	31 (9.4)	19 (5.6)	68 (6.8)	140 (7.0)
Stroke	37 (11.2)	36 (10.9)	43 (12.7)	125 (12.4)	241 (12.0)
Transient ischaemic attack	16 (4.9)	12 (3.6)	17 (5.0)	53 (5.3)	98 (4.9)
Peripheral artery disease	89 (27.1)	109 (33.0)	106 (31.4)	302 (30.0)	606 (30.3)
eGFR (CKD-EPI, mL/min per 1.73 m^2), mean ± SD	80.28 ± 22.66	81.43 ± 20.44	81.55 ± 21.22	81.47 ± 20.78	81.28 ± 21.11
eGFR category (CDK-EPI), n (%)					
<60 mL/min per 1.73 m^2	62 (18.8)	56 (17.1)	58 (17.2)	166 (16.6)	342 (17.1)
≥60 mL/min per 1.73 m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Macroalbuminuria	25 (7.7)	33 (10.3)	24 (7.1)	79 (8.1)	161 (8.2)
Microalbuminuria	76 (23.5)	97 (30.4)	103 (30.6)	270 (27.6)	546 (27.9)
Baseline antihyperglycaemic medications, n (%)					
Metformin alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SU	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SGLT2i	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SU plus SGLT2i	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SU alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SGTL2i alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SU + SGLT2i	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BMI: body mass index; bpm: beats per minute; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration; CV: cardiovascular; eGFR: estimated glomerular filtration rate; FSG: fasting serum glucose; Hb: haemoglobin; HbA1c: glycated haemoglobin; SD: standard deviation; SGLT2i: sodium glucose cotransporter 2 inhibitor; SU: sulfonylurea; TZP: tirzepatide.
Source: Del Prato *et al*, 2021;[85] SURPASS-4 CSR.[92]

B.2.3.5 SURPASS-5

Trial design

SURPASS-5 is a phase 3, international, multicentre, randomised, double-blind, parallel-group, 40-week, placebo-controlled study that investigated the effects of treatment with tirzepatide 5, 10 and 15 mg compared with placebo in patients with T2D, as an add-on to basal insulin glargine (titrated to a fasting blood glucose of <100 mg/dL [<5.6 mmol/L]) with or without metformin.

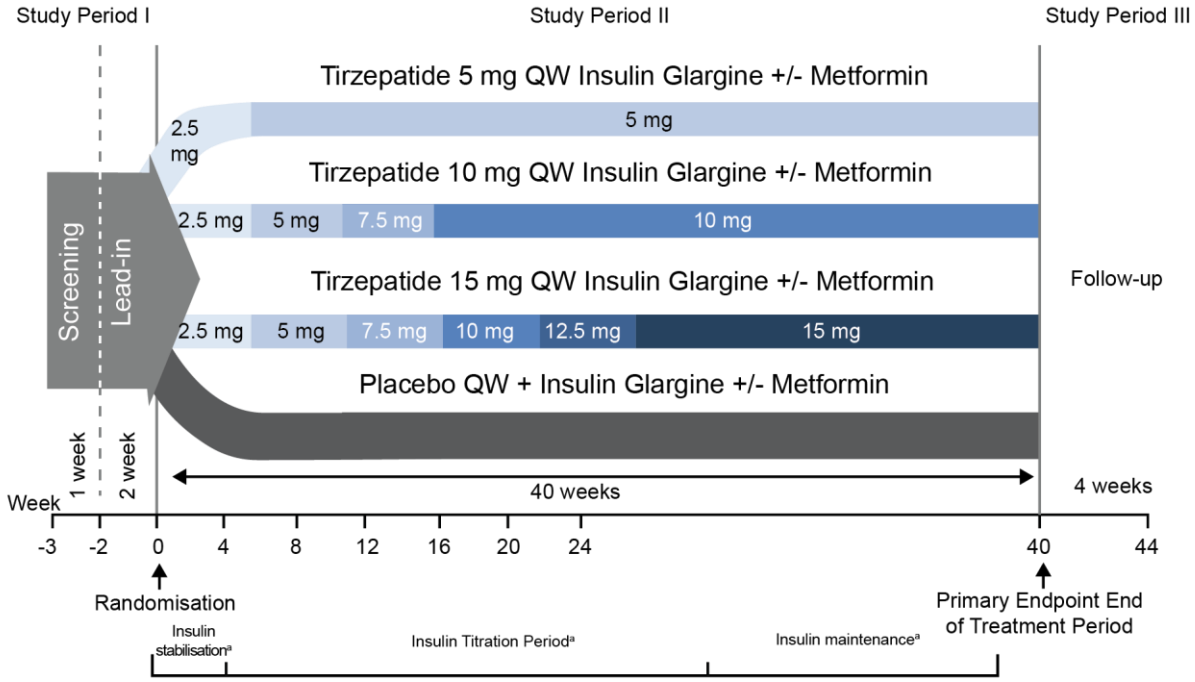
The trial has three study period:

- Period I: screening and lead-in period lasting three weeks
- Period II: treatment period lasting 40 weeks
- Period III: safety follow-up period lasting four weeks

The primary endpoint was change in HbA1c from baseline to 40 weeks for tirzepatide 10 mg and 15 mg. Key secondary endpoints were change in HbA1c from baseline to 40 weeks for tirzepatide 5 mg, change in body weight from baseline to 40 weeks and achievement of the HbA1c targets of <7.0% (<53 mmol/mol) for all tirzepatide doses, and achievement of HbA1c <5.7% (<39 mmol/mol) for tirzepatide 10 mg and 15 mg.

A summary of the trial design of SURPASS-5 is presented in Figure 6.

Figure 6: Study design of SURPASS-5



[a]Insulin Stabilization Period: first 4 weeks after randomisation, with restricted insulin dose adjustments. Insulin Titration Period: Weeks 4 to 40 (end of treatment/end of study), with unrestricted insulin dose adjustments. Maintenance Period: Weeks 24 to 40 (end of treatment/end of study), the period when insulin glargine dose was expected to be stable.
Abbreviations: QW: once weekly.

Source: SURPASS-5 CSR. [92]

Trial methodology

A summary of the methodology of SURPASS-5 is presented in Table 15. A summary of the pre-planned subgroups is presented in Section B.2.7.

Table 15. Summary of the methodology of SURPASS-5

Trial name	SURPASS-5 (NCT04039503)
Location	45 sites in 8 countries (Czech Republic, Germany, Japan, Poland, Slovakia, Spain, and the USA [including Puerto Rico])
Trial design	Phase 3, international, multicentre, randomised, double-blind, 40-week, study to assess the efficacy and safety of tirzepatide, compared with placebo, for the treatment of patients with T2D as an add-on to titrated basal insulin glargine or without metformin

Eligibility criteria for participants	**Eligibility criteria**
	• ≥18 years of age
	• On stable doses of once-daily insulin glargine (>0.25 U/kg/day or >20 U/day) with or without metformin (≥1500 mg/day) for 3 months prior to Visit 1
	• HbA1c of ≥7.0% (≥53 mmol/mol) to ≤10.5% (≤91 mmol/mol) at Visit 1
	• Stable weight for 3 months prior to Visit 1
	• BMI of at least 23 kg/m^2 at Visit 1
	• Required further insulin glargine dose increase at Visit 3 per the treat-to-target algorithm based on the SMBG data collected during the prior week
	
	Exclusion criteria
	• Type 1 diabetes mellitus
	• History of
	  ○ Proliferative diabetic retinopathy,
	  ○ Diabetic maculopathy, or
	  ○ Non-proliferative diabetic retinopathy that requires acute treatment
	• Had treatment with any glucose-lowering agent(s) other than stated in the inclusion criteria in a period of 3 months prior to Visit 1 and between Visit 1 and Visit 3 (randomisation)
	• Treated with prescription drugs that promote weight loss within 3 months prior to Visit 1 and/or between study entry (Visit 1) and randomisation (Visit 3)

Intervention	After confirmation of the eligibility criteria, patients were randomised 1:1:1:1 to once-weekly injectable tirzepatide 5 mg, 10 mg, 15 mg, or injectable placebo. Assignment to treatment group was determined by a computer-generated random sequence using an IWRS.
Method of study drug administration	**Tirzepatide** • Tirzepatide doses: 5 mg, 10 mg, or 15 mg once weekly • The dosing algorithm for tirzepatide is summarised in Section B.2.3.1 **Placebo** • Injectable placebo was administered once weekly

Permitted and disallowed concomitant medication

The following concomitant medications were permitted during the study:

- Insulin glargine was to be injected once daily, always at the same time of day, ideally at bedtime. Patients were instructed to assess and adjust insulin glargine doses to a target FBG of <100 mg/dL (5.6 mmol/L) once per week according to the treat-to-target algorithm summarised below

Titration of insulin glargine

Median fasting blood glucose		Adjustment of insulin glargine if dose is <20 units	Adjustment of insulin glargine if dose is ≥20 units
mg/dL	mmol/L		
≤70	≤3.9	Decrease by 1 to 2 units	Decrease by 2 to 4 units
71 to 100	4.0 to 5.5	No adjustment	No adjustment
101 to 119	5.6 to 6.6	Increase by 1 unit	Increase by 2 units
120 to 139	6.7 to 7.7	Increase by 2 units	Increase by 4 units
140 to 179	7.8 to 9.9	Increase by 3 units	Increase by 6 units
≥180	≥10.0	Increase by 4 units	Increase by 8 units

Source: SURPASS-5 CSR.[89]

- Patients on metformin were required to continue using the same dose of metformin throughout the treatment period; post-randomisation, patients on metformin could temporarily discontinue metformin in certain situations, in line with the product(s) labelling for each respective country (e.g. severe dehydration, elective surgery, or need for radiologic examination involving IV iodinated contrast dye)
- Post-randomisation, patients were permitted to use concomitant medications that they required during the study, such as antihypertensives or lipid-lowering medications, unless identified as having the potential to interfere with the assessment of efficacy and safety characteristics of the study treatments
- Investigators were allowed to prescribe medications such as antiemetic or antidiarrheal medications to mitigate GI symptoms and manage intolerable GI AEs after patients started the study drug, per local country availability and individual patient needs
- Antihyperglycaemic medications other than study drug were allowed only in specific circumstances
    - For patients who required permanent discontinuation of study drug, but remain in the study
    - As rescue therapy after randomisation due to severe, persistent hyperglycaemia
    - During the safety follow-up period
- Short-term treatment with a non-study insulin for <14 days was allowed in certain clinical situations (for example, elective surgery, during hospitalisation, hyperosmolar states)

The following medications were prohibited during the study:

	• GLP-1 RAs • DPP-4 inhibitors • Pramlintide • Other basal insulins
Primary outcomes	Mean change in HbA1c values from baseline to 40 weeks.
Secondary and exploratory outcomes	**Key secondary efficacy endpoints (controlled for type 1 error)** • Mean CfB in HbA1c for tirzepatide 5 mg • Mean CfB in body weight for all tirzepatide doses • Mean CfB in FSG for all tirzepatide doses • Proportion of patients achieving a target HbA1c <7% (53 mmol/mol) for all tirzepatide doses • Proportion of patients achieving HbA1c <5.7% (39 mmol/mol) for tirzepatide 10 mg and 15 mg **Additional secondary efficacy endpoints (not controlled for type 1 error; for all tirzepatide doses unless elsewhere specified)** • Proportion of patients achieving a target HbA1c of ≤6.5% (48 mmol/mol) • Proportion of patients achieving HbA1c <5.7% (39 mmol/mol) for tirzepatide 5 mg • Mean CfB in 7-point SMBG profiles • Proportion of patients who achieved weight loss ≥5%, ≥10% and ≥15% • Mean CfB in daily mean insulin glargine dose **Tertiary or exploratory efficacy endpoints (for all tirzepatide doses)** • Mean CfB in lipids (total cholesterol, HDL, LDL, VLDL, and triglycerides) • Mean CfB in waist circumference • Mean CfB in BMI • Mean CfB in patient-reported outcomes, including APPADL, DTSQs/DTSQc, and EQ-5D-5L scores **Safety assessments** • AEs • Patient diaries • Concomitant medications • Dilated fundoscopic examinations were performed at baseline for all patients; follow-up dilated fundoscopic examinations were performed as deemed appropriate by the investigator • Vital signs • ECGs • Laboratory tests, including hepatic safety monitoring
Duration of study and follow-up	The study was initiated on 30th August 2019 and completed on 13th January 2021. The approximately 4-week safety follow-up period occurred after the last treatment visit for patients who either: • Completed the entire treatment period • Discontinued early and performed an early termination (ET) visit

	During the safety follow-up, patients did not receive study treatment and were treated with another glucose-lowering intervention decided upon by the investigator.

Abbreviations: AE: adverse event; ALT: alanine transaminase; APPADL: ability to perform physical activities of daily living; BMI: body mass index; CDK-EPI: chronic Kidney Disease-Epidemiology; CfB: change from baseline; DPP-4: dipeptidyl-peptidase 4; DTSQ(c/s): diabetes treatment satisfaction questionnaire (change/status); ECG: electrocardiogram; eGFR: estimated glomerular filtration rate; EQ-5D-5L: EuroQol-5 dimension-5 level descriptive system; ET: early termination; FSG: fasting serum glucose; GI: gastrointestinal; GLP-1 RA: glucagon-like peptide-1 receptor agonist; HbA1c: glycated haemoglobin; HDL: high density lipoprotein; IW-SP: impact of weight on self-perception; IWQOL-Lite-CT: impact of weight on quality of life lite clinical trials version; IWRS: Interactive Web Response System; LDL: low density lipoprotein; MTC: medullary thyroid cancer; NAFLD: non-alcoholic fatty liver disease; OUS: outside the USA; SMBG: self-monitored blood glucose; T2D: type 2 diabetes; U: units; ULN: upper limit of normal; USA: United States of America; VLDL: very low density lipoprotein.
Source : Dahl *et al*, 2021 ;[84] SURPASS-5 CSR.[89]

Baseline characteristics

Baseline demographics, disease characteristics and a summary of prior therapies of the mITT population of patients with T2D included in the final analysis of SURPASS-5 are presented in Table 16.

Table 16: Baseline characteristics of patients included in the SURPASS-5 trial

Characteristics	TZP 5 mg (N=116)	TZP 10 mg (N=119)	TZP 15 mg (N=120)	Placebo (N=120)	Overall population (N=475)
Demographics					
Age (years), mean ± SD	61.5 ± 9.8	60.4 ± 10.2	60.5 ± 9.9	60.0 ± 9.6	60.6 ± 9.9
Female, n (%)	55 (47.4)	47 (39.5)	55 (45.8)	54 (45.0)	211 (44.4)
Race, n (%)					
White	95 (81.9)	94 (79.0)	94 (78.3)	97 (80.8)	380 (80.0)
American Indian or Alaska native	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Asian	20 (17.2)	21 (17.6)	22 (18.3)	22 (18.3)	85 (17.9)
Black or African American	1 (0.9)	2 (1.7)	3 (2.5)	0	6 (1.3)
Multiple	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weight (kg), mean ± SD	95.8 ± 19.8	94.5 ± 22.2	96.3 ± 22.8	94.1 ± 21.8	95.2 ± 21.6
BMI (kg/m^2), mean ± SD	33.6 ± 5.9	33.4 ± 6.2	33.4 ± 5.9	33.2 ± 6.3	33.4 ± 6.1
BMI category, n (%)					
<30 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
30 to <35 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥35 kg/m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease Characteristics					
Duration of diabetes	14.1 ± 8.1	12.6 ± 6.2	13.7 ± 7.5	12.9 ± 7.4	13.3 ± 7.3

Characteristics	TZP 5 mg (N=116)	TZP 10 mg (N=119)	TZP 15 mg (N=120)	Placebo (N=120)	Overall population (N=475)
(years), mean ± SD					
HbA1c (%), mean ± SD	8.30 ± 0.88	8.36 ± 0.83	8.23 ± 0.86	8.37 ± 0.84	8.31 ± 0.85
HbA1c (mmol/mol), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HbA1c category, n (%)					
≤8.0% (69 mmol/mol)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>8.0% (69 mmol/mol)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
FSG (mg/dL), mean ± SD	162.9 ± 53.9	162.3 ± 52.0	160.3 ± 54.2	164.1 ± 45.0	162.4 ± 51.3
FSG (mmol/L), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Systolic blood pressure (mm Hg), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Diastolic blood pressure (mm Hg), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
History of CV disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
eGFR (CKD-EPI, mL/min per 1.73 m^2), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
eGFR category (CKD-EPI), n (%)					
<60 mL/min per 1.73 m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥60 mL/min per 1.73 m^2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment use					
Dose of insulin glargine (IU/day)					
Mean ± SD	39.1 ± 25.4	34.7 ± 15.4	40.5 ± 29.1	36.3 ± 18.0	37.6 ± 22.7
Median (minimum, maximum)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dose of insulin glargine (IU/kg/day)					
Mean ± SD	0.4 ± 0.3	0.4 ± 0.2	0.4 ± 0.3	0.4 ± 0.2	0.4 ± 0.2

Characteristics	TZP 5 mg (N=116)	TZP 10 mg (N=119)	TZP 15 mg (N=120)	Placebo (N=120)	Overall population (N=475)
Median (minimum, maximum)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin use, n (%)	99 (85.3)	99 (83.2)	97 (80.8)	99 (82.5)	394 (82.9)
Metformin dose (mg/day), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BMI: body mass index; bpm: beats per minute; CKD-EPI: chronic Kidney Disease-Epidemiology; CV: cardiovascular; eGFR: estimated glomerular filtration rate; FSG: fasting serum glucose; HbA1c: glycated haemoglobin; SD: standard deviation; TZP: tirzepatide.
Source: Dahl *et al,* 2021;[84] SURPASS-5 CSR.[89]

B.2.4 *Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence*

The study populations and statistical analysis methods used in each of the SURPASS 2–5 trials are summarised below. Participant flow (CONSORT) diagrams for each of the SURPASS 2–5 trials are presented in Appendix D.9.

B.2.4.1 Study populations

SURPASS-2, -3 and -5

The same study population definitions were used in the SURPASS-2, -3 and -5 trials (Table 17), and the number of patients in the analysis sets of these trials is summarised in Table 18. The study populations definitions differ in SURPASS-4 and are therefore defined separately below.

Table 17: Trial populations used for the analysis of outcomes of SURPASS-2, -3 and -5

Analysis Set	Definition
Modified intent-to-treat (mITT) population	All randomly assigned patients who took at least 1 dose of study drug. In the event of a treatment error, patients were analysed according to the treatment they were randomised to
Full Analysis Set (FAS)	All available data obtained during Study Period II from the mITT population, excluding patients who discontinued study drug due to inadvertent enrolment, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication
Safety population (SS)	Same as mITT population
Efficacy analysis set	Data obtained during Study Period II from the mITT population, excluding patients who discontinued study drug due to inadvertent enrolment, and data after initiating rescue antihyperglycaemic medication or prematurely stopping study drug (last dose + 7 days)
Safety analysis set	All available data obtained during Study Periods II and III from the mITT population, regardless of adherence to study drug or initiation of new antihyperglycaemic medication

Source: SURPASS-2 CSR;[90]; SURPASS-3 CSR; [91] SURPASS-5 CSR.[89]

Table 18: Number of patients in the analysis sets of SURPASS-2, -3 and -5

Analysis set, n	SURPASS-2	SURPASS-3	SURPASS-5
Modified intent-to-treat (mITT) population	1,878	1,437	475
Full Analysis Set (FAS)	1,876	1,435	471
Safety population (SS)	1,878	1,437	475
Efficacy analysis set	1,876	1,435	471
Safety analysis set	1,878	1,437	475

Source: SURPASS-2 CSR;[90]; SURPASS-3 CSR; [91] SURPASS-5 CSR.[89]

SURPASS-4

The definitions of the study populations in the SURPASS-4 trial, presented in Table 19, are different to those used in SURPASS-2, -3 and -5, due to SURPASS-4 including a variable treatment period of 52 to 104 weeks.

Table 19: Trial populations used for the analysis of outcomes of SURPASS-4

Analysis Set	Definition
Modified intent-to-treat (mITT) population (n=1995)	All randomly assigned patients who took at least one dose of study drug. In the event of a treatment error, patients were analysed according to the treatment they were randomised to
Full Analysis Set (FAS; n=1989)	All available data obtained during Study Periods II and III from the mITT population, excluding patients who discontinued study drug due to inadvertent enrolment, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication
Safety population (SS; n=1995)	Same as mITT population
Efficacy analysis set (n=1989)	Data obtained during Study Period II from the mITT population, excluding patients who discontinued study drug due to inadvertent enrolment, and data after initiating rescue antihyperglycaemic medication or prematurely stopping study drug (last dose + 7 days)
Safety analysis set (n=1995)	All available data obtained during Study Periods II, III and IV from the mITT population, regardless of adherence to study drug or initiation of new antihyperglycaemic medication

Source: SURPASS-4 CSR.[92]

B.2.4.2 Statistical methods

This submission will present the efficacy estimand to align with the results used within the cost-effectiveness model (CEM) and NMA. For all SURPASS trials, the efficacy estimand was conducted using the efficacy analysis set and assessed on-treatment efficacy using data up to the time of initiating rescue therapy for severe persistent hyperglycaemia. The treatment-regimen estimand was conducted using the full analysis set and assessed efficacy using all data irrespective of adherence to study drug or introduction of rescue therapy for severe persistent hyperglycaemia. Both estimands were provided during regulatory submission, but the efficacy estimand data were considered the primary source within the submission to the European

Medicines Agency (EMA) and therefore the MHRA, while the treatment-regimen estimand data were preferred during the submission to the US Food and Drugs Administration (FDA).

SURPASS-2

The statistical methods for the primary analysis of SURPASS-2 are presented in Table 20.

Table 20: Statistical methods for the primary analysis of SURPASS-2

Hypothetical objective	Primary objectives of the study were to demonstrate that tirzepatide once weekly (QW) 10 mg and/or 15 mg was noninferior to semaglutide 1 mg in HbA1c change from baseline to 40 weeks. A key secondary objective was to demonstrate that tirzepatide QW 10 mg and/or 15 mg were superior to semaglutide 1 mg in HbA1c change from baseline at 40 weeks.
Statistical analysis	The analysis was conducted utilising HbA1c data in the EAS from baseline through the 40 week visit with the aid of a MMRM. REML was used to obtain model parameter estimates and the Kenward-Roger option was used to estimate denominator degrees of freedom. The response variable of the MMRM model was the primary endpoint and model terms included treatment, visit, treatment by visit interaction, pooled country as fixed effects, and baseline HbA1c as a covariate. An unstructured covariance structure was used to model the within-patient errors. The resulting least squares mean estimate of mean change from baseline in HbA1c was summarised by visit and by study treatment. The type 1 error rate was strongly controlled for evaluation of the primary and key secondary endpoints with a graphical approach. This procedure controlled the family-wise type 1 error rate at a 2-sided alpha level of 0.05. With the aid of the MMRM analysis, 2-sided 97.5% confidence intervals (CI) for mean change in HbA1c from baseline to the 40-week visit between 10 mg tirzepatide QW and 1 mg semaglutide QW, as well as between 15 mg tirzepatide QW and 1 mg semaglutide QW were derived and summarised. When the upper limit of the CI was ≤0.3%, then the respective dose of tirzepatide was declared noninferior to semaglutide relative to change in HbA1c from baseline and testing of superiority could happen. All results for superiority (key secondary endpoints) were presented with 2-sided 95% confidence intervals and p-values. All results for other secondary endpoints are reported as point estimates and p-values testing for difference but have not been adjusted for multiplicity.
Sample size, power calculation	Approximately 1872 patients (468 per group) were randomly assigned in a 1:1:1:1 ratio to tirzepatide 5 mg, 10 mg, 15 mg QW, or semaglutide 1 mg QW. Patient randomization was stratified based on country and baseline HbA1c (≤8.5% or >8.5% [69 mmol/mol]). The trial was powered to assess noninferiority of tirzepatide 10 mg and/or tirzepatide 15 mg QW to semaglutide 1 mg QW, relative to the primary endpoint: mean change in HbA1c from baseline to 40 weeks. The power was assessed based on the following assumptions: • each of the 10 mg and 15 mg tirzepatide QW treatment groups was tested in parallel against semaglutide 1 mg QW at 2-sided 0.025 significance level • use of 2-sample *t* test utilizing HbA1c data collected before initiation of any rescue medication or premature treatment discontinuation with no more than 28% of patients initiating any rescue medication or premature treatment discontinuation in each

	treatment group • no difference between tirzepatide doses and semaglutide 1 mg relative to the primary endpoint • a noninferiority margin of 0.3%, and • a common standard deviation (SD) of 1.1%. On the basis of these assumptions, randomly assigning approximately 1872 patients to the 4 treatments using a 1:1:1:1 ratio provided at least 90% power to demonstrate noninferiority of tirzepatide 10 mg and/or 15 mg QW doses to semaglutide 1 mg QW, relative to the primary endpoint for the "efficacy" estimand. Furthermore, this sample size ensured 90% power to demonstrate noninferiority for the "treatment-regimen" estimand conducted using an analysis of covariance (ANCOVA) utilizing all available HbA1c data at 40 weeks. Missing data were imputed with a conservative multiple imputation if SD were to increase up to 1.3% due to the inclusion of data on rescue medications, inclusion of data after premature treatment discontinuation, and imputation of the missing data.
Data management, patient withdrawals	End of study participation for a patient was the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow up visit (Visit 801). For patients considered to be lost to follow-up, end of study participation was the date of lost to follow-up reported by the investigator. Patient data included in the database after the last date of study participation (date of death, date of ET or date of safety follow-up) were excluded from statistical analysis. Listings of such data may be provided. For the 'efficacy estimand', no imputation of missing data was used.

Abbreviations: ET: early termination; HbA1c: glycated haemoglobin; MMRM: mixed model for repeated measures; REML: restricted maximum likelihood; QW: once-weekly.
Source: SURPASS-2 CSR.[90]

SURPASS-3

The statistical methods for the primary analysis of SURPASS-3 are presented in Table 21.

Table 21: Statistical methods for the primary analysis of SURPASS-3

Hypothetical objective	The primary objective of the study was to demonstrate that once-weekly (QW) tirzepatide 10 mg and/or 15 mg were noninferior to insulin degludec for change from baseline in haemoglobin (HbA1c) at 52 weeks. A key secondary objective was to demonstrate that tirzepatide QW 10 mg and/or 15 mg were superior to insulin degludec in HbA1c change from baseline at 52 weeks.
Statistical analysis	The analysis was conducted utilising HbA1c data in the EAS from baseline through the 52 week visit with the aid of a MMRM. REML was used to obtain model parameter estimates and the Kenward-Roger option was used to estimate the denominator degrees of freedom. The response variable of the MMRM model was the primary measure and model terms included treatment, visit, treatment-by-visit interaction, country/pooled country, and baseline concomitant oral antihyperglycemic medication use (metformin alone versus metformin plus an SGLT-2i) as fixed effects and baseline HbA1c as a covariate. An unstructured covariance structure was used to model the within-patient errors. The resulting LSM estimates of mean change from baseline in HbA1c were plotted by visit and by study treatment. With the aid of the MMRM analysis, 2-sided 97.5% CI for mean change in HbA1c from baseline to the 52-week visit for (10 mg tirzepatide – insulin degludec) as well as for (15 mg tirzepatide – insulin degludec) was derived. When the upper limit of the CI was ≤0.3%, then the respective dose of tirzepatide (10 mg and/or 15 mg) was declared noninferior to

	insulin degludec relative to change in HbA1c from baseline and testing of superiority could happen. All results for superiority (key secondary endpoints) were presented with 2-sided 95% confidence intervals and p-values. All results for other secondary endpoints are reported as point estimates and p-values testing for difference but have not been adjusted for multiplicity.
Sample size, power calculation	Approximately 1420 patients (355 per group) were randomly assigned in a 1:1:1:1 ratio to tirzepatide 5, 10, 15 mg QW, or insulin degludec. Patient randomisation was stratified based on country, baseline concomitant oral medication (metformin alone or metformin plus SGLT-2i), and baseline HbA1c (≤8.5% or >8.5% [69 mmol/mol]). Although the primary objective of the trial was to demonstrate that once-weekly 10 and/or 15 mg tirzepatide doses were noninferior to titrated insulin degludec relative to mean change in HbA1c from baseline (using a 0.3% noninferiority boundary), the study was powered to assess superiority of tirzepatide 10 and/or tirzepatide 15 mg QW to insulin degludec, relative to the primary endpoint: mean change in HbA1c from baseline to 52 weeks. The power was assessed based on the following assumptions: • each of the 10 and 15 mg tirzepatide QW treatment groups were tested in parallel against insulin degludec at a 2-sided 0.025 significance level • use of a 2-sample t test utilizing HbA1c data collected before initiation of any rescue medication or premature treatment discontinuation with no more than 28% of patients initiating any rescue medication or prematurely discontinuing treatment in each treatment group • a 0.35% greater mean reduction in HbA1c from baseline for 10 and 15 mg of tirzepatide compared with insulin degludec • a superiority margin of 0.05% • a common standard deviation of 1.1% On the basis of these assumptions, randomly assigning approximately 1420 patients to the 4 treatments using a 1:1:1:1 ratio was required to ensure at least 90% power to demonstrate superiority of tirzepatide 10 mg and/or 15 mg QW doses to insulin degludec, relative to the primary endpoint for the “efficacy” estimand.
Data management, patient withdrawals	End of study participation for a patient was the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (Visit 801). For patients considered to be lost-to-follow-up, end of study participation was the date of lost-to-follow-up reported by the investigator. Patient data included in the database after the last date of study participation (date of early termination or date of safety follow-up) were excluded from statistical analysis. Listings of such data may be provided.

Abbreviations: CI: confidence interval; ET: early termination; HbA1c: glycated haemoglobin; LSM: least squares mean; MMRM: mixed model for repeated measures; REML: restricted maximum likelihood; QW: once-weekly.
Source: SURPASS-3 CSR.[91]

SURPASS-4

The statistical methods for the primary analysis of SURPASS-4 are presented in Table 22.

Table 22: Statistical methods for the primary analysis of SURPASS-4

Hypothetical objective	Primary objectives of the study were to demonstrate that QW tirzepatide 10 mg and/or 15 mg was noninferior to insulin glargine for change from baseline in HbA1c at 52 weeks. A key secondary objective was to demonstrate that tirzepatide QW 10 mg and/or 15 mg were superior to insulin glargine in HbA1c change from baseline at 52 weeks.
Statistical analysis	The analysis was conducted utilizing HbA1c data in EAS from baseline through the 52-week visit with the aid of a mixed model for repeated measures (MMRM). Restricted maximum likelihood (REML) was used to obtain model parameter estimates and Kenward-Roger option was used to estimate denominator degrees of freedom. The response variable of the MMRM model was the primary measure and model terms included treatment, visit, treatment by visit interaction, country/pooled country, and SGLT2 inhibitor use at baseline as fixed effects, and baseline HbA1c as a covariate. An unstructured covariance structure was used to model the within-patient errors. Resulting Least Squares Mean (LSM) estimate of mean change from baseline in HbA1c was plotted by visit and by study treatment. With the aid of the MMRM analysis, 2-sided 97.5% Confidence Interval (CI) for mean change in HbA1c from baseline to 52-week visit for (10 mg tirzepatide – insulin glargine) as well as for (15 mg tirzepatide – insulin glargine) was derived. When the upper limit of the CI was ≤0.3%, then the respective dose of tirzepatide (10 mg and/or 15 mg) was declared noninferior to insulin glargine relative to change in HbA1c from baseline and testing of superiority could happen. All results for superiority (key secondary endpoints) were presented with 2-sided 95% confidence intervals and p-values. All results for other secondary endpoints are reported as point estimates and p-values testing for difference but have not been adjusted for multiplicity.
Sample size, power calculation	Patients were randomised in a 1:1:1:3 ratio to 5 mg tirzepatide, 10 mg tirzepatide, 15 mg tirzepatide, and insulin glargine to optimise CV risk assessment of tirzepatide. Although the primary objective was to establish noninferiority, sample size selection was guided by the objective of establishing superiority of each tirzepatide dose to insulin glargine relative to the reduction in mean HbA1c change from baseline at 52 weeks from randomisation irrespective of adherence to investigational product or introduction of rescue therapy for persistent severe hyperglycaemia (“treatment-regimen” estimand). The sample size determination assumed that evaluation of superiority of 10 mg tirzepatide and 15 mg tirzepatide to insulin glargine would be conducted in parallel, each at a 2-sided significance level of 0.025 using a 2-sample t-test. Additionally, a 0.30% superior mean HbA1c reduction from baseline at 52 weeks from randomization for 10 mg tirzepatide and 15 mg tirzepatide compared to insulin glargine and a common standard deviation (SD) of 1.3% were assumed for statistical power calculations. Under the assumptions above, randomizing 1878 patients in a 1:1:1:3 ratio to 5 mg tirzepatide (313 patients), 10 mg tirzepatide (313 patients), 15 mg tirzepatide (313 patients), and insulin glargine (939 patients) provided 90% power to demonstrate superiority of each tirzepatide dose to insulin glargine. The chosen sample size and randomization ratio also provided >90% power to establish superiority of 10 mg tirzepatide and 15 mg tirzepatide doses to insulin glargine in absence of confounding effects of rescue therapy for persistent severe hyperglycaemia (“efficacy” estimand). For comparison of each tirzepatide dose to insulin glargine using data

	collected prior to the initiation of any rescue medication or premature treatment discontinuation, conducted in parallel using a 2-sample t-test, each at a 2-sided significance level of 0.025, a 0.30% superior mean HbA1c reduction compared to insulin glargine, a common SD of 1.1%, and no more than 28% initiate of any rescue antihyperglycemic medication or prematurely discontinue study drug by 52 weeks were assumed. The trial was designed to contribute toward a meta-analysis across Phase 3 trials demonstrating that tirzepatide treatment is not associated with excessive CV risk. The anticipated treatment allocation of pooled tirzepatide versus pooled comparator was 3:1 in other tirzepatide Phase 3 clinical trials. The primary measure of CV risk is the hazard rate of CEC that confirmed 4-component MACE-4: CV death, myocardial infarction (MI), stroke, and hospitalization for unstable angina. Under the assumption of no increase or decrease in CV risk with tirzepatide compared to pooled comparator, approximately 133 patients with MACE-4 events were required to have 90% power to ensure that the upper 95% confidence limit for the hazard ratio is less than 1.8 in the meta-analysis. Assuming 33 to 40 patients with MACE-4 events per 1000 patient years of exposure and 2% reduced exposure due to lost to follow-up, patients followed for an average of 18 months were expected to result in 90 to 110 patients with MACE-4 events.
Data management, patient withdrawals	End of study participation for a patient was the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (visit 801). For patients considered to be lost-to-follow-up, end of study participation was the date of lost-to-follow-up reported by the investigator. Patient data included in database after the last date of study participation (date of death, date of early termination, or date of safety follow-up) was excluded from statistical analysis. Listing of such data may be provided. For the 'efficacy estimand', no imputation of missing data was used.

Abbreviations: CI: confidence interval; ET: early termination; HbA1c: glycated haemoglobin; LSM: least squares mean; MMRM: mixed model for repeated measures; REML: restricted maximum likelihood; QW: once-weekly.
Source: SURPASS-4 CSR.[92]

SURPASS-5

The statistical methods for the primary analysis of SURPASS-5 are presented in Table 23.

Table 23: Statistical methods for the primary analysis of SURPASS-5

Hypothetical objective	The primary objectives of the study were to demonstrate superiority of once-weekly (QW) tirzepatide 10 mg and/or 15 mg versus placebo when added to titrated basal insulin glargine, with or without metformin, with respect to mean change in HbA1c from baseline to 40 weeks.
Statistical analysis	The primary analysis relative to the "efficacy" estimand was conducted using HbA1c data in the EAS from baseline through the 40-week visit with the aid of a mixed-model repeated measure (MMRM). Restricted maximum likelihood (REML) was used to obtain model parameter estimates and the Kenward-Roger option was used to estimate denominator degrees of freedom. The response variable of the MMRM model was the primary measure and model terms of interest included treatment, visit, treatment-by-visit interaction, country/pooled country, and baseline metformin use (Yes/No) as fixed effects, and baseline HbA1c as a covariate. An unstructured covariance matrix was used to model the within-patient errors. The resulting least squares mean (LSM) estimate of mean change from baseline in HbA1c was summarized by visit and by study treatment.

	With the aid of the MMRM analysis, p-values, and 2-sided 95% confidence intervals (CIs) for mean change in HbA1c from baseline to the 40-week visit were derived and summarized for the 5 mg, 10 mg, and 15 mg doses of tirzepatide compared to placebo.
Sample size, power calculation	The trial was powered to assess superiority of tirzepatide 10 mg and 15 mg relative to the primary endpoint (mean change in HbA1c from baseline to 40 weeks). The power was assessed based on the following assumptions: • each of the 10 and 15 mg tirzepatide treatment groups were tested in parallel against placebo at a 2-sided 0.025 significance level • use of a 2 sample t test utilising HbA1c data collected before initiation of any rescue medication or premature treatment discontinuation with no more than 28% of subjects initiating rescue medication or prematurely discontinuing treatment in each treatment group • 0.6% greater mean reduction in HbA1c from baseline for 10 and 15 mg tirzepatide compared with placebo • 1:1:1:1 randomization, and • a common standard deviation (SD) of 1.1% On the basis of these assumptions, a sample size of 472 subjects was required to ensure at least 90% power to demonstrate that tirzepatide 10 mg and/or 15 mg were superior to placebo relative to the primary endpoint. Furthermore, this sample size ensured 90% power for the superiority evaluation conducted using an analysis of covariance (ANCOVA) utilising all available HbA1c data at 40 weeks with missing data imputed with a conservative multiple imputation method (as described below), provided a 0.6% greater mean reduction in HbA1c from baseline for 10 and 15 mg tirzepatide compared with placebo and SD increased to no more than 1.3% due to the inclusion of data on rescue medications and after premature treatment discontinuation and imputation of missing data.
Data management, patient withdrawals	The end of study participation for a patient was the earliest of date of death, date of withdrawal from further participation in the study, or date of the safety follow-up visit (Visit 801). For patients considered to be lost to follow-up, end of study participation was the date of lost to follow-up reported by the investigator. Patient data included in the database after the last date of study participation (date of death, date of early termination or date of safety follow-up) were excluded from statistical analyses. A listing of such data may be provided. For the 'efficacy estimand', no imputation of missing data was used.

Abbreviations: CI: confidence interval; ET: early termination; HbA1c: glycated haemoglobin; LSM: least squares mean; MMRM: mixed model for repeated measures; REML: restricted maximum likelihood; QW: once-weekly.
Source: SURPASS-5 CSR.[89]

Endpoints controlled for type 1 error

The endpoints controlled for type 1 error (i.e. a false positive result) in SURPASS 2−5 are summarised in Table 24. All endpoints listed are secondary endpoints in all trials, whether controlled for type 1 error or not, unless elsewhere specified. Exploratory endpoints were not controlled for type 1 error.

Table 24: Endpoints controlled for type 1 error in SURPASS 2−5

	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
TZP 5 mg				

HbA1c CfB	✓	✓	✓	✓
HbA1c <7.0%	✓	✓	✓	✓
HbA1c <5.7%				
Body weight CfB	✓	✓	✓	✓
FSG CfB*				✓
TZP 10 mg and 15 mg				
HbA1c CfB*	✓	✓	✓	✓
HbA1c <7.0%	✓	✓	✓	✓
HbA1c <5.7%	✓			✓
Body weight CfB	✓	✓	✓	✓
FSG CfB**				✓

Endpoints not listed are not controlled for type 1 error in any of SURPASS 2−5. *Primary endpoint; **Exploratory endpoint in all other SURPASS trials.
Abbreviations: CfB: change from baseline; FSG: fasting serum glucose; HbA1c: glycated haemoglobin; TZP: tirzepatide.

B.2.5 *Critical appraisal of the relevant clinical effectiveness evidence*

Quality assessments of the SURPASS trials were conducted using the Cochrane risk of bias assessment tool and the CRD tool. The trials identified in the SLR were assessed using the same tools.

A summary of the quality assessments is presented in Table 25; the full version of this quality assessment and the quality assessments for the remaining SURPASS trials and trials identified in the SLR are presented in Appendix D.10.

Table 25: Assessment of quality and risk of bias in the SURPASS trials

Criteria	Risk of bias			
	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Was randomisation carried out appropriately?	Yes	Yes	Yes	Yes
	Patients were randomly assigned 1:1:1:1 to the treatment groups. Assignment to treatment group was determined by a computer-generated random sequence using an IWRS.	Assignment to treatment group was determined by a computer-generated random sequence using the Eli Lilly and Company interactive web-response system. This system is externally validated and compliant with the Code of Federal Regulations 21 part 11.	Participants were randomly assigned (1:1:1:3), by the Eli Lilly and Company computer-generated random sequence using an interactive web-response system to receive tirzepatide or glargine	Assignment to treatment group was determined by a computer-generated random sequence using an IWRS.
Was the concealment of treatment allocated adequate?	Yes	Yes	Yes	Yes
	Treatment group assignment was determined by computer-generated random sequence using an IWRS.	Assignment to treatment group was determined by a computer-generated random sequence using the Eli Lilly and Company interactive web-response system	Computer-generated random sequence using an interactive web-response system to receive tirzepatide or glargine	Computer-generated random sequence was used
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes	Yes
	As stated in the paper- The demographic and clinical characteristics were similar across the groups	Paper states that baseline demographics and clinical characteristics were similar across the tirzepatide and insulin degludec groups.	As seen in the baseline characteristics table	Baseline demographics and clinical characteristics were similar across treatment groups
Were the care providers, participants and outcomes assessors blind to	No	No	No	Yes
	Open-label	Open-label	Open-label	Double-blind

treatment allocation?				
Were there any unexpected imbalanced in drop-outs between groups?	No	No	No	No
	All dropouts accounted for	All dropouts accounted for	All dropouts accounted for	No unexpected imbalances in drop-outs
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No	No
	All outcomes in method section were reported	All specified outcomes reported	All outcomes in method section were reported	All specified outcomes reported
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	No	No	No
	Multiple imputation	mITT was used	mITT was used	mITT was used

Adapted from Systematic reviews: CRD's guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination).[93]
Abbreviations: CRD: centre for reviews and dissemination; IWRS: interactive web-response system; mITT: modified intention-to-treat.

B.2.6 [redacted] *Clinical effectiveness results of the relevant studies*

Brief summary of clinical effectiveness results

- Across all studies, treatment with tirzepatide at all doses demonstrated statistically significant reductions in HbA1c from baseline to the primary endpoint (Week 40 or 52) compared to either placebo or active control treatment (semaglutide, insulin degludec and insulin glargine). In a sub-population in SURPASS-4 these effects were sustained for up to 2 years

HbA1c change from baseline, % (mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	−2.1% (−22.8)	2.4% (−25.9)	−2.5% (−26.9)	−1.9% (−20.3)
SURPASS-3 (vs insulin degludec)	−1.9% (−21.1)	−2.5% (−24.0)	−2.4% (−26.0)	−1.3% (−14.6)
SURPASS-4 (vs insulin glargine)	−2.2% (−24.5)	−2.4% (26.6)	−2.6% (−28.2)	−1.4% (−15.7)
SURPASS-5 (vs placebo)	−2.2% [redacted]	−2.6% [redacted]	−2.6% [redacted]	−0.9% [redacted]

Source: Frías *et al*, 2021;[86] Ludvik *et al,* 2021;[87] Del Prato *et al,* 2021;[85] SURPASS-5 CSR.[89]

- Significantly higher proportions of patients achieved an HbA1c target of <7.0% (<53 mmol/mol) on all three doses of tirzepatide at the primary endpoint compared with either placebo or the active comparator in all four of the SURPASS trials presented. Similarly, significantly higher proportions of patients on all tirzepatide doses achieved the more stringent HbA1c improvements of ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol)

Patients achieving HbA1c <7.0% (<53 mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	85.5%	88.9%	92.2%	81.1%
SURPASS-3 (vs insulin degludec)	82.4%	89.7%	92.6%	61.3%
SURPASS-4 (vs insulin glargine)	81.0%	88.2%	90.7%	50.7%
SURPASS-5 (vs placebo)	93.0%	97.4%	94.0%	33.9%

Source: Frias *et al*, 2021;[86] Ludvik *et al,* 2021;[87] Del Prato *et al,* 2021;[85] Dahl *et al*, 2021.[84]

- Across all studies, tirzepatide at all doses, was associated with significant reduction in body weight from baseline, and had significantly greater proportions of patients achieve body weight reductions of ≥5%, ≥10%, and ≥15%, compared to placebo and all active comparators studied

Body weight change from baseline, % (kg)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	[redacted] (−7.8)	[redacted] (−10.3)	[redacted] (−12.4)	[redacted] (−6.2)
SURPASS-3 (vs insulin degludec)	[redacted] (−7.5)	[redacted] (−10.7)	[redacted] (−12.9)	[redacted] (2.3)
SURPASS-4 (vs insulin glargine)	[redacted] (−7.1)	[redacted] (−9.5)	[redacted] (−11.7)	[redacted] (1.9)
SURPASS-5 (vs placebo)	[redacted] (−6.2)	[redacted] (−8.2)	[redacted] (−10.9)	[redacted] (1.7)

Source: SURPASS-2 CSR;[90] Frias *et al*, 2021;[86] SURPASS-3 CSR;[91] Ludvik *et al,* 2021;[87] SURPASS-4 CSR;[92] Del Prato *et al,* 2021;[85] SURPASS-5 CSR;[89] Dahl *et al*, 2021.[84]

- Across all studies, decreases were seen in tirzepatide-treated patients from baseline to the primary endpoint in the following lipid parameters: triglycerides, total cholesterol, LDL-C and VLDL-C. Increases from baseline were also observed for HDL-C in SURPASS-2–5
- Across the studies, from baseline to the primary endpoint (Week 40 or 52), treatment with tirzepatide demonstrated significant improvements in the ability to engage in activities of normal daily living and a significant reduction in the impact of weight on function and daily activities

Please note, data on systolic blood pressure changes seen with tirzepatide treatment were collected and analysed as a safety endpoint; these results are therefore presented in Section B.2.9.

B.2.6.1 SURPASS-2

Population: Patients with T2D, who have inadequate glycaemic control with metformin monotherapy and had not been treated with any other OADs during the 3 months prior to the start of the study
Comparator: Injectable semaglutide 1 mg administered once weekly

B.2.6.1.1 Glycaemic control

Primary endpoint: change in HbA1c from baseline – tirzepatide superior to semaglutide

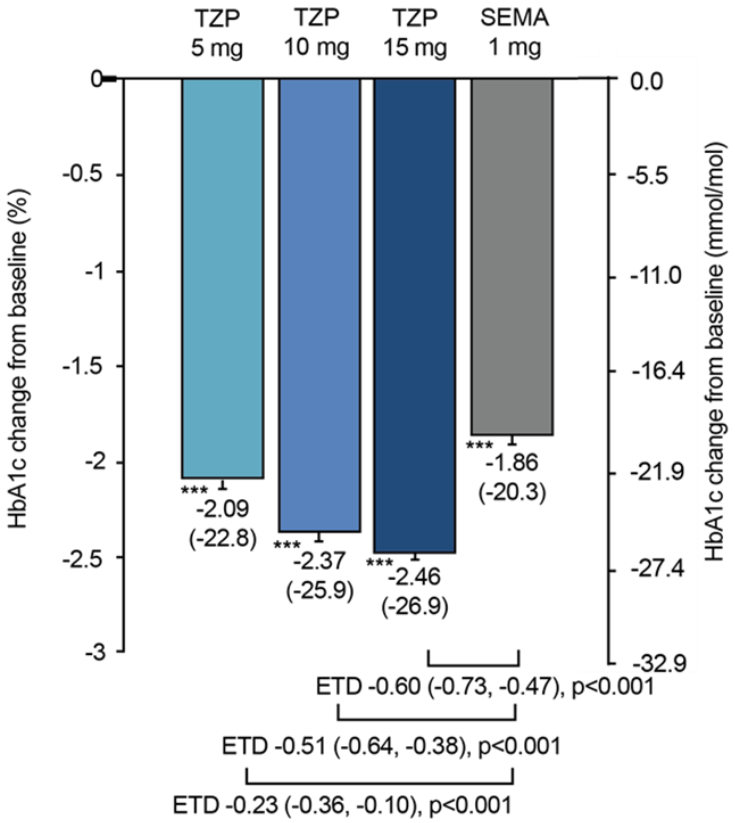
Patients on all three doses of tirzepatide (5 mg, 10 mg and 15 mg) achieved significantly greater reduction in HbA1c from baseline to 40 weeks, compared to patients on semaglutide 1 mg (Table 26; Figure 7).

Table 26: Change in HbA1c from baseline to 40 weeks; SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	470	469	469	468
HbA1c, %				
Baseline	8.33	8.31	8.25	8.24
Change from baseline to 40 weeks	−2.09*	−2.37*	−2.46*	−1.86*
Change difference from SEMA (95% CI) to 40 weeks	−0.23** (−0.36, −0.10)	−0.51** (−0.64, −0.38)	−0.60** (−0.73, −0.47)	n/a
HbA1c, mmol/mol				
Baseline	67.5	67.3	66.7	66.6
Change from baseline to 40 weeks	−22.8*	−25.9*	−26.9*	−20.3*
Change difference from SEMA (95% CI) to 40 weeks	[redacted]**	[redacted]**	[redacted]**	[redacted]

*p<0.001 versus baseline; **p<0.001 versus semaglutide 1 mg for superiority.
Abbreviations: CI: confidence interval; HbA1c: glycated haemoglobin; SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS-2 CSR;[90] Frias *et al*, 2021.[86]

Figure 7: Change in HbA1c from baseline to 40 weeks; SURPASS-2



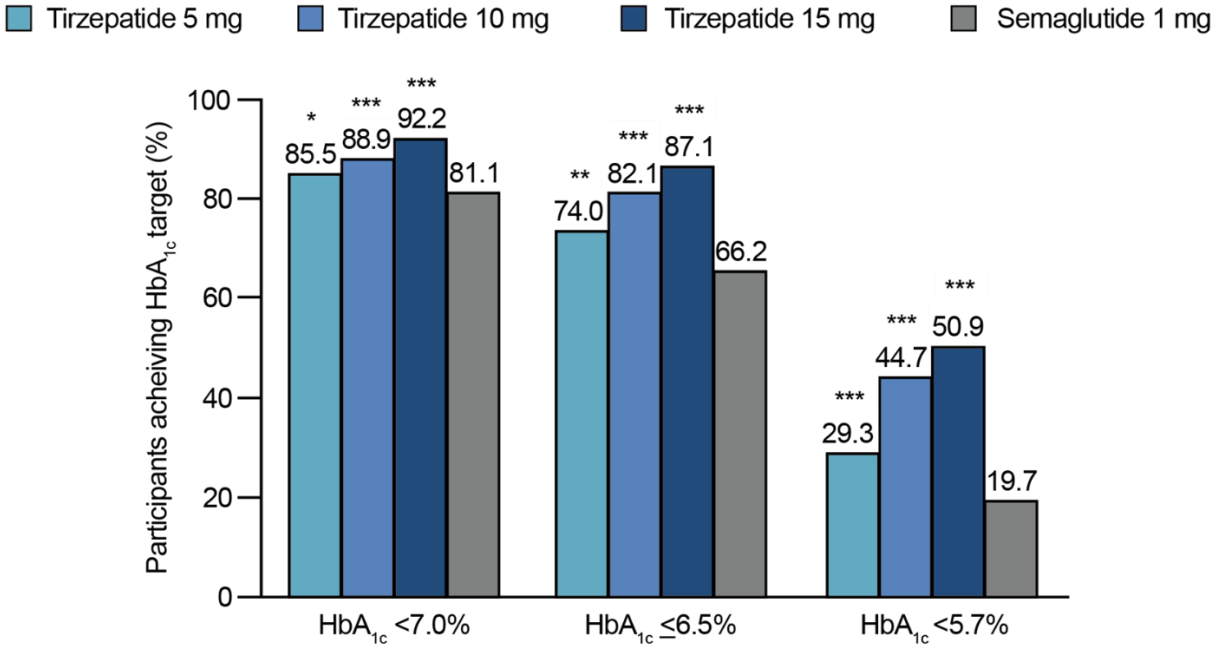
***p<0.001 versus baseline.
Abbreviations: ETD: estimated treatment difference; HbA1c: glycated haemoglobin; SEMA: semaglutide; TZP: tirzepatide.
Source: Frias *et al*, 2021.[86]

Proportion of patients achieving HbA1c targets at 40 weeks – tirzepatide superior to semaglutide

Significantly higher proportions of patients achieved the HbA1c target of <7.0% (<53 mmol/mol) at 40 weeks on all three tirzepatide doses compared to semaglutide 1 mg (Figure 8). Between 85.5% and 92.2% of patients on tirzepatide achieved HbA1c <7.0% at 40 weeks, compared to 81.1% of patients on semaglutide 1 mg.

Further, significantly higher proportions of patients on all tirzepatide doses achieved the HbA1c target of <5.7% (<39 mmol/mol) at 40 weeks compared to semaglutide 1 mg (Figure 8); between 29.3% and 50.9% of patients treated with tirzepatide achieved HbA1c <5.7% at 40 weeks, compared to 19.7% on semaglutide 1 mg. The comparisons of the proportion of patients achieving HbA1c ≤6.5% and <5.7% were not controlled for Type 1 error for all tirzepatide doses and tirzepatide 5 mg, respectively (Table 24).

Figure 8: Proportion of patients who achieved HbA1c targets at 40 weeks; SURPASS-2



HbA1c ≤6.5% comparisons not controlled for type 1 error for all doses of tirzepatide; HbA1c <5.7% comparisons not controlled for type 1 error for tirzepatide 5 mg. Significance shown only for endpoints controlled for type 1 error.
*p<0.05; **p<0.01 versus semaglutide 1 mg;***p<0.001 versus semaglutide 1 mg.
Abbreviations: HbA1c: glycated haemoglobin.
Source: Frias *et al*, 2021.[86]

B.2.6.1.2 Body weight

Change in body weight from baseline – tirzepatide superior to semaglutide

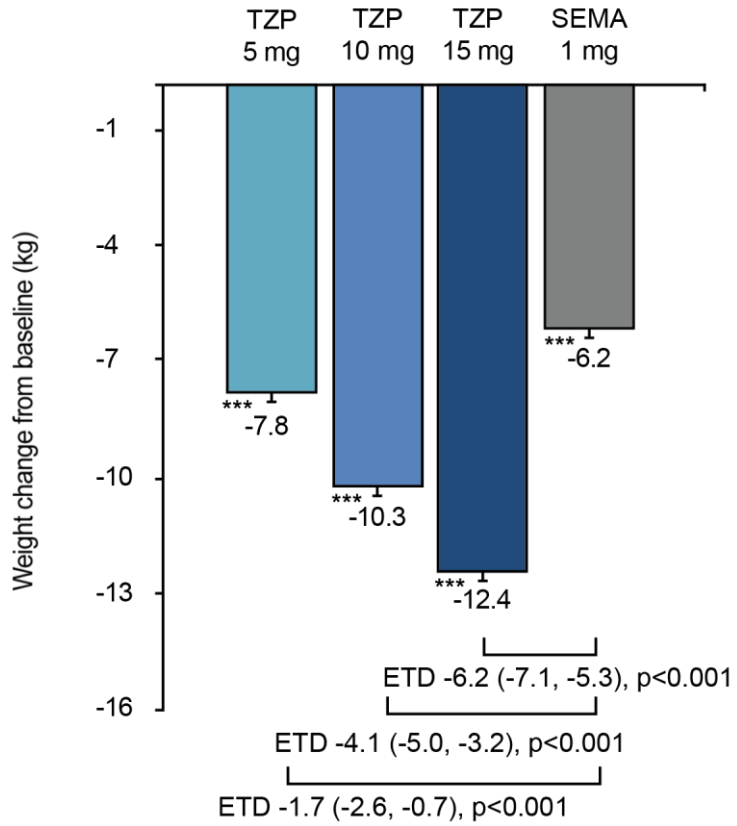
Patients on all three tirzepatide doses achieved significantly greater reductions in body weight at 40 weeks compared to patients on semaglutide 1 mg. Patients on tirzepatide 15 mg achieved a change from baseline in body weight twice that achieved by patients on semaglutide 1 mg at 40 weeks (Table 27; Figure 9).

Table 27: Change in body weight from baseline over time; SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	470	469	469	468
Weight, kg				
Baseline	92.6	94.6	93.9	93.8
Change from baseline to 40 weeks	−7.8*	−10.3*	−12.4*	−6.2*
Change difference from SEMA (95% CI) at 40 weeks	−1.7** (−2.6, −0.7)	−4.1** (−5.0, −3.2)	−6.2** (−7.1, −5.3)	n/a

*p<0.001 versus baseline; **p<0.001 versus semaglutide 1 mg.
Abbreviations: CI: confidence interval; HbA1c: glycated haemoglobin; SEMA: semaglutide; TZP: tirzepatide.
Source: Frias *et al*, 2021.[86]

Figure 9: Change in body weight from baseline over time; SURPASS-2

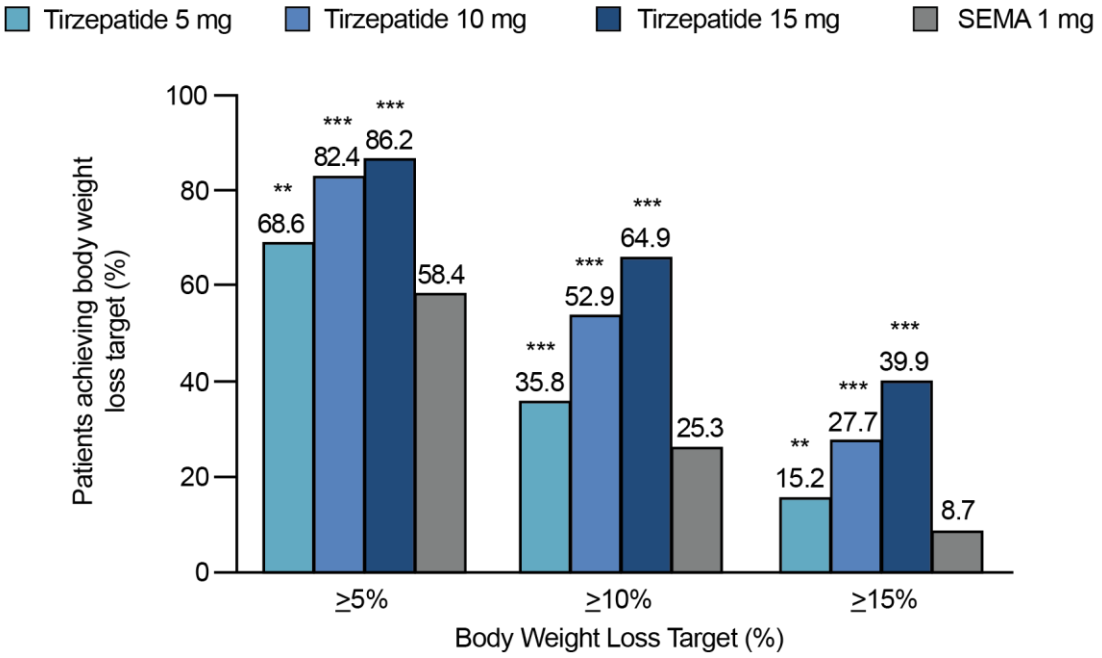


***p<0.001 versus baseline.
Abbreviations: ETD: estimated treatment difference; SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS-2 CSR.[90]

Proportion of patients achieving weight loss targets at 40 weeks – tirzepatide superior to semaglutide

Significantly higher proportions of patients on all three tirzepatide doses achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% from baseline to 40 weeks compared to semaglutide 1 mg (not controlled for type 1 error). The proportion of patients on tirzepatide 10 mg and 15 mg achieving ≥10% and ≥15% body weight reductions at Week 40 was over twice that for semaglutide 1 mg (Figure 10).

Figure 10: Proportion of patients achieving body weight loss targets at 40 weeks; SURPASS-2



Not controlled for type 1 error. ** p<0.01 versus semaglutide 1 mg. ***p<0.001 versus semaglutide 1 mg.
Abbreviations: SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS-2 CSR.[90]

Change in BMI from baseline – tirzepatide treatment led to greater BMI reduction vs semaglutide

In a pre-specified exploratory endpoint, patients on all three doses of tirzepatide achieved significantly reduced BMI from baseline to 40 weeks compared with patients on semaglutide 1 mg, with the largest reduction observed for the tirzepatide 15 mg group (Table 28).

Table 28: Change in BMI from baseline to 40 weeks: SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	470	469	469	468
BMI				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from SEMA (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Not controlled for type 1 error. *p<0.001 versus baseline; ** p<0.001 versus semaglutide 1 mg
Abbreviations: BMI: body mass index; CI: confidence interval; mITT: modified intent-to-treat; N: number of patients who were randomized and received at least 1 dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; n/a: not applicable; SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS-2 CSR.[90]

B.2.6.1.3 Lipids

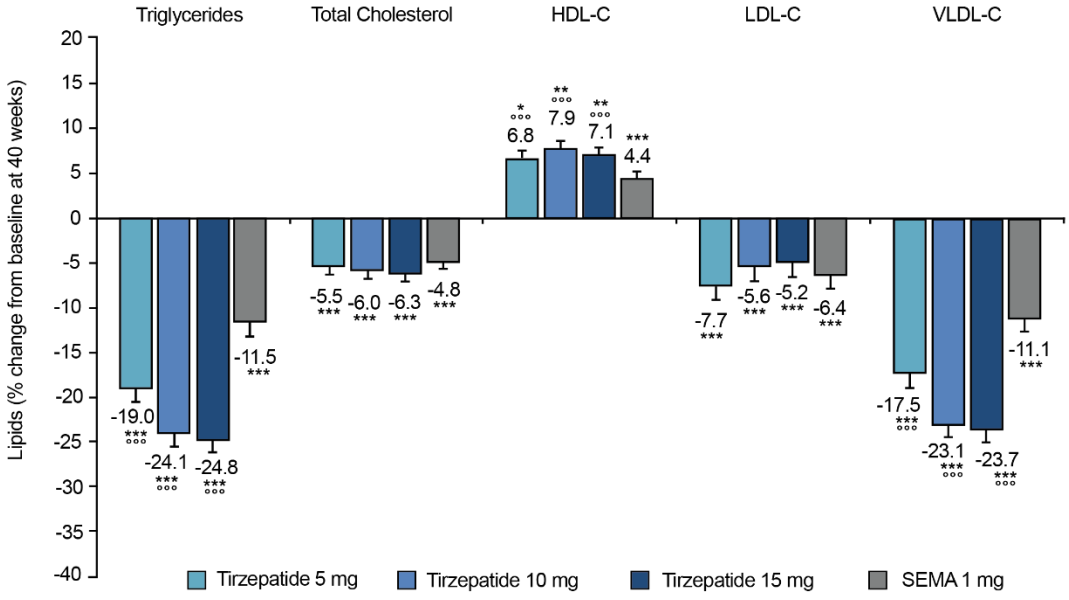
In a pre-specified exploratory endpoint, patients on all three doses of tirzepatide achieved greater reductions in triglycerides and VLDL-C and greater increases in HDL-C compared with patients on semaglutide 1 mg at 40 weeks, with the greatest changes observed in the tirzepatide 15 mg group. There were no significant differences in LDL and total cholesterol between any of the tirzepatide groups and the semaglutide 1 mg group (Table 29; Figure 11).

Table 29: Change in lipids from baseline to 52 weeks; SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	470	469	469	468
Triglycerides				
Baseline (mg/dL)	165.9	167.4	163.6	165.2
Change from baseline to 40 weeks (mg/dL)	−31.4	−40.0	−41.1	−19.1
Total cholesterol				
Baseline (mg/dL)	171.5	171.3	168.6	170.9
Change from baseline to 40 weeks (mg/dL)	−9.4	−10.2	−10.7	−8.2
HDL-C				
Baseline (mg/dL)	42.9	42.7	42.9	42.7
Change from baseline to 40 weeks (mg/dL)	2.9	3.4	3.0	1.9
LDL-C				
Baseline (mg/dL)	88.2	88.4	86.4	88.2
Change from baseline to 40 weeks (mg/dL)	−6.7	−4.9	−4.5	−5.6
VLDL-C				
Baseline (mg/dL)	32.5	32.8	32.3	32.7
Change from baseline to 40 weeks (mg/dL)	−5.7	−7.5	−7.7	−3.6

mITT population. The estimated means are shown. MMRM analysis on log-transformed data then converted back to original scale. Not controlled for type 1 error.
Abbreviations: CI: confidence interval; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; mITT: modified intent-to-treat; MMRM: mixed model repeated measures; N: number of patients who were randomized and received at least 1 dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; n/a: not applicable; TZP: tirzepatide; VLDL-C: very-low-density lipoproteins cholesterol.
Source: SURPASS-2 CSR; [90] Frias et al (2021).[86]

Figure 11: Percentage change in lipid parameters from baseline to 40 weeks; SURPASS-2



The Estimated means ± standard errors are shown. Not controlled for type 1 error. *** p<0.001 versus baseline; *p<0.05 versus semaglutide 1 mg;** p<0.01 versus semaglutide 1 mg; °°°p<0.001 versus semaglutide 1 mg.
Abbreviations: HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; SEMA: semaglutide; TZP: tirzepatide; VLDL-C: very-low-density lipoproteins cholesterol.
Source: SURPASS-2 CSR.[90]

B.2.6.1.4 Patient-reported outcomes

APPADL: The APPADL questionnaire contains seven items that assess how difficult it is for patients to engage in activities of normal daily life, such as walking, standing and climbing stairs. Items are scored on a 5-point numeric rating scale (5 = “not at all difficult”; 1 = “unable to do”); scores are then linearly transformed to a 0–100 range. Higher APPADL raw and transformed scores correspond to a better self-reported ability to perform physical activities of daily living.

IWQOL-Lite-CT: The IWQOL-Lite-CT is a measure of weight-related functioning in populations commonly targeted for clinical trials. The IWQOL-Lite-CT is a 20-item, measure with 2 primary domains: physical (7 items, with physician function comprising 5 of the 7 items) and psychosocial (13 items). Higher transformed scores correspond to better HRQoL and functioning.

APPADL – tirzepatide 15 mg superior to semaglutide for improvements in daily living

Total raw and transformed APPADL scores, for each of the three tirzepatide groups and the semaglutide 1 mg group, significantly improved from baseline to 40 weeks indicating better ability to perform physical activities of daily living. Tirzepatide 15 mg achieved a significantly greater improvement in APPADL scores from baseline to 40 weeks compared to semaglutide 1 mg (Table 30). The comparisons of APPADL scores were not controlled for Type 1 error.

Table 30: Summary of APPADL transformed scores by treatment group; SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	422	403	400	416
Baseline	[redacted]	[redacted]	[redacted]	[redacted]

Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from semaglutide 1 mg at 40 weeks (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

mITT population. Change from baseline calculated using ANCOVA, LOCF; only the non-missing post-baseline observation prior to rescue or study drug discontinuation was carried forward. Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.05 versus semaglutide 1 mg.
Abbreviations: ANCOVA: analysis of covariance; APPADL: Ability to Perform Physical Activities of Daily Living; CI: confidence interval; LOCF: last observation carried forward; mITT: modified intent-to-treat; n: number of patients in the mITT efficacy analysis set with baseline and at least 1 postbaseline value; n/a: not applicable; SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS-2 CSR.[90]

IWQOL-Lite-CT – tirzepatide treatment achieved greater improvements in physical function associated with weight vs semaglutide

In a pre-specified, exploratory analysis, IWQOL-Lite-CT total scores of each of the three tirzepatide groups and the semaglutide 1 mg group improved from baseline to 40 weeks, demonstrating improvements in overall HRQoL and functioning associated with weight. Tirzepatide 10 mg and 15 mg achieved greater improvements in IWQOL-Lite-CT total scores from baseline to 40 weeks, compared to semaglutide 1 mg. The IWQOL-Lite-CT total scores are presented in the SURPASS-2 CSR.

The IWQOL-Lite-CT Physical Function scores of each of the three tirzepatide groups and the semaglutide 1 mg group improved from baseline to 40 weeks, demonstrating improvements in the physical impacts of weight on patients completing daily activities. The improvement in IWQOL-Lite-CT Physical Function score for all three tirzepatide groups was greater than that of the semaglutide 1 mg group (Table 31).

Table 31: Summary of IWQOL-Lite-CT Physical Function scores by treatment group; SURPASS-2

	TZP 5 mg	TZP 10 mg	TZP 15 mg	SEMA 1 mg
N	417	401	395	414
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks (LOCF)	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from semaglutide 1 mg at 40 weeks (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

mITT population. Change from baseline calculated using ANCOVA, LOCF; only the non-missing post-baseline observation prior to rescue or study drug discontinuation was carried forward. Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.05 versus semaglutide 1 mg. *p<0.001 versus baseline; **p<0.05 versus semaglutide 1 mg; ***p<0.001 versus semaglutide 1 mg.
Abbreviations: ANCOVA: analysis of covariance; CI: confidence interval; IWQOL-Lite-CT: Impact of Weight on Quality of Life-Lite-Clinical Trials Version; LOCF: last observation carried forward; mITT: modified intent-to-treat; n: number of patients in the mITT efficacy analysis set with baseline and at least 1 postbaseline value; n/a: not applicable; TZP: tirzepatide.

Source: SURPASS-2 CSR.[90]

B.2.6.2 SURPASS-3

Population: Patients with T2D, who had inadequate glycaemic control on stable doses of metformin with or without an SGLT2i
Comparator: titrated insulin degludec to a target fasting blood glucose of <90 mg/dL (5.0 mmol/L)

B.2.6.2.1 Glycaemic control

Primary endpoint: change in HbA1c from baseline – tirzepatide superior to insulin degludec

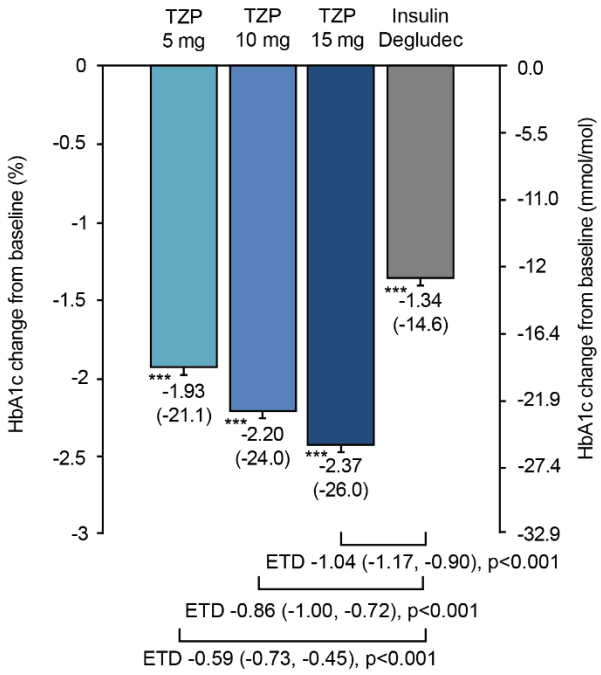
All three doses of tirzepatide achieved significantly greater reduction in HbA1c from baseline to 52 weeks, compared to insulin degludec (Table 32; Figure 12). Within the insulin degludec arm, most of the titration occurred during the first half of the study and a mean dose of 48.8 U per day (SD 30.4; 0.5 U/kg/day [0.3]) was reached at week 52, a similar dose to that seen in previous studies of insulin degludec.[94, 95]

Table 32: Change in HbA1c from baseline to 52 weeks; SURPASS-3

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin degludec
N	358	360	358	359
HbA1c, %				
Baseline	8.17	8.19	8.21	8.13
Change from baseline to 52 weeks	−1.93*	−2.20*	−2.37*	−1.34*
Change difference from insulin degludec (95% CI) at 52 weeks	−0.59** (−0.73, −0.45)	−0.86** (−1.00, −0.72)	−1.04** (−1.17, −0.90)	n/a
HbA1c, mmol/mol				
Baseline	65.8	66.0	66.3	65.4
Change from baseline to 52 weeks	−21.1*	−24.0*	−26.0*	−14.6*
Change difference from insulin degludec (95% CI) at 52 weeks	−6.4** (−7.9, −4.9)	−9.4** (−10.9, −7.9)	−11.3** (−12.8, −9.8)	n/a

*p<0.001 versus baseline; ** p<0.001 versus insulin degludec for superiority.
Abbreviations: CI: confidence interval; HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: Ludvik *et al,* 2021.[87]

Figure 12: Change in HbA1c from baseline to 52 weeks: SURPASS-3

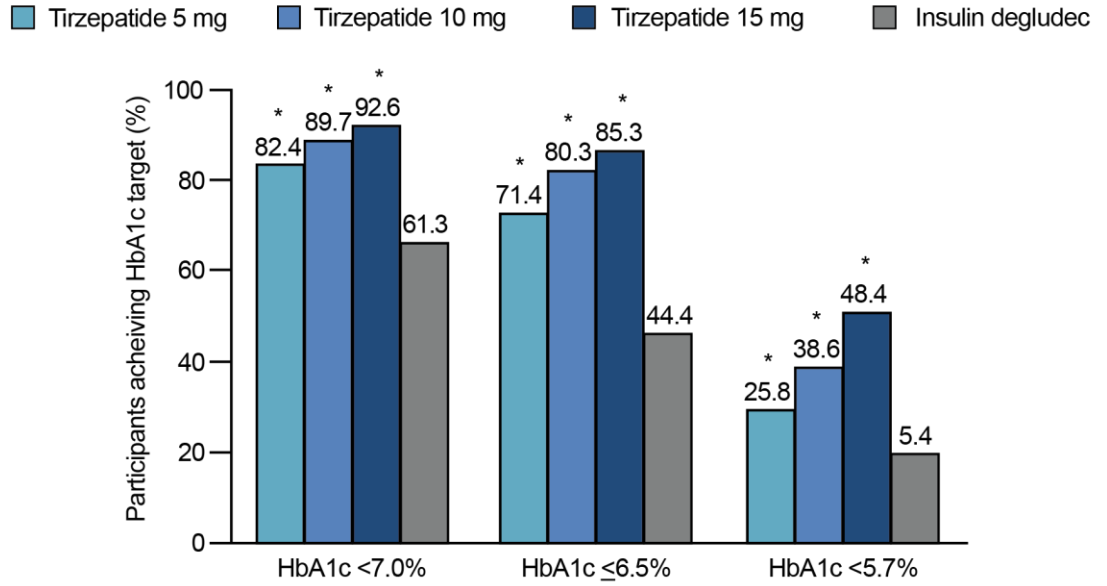


The least-squares means ± standard errors are shown. Values in parentheses are in mmol/mol. ***p<0.001 versus baseline.
Abbreviations: CI: confidence interval; ETD: estimated treatment difference; HbA1c: glycated haemoglobin.
Source: SURPASS-3 CSR.[91]

Proportion of patients achieving HbA1c targets at 40 weeks – tirzepatide superior to insulin degludec

Significantly higher proportions of patients achieved the HbA1c target of <7.0% (<53 mmol/mol) at 52 weeks on all three tirzepatide doses compared to insulin degludec. For tirzepatide-treated patients, between 82.4% and 92.6% of patients achieved HbA1c <7.0% (<53 mmol/mol) at 52 weeks, compared to 61.3% on insulin degludec, at 52 weeks (Figure 13). The proportion of patients achieving HbA1c <5.7% on tirzepatide 15 mg was over twice that on insulin degludec at 52 weeks (Figure 13). The comparisons of the proportion of patients achieving HbA1c ≤6.5% and <5.7% were not controlled for Type 1 error for all tirzepatide doses (Table 24).

Figure 13: Proportion of patients who achieved HbA1c targets at 52 weeks: SURPASS-3



HbA1c ≤6.5% and <5.7% comparisons not controlled for type 1 error for all doses of tirzepatide. *p<0.0001 vs insulin degludec at Week 52.
Abbreviations: HbA1c: glycated haemoglobin.
Source: Ludvik *et al,* 2021.[87]

B.2.6.2.2 Body Weight

Change in body weight from baseline – tirzepatide superior to insulin degludec

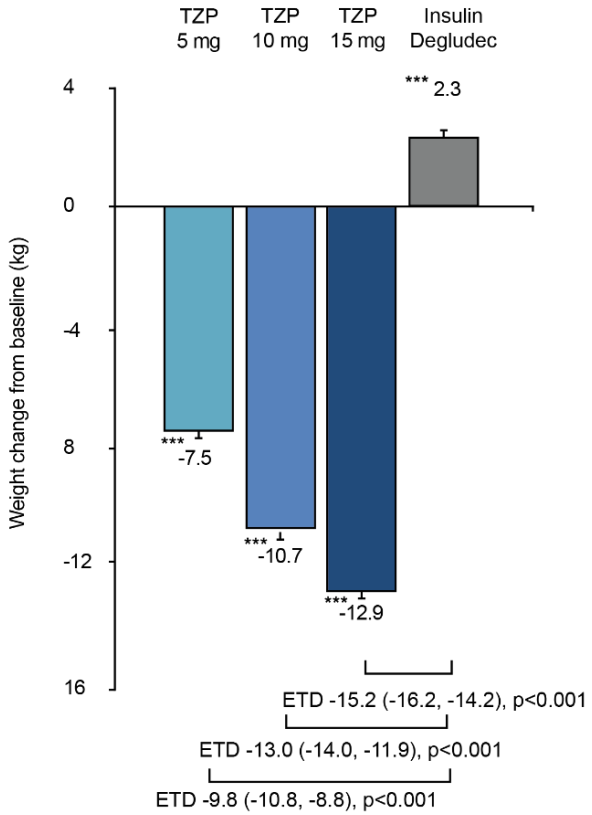
Patients on all three tirzepatide doses achieved significantly greater reductions in body weight at 52 weeks compared to patients on insulin degludec, with the greatest change in body weight from baseline observed for patients on tirzepatide 15 mg. Conversely, patients in the insulin degludec group showed an increase in mean body weight (Table 33; Figure 14).

Table 33: Change in body weight from baseline to 52 weeks: SURPASS-3

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin degludec
N	358	360	358	359
Weight, kg				
Baseline	94.5	94.3	94.9	94.2
Change from baseline to 52 weeks	−7.5*	−10.7*	−12.9*	2.3*
Change difference from insulin degludec (95% CI) at 52 weeks	−9.8** (−10.8, −8.8)	−13.0** (−14.0, −11.9)	−15.2** (−16.2, −14.2)	n/a

*p<0.001 versus baseline; **p<0.001 versus insulin degludec for the mean change difference.
Abbreviations: CI: confidence interval; TZP: tirzepatide.
Source: Ludvik *et al,* 2021.[87]

Figure 14: Change in body weight from baseline over time; SURPASS-3

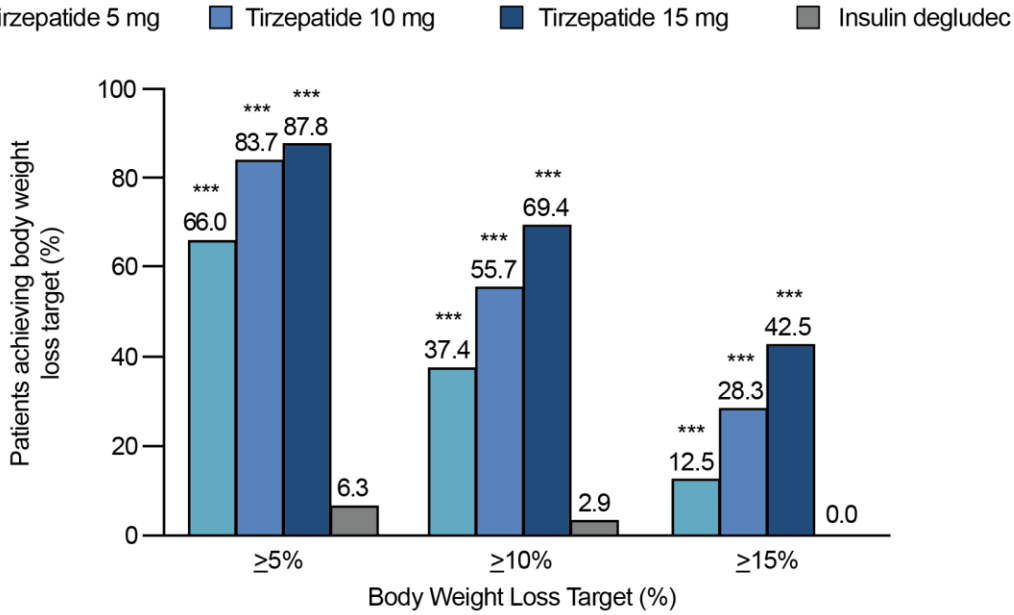


The least-squares means ± standard errors are shown. †††p-Value <0.001 versus baseline.
Abbreviations: ETD: estimated treatment difference; TZP: tirzepatide.
Source: SURPASS-3 CSR.[91]

Proportion of patients achieving weight loss targets at 52 weeks – tirzepatide superior to insulin degludec

Significantly higher proportions of patients achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% from baseline to 52 weeks in all three tirzepatide groups compared with insulin degludec (not controlled for type 1 error). Only 6.3% of patients on insulin degludec achieved a body weight reduction of ≥5%, compared to between 66.0% and 87.8% of patients on tirzepatide. No patients on insulin degludec achieved a body weight reduction of ≥15%, compared to between 12.5% and 42.5% of patients on tirzepatide (Figure 15).

Figure 15: Proportion of patients achieving body weight loss targets at 52 weeks; SURPASS-3



Not controlled for type 1 error. ***p<0.001 versus insulin degludec.
Abbreviations: TZP: tirzepatide.
Source: SURPASS-3 CSR.[91]

Change in BMI from baseline – tirzepatide treatment led to greater BMI reduction vs insulin degludec

In a pre-specified exploratory endpoint, patients on all three doses of tirzepatide achieved larger reductions in BMI from baseline to 52 weeks compared with patients on insulin degludec, with the largest reduction observed for the tirzepatide 15 mg group. Conversely, patients in the insulin degludec treatment group showed an increase in BMI (Table 34).

Table 34: Change in BMI from baseline to 52 weeks: SURPASS-3

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin degludec
N	358	360	358	359
BMI				
Baseline	33.6	33.5	33.7	33.4
Change from baseline to 52 weeks	−2.7*	−3.8*	−4.6*	0.8*
Change difference from insulin degludec (95% CI) at 52 weeks	−3.6** (−3.9, −3.2)	−4.7** (−5.0, −4.3)	−5.5** (−5.8, −5.1)	n/a

Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.001 versus insulin degludec.

Abbreviations: BMI: body mass index; CI: confidence interval; mITT: modified intent-to-treat; N: number of patients who were randomized and received at least 1 dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; n/a: not applicable; TZP: tirzepatide.
Source: SURPASS-3 CSR.[91]

B.2.6.2.3 Lipids

In a pre-specified exploratory endpoint, all treatment groups had reduced triglycerides, total cholesterol, LDL-C and VLDL-C and increased HDL-C compared to baseline at 52 weeks. Patients on tirzepatide 10 mg and 15 mg had greater reductions in triglycerides and VLDL-C at 52 weeks compared with patients on insulin degludec. Additionally, patients on all three doses of tirzepatide had greater increases in HDL-C at 52 weeks, compared with patients on insulin degludec. Full results are presented in Appendix M.

B.2.6.2.4 Patient-reported outcomes

APPADL – tirzepatide superior to insulin degludec for improvements in daily living

Total raw and transformed APPADL scores were significantly improved from baseline to 52 weeks for patients on all three doses of tirzepatide compared with insulin degludec, indicating better daily living (Table 35). The comparisons of APPADL scores were not controlled for Type 1 error.

Table 35: Summary of APPADL transformed scores by treatment group; SURPASS-3

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin degludec
n	312	297	297	320
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from insulin degludec (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

mITT population. Change from baseline calculated using ANCOVA, LOCF; only the non-missing post-baseline observation prior to rescue or study drug discontinuation was carried forward. Not controlled for type 1 error.
*p<0.001 versus baseline; **p<0.01 versus insulin degludec; ***p<0.001 versus insulin degludec.
Abbreviations: ANCOVA: analysis of covariance; APPADL: Ability to Perform Physical Activities of Daily Living; CI: confidence interval; LOCF: last observation carried forward; mITT: modified intent-to-treat; n: number of patients in the mITT efficacy analysis set with baseline and at least 1 postbaseline value; n/a: not applicable; TZP: tirzepatide.
Source: SURPASS-3 CSR.[91]

B.2.6.3 SURPASS-4

Population: Patients with T2D with high CV risk, who had inadequate glycaemic control on stable doses of at least 1 and no more than 3 OADs, including of metformin, an SGLT2i and/or an SU
Comparator: insulin glargine titrated to a target fasting blood glucose of <100 mg/dL (5.6 mmol/L)

B.2.6.3.1 Glycaemic control

Primary endpoint: change in HbA1c from baseline – tirzepatide superior to insulin glargine

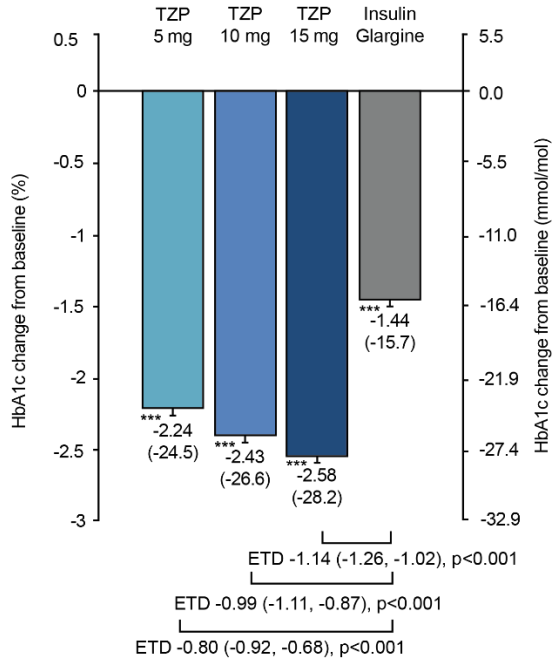
Patients on all three doses of tirzepatide achieved significantly greater reduction in HbA1c from baseline to 52 weeks, compared to patients on insulin glargine (Table 36; Figure 16). Additionally, for patients treated for longer than 52 weeks, the reduction in HbA1c appeared to be sustained up to 104 weeks for all tirzepatide groups.[92] Within the insulin glargine arm, the mean insulin glargine dose was 43·5 U per day (SD: 24.96; 0.5 U/kg/day) at week 52 and 47·0 U per day (SD: 22.69; 0.5 U/kg/day) at week 104. The doses of glargine used and the fasting glucose achieved suggest that the glargine titration algorithm was followed appropriately.

Table 36: Change in HbA1c from baseline to 52 weeks; SURPASS-4

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin glargine
N	326	321	334	978
HbA1c, %				
Baseline	8.52	8.60	8.52	8.51
Change from baseline to 52 weeks	−2.24*	−2.43*	−2.58*	−1.44*
Change difference from insulin glargine (95% CI) at 52 weeks	−0.80** (−0.92, −0.68)	−0.99** (−1.11, −0.87)	−1.14** (−1.26, −1.02)	n/a
HbA1c, mmol/mol				
Baseline	69.6	70.5	69.6	69.5
Change from baseline to 52 weeks	−24.5*	−26.6*	−28.2*	−15.7*
Change difference from insulin glargine (95% CI) at 52 weeks	−8.8** (−10.1, −7.4)	−10.9** (−12.3, −9.6)	−12.5** (13.8, −11.2)	n/a

*p<0.001 versus baseline; **p<0.001 versus insulin glargine.
Abbreviations: CI: confidence interval; HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: Del Prato *et al,* 2021. [85]

Figure 16: Change in HbA1c from baseline to 52 weeks; SURPASS-4

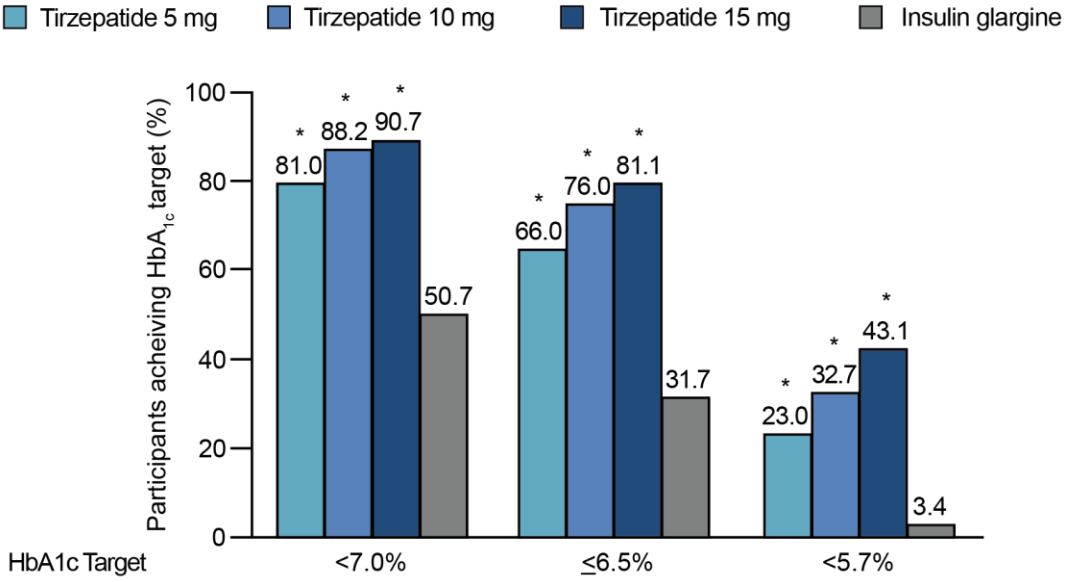


The least-squares means ± standard errors are shown. Values in parentheses are in mmol/mol. ***p<0.001 versus baseline.
Abbreviations: CI: confidence interval; ETD: estimated treatment difference; HbA1c: glycated haemoglobin.
Source: SURPASS-4 CSR.[92]

Proportion of patients achieving HbA1c targets at 52 weeks – tirzepatide superior to insulin glargine

Significantly higher proportions of patients achieved the HbA1c target of <7.0% (<53 mmol/mol) at 52 weeks on all three tirzepatide doses compared to insulin glargine: 81–91% of tirzepatide-treated participants versus 51% on insulin glargine. The proportion of patients meeting the more stringent target of HbA1c ≤6.5% for all doses of tirzepatide was over twice that of insulin glargine (66–81% of tirzepatide-treated participants versus 32% on insulin glargine). Similarly, 23–43% of tirzepatide-treated participants versus 3% on insulin glargine achieved a HbA1c of <5.7% (<39 mmol/mol) (Figure 17). The comparisons of the proportion of patients achieving HbA1c ≤6.5% and <5.7% were not controlled for Type 1 error for all tirzepatide doses (Table 24).

Figure 17: Proportion of patients who achieved HbA1c targets at 52 weeks; SURPASS-4



HbA1c ≤6.5% and <5.7% comparisons not controlled for type 1 error for all doses of tirzepatide. *p<0.0001 vs insulin degludec.
Abbreviations: HbA1c: glycated haemoglobin.
Source: SURPASS-4 CSR.[92]

B.2.6.3.2 Body Weight

Change in body weight from baseline – tirzepatide superior to insulin glargine

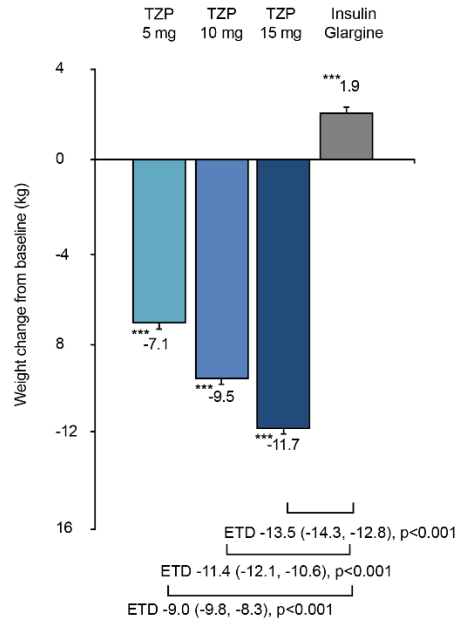
Patients on all three doses of tirzepatide achieved significantly greater reductions in body weight at 52 weeks compared to patients on insulin glargine, with the largest reduction observed for the tirzepatide 15 mg group (Table 37; Figure 18). In comparison, patients in the insulin glargine group showed an increase in mean body weight. Additionally, for patients treated for longer than 52 weeks, the reduction in body weight appeared to be sustained up to 104 weeks for all tirzepatide groups.[92]

Table 37: Change in body weight from baseline to 52 weeks: SURPASS-4

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin glargine
N	326	321	334	978
Weight, kg				
Baseline	90.3	90.7	90.0	90.3
Change from baseline to 52 weeks	−7.1*	−9.5*	−11.7*	1.9
Change difference from insulin glargine (95% CI) at 52 weeks	−9.0** (−9.8, −8.3)	−11.4** (−12.1, −10.6)	−13.5** (−14.3, −12.8)	n/a

*p<0.001 versus baseline; **p<0.001 versus insulin glargine.
Abbreviations: CI: confidence interval; TZP: tirzepatide.
Source: Del Prato *et al,* 2021. [85]

Figure 18: Change in body weight from baseline to 52 weeks; SURPASS-4

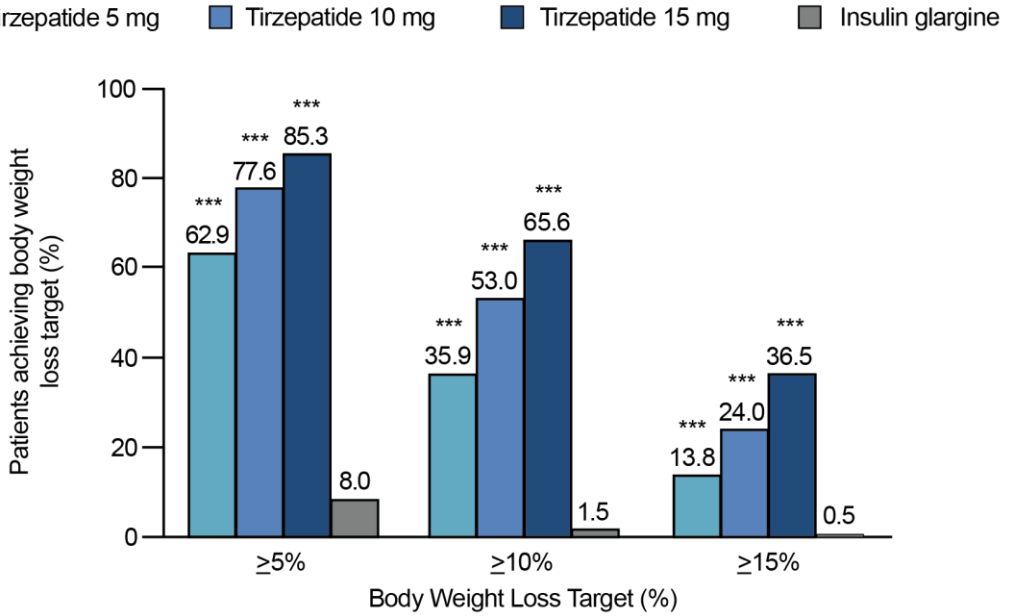


The least-squares means ± standard errors are shown. †††p-Value <0.001 versus baseline.
Abbreviations: ETD: estimated treatment difference; TZP: tirzepatide.
Source: SURPASS-4 CSR.[92]

Proportion of patients achieving weight loss targets – tirzepatide superior to insulin glargine

Significantly higher proportions of patients in all three tirzepatide groups achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% from baseline to 52 weeks compared to insulin glargine (not controlled for type 1 error). Only 8.0% of patients on insulin glargine achieved a body weight reduction of ≥5%, compared to between 62.9% and 85.3% of patients on tirzepatide (Figure 19).

Figure 19: Proportion of patients achieving weight loss targets at 52 weeks; SURPASS-4



Not controlled for type 1 error. ***p<0.001 versus insulin degludec.
Abbreviations: TZP: tirzepatide.
Source: SURPASS-4 CSR.[92]

Change in BMI from baseline – tirzepatide treatment led to greater BMI reduction vs insulin glargine

In a pre-specified exploratory analysis, patients on all three doses of tirzepatide achieved significantly reduced BMI from baseline to 52 weeks compared to patients on insulin glargine, with the largest reduction observed for the tirzepatide 15 mg group (Table 38). Conversely, an average increase in BMI was reported for patients on insulin glargine. Additionally, for patients treated for longer than 52 weeks, the change in BMI observed at 52 weeks was sustained without substantial further reductions at 78 and 104 weeks for all tirzepatide groups.[92]

Table 38: Change in BMI from baseline to 52 weeks; SURPASS-4

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin glargine
N	328	326	337	998
BMI				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 52 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from insulin glargine (95% CI) at 52 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.001 versus insulin glargine.
Abbreviations: BMI: body mass index; CI: confidence interval; N: number of patients who were randomized and received at least 1 dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; n/a: not applicable; TZP: tirzepatide.
Source: SURPASS-4 CSR.[92]

B.2.6.3.3 Lipids

In a pre-specified exploratory analysis, patients on all three doses of tirzepatide achieved greater reductions in triglycerides, VLDL-C, LDL, and total cholesterol, and greater increases in HDL-C compared with patients on insulin glargine at 52 weeks, with the greatest changes observed in the tirzepatide 15 mg group. Full results are presented in Appendix M.

B.2.6.3.4 Patient-reported outcomes

APPADL – tirzepatide superior to insulin glargine for improvements in daily living

Total raw and transformed APPADL scores for each of the three tirzepatide groups significantly improved from baseline to 52 weeks, indicating better daily living; conversely, the APPADL scores for the insulin glargine group decreased at 52 weeks. All tirzepatide doses achieved a significant improvement in APPADL scores compared to insulin glargine at 52 weeks (Table 39). The comparisons of APPADL scores were not controlled for Type 1 error.

Table 39: Summary of APPADL transformed scores at 52 weeks; SURPASS-4

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Insulin glargine
n	285	291	296	903
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 52 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from insulin glargine at 52 weeks (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.05 versus insulin glargine.
Abbreviations: APPADL: Ability to Perform Physical Activities of Daily Living; CI: confidence interval; n: number of patients in the mITT efficacy analysis set with baseline and at least 1 postbaseline values; n/a: not applicable; TZP: tirzepatide.
Source: SURPASS-4 CSR.[92]

B.2.6.4 SURPASS-5

Population: Patients with T2D, with background therapy of insulin glargine with or without metformin **Comparator**: placebo

B.2.6.4.1 Glycaemic control

Primary endpoint: change in HbA1c from baseline – tirzepatide superior to placebo

Patients on all three doses of tirzepatide (5 mg, 10 mg, 15 mg) achieved significantly greater reduction in HbA1c from baseline to 40 weeks, compared to patients on placebo (Table 40; Figure 20).

Table 40: Change in HbA1c from baseline to 40 weeks; SURPASS-5

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Placebo
N	116	118	118	119
HbA1c, %				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	−2.23*	−2.59*	−2.59*	−0.93*
Change difference from placebo (95% CI) at 40 weeks	−1.30** (−1.52, −1.07)	−1.66** (−1.88, −1.43)	−1.65** (−1.88, −1.43)	n/a
HbA1c, mmol/mol				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	[redacted]*	[redacted]*	[redacted]*	[redacted]*
Change difference from placebo (95% CI) at 40 weeks	[redacted]** [redacted]	[redacted]** [redacted]	[redacted]** [redacted]	[redacted]

*p<0.001 versus baseline; **p<001 versus placebo.
Abbreviations: CI: confidence interval; HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: Dahl *et al*, 2021; [84] SURPASS-5 CSR.[89]

Figure 20: Change in HbA1c from baseline to 40 weeks; SURPASS-5

The least-squares means ± standard errors are shown. Values in parentheses are in mmol/mol. ***p<0.001 versus baseline.
Abbreviations: CI: confidence interval; ETD: estimated treatment difference; HbA1c: glycated haemoglobin.
Source: SURPASS-5 CSR.[89]

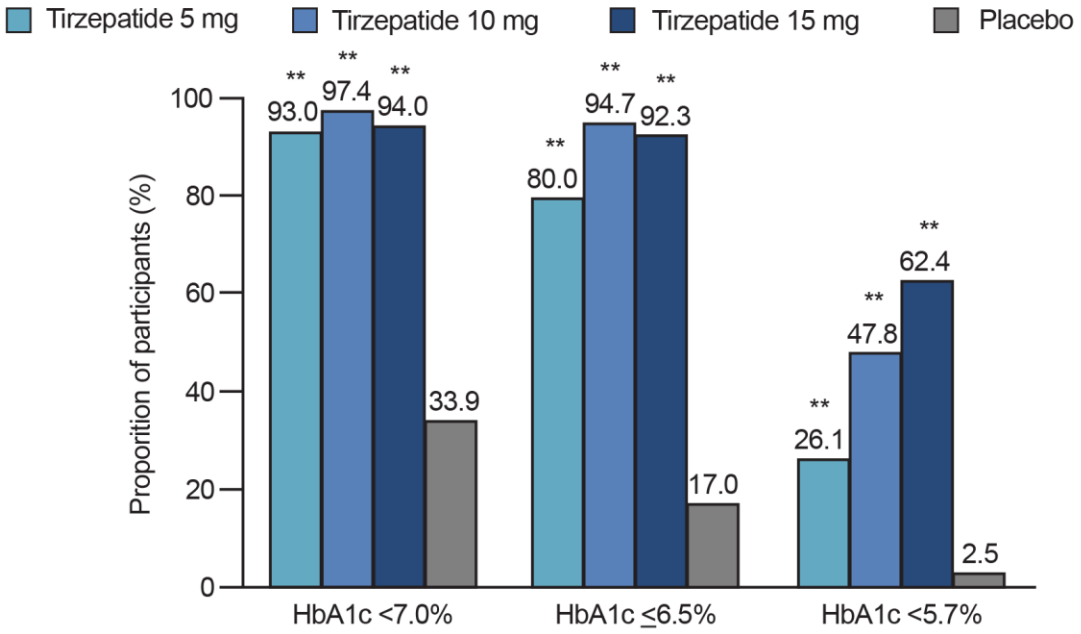
Proportion of patients achieving HbA1c targets at 40 weeks – tirzepatide superior to placebo

Significantly higher proportions of patients achieved the HbA1c target of <7.0% (<53 mmol/mol) at 40 weeks on all three tirzepatide doses compared to placebo. Between 93.0% and 97.4% of patients

on tirzepatide achieved HbA1c <7.0% at 40 weeks, compared to 33.9% of patients on placebo (Figure 21).

Similarly, higher proportions of patients on all tirzepatide doses achieved the more stringent HbA1c targets of ≤6.5% and <5.7% (Figure 21). The comparison of the proportion of patients achieving HbA1c ≤6.5% was not controlled for Type 1 error for any tirzepatide doses and for <5.7% was not controlled for Type 1 error for tirzepatide 5 mg (Table 24).

Figure 21: Proportion of patients reaching HbA1c targets at 40 weeks; SURPASS-5



HbA1c ≤6.5% comparisons not controlled for type 1 error for any doses of tirzepatide; HbA1c <5.7% comparisons not controlled for type 1 error for tirzepatide 5 mg. **p<0.001 versus placebo.
Abbreviations: HbA1c: glycated haemoglobin.
Source: Dahl *et al*, 2021.[84]

B.2.6.4.2 Mean insulin glargine dose

There was a significant increase in daily mean insulin glargine dose from baseline to 40 weeks for patients receiving placebo, and no significant difference from baseline to 40 weeks for patients on all tirzepatide doses (Table 41). The comparisons of daily mean insulin glargine dose were not controlled for Type 1 error for all tirzepatide doses.

Table 41: Summary of change in mean insulin glargine dose from baseline to 40 weeks; SURPASS-5

	TZP 5 mg (N=116)	TZP 10 mg (N=118)	TZP 15 mg (N=118)	Placebo (N=119)
IU/day				
Baseline	34.3	32.0	35.0	32.9

Change from baseline to 40 weeks	4.4	2.7	−3.8	25.1
Percent change from baseline to 40 weeks	13.0	8.1	−11.4	75.0*
IU/kg/day				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Percent change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.01 versus placebo.
Abbreviations: CI: confidence interval; IU: International Units; MMRM: mixed model repeated measures; N: number of patients who were randomized and received at least one dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; TZP: tirzepatide.
Source: Dahl et al (2021);[84] SURPASS-5 CSR.[89]

B.2.6.4.3 Body weight

Change in body weight from baseline – tirzepatide superior to placebo

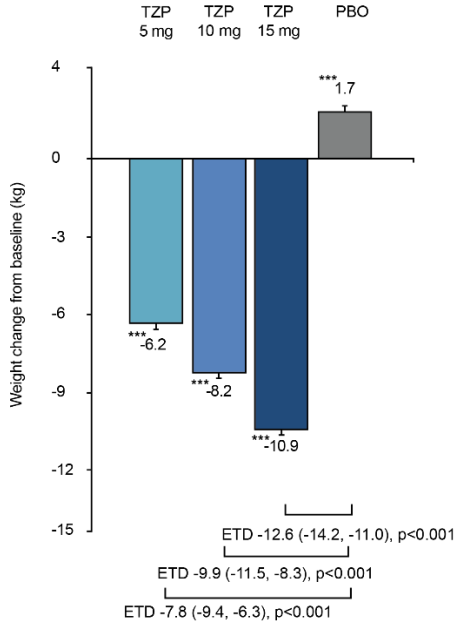
Patients on all three tirzepatide doses achieved significant reductions in body weight from baseline to 40 weeks, compared to patients on placebo (Table 42; Figure 22).

Table 42: Change in body weight from baseline to 40 weeks; SURPASS-5

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Placebo
N	116	118	118	119
Weight, kg				
Baseline*	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	−6.2*	−8.2*	−10.9*	1.7**
Change difference from placebo (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

*p<0.001 versus baseline; **p<0.01 versus baseline; *** p<0.001 versus placebo for the mean change difference.
Abbreviations: CI: confidence interval; TZP: tirzepatide.
Source: Dahl et al, 2021; [84] SURPASS-5 CSR.[89]

Figure 22: Change in body weight from baseline to 40 weeks; SURPASS-5

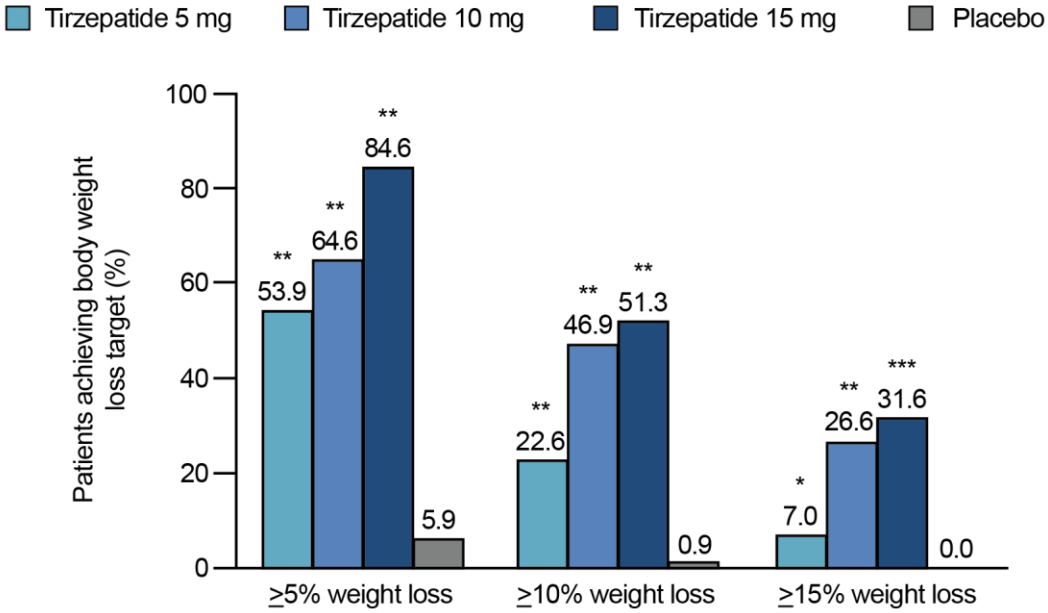


The least-squares means ± standard errors are shown. *** p-Value <0.001 versus baseline.
Abbreviations: ETD: estimated treatment difference; TZP: tirzepatide.
Source: SURPASS-5 CSR.[89]

Proportion of patients achieving weight loss targets at 40 weeks – tirzepatide superior to placebo

Significantly higher proportions of patients on all three tirzepatide doses achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% from baseline to 40 weeks compared to placebo (not controlled for type 1 error). Only 5.9% of patients on placebo achieved a body weight reduction ≥5%, compared to between 53.9% and 84.6% of patients on tirzepatide (Figure 23).

Figure 23: Proportion of patients reaching body weight loss goals at 40 weeks; SURPASS-5



Not controlled for type 1 error. *p<0.5 versus placebo, **p<0.01 versus placebo, ***p<0.001 versus placebo.
Source: Dahl *et al*, 2021.[84]

Change in BMI from baseline – tirzepatide treatment led to greater BMI reduction vs placebo

In a pre-specified exploratory endpoint, patients on all three doses of tirzepatide achieved significantly reduced BMI from baseline to 40 weeks compared with patients on placebo, with the largest reduction observed for the tirzepatide 15 mg group. Conversely, an average increase in BMI was observed for patients on placebo (Table 43).

Table 43: Change in BMI from baseline to 40 weeks: SURPASS-5

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Placebo
N	116	118	118	119
BMI				
Baseline	33.6	33.5	33.4	33.3
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from placebo (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.005 versus baseline; ***p<0.001 versus placebo
Abbreviations: BMI: body mass index; CI: confidence interval; mITT: modified intent-to-treat; N: number of patients who were randomized and received at least 1 dose of study drug, excluding patients who discontinued study drug due to inadvertent enrolment; n/a: not applicable; TZP: tirzepatide.
Source: Dahl et al (2021);[84] SURPASS-5 CSR.[89]

B.2.6.4.4 Lipids

In a pre-specified exploratory endpoint, patients on all three doses of tirzepatide had significant reductions in triglycerides, total cholesterol, LDL-C, and VLDL-C compared with patients on placebo

at 40 weeks, with the greatest changes observed in the tirzepatide 15 mg group. There was no difference in HDL-C between any of the tirzepatide groups and the placebo group. Full results are presented in Appendix M.

B.2.6.4.5 Patient-reported outcomes

APPADL – tirzepatide 10 mg and 15 mg superior to placebo for improvements in daily living

Total raw and transformed APPADL scores for the 10 mg and 15 mg tirzepatide groups significantly improved from baseline to 40 weeks, indicating better daily living. No significant changes were seen in the tirzepatide 5 mg group and the placebo group from baseline to 40 weeks. The tirzepatide 10 mg and 15 mg groups achieved significant improvements in APPADL scores compared to the placebo group (Table 44). The comparisons of APPADL scores were not controlled for Type 1 error.

Table 44: Summary of APPADL transformed scores by treatment group; SURPASS-5

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Placebo
n	106	106	98	111
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from placebo 1mg at 40 weeks (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

The least-squares means are shown. Not controlled for type 1 error. *p<0.001 versus baseline; **p<0.05 versus placebo.
Abbreviations: ANCOVA: analysis of covariance; APPADL: Ability to Perform Physical Activities of Daily Living; CI: confidence interval; LOCF: last observation carried forward; mITT: modified intent-to-treat; n: number of patients in the mITT efficacy analysis set with baseline and at least 1 postbaseline value; n/a: not applicable; TZP: tirzepatide.
Source: SURPASS-5 CSR.[89]

B.2.7 *Subgroup analysis*

Pre-planned subgroup analyses were conducted to investigate potential treatment effect modifiers. The results of the subgroup analyses found no evidence treatment effect modifiers in any of the parameters tested, and that the SURPASS trials are generalisable to the patient population of interest to this submission.

A summary of the subgroup analyses conducted in each trial is presented in Table 45.

Table 45: Subgroup analyses in the SURPASS trials

Trial	Subgroup analyses
SURPASS-2	Subgroup analyses were conducted for CfB in HbA1c at 40 weeks, and CfB in body weight at 40 weeks. Both included the patient characteristics of: • Age (<65 vs. ≥65 years, age group 1) • Age (<75 vs. ≥75 years, age group 2) • Race • Ethnicity • Sex • Geographic region (US vs. OUS) • Duration of diabetes (<median vs. ≥median, duration of diabetes group 1) • Duration of diabetes (≤5 years vs. >5 to ≤10 years vs. >10 years, duration of diabetes group 2) • Baseline HbA1c (≤8.5% vs. >8.5%) • Baseline eGFR (<60 ml/min/1.73 m^2 vs. ≥60 ml/min/1.73 m^2) • Baseline BMI (<27 kg/m^2 vs. ≥27 kg/m^2, baseline BMI Group 1) • Baseline BMI (<30 kg/m^2 vs. ≥30 to <35 kg/m^2 vs. ≥35 kg/m^2, baseline BMI Group 2)
SURPASS-3	Subgroup analyses were conducted for CfB in HbA1c at 52 weeks, and CfB in body weight at 52 weeks. Both included the patient characteristics of: • Age (<65 vs. ≥65 years, age group 1) • Age (<75 vs. ≥75 years, age group 2) • Race • Ethnicity • Sex • Geographic region (US vs. OUS) • Duration of diabetes (<median vs. ≥median, duration of diabetes group 1) • Duration of diabetes (≤5 years vs. >5 to ≤10 years vs. >10 years, duration of diabetes group 2) • Baseline HbA1c (≤8.5% vs. >8.5%) • Baseline eGFR (<60 ml/min/1.73 m2 vs. ≥60 ml/min/1.73 m2) • Baseline BMI (<27 kg/m2 vs. ≥27 kg/m2, baseline BMI Group 1) • Baseline BMI (<30 kg/m2 vs. ≥30 to <35 kg/m2 vs. ≥35 kg/m2, baseline BMI Group 2) • Prior use of OAD (yes vs no)
SURPASS-4	Subgroup analyses were conducted for CfB in HbA1c at 52 weeks, and CfB in body weight at 52 weeks. Both included the patient characteristics of: • Age (<65 vs. ≥65 years, age group 1) • Age (<75 vs. ≥75 years, age group 2) • Baseline BMI (<27 kg/m2 vs. ≥27 kg/m2, baseline BMI Group 1) • Baseline BMI (<30 kg/m2 vs. ≥30 to <35 kg/m2 vs. ≥35 kg/m2, baseline BMI Group 2) • Duration of diabetes (<median vs. ≥median, duration of diabetes group 1)

	• Duration of diabetes (≤5 years vs. >5 to ≤10 years vs. >10 years, duration of diabetes group 2) • Baseline EGFR (<60 ml/min/1.73 m2 vs. ≥60 ml/min/1.73 m2) • Ethnicity • Baseline HbA1c (≤8.5% vs. >8.5%) • Baseline OAD use (metformin alone, metformin + SU, metformin + SGLT2i, metformin + SU + SGLT2i, other) • Race • Geographic region (US vs. OUS) • Sex
SURPASS-5	Subgroup analyses were conducted for CfB in HbA1c at 40 weeks, as well as CfB in body weight at 40 weeks, both using the following patient characteristics: • Age (<65 vs. ≥65 years) • Age (<75 vs. ≥75 years) • Race • Ethnicity • Sex • Geographic region (US vs. OUS) • Duration of diabetes (<median vs. ≥median, duration of diabetes group 1) • Duration of diabetes (≤5 years vs. >5 to ≤10 years vs. >10 years, duration of diabetes group 2) • Baseline HbA1c (≤8.5% vs. >8.5%) • Baseline eGFR (<60 ml/min/1.73 m^2 vs. ≥60 ml/min/1.73 m^2) • Baseline BMI (<27 kg/m^2 vs. ≥27 kg/m^2, baseline BMI Group 1) • Baseline BMI (<30 kg/m^2 vs. ≥30 to <35 kg/m^2 vs. ≥35 kg/m^2, baseline BMI Group 2) • Baseline use of metformin (yes vs. no)

Abbreviations: BMI: body mass index; CfB: change from baseline; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; OAD: oral antidiabetic drug; OUS: outside the USA; US: United States;
Sources: SURPASS-2 CSR;[90] SURPASS-3 CSR;[91]SURPASS-4 CSR;[92] SURPASS-5 CSR.[89]

Details of the subgroup analyses for SURPASS 2–5 are available in the CSRs Section 5.1.[89-92] Overall, analyses of change from baseline in both HbA1c and body weight were generally consistent with the primary results in all of the SURPASS 2–5 trials, with the treatment difference favouring all three doses of tirzepatide compared with the comparator in the majority of subgroups..

B.2.8 *Meta-analysis*

No efficacy meta-analyses were conducted for this submission; however, an NMA was conducted and is presented in Section B.2.9. A meta-analysis of cardiovascular safety data from the SURPASS trial programme was conducted and is presented in Section B.2.9.

B.2.9 *Indirect and mixed treatment comparisons*

- As it is not feasible to conduct RCTs versus all relevant comparators in all clinical settings, an NMA has been conducted to assess the relative efficacy and safety of tirzepatide versus GLP-1 RAs available in NHS practice; the results of this NMA inform clinical inputs within the cost-effectiveness model

Methods and study inclusion

- The network was defined to align with SURPASS-2 and 3 trials and included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice
- Evidence from RCTs was identified in the clinical SLR presented in Section B.2.1. Of the 205 studies included in the SLR, a total of 72 were eligible for inclusion in the NMA (53 in the main analysis and 19 in sensitivity analyses only)
    - The main reasons for exclusion of studies being incorrect study populations (e.g. cardiovascular outcomes trials [CVOTs], renal impairment population, patients with severe insulin resistance, etc.), studies not reporting data at relevant timepoints and evaluation of comparators that were not of interest
- A range of efficacy and safety endpoints were included in the main analysis of the NMA, including change from baseline in HbA1c, weight and BMI

Results

- Overall, limited concerns with regards to inconsistency and heterogeneity were identified. No concerns regarding inconsistency were identified for continuous or binary endpoints. However, heterogeneity was identified for some outcomes, with the SURPASS-2 and 3 trials contributing to heterogeneity in change from baseline in HR, LDL and total cholesterol, and other treatments (excluding tirzepatide) contributing to heterogeneity in change from baseline HbA1c, weight, diastolic blood pressure (DBP) and HDL. Nevertheless, considering the number of studies in the network, only a minority of studies contributed to the heterogeneity
- As GLP-1 RAs and tirzepatide exhibit a dose-response relationship in terms of efficacy and gastrointestinal side-effects, when interpreting the NMA comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps
- For HbA1c change from baseline, all three doses of tirzepatide demonstrated a statistically significantly greater reduction in HbA1c from baseline compared to all GLP1-RAs within the same recommended maintenance dose step
- For body weight change from baseline, all three doses of tirzepatide demonstrated a significantly greater reduction in body weight from baseline compared to all GLP-1 RAs within the same recommended maintenance dose step
- All doses of tirzepatide demonstrated significantly greater reductions in BMI compared to all other GLP-1 RAs within the same recommended maintenance dose step; although, BMI data for studies in the main analyses were limited
- Sensitivity analyses were conducted to assess the robustness of the findings of the main analyses and were largely consistent with the main analyses

Conclusions

- The NMA provides robust results that are generalisable to UK clinical practice. Baseline characteristics were largely consistent across the included treatment arms and as such, the results are likely to be robust with minimal impact from prognostic variables. In addition, results of the sensitivity analyses demonstrate the robustness of the results of the main analyses. Limited concerns with regards to inconsistency and heterogeneity were identified
- For glycaemic and weight loss outcomes, tirzepatide demonstrated statistically significant

improved outcomes when compared to all GLP-1 RAs at the same recommended maintenance dose step

The SURPASS trials provided direct head-to-head data on the efficacy and safety of tirzepatide versus a wide range of comparators, however, it was not feasible to conduct RCTs versus all relevant comparators in all clinical settings. As such, an NMA has been conducted to assess the relative efficacy and safety of tirzepatide versus GLP-1 RAs available in NHS practice .

For the analyses, the network was defined to align with SURPASS-2 and 3 trials through focussing only studies conducted in patients with one to two OADs, as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. A range of efficacy and safety endpoints were included in the main analysis of the NMA, including change from baseline in HbA1c, weight and BMI. Overall, this NMA provides robust results on the comparative efficacy and safety of tirzepatide 5 mg, 10 mg and 15 mg versus relevant GLP-1 RAs at the second and third line of treatment for T2D that can be considered generalisable to the use of tirzepatide as as a more efficacious option whenever GLP-1 RAs would otherwise be considered.

B.2.9.1 Identification of comparator studies

This NMA was based on evidence from the RCTs identified in the clinical SLR, which was conducted in October 2021; see Section B.2.1 and Appendix D for further details of the SLR. Of the 205 studies included in the SLR, a total of 72 were eligible for inclusion in the network (53 in the main analysis and 19 in sensitivity analyses only). Further details of the studies included in the NMA as well as those that were excluded alongside the reasons for their exclusion are provided in Section B.2.9.6. The included studies were assessed for risk of bias using the Cochrane risk of bias assessment tool and the CRD tool and responses were consolidated. The risk of bias assessment for all studies identified by the SLR included in the NMA is presented in Appendix D.

B.2.9.2 Comparators

The SmPC for each comparator treatment was used to identify the licensed doses at the time the NMA was undertaken; these are likely to reflect the different treatment options received by patients with T2D in real-world clinical practice. Based on this, treatment arms using the licensed doses were identified from relevant comparator trials.

The following comparators were included in the NMA (in alphabetical order):

- dulaglutide 0.75 mg QW
    - This dose is currently only licensed as monotherapy and as a starting dose for patients who may be considered more vulnerable, therefore only relevant to a sub-population in UK clinical practice. Results from the NMA are presented although no formal comparisons have been made
- dulaglutide 1.5 mg QW
- dulaglutide 3.0 mg QW
- dulaglutide 4.5 mg QW

- exenatide 5 µg BID (pre-filled pen)
- exenatide 10 µg BID (pre-filled pen)
- exenatide 2.0 mg QW (powder and solvent for prolonged-release suspension for injection)
- lixisenatide 20 µg once daily (QD)
- liraglutide 1.2 mg QD
- liraglutide 1.8 mg QD
- semaglutide 0.5 mg QW (injectable)
- semaglutide 1.0 mg QW (injectable)
- semaglutide 2.0 mg QW (injectable)
    - This dose is not currently available in UK clinical practice but is included in the NMA results
- semaglutide 7.0 mg QD (oral)
- semaglutide 14 mg QD (oral)
- Placebo

As GLP-1 RAs and tirzepatide exhibit a dose-response relationship in terms of efficacy and gastrointestinal side-effects, when interpreting the NMA comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps. Therefore, comparisons were made as per Table 46.

Table 46: Overview of comparators and doses

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide 5 mg	Dulaglutide 1.5 mg Semaglutide 0.5 mg Oral Semaglutide 7 mg Liraglutide 1.2 mg
Tirzepatide 10 mg	Dulaglutide 3.0 mg Semaglutide 1.0 mg Oral Semaglutide 14 mg Liraglutide 1.8 mg
Tirzepatide 15 mg	Dulaglutide 4.5 mg Semaglutide 1.0 mg Oral Semaglutide 14 mg Liraglutide 1.8 mg

Furthermore, to link a network of evidence in the NMA, studies including non GLP-1 RA treatment arms (such as basal insulin, bolus insulin, premixed insulin, dipeptidyl-peptidase 4 inhibitors [DPP-4i], sulfonylurea [SU], thiazolidinediones [TZD], sodium-glucose cotransporter-2 inhibitors SGLT-2i

and placebo) were considered. The use of non-comparator treatment arms to link the network of evidence in the NMA is common practice and this approach is supported by published literature.[96]

B.2.9.3 Reference Treatments

The reference treatments in the analysis were placebo, tirzepatide 5 mg, 10 mg and 15 mg, with the results being presented as treatment relative to tirzepatide 5 mg, 10 mg and 15 mg.

B.2.9.4 Endpoints

The following efficacy endpoints were included in the main analysis of the NMA:

- Change from baseline in HbA1c (%)
- Change from baseline in weight (kg)
- Change from baseline in body mass index (BMI; kg/m^2)
- Change from baseline in low-density lipoprotein (mmol/L)
- Change from baseline in high-density lipoprotein (mmol/L)
- Change from baseline in eGFR (mL/min/1.73 m^2)

The following safety endpoints were also included:

- Change from baseline in systolic blood pressure
- Proportion of patients experiencing nausea

B.2.9.5 Methods

The code used to conduct the NMA is presented in Appendix D.8.

B.2.9.5.1 Analysis population

The analysis population was defined to align with SURPASS-2 and 3 trials, and included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. More specifically, the population included studies including patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin. Trials with an unknown proportion of patients on metformin background therapy and trials including patients on ≥3 OADs were excluded from the main analysis. These trials were included in the sensitivity analyses described in Section B.2.9.7.3 (results presented in Appendix D.8).

B.2.9.5.2 Analysis time window

The duration of dose escalation employed to reach the tirzepatide target dose in the SURPASS trials is longer (0–20 weeks) than the corresponding durations used for the comparators in the comparator

studies, which ranged from 0–12 weeks. In addition, patients in the tirzepatide 15 mg treatment arm in the SURPASS trials received the intended dose (15 mg) for 20 weeks after the dose escalation period. Since most comparator studies had a duration between 22 and 30 weeks (and all comparator studies reported on at least one outcome of interest between 20 and 28 weeks), the endpoints were analysed at 26 ± 4 (22–30) weeks for comparator data, compared to tirzepatide data at Week 40. For SURPASS-4 (inclusion of study as per sensitivity analysis in Section B.2.9.5.3), no visit was conducted at week 40, so data at week 42 was used instead.

The time window of 26 ± 4 weeks allows a balanced approach to utilising data obtained from the dose escalation of tirzepatide and data available from the comparators. It was assumed that the level of response to treatment within 4 weeks of the target week was unlikely to vary considerably, so data for tirzepatide at week 40 may be considered comparable to data at 26 ± 4 weeks in comparator trials. Sensitivity analyses were performed to account for other possible analysis time windows (Appendix D.8).

For AEs, the assessment timepoint used in studies often included a safety follow-up period (4–5 weeks). Moreover, in the SURPASS trials, AEs were assessed in the time interval between baseline and Week 44 (4-week safety follow-up). Therefore, analysis of AEs allowed for the inclusion of comparator studies with safety windows ending outside the analysis window (26 ± 4 weeks).

B.2.9.5.3 Sensitivity analysis

To assess the robustness of the findings of the main analysis of the NMA, sensitivity analyses were planned considering the key variations in study population, network definition, time window reported across trials, and the inclusion of a Phase 2 study. Sensitivity analyses were conducted for the endpoints change from baseline in HbA1c (%), weight (kg) and BMI (kg/m^2). These endpoints were chosen because these are critical clinical endpoints in the management of diabetes and are also important endpoints for the cost-effectiveness analyses of a treatment in T2D. The following sensitivity analyses were conducted:

- Consideration of Asian population studies
- Inclusion of Phase 2 tirzepatide study
- Modification of network definition: Studies including patients with unclear proportion of metformin as background therapy and studies including patients on a background therapy of three OADs (e.g., SURPASS-4) were included in this sensitivity analysis
- Exclusion of studies with insulin glargine as treatment arm
- Analysis time windows
- Different analyses timepoints and windows
- Model-based NMA for continuous outcomes, as described in Pedder 2019[97]
- A meta-regression adjusting for number of OADs for change from baseline in HbA1c (%) and weight (kg) was also conduced

Further details on the methods involved in the sensitivity analyses are presented in Appendix D.8.

B.2.9.5.4 Statistical approach

The NMA was conducted in Just Another Gibbs Sampler (JAGS) version 4.2.0 software via R. A two-stage analytical approach was used for this NMA, as outlined below:

- Frequentist meta-analysis (MA) was conducted to assess heterogeneity and understand the data
- NMA was conducted using Bayesian Mixed Treatment Comparisons as described in the National Institute for Health and Care Excellence Decision Support Unit (NICE DSU) technical support documents (TSD)

Bayesian model specification

The Bayesian NMA models were computed using a Markov chain Monte Carlo (MCMC) simulation method. This method involved drawing samples for each parameter in the model, repeatedly, such that in the long run, the samples of values converge to the posterior distribution. Simulations prior to convergence are ones obtained through the burn-in period, which are then discarded. Additional simulations were run after the burn-in period to ensure that there are no convergence issues in the posterior distributions obtained. Additional adjustment such as increasing the thinning and number of burn-in simulations from the default values highlighted below were applied to address convergence and autocorrelation issues.

If the Bayesian model still did not converge, a frequentist NMA model, based on the method proposed by Rücker 2012,[98] was conducted as part of the sensitivity analysis.

Initial Values:

Markov Chain Monte Carlo Setting:

- Burn-in simulations: 20,000
- Sample: 100,000
- Thin: 5

If autocorrelation was detected in the diagnostic plots, thinning was also used.

Model Specifications:

- Data distribution:
  - “binomial logit” for binary data
  - “normal” for continuous data
- Relative treatment effect model: both “fixed” and “random” (if valid for a specified endpoint) effects were run, one at a time
- Baseline model: “separate model” as per Dias 2013 recommendation[99]

- •Meta-regression, adjusting for OAD, was considered as an exploration of heterogeneity. Additionally, meta-regression, adjusting for analysis time window and baseline covariates, such as HbA1c and weight, resulted in convergence issues
- Missing covariate option: was set up as “exclude studies with missing covariate”. If imputing missing data was deemed necessary, imputation was dealt within the dataset

B.2.9.5.5 Inconsistency and heterogeneity

For tests of heterogeneity, an I^2 statistic of >60% indicated substantial heterogeneity. No concerns regarding heterogeneity were defined as I^2 statistic < 40%, or I^2 statistic 40%–60% and Cochrane Q test p-value > 0.1.

For tests of inconsistency, a DIC for the consistency model lower than the DIC for the inconsistency model (unrelated mean effect model) indicated no concerns with inconsistency. If the DIC for the consistency model was greater than the DIC for the inconsistency model (by a margin of 2), this suggested concerns regarding inconsistency.

Results relating to heterogeneity and inconsistency are reported in Section B.2.9.7.

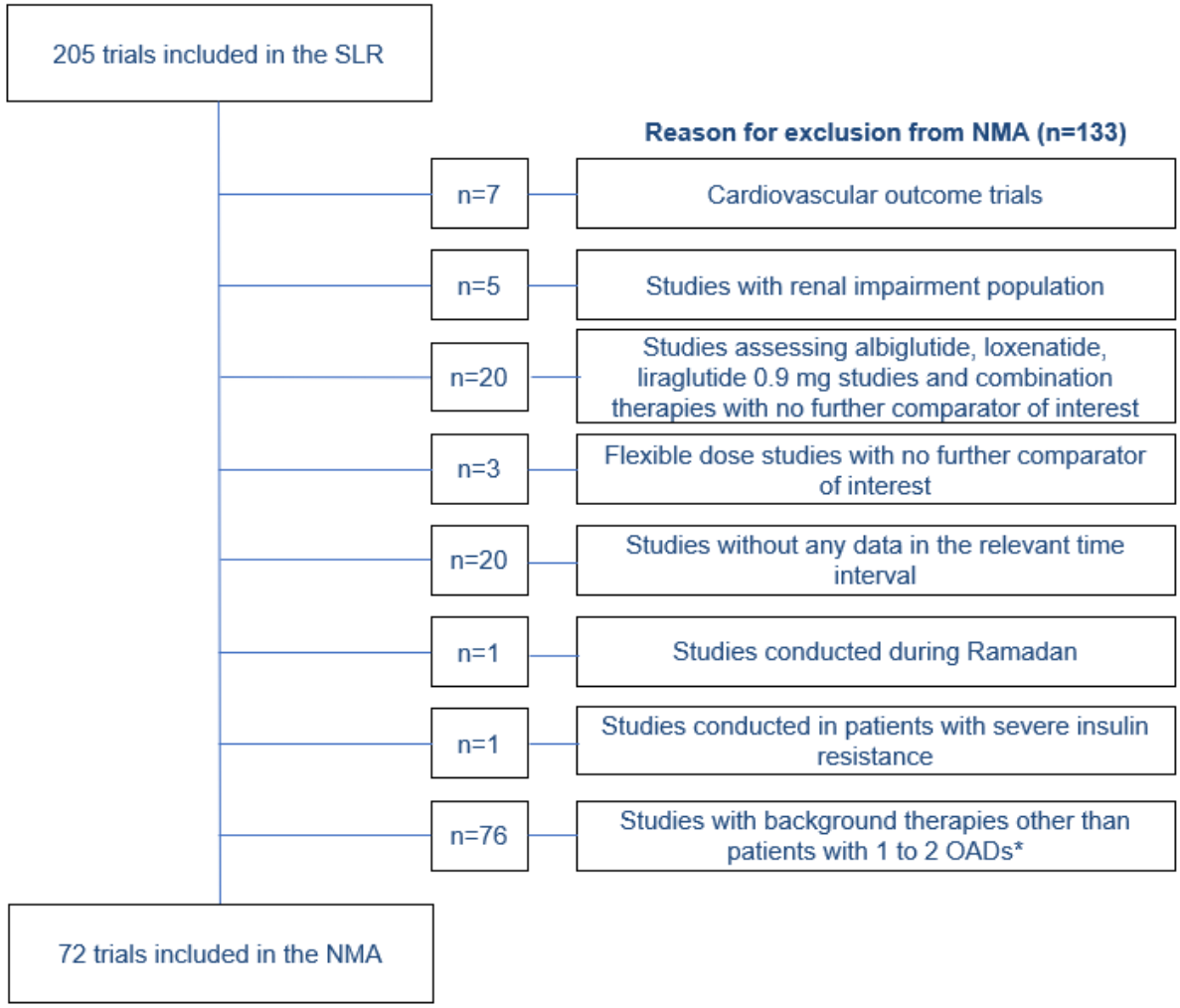
B.2.9.6 Feasibility Assessment

To ensure that the network was consistent in terms of study design and study population, and would therefore provide the most robust networks, an assessment of the studies included in the NMA was conducted. The key decisions relating to the feasibility assessment are summarised below.

Summary of studies included in the NMA

The 205 studies included in the SLR were assessed for inclusion in the NMA of which 133 studies were excluded. The reasons for exclusion of studies are detailed in Figure 3. Thus, a total of 72 studies were eligible for inclusion in the NMA.

Figure 3: Reasons for exclusion of studies in NMA



*Defined as included studies including patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin.
Abbreviations: NMA: network meta-analysis; OAD: oral antidiabetic drug; SLR: systematic literature review.

Rationale to exclude certain studies in NMA

The details of studies or treatment arms within a study that were excluded from the analysis are presented below.

Cardiovascular outcomes trials (CVOTs)

A total of seven CVOTs were identified in the SLR. CVOTs are designed to examine the CV safety and efficacy of anti-hyperglycaemic drugs.[100-106] The CVOTs generally assess the impact of the therapeutic intervention on a set of composite CV endpoints, termed MACE, of CV mortality, non-fatal myocardial infarction (MI), and non-fatal stroke (3P-MACE), and, in some studies, an additional endpoint of hospitalization for unstable angina (4P-MACE)[107]. Since the purpose of CVOTs is to assess the CV effects of antihyperglycemic agents over an extended period of time, CVOTs have been designed with a mean follow-up of at least 14 months. The patient population across CVOTs included patients with T2D, but also included patients with pre-diabetes and glucose intolerance, as

well as patients with established chronic kidney disease (CKD). Furthermore, the background therapies included in the CVOTs include OAD as well as injectables. Moreover, while changes in glycaemic levels are reported in CVOTs, they were not designed to assess glycaemic efficacy. Collectively, these points suggest that CVOTs are not comparable with the SURPASS program or usual glycaemic control trials with regard to study design (study length and background therapy). Consequently, CVOTs were excluded from all analyses.

Renal disease

The SURPASS programme evaluated patients with varying degrees of renal impairment, which has shown no significant effect on overall efficacy or safety results for TZP. However, in order to make the populations from the various studies as generalizable as possible, any renal impairment (e.g. <45 mL/min or S4) cut-offs were excluded from this NMA. Consequently, studies in a specific population of patients with renal impairment (stage 3 or 4 CKD or macroalbuminuria) were excluded from the NMA.[108-112]

Comparators not of interest

In concordance with the assessment of tirzepatide for relevance in UK clinical practice, the focus of this NMA was on the treatment options that are available in that setting. Among the treatments that were considered eligible for the SLR, albiglutide, loxenatide and liraglutide 0.9 mg are not available in the UK. Furthermore, oral semaglutide 3.0 mg is not available as a maintenance dose in the UK. Consequently, studies that included albiglutide (all strengths), loxenatide (all strength), liraglutide 0.9 mg, or oral semaglutide 3.0 mg, were excluded completely from the NMA if there were no other comparators of interest included.

Flexible dose

Three studies assessed the glycaemic impact of flexible doses of liraglutide, oral semaglutide and exenatide in patients with T2D.[113-115] Although these studies were included in the SLR, none of the maintenance doses approved in the UK include a flexible dose. The flexible dosing employed in these studies also prevent the direct comparison of these studies with the SURPASS program. Furthermore, these studies did not include any treatment arms that were relevant to the aim of comparing these studies to the SURPASS program. Consequently, these three studies were excluded from the NMA.

No data in the analysis interval

Studies not reporting any data on the relevant endpoints in the analysis time interval were excluded from the NMA.

Combination therapy

The EXENDA study was designed to compare the impact of a combination of exenatide and dapagliflozin versus a combination of dapagliflozin and placebo.[116] In this study, patients were randomized to receive dapagliflozin at baseline, and this, dapagliflozin, acts as a study treatment and not background therapy. This study design makes EXENDA dissimilar to the study design of the SURPASS trials and consequently, the EXENDA study was not included in the NMA.

Studies conducted during Ramadan

The LIRA-RAMADAN study was conducted to evaluate the effect of a combination of liraglutide 1.8 mg or sulfonylureas with metformin in patients with T2D during the season of Ramadan.[117] The primary endpoint evaluated the change in fructosamine, which is a short-term biomarker of glycaemic control. The season of Ramadan involves atypical dietary patterns with individuals fasting during the day and feasting after sunset. Thus, the results obtained from participants of this study may not be comparable to the SURPASS program or other glycaemic control studies, especially with regards to weight, BMI and HbA1c change. Consequently, this study was not included in the NMA.

Patients with severe insulin resistance

Distiller, 2014,[118] evaluated the efficacy of a combination of U-500 and metformin versus a combination of U-500, metformin, and exenatide in patients who were severely insulin-resistant. Moreover, the patient population in the study also demonstrated very high BMI. Collectively, the very high insulin dose and severe insulin resistance and high BMI among the participants prevent the comparison of this study with other studies on glycaemic control. As a result, the intervention and comparator arm do not fit the eligibility criteria specified in our PICOTS, therefore the study was excluded from the NMA.

Rationale to include certain studies in NMA

Studies with 2 arms of lixisenatide 20 µg[119, 120]

Two studies included two treatment arms of same dose lixisenatide 20 µg. The treatment arms data needed to be pooled; however, individual patient level data were not available in either of the publications. Therefore, a decision regarding the selection of one treatment arm for each of these two trials was made and described below:

- GetGoal F1[119] includes one lixisenatide 20 µg arm with one-step dose escalation (10 µg once daily for 2 weeks, then 20 µg once daily) and another arm with two steps dose escalation (10 µg once daily for 1 week, 15 µg once daily for 1 week, then 20 µg once daily). The one step treatment arm in this trial was considered in the analysis as this is the dose escalation scheme described in the EMA label.
- GetGoal M includes one lixisenatide 20 µg arm with a morning injection and one arm with an evening injection.[120] Morning injection treatment arm was included in analysis as breakfast time was considered as the most common use of the injectable (most likely that patients would inject in the morning).

Comorbidities

Studies including patients with comorbidities such as CVD/CV high risk, obesity, non-alcoholic fatty liver disease, and other comorbidities were included in the NMA.

Baseline characteristics

Studies were included in the NMA irrespective of baseline mean age, female proportion, mean BMI, mean body weight, diabetes duration, mean HbA1c and trial duration (>104 weeks).

Characteristics of included studies

A total of 72 studies were included in the NMA. Mean age in treatment arms of the included studies ranged from 42.7 years to 63.8 years. Mean HbA1c ranged from 7.4% to 10.3%. Mean body weight ranged from 67 kg to 101.9 kg. A total of 27 studies included greater than 50% of females, proportion of females ranged from 22.2% to 65.2%. Duration of diabetes ranged from 0.63 years to 9.9 years in two studies. Treatment duration ranged from 24 weeks in 11 studies to 156 weeks in 2 studies. The baseline characteristics across the included study arms are summarised in the sections below.

Trial design characteristics

A summary of key trial design features for each of the 53 studies included in the main analysis are presented in Table 47 and Table 48, as well as Figure 24 to Figure 26. Based on the available data, key trial design features were largely consistent across the included studies. The majority of studies were single, double or triple-blind (29/53; 55%), although a large minority were open-label (22/53, 42%).

Table 47: Summary of type of background therapy received across included studies

Background therapy received	Number of studies
Metformin monotherapy	25
Metformin alone or metformin + SU	15
Metformin, SU, glitazones	1
Metformin alone or metformin + glitazones	6
Metformin alone or metformin + SGLT-2 inhibitors	4
SGLT-2i alone or SGLT-2i + metformin or SU	1
Metformin alone or metformin + SU, DPP-4i, SGLT-2i or glinides	1

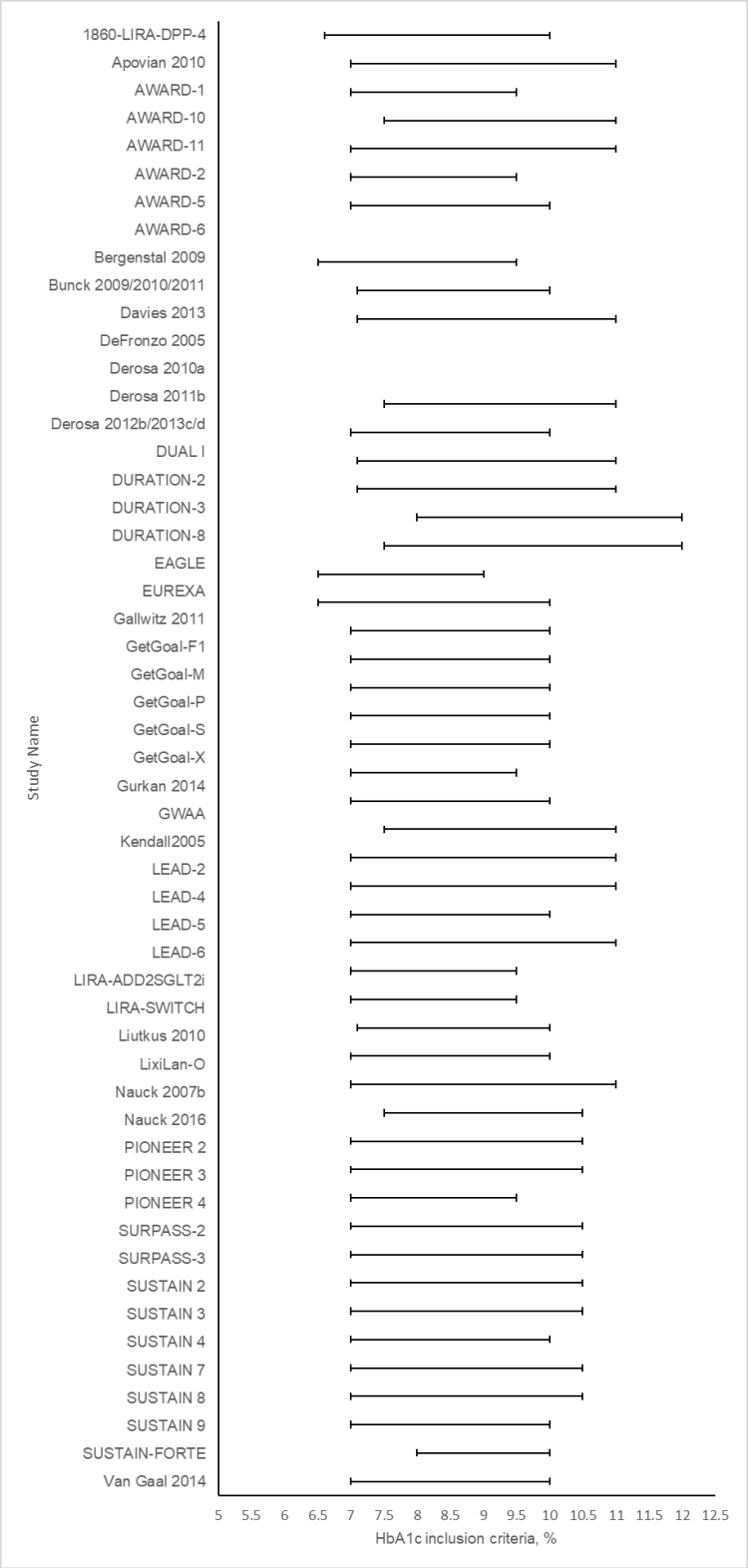
Abbreviations: DPP-4 dipeptidyl peptidase-4 inhibitors; SU: sulphonyl urea; SGLT-2i: sodium glucose cotransport-2 inhibitors.

Table 48: Summary of blinding status across included studies

Blinding status	Number of studies
Single-blind	2
Double-blind	26
Triple-blind	1
Open label	22
Mixed*	1
Not reported	1

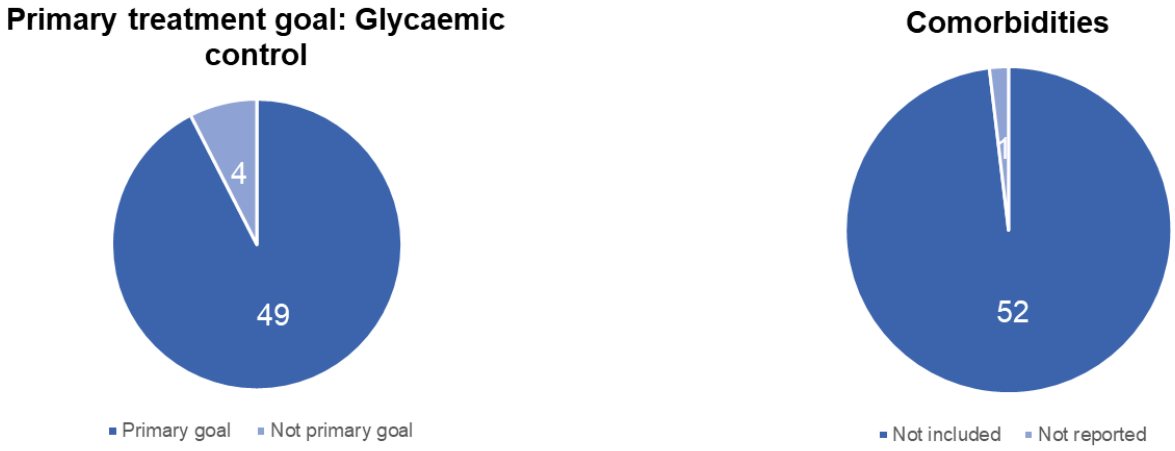
Footnotes: *Mixed trials included both double-blind and open-label design.

Figure 24: Summary of lower and upper bound for HbA1c inclusion criteria in each study included in the main analyses



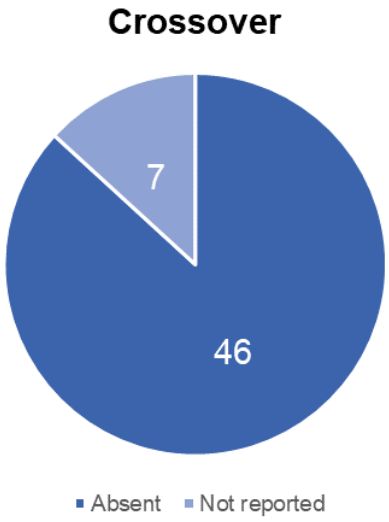
Footnotes: Derosa 2010, Derosa 2011b and Bergenstal 2011 did not report an upper limit for HbA1c. As such, these studies are not presented on the above figure. The lower HbA1c limit for these 3 trials was 7%.
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 25: Summary of primary treatment goal (glycaemic control) and comorbidities included in each study included in the main analysis



Footnotes: The two treatment arms that included comorbidities specified 'Obese'.

Figure 26: Summary of crossover in each study included in the main analysis



Baseline characteristics

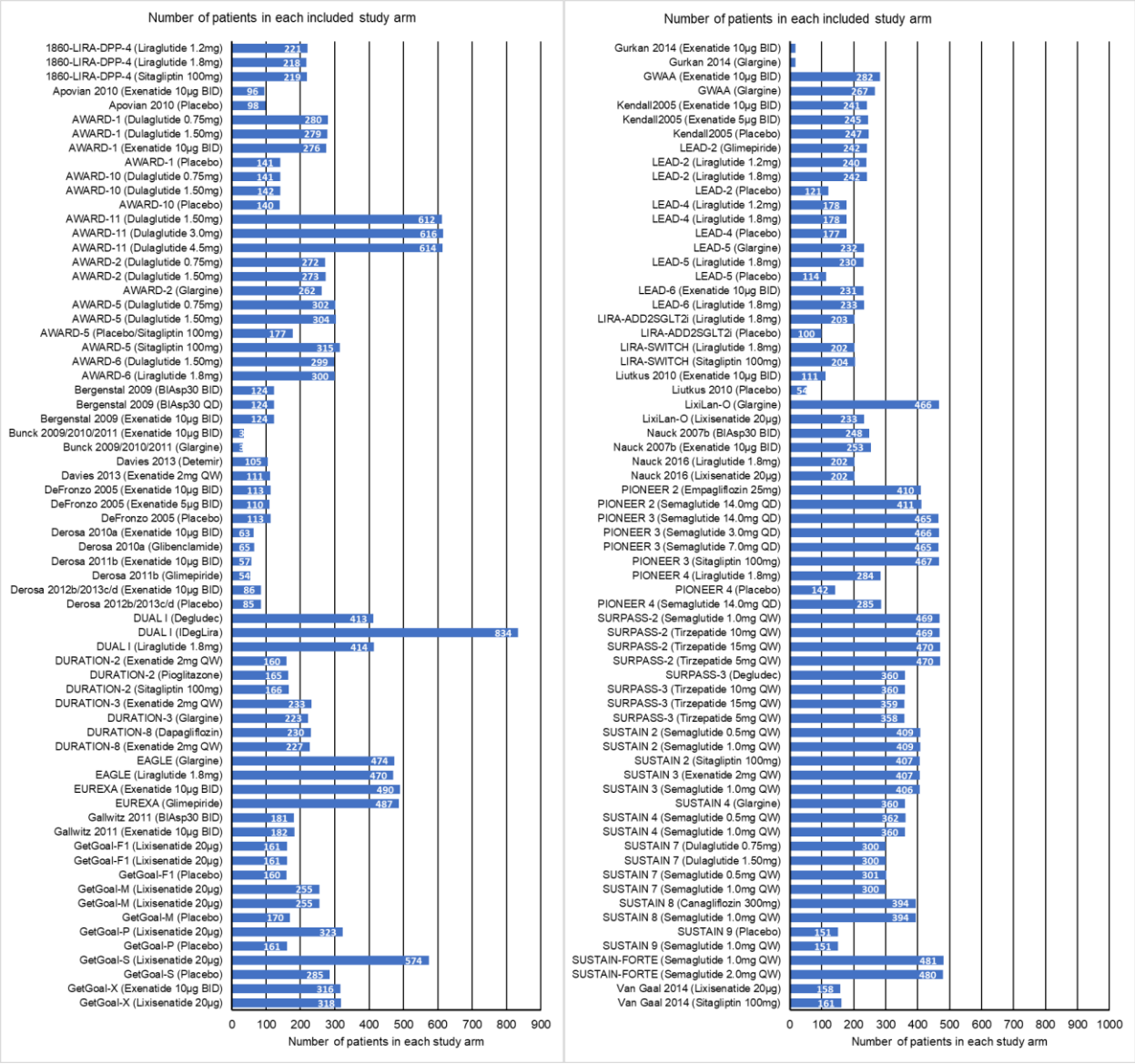
A summary of the baseline characteristics for each of the study arms included in the NMA are presented in Figure 27 to Figure 34 below, while a summary of the mean and ranges across the baseline characteristics of the included studies is presented in Table 49. Baseline characteristics were largely consistent across the included treatment arms. Baseline HbA1c, baseline weight and number of OADs were identified as potential treatment effect modifying variables and therefore, a meta-regression was conducted where feasible to adjust for these (Section B.2.9.5.3).

Table 49: Summary of baseline characteristics across the study arms

Baseline characteristics	Mean value	Minimum value	Maximum value
Number of patients	264.7	17.0	834.0
Proportion of female patients, %	47.4	31.0	70.0
Mean age, years	55.9	42.7	59.8
Mean baseline weight, kg	91.9	80.2	101.9
Mean baseline BMI, kg/m^2	32.75	28.4	36.8
Mean baseline HbA1c, %	8.3	7.4	10.3
Mean baseline duration of diabetes, years	7.6	0.6	10.1
Mean treatment duration, weeks	46.5	24.0	156.0

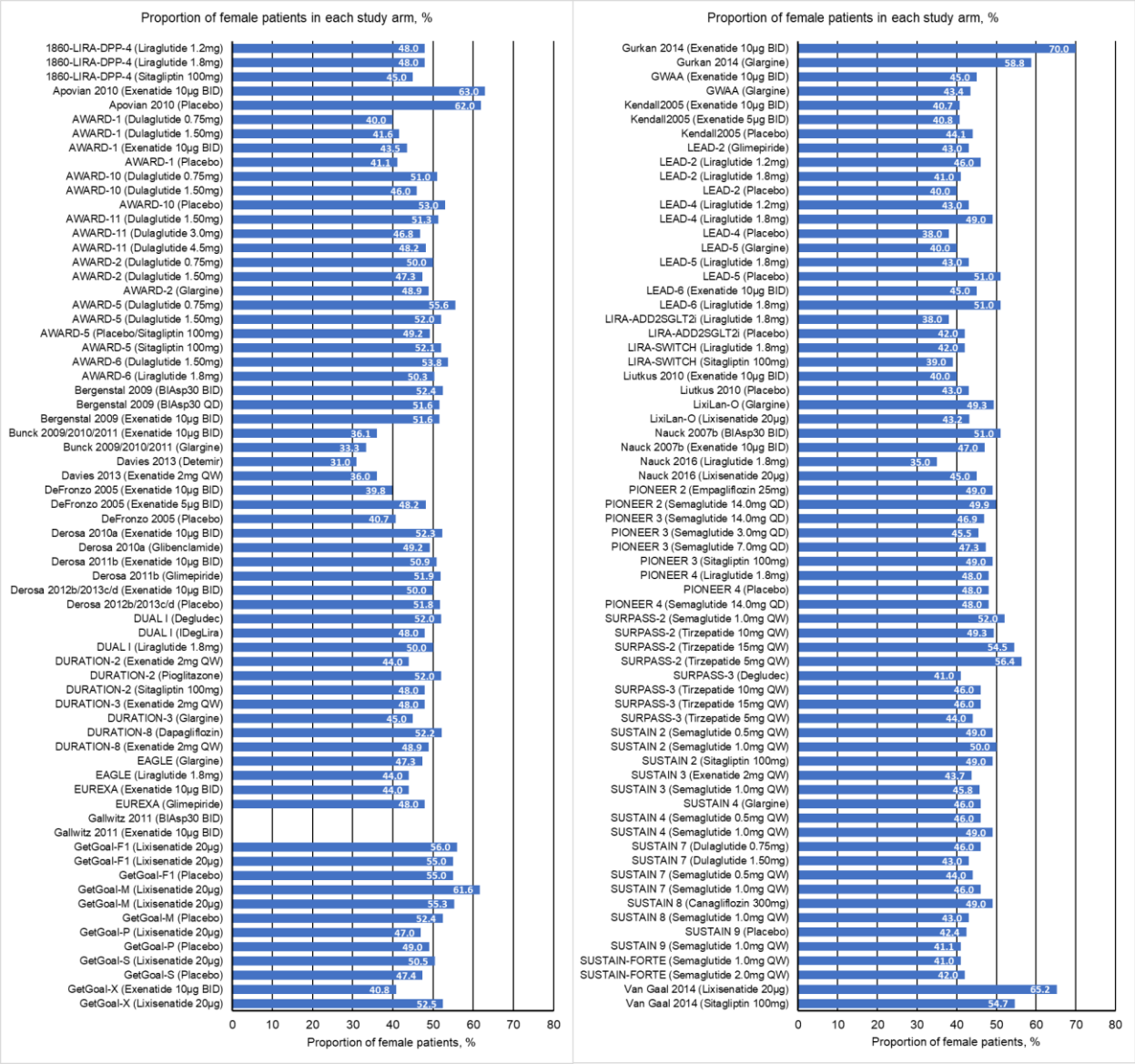
Abbreviations: BMI: body mass index; HbA1c: haemoglobin A1c.

Figure 27: Summary of the number of patients included in each study arm included in the main analysis



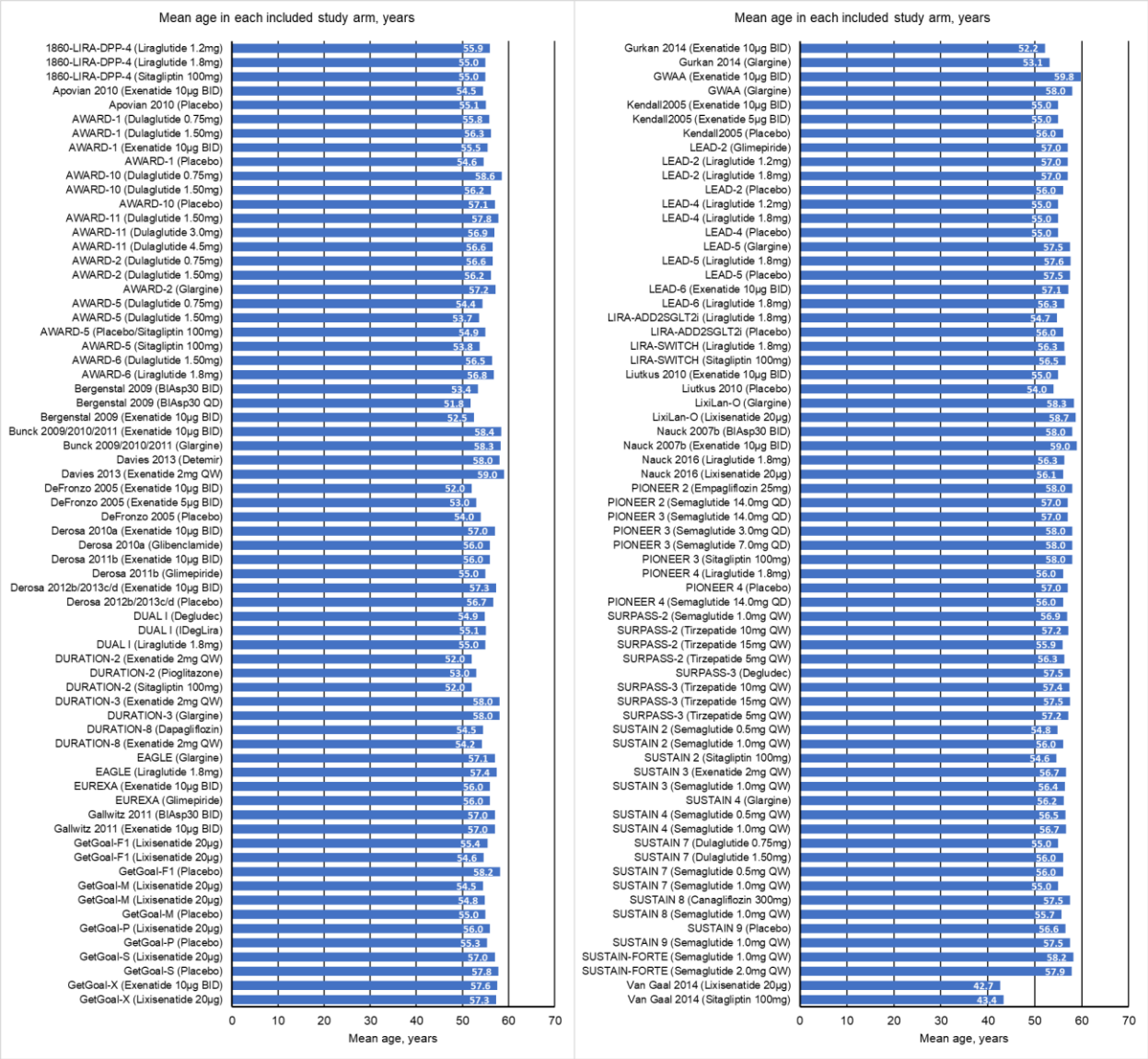
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 28: Summary of the proportion of female patients in each study arm for each study included in the main analysis



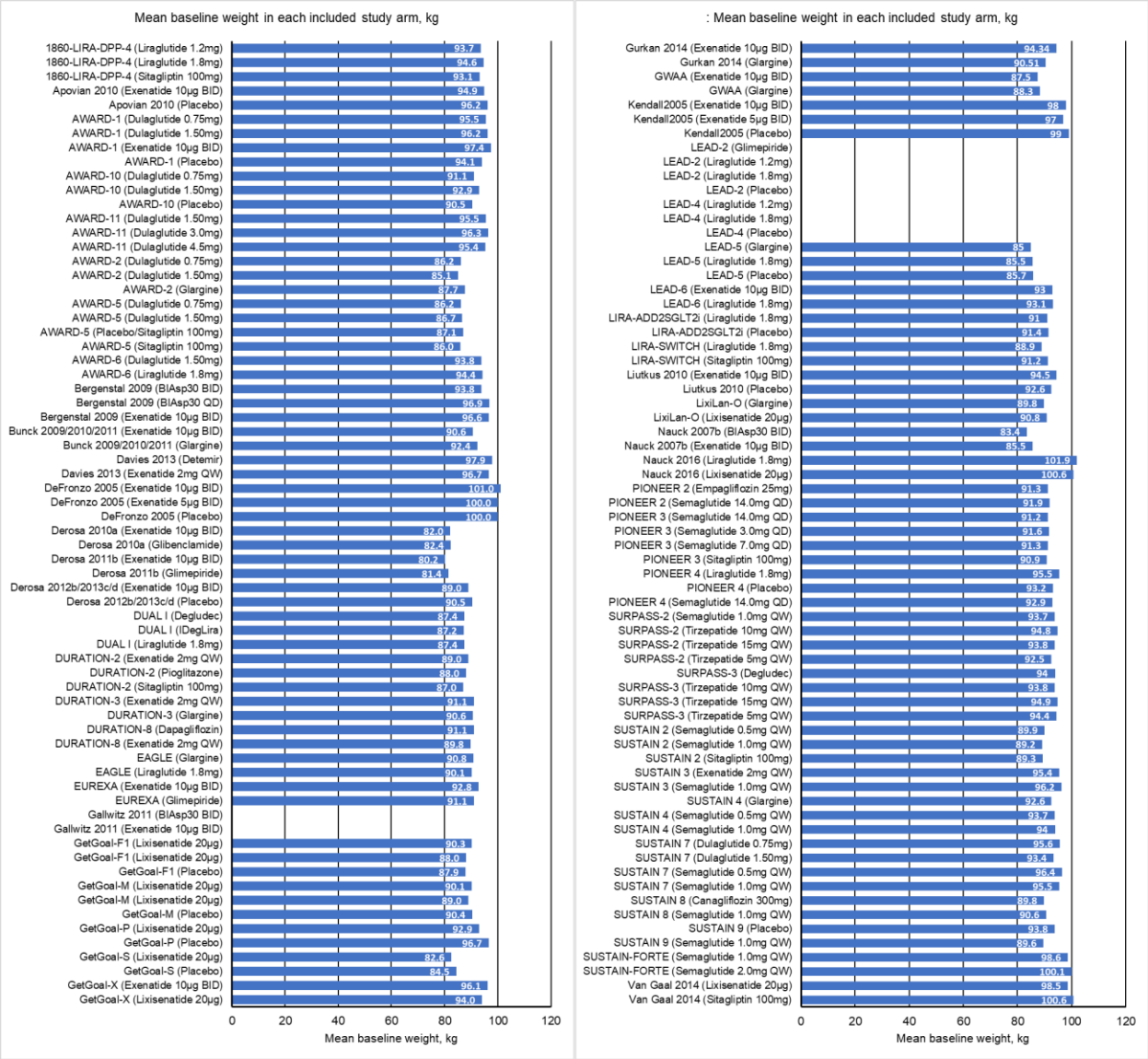
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 29: Summary of the mean age (years) in each study arm for each study included in the main analyses



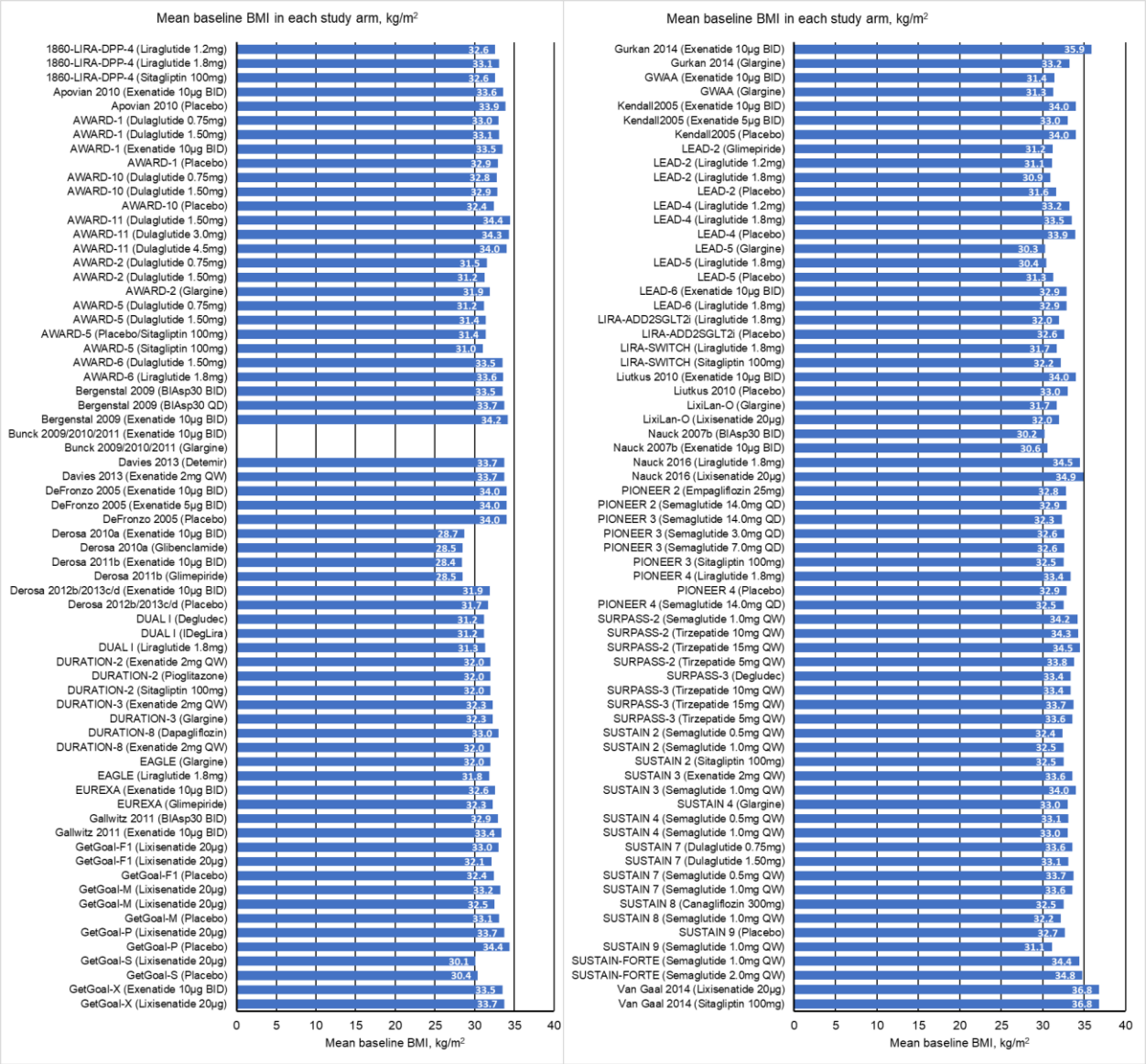
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 30: Summary of the mean baseline weight (kg) in each study arm for each study included in the main analyses



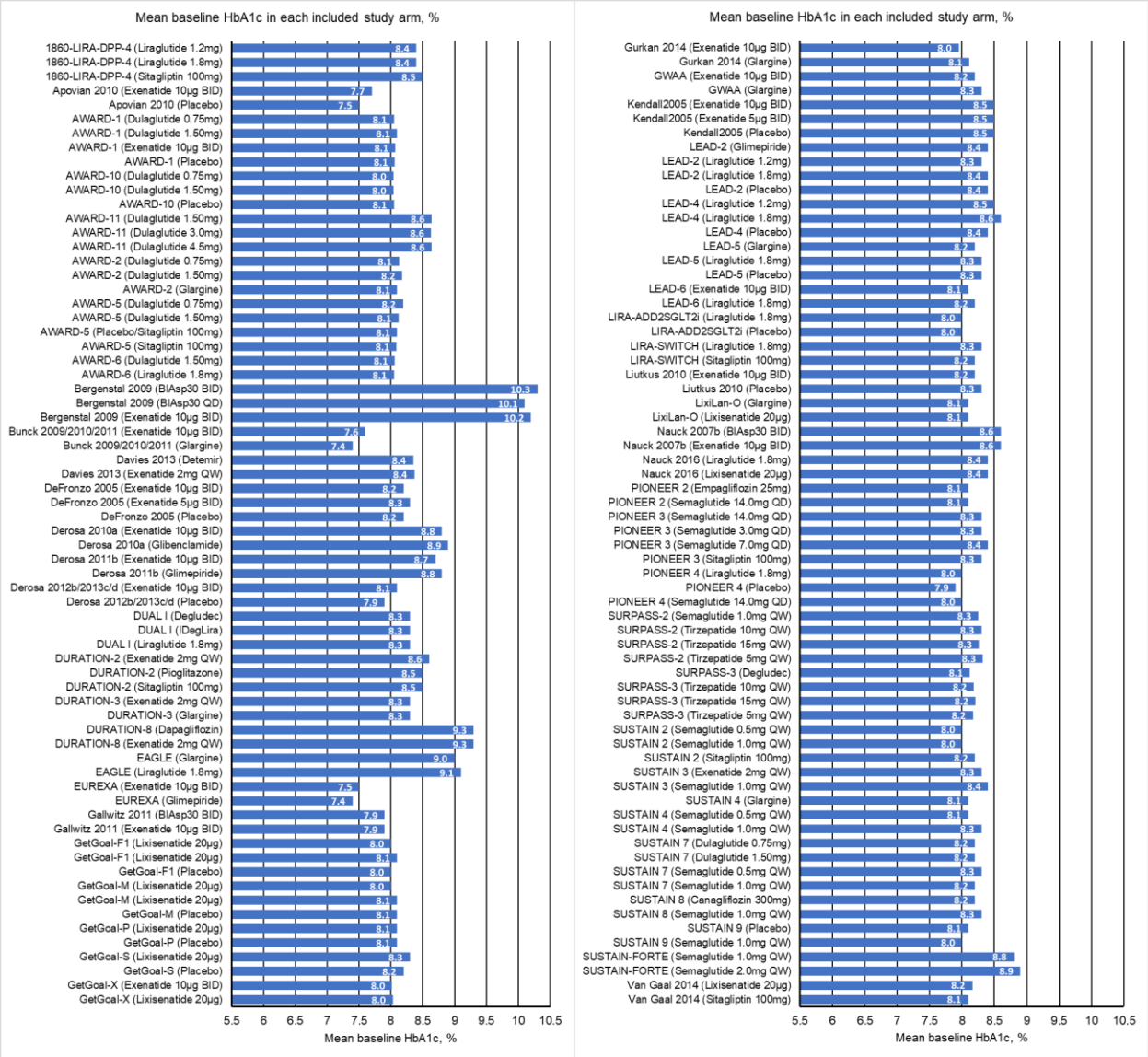
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 31: Summary of the mean BMI (kg/m^2) in each treatment arm for each study included in the main analyses



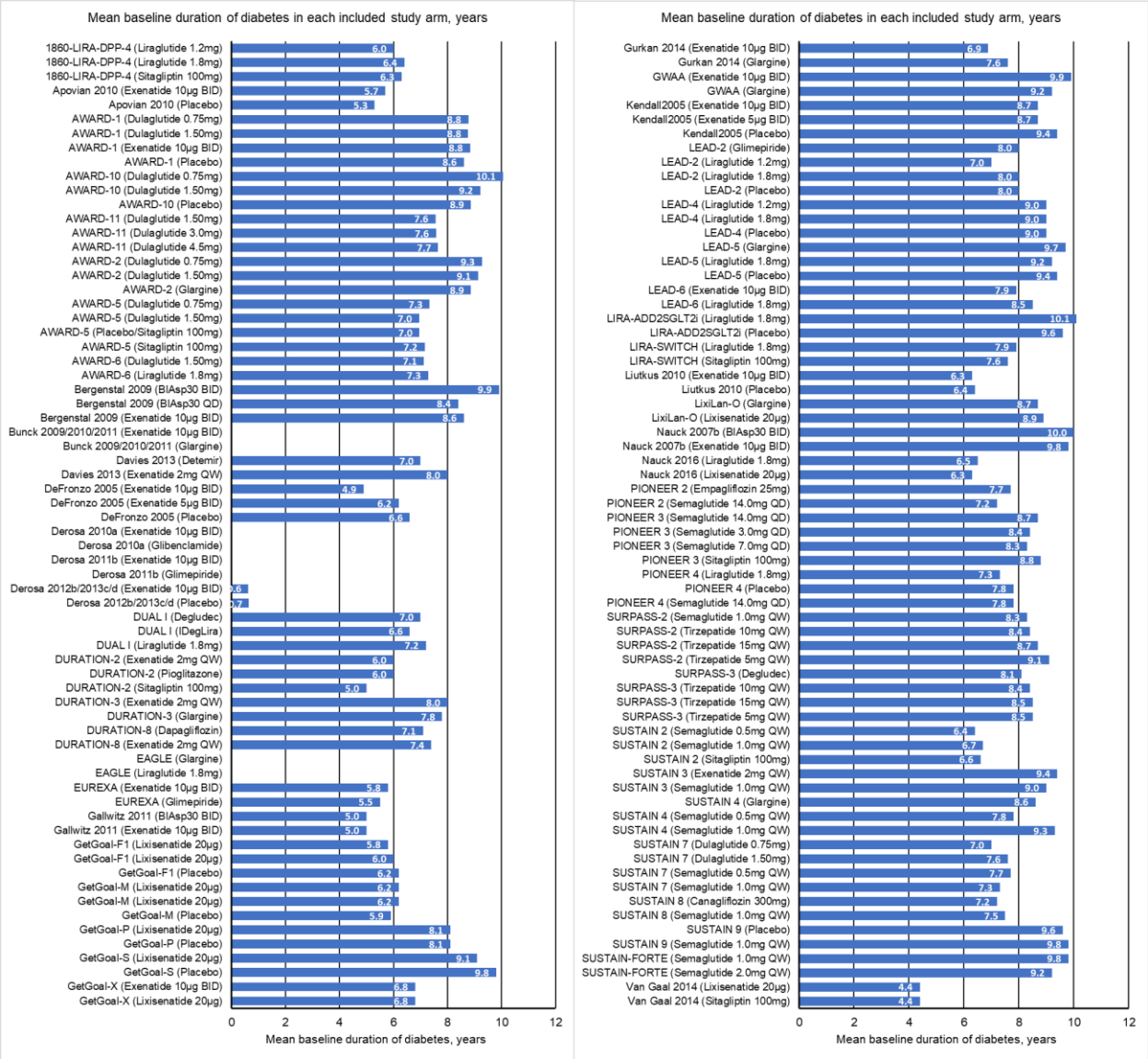
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 32: Summary of the mean baseline HbA1c (%) in each study arm for each study included in the main analyses



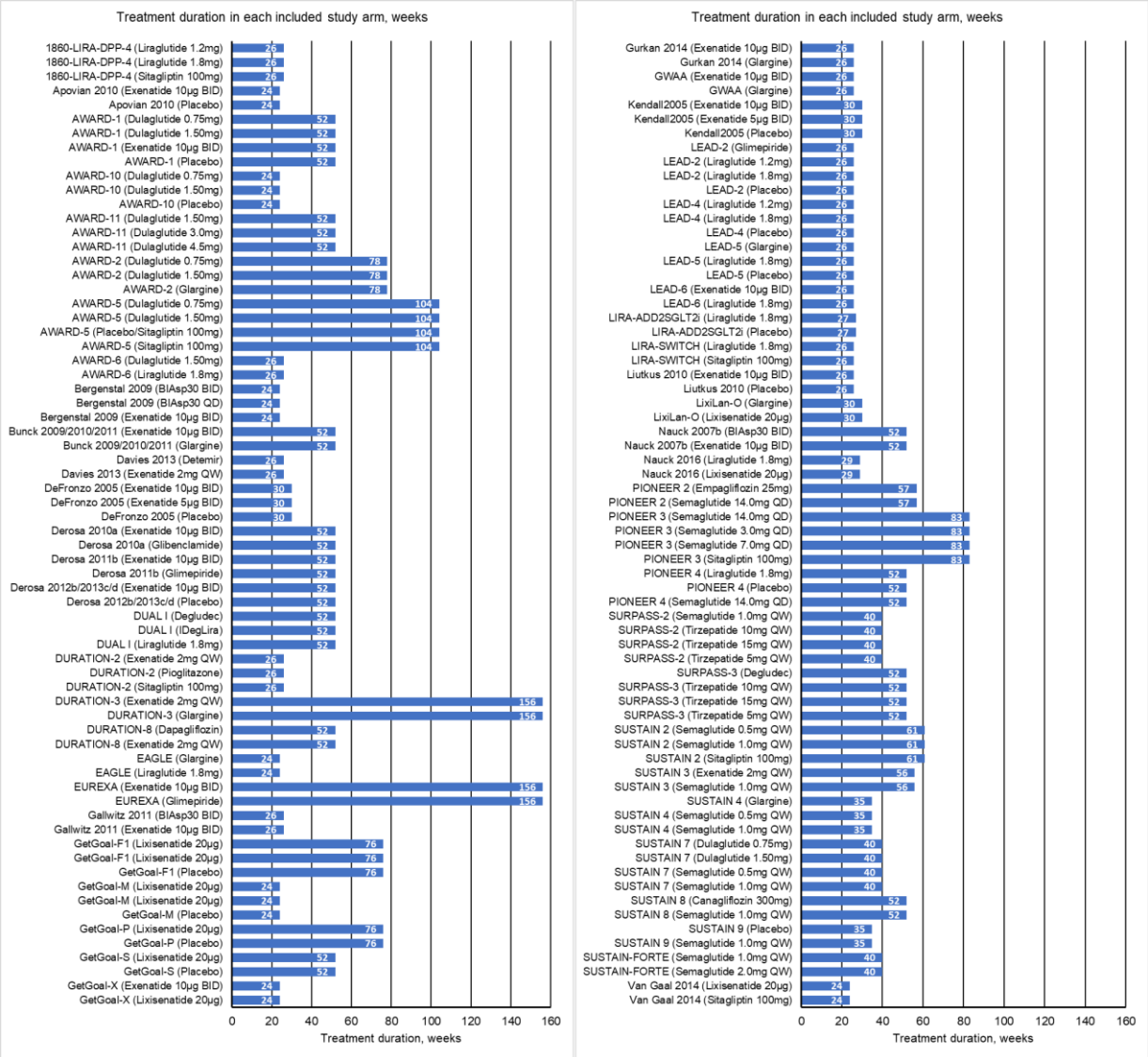
Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 33: Summary of the mean baseline duration of diabetes (years) in each study arm for each study included in the main analyses



Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Figure 34: Summary of the treatment duration (weeks) in each study arm for each study included in the main analyses

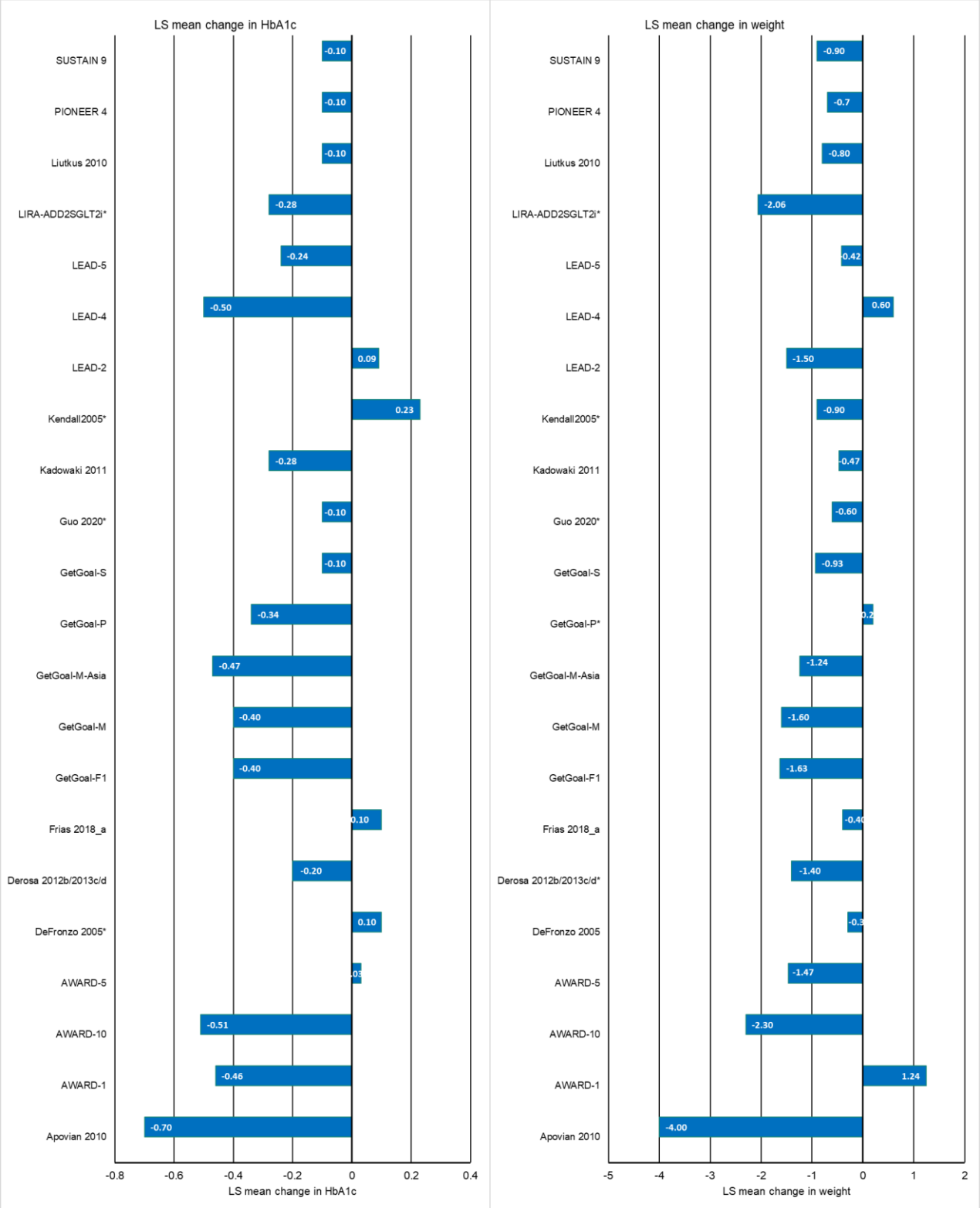


Abbreviations: BID: twice daily; QD: once daily; QW: once weekly.

Placebo outcomes

A summary of (LS) mean change in HbA1c and weight for placebo treatments arms in the main analyses are presented in Figure 35. (LS) mean change from baseline in HbA1c demonstrated some variation across placebo arms, ranging from −0.70% (Apovian 2010) to 0.23% (Kendall 2005). (LS) mean change from baseline in weight also demonstrated some variation across placebo arms, ranging from −4.00 kg (Apovian 2010) to 1.24 kg (AWARD-1). Such variation is likely to reflect the different combination therapies used in the included trials. For example, in Apovian (2010) all treatments were combined with starting a lifestyle modification programme which would be expected to have a positive effect on T2D alone, however, as the NMA is based on relative treatment effects, such differences were not judged to invalidate the assumption of treating all placebo arms as one node.

Figure 35: Summary of placebo outcomes for HbA1c and weight for treatment arms

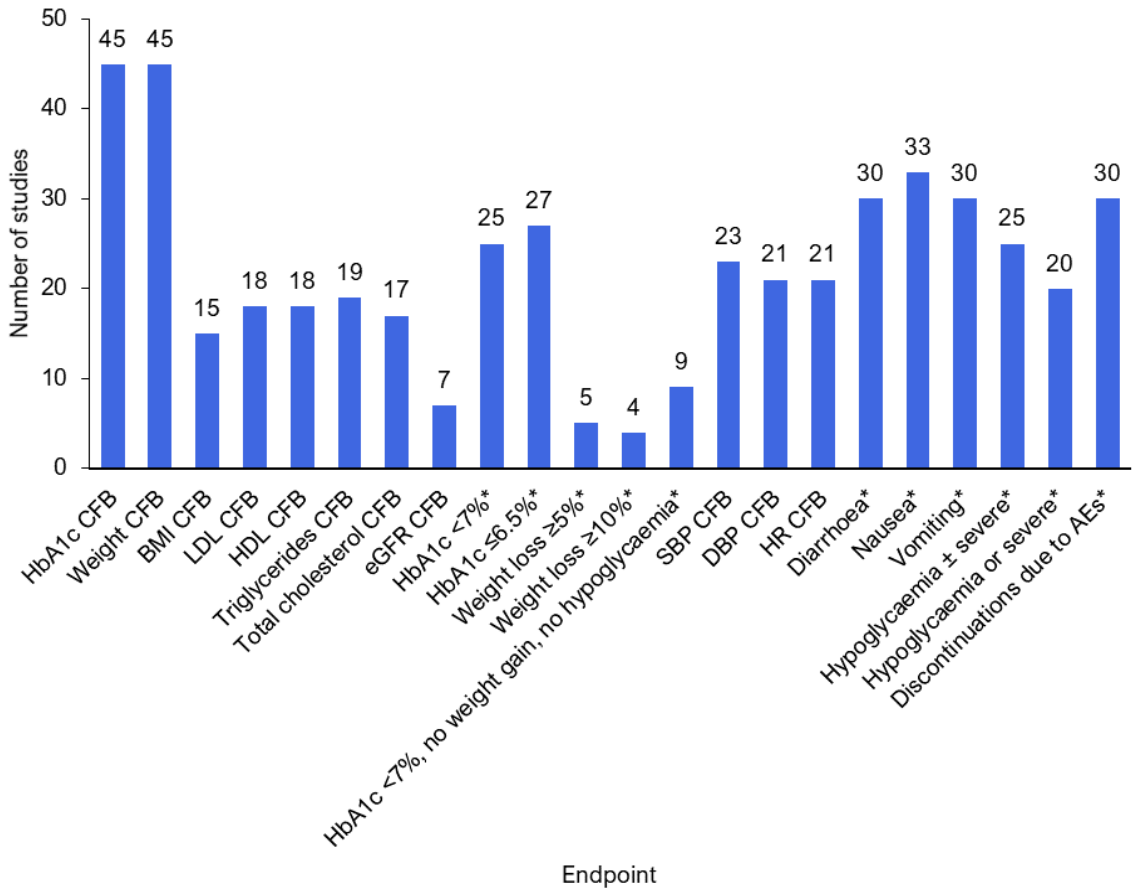


*LS mean data were not reported in the publication, so mean data are presented.
Abbreviations: LS: least squares.

B.2.9.6.1 Availability of Endpoint Data

The number of studies included in the analysis according to endpoint is shown in Figure 36.

Figure 36: Graph showing number of studies included in NMA according to endpoint



*Proportion of patients experiencing event.
Abbreviations: AE: adverse event; BMI: Body Mass Index; CFB: change from baseline; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HbA1c: haemoglobin A1c; HR: heart rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; TC: total cholesterol; TG: triglycerides.

B.2.9.7 Results

Interpretation

Both continuous (change from baseline in HbA1c, body weight, BMI, etc.) and binary endpoints (proportion of patients experiencing nausea, proportion of patients with at least one episode of hypoglycaemia with BG <54 mg/dL [3.0 mmol/L] or severe hypoglycaemia, etc.) were assessed in the NMA; these endpoints were interpreted differently.

For the binary endpoints, odds ratios (ORs) were estimated in each analysis. The OR represents the increase or decrease in the odds of an event occurring in one group compared with another. An OR >1 indicates greater odds for the treatment arm compared to the control arm and similarly an OR

between 0 and 1 indicates a reduction in odds for the treatment arm compared to the control arm. Where both the upper and lower bounds of the credible intervals (CrI) around the OR are either >1 or less than <1, a significantly greater increase or reduction, respectively, in the odds of the event for the treatment arm compared to the control arm is indicated. Where the CrI cross 1, this indicates a lack of statistically significant difference in odds between the two arms.

For the continuous endpoints, standardized median differences and 95% CrI were estimated for each treatment versus placebo and comparators. Median differences below 0 indicate greater reduction in the outcome with the treatment versus the comparator; values above 0 indicate lower reduction in the outcome with the treatment versus the comparator.

Results from an NMA cannot be interpreted in the same way as results from a clinical trial, because an NMA synthesizes both direct and indirect effects. Within the Bayesian framework NMA, significance of a treatment effect is determined by the 95% credible interval (CrI), which represents a 95% probability that the true treatment effect lies within this interval. This interpretation differs in the Frequentist NMA, where the 95% confidence interval can be interpreted as follows: if the analysis was repeated many times, in the long run there is a 95% probability that the true value lies within the 95% confidence interval.

Model convergence

For each analysis in which auto-correlation or poor convergence was observed, the number of simulations was increased; thus, the number of simulations differed between analyses. However, this did not always affect the median posterior estimates. The main impact of increasing the number of simulations were on the CrIs for each parameter. The number of simulations for each analysis is provided in Table 50.

Table 50: Number of simulations and parameter updates for main analysis for each endpoint

Endpoint	Random effects model			
	Thinning	Burn-in simulations	Sampling	Prior distribution
HbA1c (%) change from baseline	50	100,000	1,500,000	Uniform
Weight (kg) change from baseline	30	100,000	1,000,000	Uniform
Body Mass Index (kg/m^2) change from baseline	5	20,000	100,000	Uniform
Low-density lipoprotein (mmol/L) change from baseline	30	100,000	1,000,000	Inverse Gamma
High-density lipoprotein	30	100,000	1,000,000	Uniform

(mmol/L) change from baseline				
Estimated glomerular filtration rate (mL/min/1.73 m^2) change from baseline	50	100,000	1,000,000	Uniform
Systolic blood pressure (mmHg) change from baseline	30	100,000	1,000,000	Uniform
Proportion of patients experiencing nausea (any grade permitted)	30	100,000	1,000,000	Inverse Gamma

Abbreviations: HbA1c: haemoglobin A1c.

Two chains of initial values were run for each analysis. The number of burn-in simulations was observed to be adequate to ensure that the choice of initial value did not affect the posterior estimates. The choice of initial value only had a substantive impact on the models for which the chains themselves did not converge (i.e., model that showed poor convergence). These models have been described below within each analysis section.

Goodness of fit statistics

The reference treatment for the analysis was placebo. Goodness of fit statistics are provided in Table 51. The residual deviance and Deviance Information Criterion (DIC) values were observed to be lower or similar for the random effects compared with fixed effects model for all the endpoints. Hence, random effects models were selected for presentation of the results in all endpoints except for the proportion of patients reaching weight loss ≥5% and ≥10%, where fixed effects models were chosen.

Table 51: Goodness of fit statistics for all endpoints

Endpoint	Data points (N)	Residual deviance		DIC	
		Fixed-effects	Random-effects	Fixed-effects	Random-effects
HbA1c (%) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weight (kg) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Body Mass Index (kg/m^2) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Low-density lipoprotein (mmol/l) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
High-density lipoprotein (mmol/L) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Endpoint	Data points (N)	Residual deviance		DIC	
		Fixed-effects	Random-effects	Fixed-effects	Random-effects
Estimated glomerular filtration rate (mL/min/1.73 m^2) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Systolic blood pressure (mmHg) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Proportion of patients experiencing nausea (any grade permitted)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: DIC: Deviance Information Criterion; HbA1c: haemoglobin A1c.

For continuous outcomes, the pairwise results for the random effects model are presented in terms of the standardised median differences and 95% CrI for each tirzepatide dose versus placebo and comparators. For binary outcomes, the pairwise results are presented in terms of the OR and 95% CrI for each tirzepatide dose versus placebo and comparators.

Studies and treatment arms excluded from all analyses

A total of 53 studies were eligible for inclusion in the main analysis of the NMA. In eight of these, studies, a GLP-1 was compared to a treatment that did not connect with other treatments besides the study in question, so these studies did not inform the network. The comparators in these eight studies were not considered as treatments of interest (e.g., insulins, DPP-4, SGLT-2i) and hence were excluded for all endpoints in the network. In addition, the pioglitazone arm was removed from DURATION-2 study and semaglutide 3.0 mg QD arm from PIONEER 3 study because they were not treatments of interest. Other treatment arms of these two studies were included in the analysis. After removing these studies and treatment arms, a maximum of 45 studies and 23 treatments were included in the NMA.

Treatments such as insulin glargine, insulin degludec, sitagliptin and glimepiride were used as nodes to connect other treatments and inform the network. These were not considered as treatments of interest and were not included in the PICOTS. Hence the results for these treatments shown in the tables and figures should not be interpreted.

Inconsistency and heterogeneity outcomes

For continuous variables in the network, there was substantial heterogeneity in at least one of the relative comparisons for each outcome. For change from baseline in HR, LDL and total cholesterol, SURPASS-2 and/or SURPASS-3 trials contributed to the heterogeneity. For change from baseline HbA1c, weight, SBP and HDL, comparisons between placebo and other treatments (apart from any tirzepatide doses) contributed to the heterogeneity. The comparisons contributing to heterogeneity observed in continuous variables in the network are presented in Table 52; any endpoints not reported in Table 52 demonstrated no concerns with heterogeneity. No concerns with inconsistency were identified for any continuous variables.

Table 52: Summary of heterogeneity in continuous variables

Outcome	Comparison contributing to heterogeneity

HbA1c	• Placebo versus Exenatide 10 mcg BID • Placebo versus Exenatide 5 mcg BID • Placebo versus Dulaglutide 0.75 mg • Placebo versus Dulaglutide 1.5 mg • Glargine versus Liraglutide 1.8 mg
Weight	• Placebo versus Exenatide 10 mcg BID • Placebo versus Dulaglutide 1.5 mg • Liraglutide 1.8 mg versus Glargine • Dulaglutide 0.75 mg versus Dulaglutide 1.5 mg
BMI	• Placebo versus Dulaglutide 1.5 mg • Placebo versus Liraglutide 1.8 mg • Dulaglutide 1.5 mg versus Dulaglutide 0.75 mg
SBP	• Placebo versus Exenatide 10 mcg BID • Placebo versus Liraglutide 1.2 mg
HDL	• Placebo versus Dulaglutide 0.75 mg
LDL	• TZP 5mg versus TZP 10 mg • TZP 5mg versus TZP 15 mg

Abbreviations: BMI: body mass index; HR: heart rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; TZP: tirzepatide; SBP: systolic blood pressure.

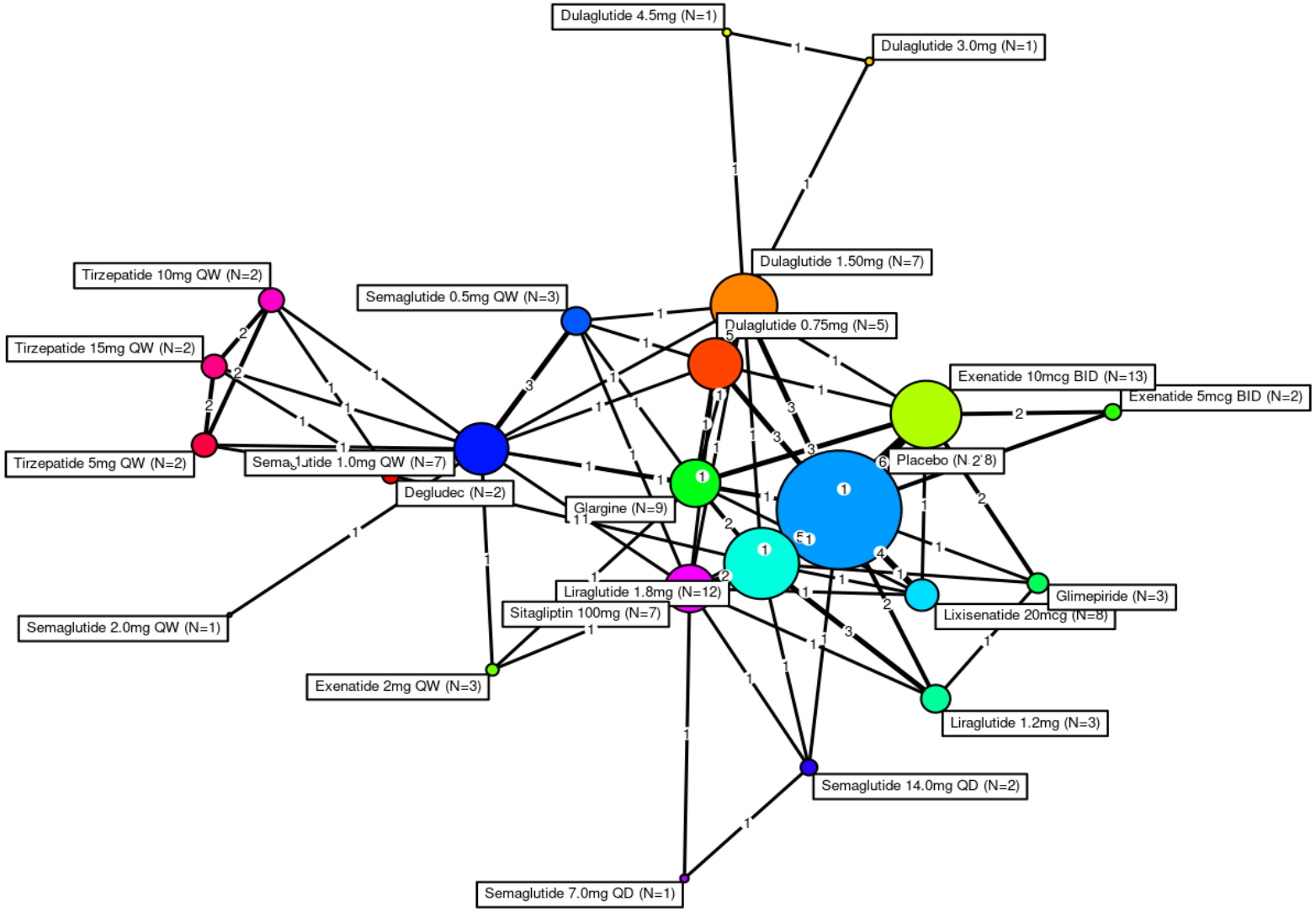
For binary variables, no concerns with heterogeneity based on I^2 statistics or Cochrane Q test were identified for the majority of endpoints, except for the analysis of nausea. For the nausea outcome, studies informing the comparison between tirzepatide 5 mg and 10 mg doses (based on SURPASS-2 and 3) and between liraglutide 1.8 mg and sitagliptin 100 mg were found to demonstrate concerns with regards to heterogeneity. No heterogeneity was identified for the comparison of tirzepatide 10 mg with tirzepatide 15 mg. No concerns regarding inconsistency were identified for any binary endpoints.

B.2.9.7.1 Main analysis results: Efficacy

HbA1c (%) Change from Baseline

The main analysis network diagram for HbA1c (%) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 37. The thickness of the lines indicates the number of studies comparing between the interventions, and the radius of the circle shows the number of studies within a given treatment arm. All 45 studies and 23 treatments (nodes) were included for this analysis.

Figure 37: Main analysis network for HbA1c (%) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Tirzepatide 5 mg showed significantly greater reductions in HbA1c (%) change from baseline compared with placebo and all GLP-1 RAs at the lowest recommended maintenance dose.

Tirzepatide 10 mg showed significant greater reductions in HbA1c (%) change from baseline compared with placebo, and all GLP-1 RAs at the intermediate recommended maintenance dose.

Tirzepatide 15 mg showed significantly greater reductions in HbA1c (%) change from baseline compared with placebo, and all GLP-1 RAs at the highest recommended maintenance dose.

Table 53: Pairwise results (median difference [95% CrI]) for HbA1c (%) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 g	[redacted]	[redacted]	[redacted]
Semaglutide 7.0 mg QD	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 38, Figure 39 and Figure 40, respectively.

Figure 38: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 39: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 10 mg versus comparators, random effects model

Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

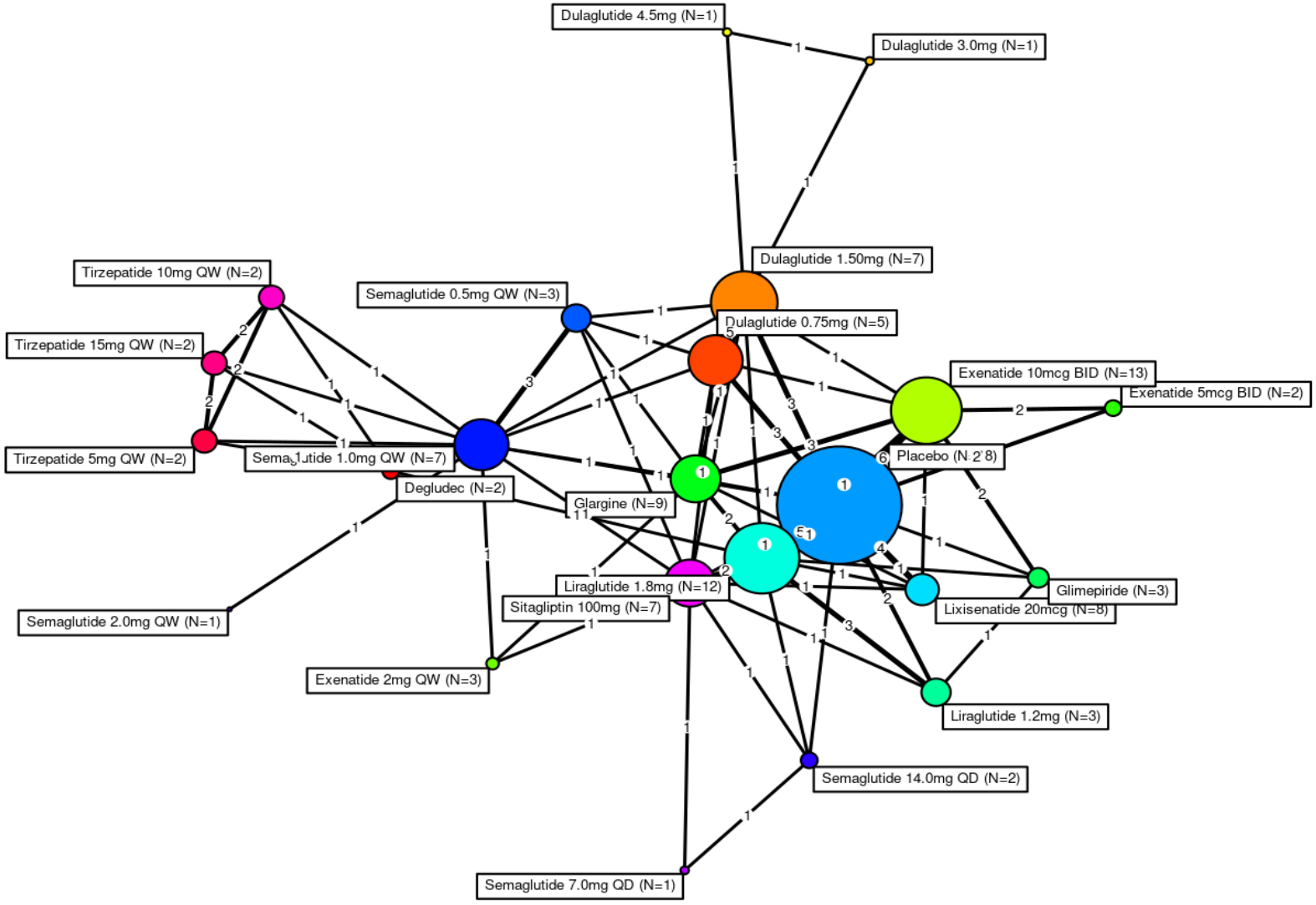
Figure 40: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 15 mg versus comparators, random effects model

Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

Weight (kg) Change from Baseline

The main analysis network diagram for weight (kg) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 41. All 45 studies and 23 treatments (nodes) were included in this analysis.

Figure 41: Main analysis network for weight (kg) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

All doses of tirzepatide showed significantly greater reductions in weight (kg) change from baseline compared with placebo and all GLP-1 RAs at the same recommended maintenance dose step (Table 54).

Table 54: Pairwise results (median difference [95% CrI]) for weight (kg) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 7.0 mg QD	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 42, Figure 43 and Figure 44, respectively.

Figure 42: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 43: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 10 mg versus comparators, random effects model

Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 44: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 15 mg versus comparators, random effects model

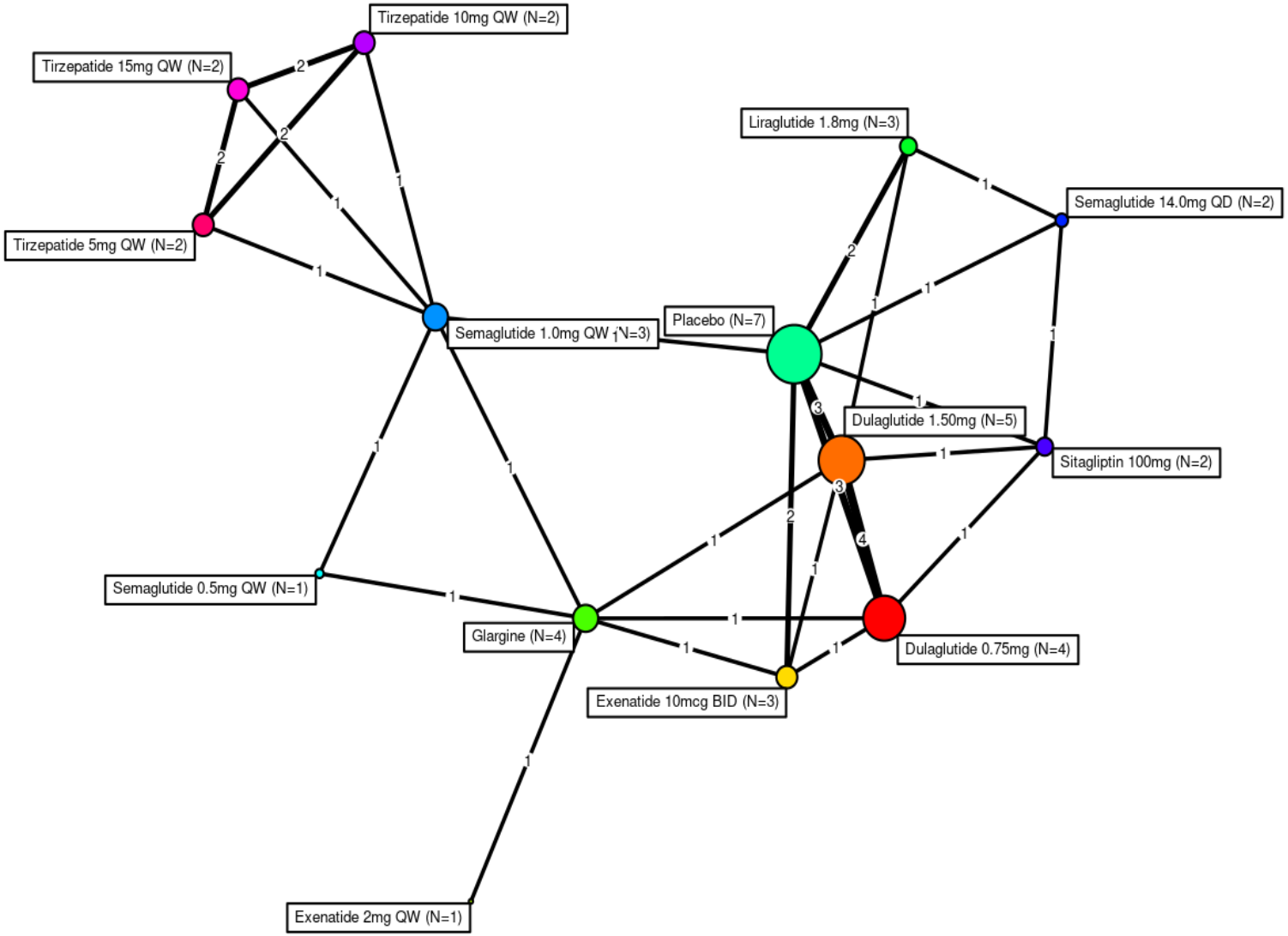
Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Body Mass Index (kg/m^2) change from baseline

The main analysis network diagram for BMI (kg/m^2) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 45. Due to limited data availability, it was not possible to include comparators such as liraglutide 1.2 mg, dulaglutide 3.0 mg, dulaglutide 4.5 mg, oral semaglutide 7.0 mg, exenatide 5 µg, and lixisenatide 20 µg in the analysis for this endpoint. The comparators in Derosa, 2011b, and the degludec arm in SURPASS-3 were not considered to be treatments of interest and they do not inform the network. Hence, Derosa, 2011b, and the degludec arm in SURPASS-3 were excluded from the analysis, in addition to the studies and treatment arms described in Section B.2.9.6.

Exenatide 2 mg QW versus detemir was assessed in only one study and hence exenatide 2 mg QW was disconnected in the network for this endpoint. Glargine and detemir were assumed to be comparable with respect to duration of action and efficacy. The efficacy data of detemir was pooled with glargine in order to include exenatide 2 mg QW in the NMA for BMI change from baseline. In total, 15 studies and 14 treatments (nodes) were included for this analysis.

Figure 45: Main analysis network for body mass index (kg/m^2) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

All the tirzepatide doses (5 mg, 10 mg and 15 mg) showed significantly greater reductions in BMI (kg/m^2) change from baseline compared with placebo and all GLP-1 RAs within the same recommended maintenance dose step included in Table 55.

Table 55: Pairwise results (median difference [95% CrI]) table for body mass index (kg/m^2) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 46, Figure 47 and Figure 48, respectively.

Figure 46: Forest plot (median difference [95% CrI]) for body mass index (kg/m^2) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 47: Forest plot (median difference [95% CrI]) for body mass index (kg/m^2) change from baseline, TZP 10 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

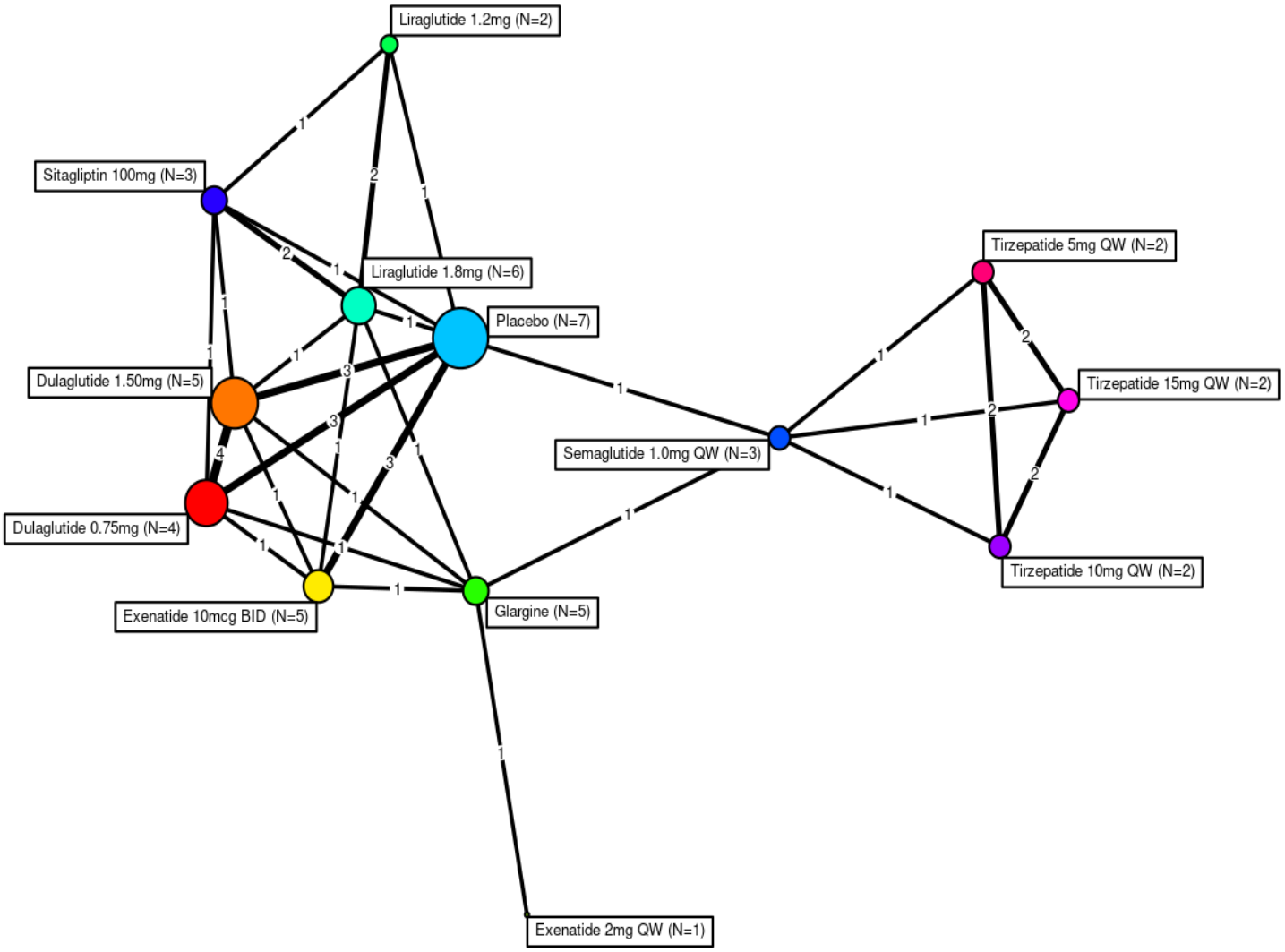
Figure 48: Forest plot (median difference [95% CrI]) for body mass index (kg/m^2) change from baseline, TZP 15 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Low-density lipoprotein (mmol/L) change from baseline

The main analysis network diagram for LDL (mmol/L) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 49. Due to limited data availability, it was not possible to include comparators, such as semaglutide 0.5 mg, dulaglutide 3.0 mg, dulaglutide 4.5 mg, oral semaglutide 7.0 mg, oral semaglutide 14.0 mg, exenatide 5 µg, and lixisenatide 20 µg, in the analysis for this endpoint. The degludec arm in SURPASS-3 was not considered a treatment of interest and does not inform the network (no connection beside the tirzepatide link). Hence, the degludec arm in SURPASS-3 was excluded from the analysis, in addition to the studies and treatment arms described in Section B.2.9.6. In total, 18 studies and 13 treatments (nodes) were included for this analysis.

Figure 49: Main analysis network for low-density lipoprotein (mmol/L) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

No significant differences in LDL (mmol/L) change from baseline were observed with tirzepatide 5 mg when compared with placebo and all GLP-1 RAs at the lowest recommended maintenance dose.

Tirzepatide 10 mg showed significant greater reductions in LDL (mmol/L) change from baseline compared with placebo. Tirzepatide 10 mg showed no statistically significant difference when compared with all GLP-1 RAs at the same recommended maintenance dose. Tirzepatide 10 mg showed similar reductions in LDL (mmol/L) change from baseline when compared with semaglutide 1.0 mg (Table 56).

Tirzepatide 15 mg showed significant greater reductions in LDL (mmol/L) change from baseline compared with placebo. No statistically significant differences were observed with tirzepatide 15 mg when compared with GLP-1 RAs at the same recommended maintenance dose.

Table 56: Pairwise results (median difference [95% CrI]) table for low-density lipoprotein (mmol/L) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 50, Figure 51 and Figure 52, respectively.

Figure 50: Forest plot (median difference [95% CrI]) for low-density lipoprotein (mmol/L) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 51: Forest plot (median difference [95% CrI]) for low-density lipoprotein (mmol/L) change from baseline, TZP 10 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

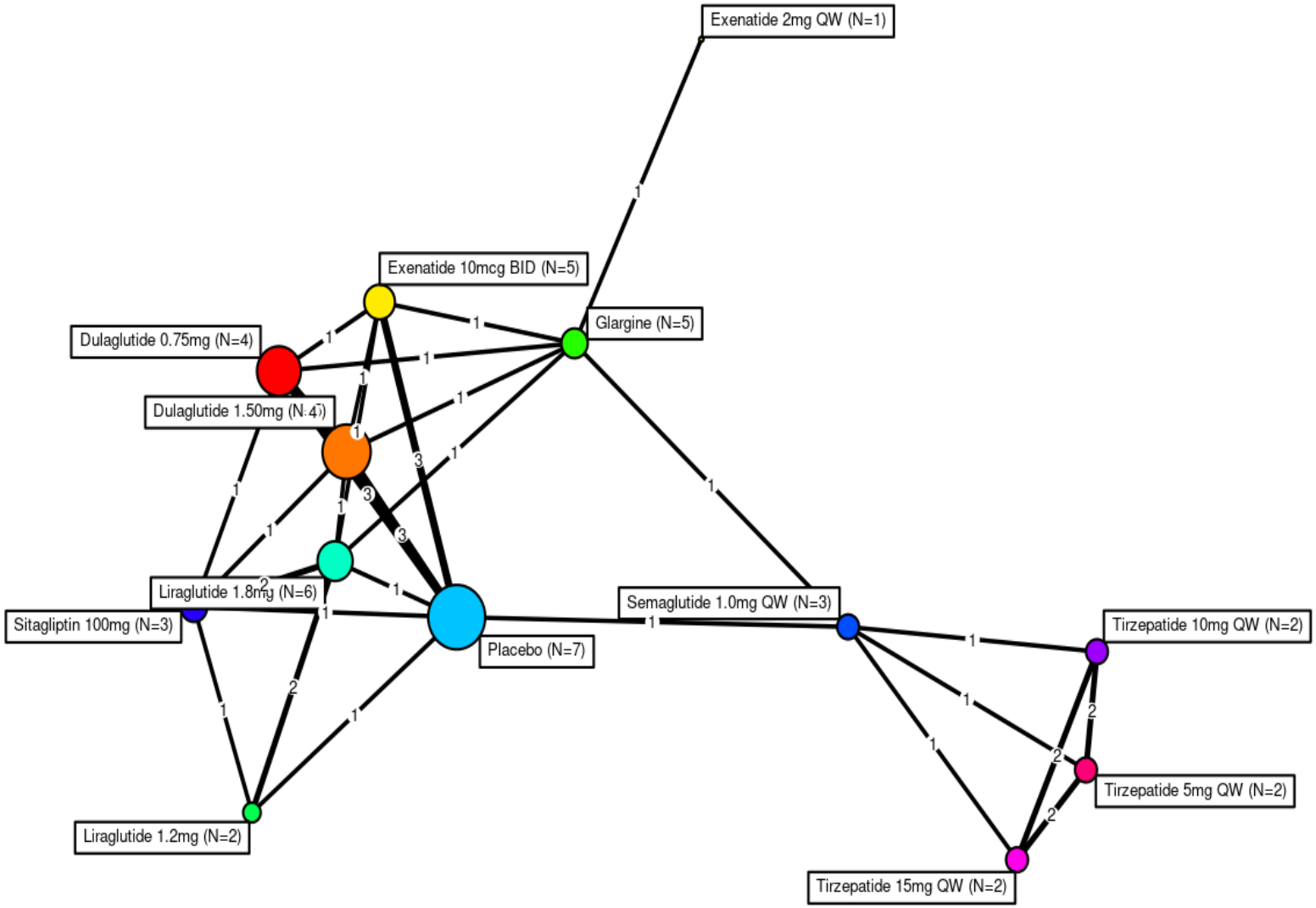
Figure 52: Forest plot (median difference [95% CrI]) for low-density lipoprotein (mmol/L) change from baseline, TZP 15 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

High-density lipoprotein (mmol/L) change from baseline

The main analysis network diagram for HDL (mmol/L) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 53. Due to limited data availability, it was not possible to include comparators, such as semaglutide 0.5 mg, dulaglutide 3.0 mg, dulaglutide 4.5 mg, semaglutide 7.0 mg, semaglutide 14.0 mg, exenatide 5 µg, and lixisenatide 20 µg, in the analysis for this endpoint. The degludec arm in SURPASS-3 was not considered to be a treatment of interest and does not inform the network (no connection beside the tirzepatide link). Hence, the degludec arm in SURPASS-3 was excluded from the analysis, in addition to the studies and treatment arms described in Section B.2.9.6. In total, 18 studies and 14 treatments (nodes) were included for this analysis.

Figure 53: Main analysis network for high-density lipoprotein (mmol/L) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Tirzepatide 5 mg showed no statistically significant difference in HDL (mmol/L) change from baseline compared with placebo, and all GLP-1 RAs at the lowest recommended maintenance dose included in Table 57.

Tirzepatide 10 mg and tirzepatide 15 mg showed statistically significant increase in HDL (mmol/L) from baseline compared with semaglutide 1.0 mg. No significant difference was observed when compared with placebo and other GLP-1 RAs at the same recommended maintenance dose steps.

Table 57: Pairwise results (median difference [95% CrI]) table for high-density lipoprotein (mmol/L) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 54, Figure 55 and Figure 56, respectively.

Figure 54: Forest plot (median difference [95% CrI]) for high-density lipoprotein (mmol/L) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 55: Forest plot (median difference [95% CrI]) for high-density lipoprotein (mmol/L) change from baseline, TZP 10 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 56: Forest plot (median difference [95% CrI]) for high-density lipoprotein (mmol/L) change from baseline, TZP 15 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

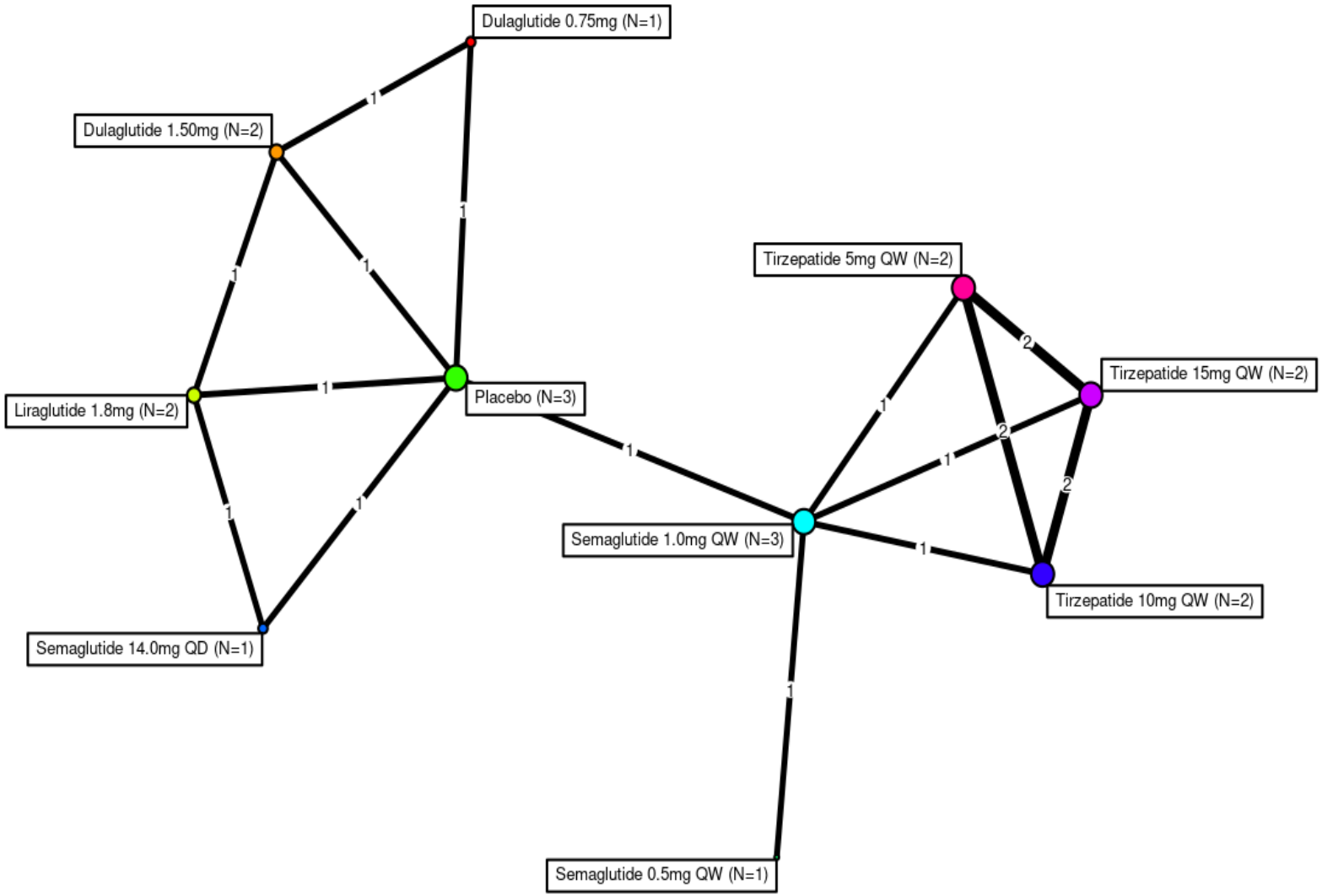
eGFR (mL/min/1.73 m^2) change from baseline

The main analysis network diagram for eGFR (mL/min/1.73 m^2) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 57. Due limited data availability, it was not possible to include comparators such as liraglutide 1.2 mg, dulaglutide 3.0 mg, dulaglutide 4.5 mg, semaglutide 7.0 mg, exenatide 2 mg, exenatide 5 µg, exenatide 10 µg and lixisenatide 20 µg in the analysis for this endpoint. The degludec treatment arm in SURPASS-3 and glargine treatment arm in SUSTAIN-4 were not considered to be treatments of interest and they do not inform the network. Hence, these treatment arms from both studies were excluded from the analysis, in addition to the studies and treatment arms described in Section B.2.9.6. In total, 7 studies and 10 treatments (nodes) were included for this analysis.

Whilst this endpoint was included in the NMA, limitations with the available data meant that there was high uncertainty within the network. More specifically, the eGFR network had a limited number of studies (seven studies in total) with variability between studies in terms of background therapies and change from baseline data for eGFR. In SUSTAIN 9 and AWARD 10 studies, all patients took SGLT2i as background therapy (with or without metformin). As a result, both studies showed improvement in eGFR from baseline in all treatment arms, including the placebo arm, with more eGFR improvement seen in the SUSTAIN 9 study. However, other studies of semaglutide (SUSTAIN 4, metformin with or without sulfonylurea) and dulaglutide (AWARD 6, metformin only) included in the network showed a decline in eGFR, in line with eGFR results reported by other comparator studies in the network. Further, the effect of placebo on renal function differed between studies (e.g. AWARD 10 and SUSTAIN 9) showing improvement in renal disease, likely due to the background of SGLT2i therapy. With SUSTAIN 9 contributing pivotally to the network due to the limited number of studies, there is high uncertainty in the network; this is demonstrated by the small

change in the eGFR results which is not significant but has high variability (e.g. wide confidence intervals). Therefore, the eGFR NMA results were not considered robust enough to draw any conclusions. Of note, however, the measured eGFR decreases in patients treated with tirzepatide and injectable semaglutide 1 mg were similar in SURPASS-2. It was therefore assumed for the cost-effectiveness modelling analysis that tirzepatide and all comparators had an equivalent effect on renal function, with changes from baseline in estimated glomerular filtration rate (eGFR) set to zero for all treatments.

Figure 57: Main analysis network for estimated glomerular filtration rate (eGFR) (mL/min/1.73 m^2) change from baseline



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Tirzepatide (5 mg, 10 mg, and 15 mg) showed no statistically significant difference for change from baseline in eGFR (mL/min/1.73 m^2) when compared with placebo and GLP-1 RAs at the lowest recommended maintenance dose (Table 58).

Table 58: Pairwise results (median difference [95% CrI]) table for estimated glomerular filtration rate (eGFR) (mL/min/1.73 m^2) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP dose (according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with tirzepatide 5 mg, 10 mg, and 15 mg for the random effects model are presented in Figure 58, Figure 59 and Figure 60, respectively.

Figure 58: Forest plot (median difference [95% CrI]) for estimated glomerular filtration rate (eGFR) (mL/min/1.73 m^2) change from baseline, TZP 5 mg versus comparators, random effects model

Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 59: Forest plot (median difference [95% CrI]) for estimated glomerular filtration rate (eGFR) (mL/min/1.73 m^2) change from baseline, TZP 10 mg versus comparators, random effects model

Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 60: Forest plot (median difference [95% CrI]) for estimated glomerular filtration rate (eGFR) (mL/min/1.73 m^2) change from baseline, TZP 15 mg versus comparators, random effects model

Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

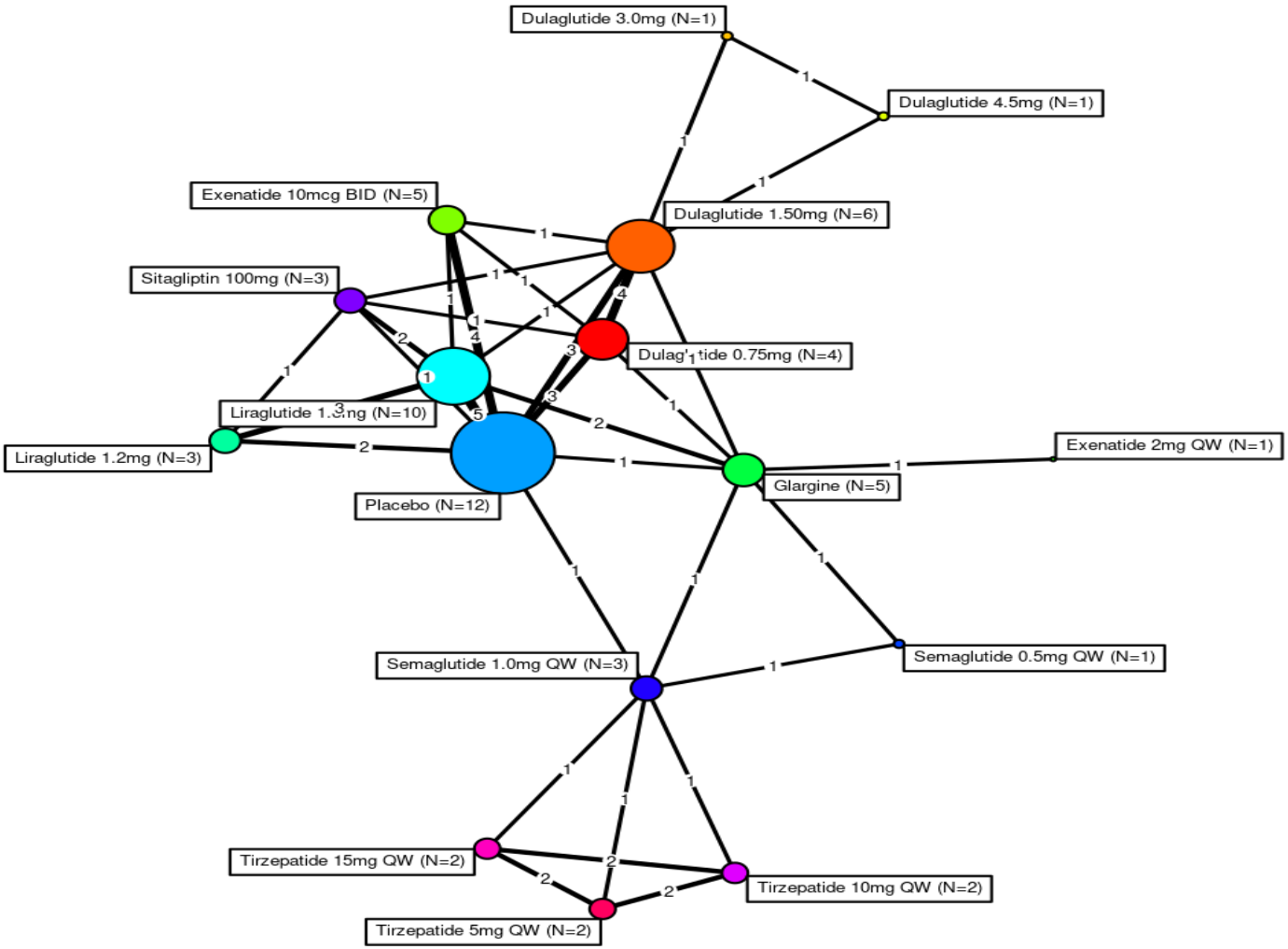
B.2.9.7.2 Main analysis results: Safety

Systolic blood pressure (mmHg) change from baseline

The main analysis network diagram for change from baseline in SBP (mmHg) at 40 weeks (TZP) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 61. Due to

limited data availability, it was not possible to include comparators such as oral semaglutide 7.0 mg, oral semaglutide 14.0 mg, exenatide 5 µg and lixisenatide 20 µg in the analysis for this endpoint. The degludec treatment arm in SURPASS-3 and glimepiride treatment arm in LEAD-2 studies were not considered to be treatments of interest and they do not inform the network. Hence, these treatment arms from both studies were excluded from the analysis in addition to the studies and treatment arms described above. In total, 23 studies and 16 treatments (nodes) were included for this analysis.

Figure 61: Main analysis network for change from baseline in systolic blood pressure (mmHg)



Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Tirzepatide 5 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo and dulaglutide 0.75 mg. No significant differences were observed with tirzepatide 5 mg compared with all other GLP-1 RAs at the lowest recommended maintenance dose (Table 59).

Tirzepatide 10 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo, liraglutide 1.8 mg, and exenatide 10.0 µg. No significant differences were observed with tirzepatide 10 mg compared with any other GLP-1 RAs at the intermediate recommended maintenance dose.

Tirzepatide 15 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo and all GLP-1 RAs at the highest recommended maintenance dose, except dulaglutide 4.5 mg.

Table 59: Pairwise results (median difference [95% CrI]) for systolic blood pressure (mmHg) change from baseline, fixed effects model; TZP 5, 10 or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header). Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the estimated treatment difference for the treatments of interest compared with TZP 5 mg, 10 mg, and 15 mg for the fixed effects model are presented in Figure 62, Figure 63, and Figure 64, respectively.

Figure 62: Forest plot (median difference [95% CrI]) for systolic blood pressure (mmHg) change from baseline, TZP 5 mg versus comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QW: once a week; TZP: tirzepatide.

Figure 63: Forest plot (median difference [95% CrI]) for systolic blood pressure (mmHg) change from baseline, TZP 10 mg vs. comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QW: once a week; TZP: tirzepatide.

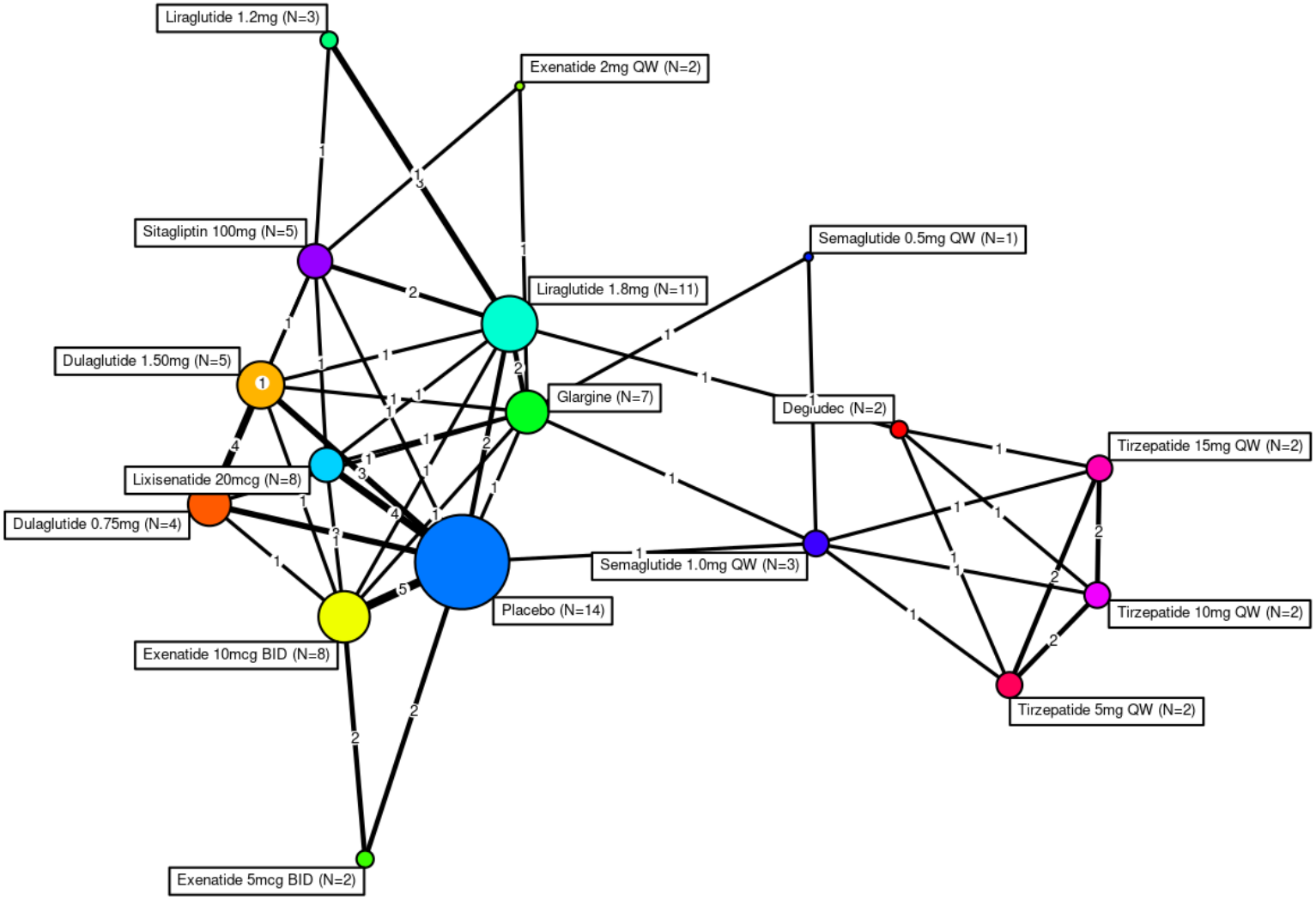
Figure 64: Forest plot (median difference [95% CrI]) for systolic blood pressure (mmHg) change from baseline, TZP 15 mg vs. comparators, random effects model

Footnotes: Glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QW: once a week; TZP: tirzepatide.

Proportion of patients experiencing nausea (any grade permitted)

The main analysis network diagram for proportion of patients experiencing nausea (any grade permitted) at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using the random effects model is shown in Figure 65. Due to limited data availability, it was not possible to include comparators such as dulaglutide 3.0 mg, dulaglutide 4.5 mg, oral semaglutide 7.0 mg, oral semaglutide 14.0 mg, and lixisenatide 20 µg in the analysis for this endpoint. In total, 33 studies and 17 treatments (nodes) were included for this analysis.

Figure 65: Main analysis network for proportion of patients experiencing nausea (any grade permitted)



Abbreviations: BID: twice a day; QW: once a week.

A significantly higher proportion of patients receiving tirzepatide 5 mg experienced nausea (proportion of patients experiencing nausea adverse event) compared with placebo. Tirzepatide 5 mg showed no significant differences in nausea compared with all GLP-1 RAs at the lowest recommended maintenance dose (Table 60).

A significantly higher proportion of patients receiving tirzepatide 10 mg experienced nausea compared with placebo. Tirzepatide 10 mg showed no significant differences in nausea compared with all other GLP-1 RAs at the intermediate recommended maintenance dose.

A significantly higher proportion of patients receiving tirzepatide 15 mg experienced nausea compared with placebo. Tirzepatide 15 mg showed no significant differences in nausea compared with all other GLP-1 RAs at the highest recommended maintenance dose.

Table 60: Pairwise results (odds ratio [95% CrI]) table for proportion of patients experiencing nausea (any grade permitted), random effects model; TZP 5, 10 or 15 mg (column) vs comparators (row)

Column vs. row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The forest plots representing the ORs for the treatments of interest compared with TZP 5 mg, 10 mg, and 15 mg for the fixed effects model are presented in Figure 66, Figure 67, and Figure 68, respectively.

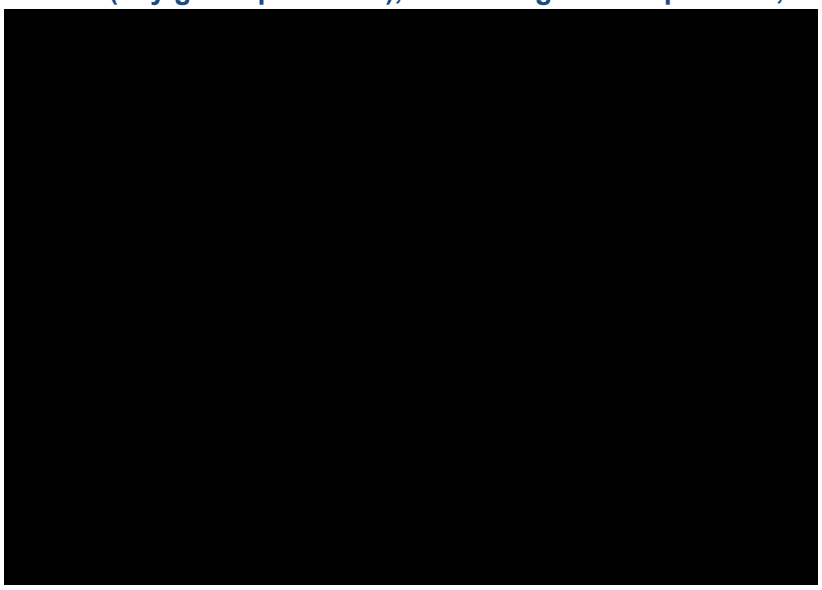
Figure 66: Forest plot (odds ratio [95% CrI]) for proportion of patients experiencing nausea (any grade permitted), TZP 5 mg vs. comparators, random effects model

Footnotes: Degludec, glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 67: Forest plot (odds ratio [95% CrI]) for proportion of patients experiencing nausea (any grade permitted), TZP 10 mg vs. comparators, random effects model

Footnotes: Degludec, glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 68: Forest plot (odds ratio [95% CrI]) for proportion of patients experiencing nausea (any grade permitted), TZP 15 mg vs. comparators, random effects model



Footnotes: Degludec, glargine and sitagliptin 100 mg were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

B.2.9.7.3 Sensitivity analysis

Results from the sensitivity analysis, conducted as per the methods described in Section B.2.9.5.3, were generally consistent with the main analysis, supporting the robustness of the base case results. Detailed results from the sensitivity analyses are presented in Appendix D.

B.2.9.8 Discussion and conclusions

B.2.9.8.1 Key Results

This NMA focussed only on studies conducted in patients with one to two OADs (i.e., second and third-line diabetes therapy), aligning with the SURPASS-2 and 3 trials and the anticipated positioning in NHS practice. Of the outcomes identified for inclusion in the analyses, analysis of the following two endpoints could not be conducted due to limited data availability: “proportion of patients experiencing at least one severe hypoglycaemic event” and “proportion of patients with at least one episode of hypoglycaemia with BG <54 mg/dL (3.0 mmol/L)”.

It is evident from the trial data that the GLP-1 RAs and tirzepatide exhibit a clear dose-response relationship in terms of efficacy and gastrointestinal side-effects. When interpreting the NMA it should be considered that patients unable to tolerate higher doses of one GLP-1 RA would not be expected to tolerate higher doses of another GLP-1 RA or tirzepatide; as such the most important comparisons are within each recommended maintenance dose step, rather than between recommended maintenance dose steps.

HbA1c change from baseline

For HbA1c change from baseline, all three doses of tirzepatide demonstrated a statistically significantly greater reduction in HbA1c from baseline compared to all GLP1-RAs within the same recommended maintenance dose step.

Results adjusted for the number of background OADs (in the meta-regression analysis) were similar to unadjusted results for all tirzepatide doses compared to all GLP-1 RAs at the same recommended maintenance dose step for HbA1c change from baseline (Appendix D.8).

The results from the sensitivity analyses were generally consistent with the main analysis for all tirzepatide doses, demonstrating that the results of the main analysis were robust to key assumptions made when undertaking the analysis. Sensitivity analyses involving the inclusion of a Phase 2 tirzepatide trial, modification of the network definition, exclusion of studies with insulin glargine as a treatment arm and allowing a broader analysis interval for comparator studies did not have a substantial impact on the results.

Some variability was observed in the sensitivity analysis including Asian population studies compared to the main analysis. In particular, a numerical reduction in HbA1c was observed for the comparisons of tirzepatide 15 mg versus dulaglutide 4.5 mg, tirzepatide 10 mg versus dulaglutide 3 mg and tirzepatide 5 mg versus exenatide 2 mg, but these reductions were not statistically significant. However, tirzepatide 5 mg had the third highest surface under these cumulative ranking curves (SUCRA) in the sensitivity analysis, which was consistent with its SUCRA in the main analysis. Additionally, tirzepatide 10 mg had the second highest SUCRA and probability of ranking best (p=0.156) and tirzepatide 15 mg had the highest SUCRA and probability of ranking best (p=0.738) in the sensitivity analysis, which were both consistent with the main analysis.

In the model-based NMA approach, tirzepatide data from Week 40 was compared with comparator data from Week 40. Results were consistent when comparing pairwise results for HbA1c change from baseline from the main analysis versus the model-based NMA approach; this demonstrates the robustness of the results.

Weight change from baseline

All three doses of tirzepatide demonstrated a significantly greater reduction in body weight from baseline compared to all GLP-1 RAs in the same recommended maintenance dose step.

Results adjusted for the number of background OADs (in the meta-regression analysis) were similar to unadjusted results for all tirzepatide doses compared to all GLP-1 RAs in the same recommended maintenance dose step for weight change from baseline.

The results from the sensitivity analyses conducted were generally consistent with the main analysis for all tirzepatide doses.

BMI change from baseline

All doses of tirzepatide demonstrated significantly greater reductions in BMI compared to all other GLP-1 RAs in the same recommended maintenance dose step; although, BMI data for studies in the main analyses were limited. Results from the sensitivity analyses were generally consistent with the main analyses for all tirzepatide doses.

CV markers

Tirzepatide at all doses was comparable to all GLP-1 RAs within the same recommended maintenance dose step in terms of the majority of CV markers (such as LDL, HDL).

Nausea

Overall, tirzepatide was comparable to the all GLP-1 RAs in the same recommended maintenance dose step, demonstrating a similar likelihood of experiencing nausea.

B.2.9.8.2 Generalisability of Results

Overall, this NMA provides robust results on the comparative efficacy and safety of tirzepatide 5 mg, 10 mg and 15 mg versus relevant GLP-1 RAs in the second and third line of treatment for T2D that can be considered generalisable to use of tirzepatide as a more efficacious option whenever GLP-1 RAs would otherwise be considered. This NMA was conducted using robust statistical methodology that is supported by published literature.

Furthermore, numerous sensitivity analyses were conducted to assess the robustness of the findings of the main analyses and these were selected based on past-precedent from published literature. The sensitivity analyses were largely consistent with the main analyses, demonstrating the robustness of the results of the main analyses.

The NMA utilises a wide range of trials that are expected to be relevant to patients in the UK, therefore producing results that are expected to be generalisable to UK clinical practice. The NMA includes reasonably homogenous trials, and the majority of key trial characteristics and baseline characteristics are consistent between treatment arms included in each network of the analysis.

Some variation was observed for the sensitivity analysis which included Asian population studies compared to the main analysis. Inclusion of Asian population studies results in some variability in results, suggesting the existence of some differences between Asian population studies, and those included in the main analysis. As such, excluding Asian population studies from the main analysis provided results that are most generalisable to UK clinical practice.

B.2.9.8.3 Strengths and Limitations

This NMA provides robust results that are generalisable to UK clinical practice, as outlined in Section B.2.9.8.2. Furthermore, comparative efficacy is presented in terms of multiple relevant efficacy outcomes, including the most critical clinical endpoints in the management of diabetes, such as change from baseline in HbA1c, weight and BMI, which are the most important endpoints to consider for cost-effectiveness analyses of a treatment in T2D.

The analysis was based on the clinical SLR described in Section B.2.1 that was conducted to identify RCTs assessing the efficacy and safety of all three doses of tirzepatide and GLP-1 RAs in adult patients (≥18 years of age) with T2D, ensuring that all relevant data were identified using a systematic approach. The SLR identified the relevant trials, and all evidence considered was from RCTs to ensure a high quality of data. As such, all studies included within the analyses were randomised trials, generally implying within-study validity of the evidence base.

Baseline characteristics were largely consistent across the included treatment arms and as such, the results are likely to be robust with minimal impact from prognostic variables. In addition, numerous sensitivity analyses were conducted to assess the robustness of the findings of the main analyses; results of the sensitivity analyses demonstrate the robustness of the results of the main analyses (Section B.2.9.7.3). The model-based NMA approach allowed for all timepoints available for change from baseline in HbA1c and weight to be able to be used, in particular to

enable a comparison at Week 40 for all comparators while a vast majority of studies end at Week 26.

Robust NMA models were fitted to the data using model specifications as recommended by the NICE DSU TSD, including both fixed and random effect models. The NMA model fit represents a further strength of these analyses; the residual deviance was approximately equal to the number of data points. Frequentist models were also run to assess the similarity to the Bayesian models. Results were generally consistent across models, demonstrating the robustness of the Bayesian NMA.

Overall, limited concerns with regards to inconsistency and heterogeneity were identified. No concerns regarding inconsistency were identified for continuous or binary endpoints. However, heterogeneity was identified for each outcome as presented in Table 52, with the SURPASS-2 and 3 trials contributing to heterogeneity in change from baseline in LDL and other treatments (excluding tirzepatide) contributing to heterogeneity in change from baseline HbA1c, weight, and HDL. Having said that, considering the number of studies in the network, only a minority of studies contributed to the heterogeneity.

As for limitations of the analyses, data availability for some endpoints was limited, meaning that comparisons between all treatments of interest could not be made for all endpoints. Although BMI is an important endpoint, this could not be fully analysed due to limited data availability across trials; however, data availability for change from baseline in weight was good. Heterogeneity across studies in follow-up time was another limitation of the analyses, with data input for TZP based on Week 40, contrasting with the Week 26 ± 4 of the comparator trials, although model-based results providing comparisons at Week 40 were consistent with the main analyses. Additionally, the duration of dose escalation employed to reach the TZP target dose in the SURPASS trials was longer (0–20 weeks) than the corresponding durations used for the comparators in the comparator studies, which ranged from 0–12 weeks, contributing to a source of heterogeneity between trials.

B.2.9.8.4 Conclusions

This NMA of GLP-1 RA and tirzepatide treatments of T2D was based on evidence from RCTs identified in an SLR. A total of 72 studies were eligible to be included in the main analysis, with the main reasons for exclusion of studies being incorrect study populations (e.g. CVOT trials, renal impairment population, patients with severe insulin resistance, etc.), studies not reporting data at relevant timepoints and evaluation of comparators that were not of interest. Of these 72 studies, 53 were included in the main analysis and 45 were included in the network. Overall, this NMA provides robust results on the comparative efficacy and safety of tirzepatide 5 mg, 10 mg and 15 mg versus GLP-1 RAs available in NHS practice.

Results of the sensitivity analyses were all largely consistent with the main analyses. A model-based NMA that allows inclusion of outcomes measured at multiple timepoints also showed similar estimates but with much narrower credible intervals. Meta-regression was conducted to adjust for the number of background OADs and the results were consistent with those observed in the base case analysis.

Overall, tirzepatide demonstrated statistically significant improved efficacy outcomes when compared with all GLP-1 RAs at the same recommended maintenance dose step. For change from baseline HbA1c, results were consistent across the network and all doses of tirzepatide

demonstrated a statistically significant greater reduction in HbA1c from baseline compared to all GLP1-RAs at the same recommended maintenance dose step. For change from baseline weight, all doses of tirzepatide demonstrated a statistically significantly greater reduction in body weight from baseline compared to all GLP-1 RAs at the same recommended maintenance dose step.

B.2.10 *Adverse reactions*

Summary of adverse events

- The safety and tolerability of tirzepatide in patients with T2D was assessed across the SURPASS trial programme, including long-term safety for up to 104 weeks in SURPASS-4 (the longest duration study)
- The SURPASS trials demonstrate that tirzepatide is generally a safe and well tolerated treatment option for people living with T2D
  - In the placebo-controlled analysis set, [redacted]% of tirzepatide-treated patients and [redacted]% of placebo-treated patients reported ≥1 TEAE
  - In the placebo-controlled analysis set, the percentage of patients reporting SAEs was similar across tirzepatide doses and placebo
  - Overall, [redacted] of tirzepatide-treated patients and [redacted] of placebo-treated patients discontinued from the study due to an AE
  - Reductions in mean sitting systolic and diastolic blood pressure (SBP/DBP) from baseline were greater in all tirzepatide dose groups compared to the placebo group through Week 40
  - A meta-analysis of positively adjudicated major adverse cardiovascular events (MACE) found that treatment with tirzepatide was not associated with excess risk for CV events in patients with T2D
- The risk of severe hypoglycaemia with tirzepatide treatment is low; [redacted] patients reported [redacted] episodes of severe hypoglycaemia across the phase 3 global studies, [redacted] of which occurred prior to reaching the maintenance dose. Of the [redacted] patients, [redacted] patients were on a background of insulin glargine or SU. Overall, the risk of severe hypoglycaemia was:
  - Higher when tirzepatide was used in combination with insulin glargine or SU, compared with other background glucose-lowering therapies studied, which has also been observed with other GLP-1 RAs
  - Similar between tirzepatide and GLP-1 RAs (semaglutide 1 mg and dulaglutide 0.75 mg)
  - Lower in tirzepatide-treated patients compared with basal insulin-treated patients
- As expected, and as seen in the GLP-1 RA class, the most common AEs were GI-related, mostly mild in severity and occurred generally within the dose-escalation phase
- The safety profile is familiar to healthcare providers in both primary and secondary care and can be readily managed by following the guidance in the SmPC and monitored via routine pharmacovigilance

The safety and tolerability of tirzepatide in patients with T2D was evaluated as an endpoint in all SURPASS trials. The safety and tolerability of tirzepatide in patients with T2D and established CVD will be evaluated in the ongoing SURPASS-CVOT trial, with expected completion in 2025.[1] A total of 19 completed phase 1, phase 2, and phase 3 studies have contributed safety data with up to 106 weeks of exposure to treatment. A total of 7,769 patients received an intervention in the phase 2 and 3 studies. Of these patients, 5415 received tirzepatide, 312 received placebo, and 2042 received an active comparator. Over the course of these investigations, the safety profile of tirzepatide has been well-characterised and robust management strategies have been developed and refined for AEs.

The primary purpose of the safety analyses is to characterize the safety of tirzepatide by identifying drug and dose effects. Integrated analysis sets were created to facilitate the assessment of the safety of tirzepatide. Because the seven phase 3 studies conducted (SURPASS 1–5 plus two Japanese studies, SURPASS J Mono and SURPASS J Combo) had the same tirzepatide treatment groups with the same dose-escalation schedules that will be proposed for the prescribing information, which were different from the phase 2 studies, it was

important to create integrated analysis sets that examined the phase 3 studies separately from the phase 2 studies. The two primary analysis sets to detect drug and dose effects, respectively, are the phase 3 Placebo-Controlled Analysis Set and the phase 3 Dose Effect Analysis Set, which are described in Table 61 and presented below.

Data on patient deaths during the trial are presented in Appendix F as, following adjudication by an external clinical endpoint committee, none were ruled as related to treatment with tirzepatide.

Table 61: Safety analysis sets

Analysis set	Studies	Time Period	Description	Treatment groups
Phase 3 placebo-controlled analysis set (N=953)	SURPASS-1, SURPASS-5	First dose of treatment to end of safety follow-up visit or date of study withdrawal	Integrated data of TZP doses compared to placebo for studies with placebo arm and same dose-escalation schedule proposed for the label	• TZP 5 mg (N=237) • TZP 10 mg (N=240) • TZP 15 mg (N=241) • TZP all doses (N=718) • Placebo (N=235)
Phase 3 Dose Effect Analysis Set (N=5,119)	SURPASS-1–5, SURPASS-J Mono, SURPASS-J Combo	First dose of treatment to end of safety follow-up visit or date of study withdrawal	Integrated data for dose comparison. Includes all studies with dose-escalation schedule proposed for the label	• TZP 5 mg (N=1,701) • TZP 10 mg (N=1,702) • TZP 15 mg (N=1,716) • TZP all doses (N=5,119)

Abbreviations: TZP: tirzepatide.

B.2.10.1 Overview of adverse events

As expected, similar to the GLP-1 RA class, the most common AEs in patients treated with tirzepatide were GI related. This side effect profile is understood by the healthcare community in both primary and secondary care and is readily managed by following the guidance in the SmPC and monitored via routine pharmacovigilance.

Placebo-controlled analysis set

In the placebo-controlled analysis set, [redacted]% of tirzepatide-treated patients and [redacted]% of placebo-treated patients reported ≥1 TEAE (Table 62). The percentage of patients reporting SAEs was similar across tirzepatide doses and placebo groups in the placebo-controlled analysis set. Overall, [redacted]% of tirzepatide-treated patients and [redacted]% of placebo-treated patients discontinued from the study due to an AE. The percentage of discontinuations from study drug due to an AE was higher in the patients treated with tirzepatide ([redacted]%) compared to placebo ([redacted]%). Additionally, there was an incremental increase in study drug discontinuation due to an AE with higher dose groups. [redacted] death (placebo) was reported. No other notable differences between tirzepatide dose groups and placebo were observed.

Table 62: Overview of adverse events (placebo-controlled analysis set)

Category[a]	n (%)	

	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)	TZP all doses vs placebo p-value
Deaths[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SAEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study due to AE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study drug due to AE[c]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TEAEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[a]Patients may be counted in more than 1 category; [b] Deaths are also included as SAEs and discontinuations due to AEs; [c]Patients were to remain in the study after permanent discontinuation of study drug and initiation of an alternative antihyperglycaemic medication so additional data could be collected; such patients may have subsequently discontinued the study for the same or a different reason.
Abbreviations: AE: adverse event; SAE: serious adverse event; TEAE: treatment-emergent adverse event; TZP: tirzepatide.

Dose effect analysis set

For the categories of TEAEs and discontinuation of study drug due to an AE, there was an incremental increase with higher dose groups (Table 63). The percentage of patients reporting SAEs and discontinuations from study due to an AE was similar across the three tirzepatide dose groups in the dose effect analysis set.

Table 63: Overview of adverse events (dose effect analysis set)

Category[a]	n (%)			
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)
Deaths[b]	[redacted]	[redacted]	[redacted]	[redacted]
SAEs	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study due to AE	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study drug due to AE	[redacted]	[redacted]	[redacted]	[redacted]
TEAEs	[redacted]	[redacted]	[redacted]	[redacted]

[a] Patients may be counted in more than 1 category; [b] Deaths are also included as SAEs and discontinuations due to AEs.
Abbreviations: AE: adverse event; SAE: serious adverse event; TEAE: treatment-emergent adverse event; TZP: tirzepatide.

B.2.10.2 Treatment emergent adverse events

Placebo-controlled analysis set

The most frequently reported TEAEs in tirzepatide-treated patients in the placebo-controlled analysis set were within the gastrointestinal disorders system organ class (SOC) with more patients treated with tirzepatide than patients treated with placebo reporting these events ([redacted]% and [redacted]%, respectively). Other frequently reported TEAEs in the placebo-controlled

analysis set were in the infections and infestations SOC (tirzepatide: [redacted]%; placebo [redacted]%) and metabolism and nutrition disorders SOC (tirzepatide: [redacted]%; placebo [redacted]%).

Common TEAEs were those reported at a frequency of at least 5%, before rounding, in any treatment group and are presented in Table 64 for the placebo-controlled analysis set .

Table 64: TEAEs occurring in at least 5% of patients in any treatment group (placebo-controlled analysis set)

Preferred Term	n (%)					TZP all doses vs placebo p-value
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)	
Nausea	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Nasopharyngitis	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Decreased appetite	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dyspepsia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vomiting	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Constipation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lipase increased	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hyperglycaemia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*p-value denotes significantly higher levels of hyperglycaemia in the placebo group compared with the TZP groups.
Abbreviations: TEAE: treatment-emergent adverse events; TZP: tirzepatide.

Dose effect analysis set

The most frequently reported TEAEs in the dose effect analysis set were within the gastrointestinal disorders SOC (5 mg, [redacted]%; 10 mg, [redacted]%; 15 mg, [redacted]%) with an incremental increase with higher dose groups. Common TEAEs were those reported at a frequency of at least 5%, before rounding, in any treatment group and are presented in Table 65 for the dose effect analysis set population.

Table 65: TEAEs occurring in at least 5% of patients in any treatment group (dose effect analysis set)

Preferred Term	n (%)			
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)
Nausea	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]
Decreased appetite	[redacted]	[redacted]	[redacted]	[redacted]
Vomiting	[redacted]	[redacted]	[redacted]	[redacted]
Dyspepsia	[redacted]	[redacted]	[redacted]	[redacted]
Constipation	[redacted]	[redacted]	[redacted]	[redacted]
Nasopharyngitis	[redacted]	[redacted]	[redacted]	[redacted]
Lipase increased	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: TEAE: treatment-emergent adverse events; TZP: tirzepatide.

TEAEs in the Gastrointestinal SOC

Placebo-controlled analysis set

A total of 336 patients (35.3%) experienced at least 1 TEAE in GI SOC, with more patients in the tirzepatide-treated groups compared to placebo group reporting these events. The most common GI-related TEAEs were nausea, diarrhoea, dyspepsia, vomiting, and constipation and generally occurred during the dose escalation period. TEAEs in the GI SOC were mostly mild in severity, as summarised in Table 66.

Table 66: Summary of TEAEs by maximum severity in the GI SOC (AS1)

Preferred Term	n (%)					TZP all doses vs placebo p-value
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)	
Patients with ≥1 GI TEAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mild	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Severe	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*Total includes one patient with a missing severity
Abbreviations: GI: gastrointestinal; SOC: system organ class; TEAE: treatment-emergent adverse events; TZP: tirzepatide.

B.2.10.3 Cardiovascular risk

Systolic and diastolic blood pressure

Placebo-controlled analysis set

Reductions in mean sitting SBP and DBP from baseline were observed in all tirzepatide dose groups and the placebo group over time. Reductions in SBP were greater in all tirzepatide dose groups compared to the placebo group at most timepoints through Week 40.

There were no notable baseline or postbaseline differences between the tirzepatide dose groups and the placebo group in the percentages of patients meeting the threshold criteria for abnormal SBP or DBP (Table 67).

Table 67: Summary of patients meeting threshold criteria for abnormal SBP and DBP at postbaseline; placebo-controlled analysis set

Threshold criteria for abnormal BP (mg Hg)	n (%)					TZP all doses vs placebo p-value
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)	
SBP						
≥140 and CFB ≥20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≤90 and CFB ≤-20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DBP						

≥90 and CFB ≥10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≤50 and CFB ≤ -10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BP: blood pressure; CFB: change from baseline; DBP: diastolic blood pressure; SBP: systolic blood pressure; TZP: tirzepatide.

Dose effect analysis set

Reductions in mean sitting SBP from baseline were greater in the tirzepatide 10 mg and 15 mg dose groups versus the 5 mg dose group from Week 16 through Week 52 and in the tirzepatide 15 mg versus the 10 mg dose group at Weeks 24 and 40.

There were no notable differences between the tirzepatide dose groups in the percentages of patients meeting the threshold criteria for abnormal SBP or DBP in the dose effect analysis set (Table 68).

Table 68: Summary of patients meeting threshold criteria for abnormal SBP and DBP at postbaseline; dose effect analysis set

Threshold criteria for abnormal BP (mg Hg)	n (%)			
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)
SBP				
≥140 and CFB ≥20	[redacted]	[redacted]	[redacted]	[redacted]
≤90 and CFB ≤-20	[redacted]	[redacted]	[redacted]	[redacted]
DBP				
≥90 and CFB ≥10	[redacted]	[redacted]	[redacted]	[redacted]
≤50 and CFB ≤ -10	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BP: blood pressure; CFB: change from baseline; DBP: diastolic blood pressure; SBP: systolic blood pressure; TZP: tirzepatide.

Pulse rate

Placebo-controlled analysis set

The mean pulse rate increased in all tirzepatide dose groups at Week 4 and remained above baseline through Week 40. There were no clinically meaningful mean changes from baseline over time in the placebo group. Mean pulse rate values for all three tirzepatide dose groups were approximately 2 bpm lower than baseline values at the time of the safety follow-up assessment.

Larger percentages of patients in the tirzepatide 10 mg and 15 mg groups met threshold criteria for abnormal pulse rate than the placebo group. There were no notable differences in the percentages of patients in the tirzepatide 5 mg group versus the placebo group meeting threshold criteria for abnormal pulse rate.

Dose effect analysis set

The mean pulse rate increased in all tirzepatide dose groups at Week 4 and remained above baseline through Week 52. The increase in mean pulse rate appeared dose dependent as it increased throughout the dose escalation and the maximum increase in the mean was observed when patients were at steady state for each dose. The mean pulse rate then decreased over

time for all three treatment groups and the difference between doses remained until the end of study treatment.

A larger percentage of patients in the tirzepatide 15 mg group versus the 5 mg and 10 mg groups, and in the 10 mg group versus the 5 mg group, met threshold criteria for abnormal pulse rate in the dose effect analysis set.

Heart rate

Consistent with the pulse rate data, incremental increases in mean ECG-derived heart rate from baseline with increasing tirzepatide dose were observed in the placebo-controlled analysis set and dose effect analysis set. No clinically meaningful differences in treatment-emergent heart rate abnormalities between placebo and tirzepatide in placebo-controlled analysis set or between tirzepatide doses in the dose effect analysis set were observed.

CV meta-analysis

A prospectively planned CV meta-analysis of the positively adjudicated MACE was executed to investigate potential differences between the pooled treatment groups. Data were analysed from seven clinical studies undertaken to investigate the efficacy and safety of tirzepatide 5, 10, and 15 mg once-weekly.

A total of 142 patients experienced the primary endpoint (adjudicated MACE-4) across all seven clinical studies and contributed to the complete analysis. Overall, a hazard ratio (HR) of 0.80 (95% CI: 0.57, 1.11) for the primary MACE-4 composite endpoint was attained when comparing pooled tirzepatide versus pooled comparators. SURPASS-4, specifically, contributed the majority of MACE-4 events for the CV safety meta-analysis (109 MACE-4 endpoints) due to the enrolment of patients with increased CV risk, and within SURPASS-4, an HR of 0.74 (95% CI, 0.51 to 1.08) was observed.

CV safety conclusions

In conclusion, treatment with all three doses of tirzepatide with exposure of up to 104 weeks, is not associated with excess risk for CV events in patients with T2D, as measured by the composite MACE-4 endpoint. The CV safety profile of tirzepatide appears to be comparable to GLP-1 RAs with regard to decreases in SBP and DBP and increases in heart rate. The ongoing CV outcomes trial (SURPASS-CVOT) will further characterize the CV profile of tirzepatide.

B.2.10.4 Retinopathy

Diabetic retinopathy is a microvascular complication of diabetes caused by damage to the retinal blood vessels. In the SURPASS trial programme, patients with a history of proliferative diabetic retinopathy, diabetic maculopathy, or non-proliferative diabetic retinopathy that required acute treatment were excluded, based on a dilated fundoscopic examination performed by a qualified eye care professional during screening to confirm eligibility. Diabetic retinopathy was evaluated through fundoscopic examinations performed when clinically indicated by any suspected adverse event of worsening retinopathy, as well as a customised MedDRA search of potential diabetic retinopathy complications.

Worsening of fundoscopic examination results, as recorded on the retinopathy eCRF, was recorded for [redacted] ([redacted]%) tirzepatide-treated patients across the SURPASS trials. No SAEs from

the SOC of eye disorders were reported in any of these [redacted] patients. A summary of potential treatment-emergent diabetic retinopathy events is presented in Table 69.

These results did not show increased risk of worsening retinopathy with tirzepatide treatment in the studied population. A dedicated addendum study to SURPASS-CVOT is ongoing to further investigate the impact of tirzepatide treatment on diabetic retinopathy progression.

Table 69: Summary of potential treatment-emergent diabetic retinopathy complications (dose-effect analysis set)

Preferred Term	n (%)			
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)
Patients with ≥1 TEAE	[redacted]	[redacted]	[redacted]	[redacted]
Diabetic retinopathy	[redacted]	[redacted]	[redacted]	[redacted]
Macular oedema	[redacted]	[redacted]	[redacted]	[redacted]
Vision blurred	[redacted]	[redacted]	[redacted]	[redacted]
Retinal detachment	[redacted]	[redacted]	[redacted]	[redacted]
Retinal vein occlusion	[redacted]	[redacted]	[redacted]	[redacted]
Retinopathy hypertensive	[redacted]	[redacted]	[redacted]	[redacted]
Visual impairment	[redacted]	[redacted]	[redacted]	[redacted]
Amaurosis fugax	[redacted]	[redacted]	[redacted]	[redacted]
Diplopia	[redacted]	[redacted]	[redacted]	[redacted]
Maculopathy	[redacted]	[redacted]	[redacted]	[redacted]
Visual acuity reduced	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: TEAE: treatment-emergent adverse event; TZP: tirzepatide.

B.2.10.5 Renal safety

In the placebo-controlled analysis set, a total of [redacted] ([redacted]%) patients receiving tirzepatide and [redacted] ([redacted]%) patient receiving placebo experienced ≥1 treatment-emergent renal event. There was no incremental increase with higher dose groups in treatment-emergent renal events, and no serious or severe treatment-emergent renal events were reported in this analysis set. A summary of the treatment-emergent renal events reported by patients in the placebo-controlled analysis set is presented in Table 70.

Across the SURPASS trials (dose-effect analysis set), [redacted]% and [redacted]% of tirzepatide-treated patients reported an adverse event within the SMQ of acute renal failure or chronic kidney disease, respectively. Overall, these data demonstrate that treatment with tirzepatide in patients with T2D does not significantly alter kidney function.

Table 70: Summary of treatment-emergent renal events (placebo-controlled analysis set)

SMQ Preferred Term	n (%)				
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)
Patients with ≥1 TEAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Acute renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal impairment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Acute kidney injury	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Chronic kidney disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Chronic kidney disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: MedDRA: Medical Dictionary for Regulatory Activities; SMQ: standardised MedDRA query; TEAE: treatment-emergent adverse event; TZP: tirzepatide.

B.2.10.6 Hypoglycaemia

Severe hypoglycaemia

Severe hypoglycaemia was defined as an episode with severe cognitive impairment requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions. These episodes may have been associated with sufficient neuroglycopenia to induce seizure or coma. Blood glucose measurements may not have been available during such an event, but neurological recovery attributable to the restoration of blood glucose (BG) to normal was considered sufficient evidence that the event was induced by a low BG concentration.

Table 71 summarizes the percentage of patients with episodes of severe hypoglycaemia in the phase 3 global studies, by background therapy. In total, of the 5,119 patients exposed to tirzepatide in global phase 3 studies, 10 (0.20%) patients reported 12 episodes of severe hypoglycaemia. Of these 10 patients who reported severe hypoglycaemia, five (0.10%) patients were on a background of insulin glargine or SU who reported one episode each. Overall, of the 12 episodes in the 10 tirzepatide-treated patients, 7 episodes were reported prior to reaching their maintenance dose. Collectively these data support the conclusion that the risk of severe hypoglycaemia with tirzepatide is low.

Table 71: Summary of severe hypoglycaemia postbaseline through the safety follow-up

Study (comparator)	Background therapy	Parameter	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 GPGL (SEMA 1 mg)	Metformin	N	470	469	470	469
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-3 GPGH (Insulin degludec)	Metformin ± SGLT-2i	N	356	360	359	358
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-4 GPGM (Insulin glargine)	Metformin ± SGLT-2i ± SU	N	329	328	338	1,000
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-5	Insulin	N	116	119	120	120

GPGI (placebo)	glargine ± metformin	n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: SEMA: semaglutide; SGLT-2i: sodium-glucose co-transporter 2 inhibitor; SU: sulfonylurea; TZP: tirzepatide.

Hypoglycaemia with blood glucose <54 mg/dL (3.0 mmol/L)

The percentage of tirzepatide-treated patients reporting hypoglycaemia with blood glucose <54 mg/dL(3.0 mmol/L) was low in patients on no background therapy ([redacted]), or background therapy of metformin ([redacted]%) alone or metformin with SGLT2i ([redacted]%). The percentage was higher in patients on background therapies of SU ([redacted]%) and insulin glargine ([redacted]%). Overall, the risk of hypoglycaemia with blood glucose <54 mg/dL (3.0 mmol/L) was higher when tirzepatide was used in combination with insulin glargine or SU as compared to other background glucose-lowering therapies studied, which has also been observed with the GLP-1 receptor agonist class.

The percentages of patients reporting hypoglycaemia with blood glucose <54 mg/dL (3.0 mmol/L) were similar in tirzepatide and placebo-treated patients. The percentage of patients reporting hypoglycaemia with blood glucose <54 mg/dL (3.0 mmol/L) was lower in tirzepatide-treated patients compared to basal insulin-treated patients, but higher in the tirzepatide 15 mg group compared with the semaglutide 1 mg group.

These data demonstrate that treatment with tirzepatide in patients with T2DM is associated with a low risk of hypoglycaemia. Severe hypoglycaemia was uncommon with tirzepatide treatment. Overall, the risk of hypoglycaemia with tirzepatide was comparable to the GLP-1 receptor agonist class.

B.2.10.7 Serious adverse events

Placebo-controlled analysis set

Overall, no important differences were observed between patients in the tirzepatide and placebo groups with respect to SAEs. The percentage of patients reporting at least one SAE was similar across tirzepatide doses and placebo groups in the placebo-controlled analysis set. The SOC with the highest percentage of SAEs was cardiac disorders, with similar percentages of patients reporting events in the tirzepatide and placebo groups. In interpreting these results, it should be considered that the two placebo-controlled trials (SURPASS-1 and -5) were relatively small in patient number and neither trial was selective for patients with increased CV risk. Results from a meta-analysis investigating CV risk are presented above (Section B.2.10.3) and demonstrate that tirzepatide is not associated with increased CV risk.

Dose effect analysis set

Overall, no important differences were observed between patients in the three tirzepatide dose groups with respect to SAEs. The percentage of patients reporting at least one SAE was similar across tirzepatide dose groups in the dose effect analysis set . The SOC with the highest percentage of SAEs was infections and infestations, with similar percentages of patients reporting events in all three tirzepatide dose groups.

The most frequently reported SAEs in tirzepatide-treated patients were:

Acute myocardial infarction (TZP 5 mg, [redacted]%; TZP 10 mg, [redacted]%; TZP 15 mg, [redacted]%)

COVID-19 pneumonia (TZP 5 mg, [redacted]%; TZP 10 mg, [redacted]%; TZP 15 mg, [redacted]%)

Coronary artery disease (TZP 5 mg, [redacted]%; TZP 10 mg, [redacted]%; TZP 15 mg, [redacted]%)

The highest percentages of patients reporting at least one SAE occurred in SURPASS-5 (TZP [all doses]: [redacted]%; placebo: [redacted]%), in which patients were on insulin glargine for background therapy, and SURPASS-4 (TZP [all doses]: [redacted]% ; insulin glargine: [redacted]%), which was conducted in patients with an increased risk of cardiovascular disease on a background regimen of 1–3 OADs (metformin, SU, SGLT-2i).

B.2.11 *Ongoing studies*

No studies are anticipated to provide additional evidence in the next 12 months for the indication being appraised. Additional data of interest for the efficacy and safety of tirzepatide to treat patients with T2D are anticipated as follows:

- SURPASS-6 is a phase 3 trial designed to compare the efficacy and safety of tirzepatide to insulin lispro (data expected to be available in 2023)[121]
- SURPASS-AP-Combo is a phase 3 trial designed to assess the efficacy and safety of tirzepatide compared to insulin glargine in patients with T2D on metformin with or without a sulfonylurea.[122] This trial has recently completed but data are not yet available. Data from this study are not relevant to this submission as the study was conducted in an Asian population
- SURPASS-CVOT is a phase 3 trial designed to assess the efficacy and safety of tirzepatide compared to dulaglutide in patients with T2D and established cardiovascular disease (expected completion 2025)[1]. A dedicated addendum study is also ongoing to further investigate the impact of tirzepatide treatment on diabetic retinopathy progression

B.2.12 *Interpretation of clinical effectiveness and safety evidence*

B.2.12.1 Principal findings from the clinical evidence base, highlighting key conclusions

Across all SURPASS studies, treatment with 5, 10, and 15 mg doses of tirzepatide demonstrated statistically significant and clinically meaningful reductions from baseline in HbA1c compared with semaglutide 1 mg, titrated insulin degludec, titrated insulin glargine and placebo. The proportion of participants achieving HbA1C <7% (<53mmol/mol), ≤6.5% (≤48 mmol/mol), and <5.7% (<39 mmol/mol) at the primary endpoint was significantly greater than for all comparators studied. These improvements were sustained up to 104 weeks in SURPASS-4. The results of SURPASS-2 also demonstrate the efficacy benefits of tirzepatide compared with a well-established GLP-1 RA and efficacious standard of care, injectable semaglutide 1 mg, and results were similar across all the SURPASS trials.

Importantly, across all studies there were significant body weight reductions seen in patients treated with all three doses of tirzepatide and significantly higher proportions of patients achieved mean body weight reductions of ≥5%, ≥10%, or ≥15% in all four of the SURPASS trials

presented when compared to either placebo or an active comparator. Considering the clear unmet need among patients with both T2D and obesity, these results indicate that tirzepatide is an important treatment option for these patients, offering both improved glycaemic control and weight loss.

Additionally, the pre-specified exploratory analyses of lipids demonstrated that patients treated with all three doses of tirzepatide achieved greater reductions in triglycerides and VLDL-C, and greater increases in HDL-C compared with patients receiving either placebo or an active comparator. Patients also reported greater improvements in patient-reported outcomes when treated with tirzepatide versus placebo or an active comparator, including the ability to perform physical activities of daily living.

Like the well-established safety profile of the GLP-1 RA class, the most frequently reported TEAEs in patients treated with tirzepatide were gastrointestinal disorders. Importantly, treatment with tirzepatide was not associated with increased CV risk. Overall, the safety profile of tirzepatide will be familiar to the healthcare community and is readily managed by following the guidance in the SmPC and monitored via routine pharmacovigilance.

The results of the NMA show that, for glycaemic and weight loss outcomes, tirzepatide demonstrated statistically significant improved outcomes when compared to all GLP-1 RAs at the same recommended maintenance dose step.

B.2.12.2 Strengths and limitations of the clinical evidence base

The global phase 3 studies (SURPASS-1 through 5) of the tirzepatide clinical development programme were designed and adequately powered to demonstrate that tirzepatide provides superior reductions of blood glucose and weight relative to comparators across the T2D population. The study population included patients along the disease continuum with regard to their duration of disease, variety of background therapies, comorbidities, and complications, similar to clinical practice. The endpoints investigated are clinically relevant and of importance to patient with T2D. All five global phase 3 studies completed the primary endpoint visit with study drug completion rates >85%, leading to limited missing data at the primary endpoint visit, allowing for robust assessment of the study objectives.

Limitations of the SURPASS trials include a lack of blinding in SURPASS-2–4 due to differences in dosing frequency, dose escalation scheme and devices between tirzepatide and the comparator. The SURPASS trials have relatively short follow-up times, so conclusions cannot yet definitively be made about long term outcomes. However, the extended follow-up for a sub-set of patients in SURPASS-4 suggests that the improvements in glycaemic control and weight loss are maintained at least until Week 104. Another limitation of the SURPASS trial programme is that none of the trials exactly match the anticipated positioning of tirzepatide in NHS England clinical practice, however the consistent significant results observed across the populations and comparators included in SURPASS trial programme suggest that this limitation is unlikely to be a material point in the appraisal.

The NMA provides robust results that are generalisable to UK clinical practice. Baseline characteristics were largely consistent across the included treatment arms and as such, the results are likely to be robust with minimal impact from prognostic variables. In addition, numerous sensitivity analyses were conducted to assess the robustness of the findings of the

main analyses; results of the sensitivity analyses demonstrate the robustness of the results of the main analyses. Limited concerns with regards to inconsistency and heterogeneity were identified.

B.2.12.3 Overall Conclusion

The clinical effectiveness evidence presented above demonstrates the tirzepatide addresses the clear unmet need for patients with T2D, offering both improved glycaemic control and weight loss alongside a familiar safety profile and will therefore be a step-change in T2D therapy for patients for whom the alternative is a GLP-1 RA.

B.3 Cost-effectiveness

PRIME T2D Model

- A literature review was performed to search for previously published health economic evaluations. Models developed prior to 2016, including UKPDS OM1 and OM2 and the IQVIA CORE Diabetes Model, have been shown to under predict cardiovascular benefits from the GLP-1 receptor class in certain situations. Therefore, a *de novo* model was developed
- The PRIME T2D model uses data exclusively from populations with T2D, meets ISPOR good modelling practice guidelines, and has been shown to project long-term patient outcomes consistent with those reported for several long-term studies, including cardiovascular outcome trials, during validation analyses
- The PRIME T2D model has also gone through Preliminary Independent Model Advice (PRIMA) review
- The model runs as a patient-level simulation and is capable of simulating treatment algorithms, risk factor progression, and projecting the cumulative incidence of macrovascular and microvascular complications as well as hypoglycaemic events
- The model can report clinical outcomes, quality-adjusted life expectancy, direct and indirect costs, along with standard measures of cost-effectiveness, including probabilistic sensitivity analysis (PSA)

Methodology

- The patient population used in the modelling analysis is intended to be representative of T2D patients in the UK who would currently be treated with GLP-1 RAs and was based on previously published information from NICE as part of evidence generation for the update of NICE Guideline NG28[123]
- Tirzepatide 5, 10 and 15 mg were compared with GLP-1 RAs currently in common use for the management of T2D in the UK
- As GLP-1 RAs and tirzepatide exhibit a dose-response relationship in terms of efficacy and gastrointestinal side-effects, when interpreting the NMA, comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps; for example, tirzepatide 5 mg is compared to the lowest recommended maintenance dose of each comparator
- Change from baseline in key risk factors including HbA1c, body weight, systolic blood pressure and serum lipid levels were informed by the NMA
- Simulated patients in the modelling analysis were assumed to intensify therapy when HbA1c levels rose above 7.5%, in line with NICE NG28. Simulated patients were assumed to switch to basal insulin therapy on intensification and remain on basal insulin therapy for the rest of the simulation
- Nausea rates for tirzepatide and all comparators were derived from the NMA and used to model the negative impact of treatment on quality of life in year 1 of the simulation. Rates of hypoglycaemia were not reported in the NMA as several studies reported zero events; therefore the rate of hypoglycaemia was set to zero for tirzepatide and all comparators in the base case analysis
- Quality-adjusted life expectancy was evaluated in the modelling analysis using an additive approach using data sourced from the literature

Cost effectiveness model results

- All three doses of tirzepatide (5, 10 and 15 mg) were associated with improvements in life expectancy and quality-adjusted life expectancy over the comparators evaluated. These benefits were driven by reductions in HbA1c and BMI associated with tirzepatide in the modelling analysis
- Tirzepatide 5 mg was associated with greater lifetime direct costs than the four comparators, with incremental costs ranging between £[redacted] and £[redacted] and ICERs ranging between £[redacted] and £[redacted] per QALY gained
- Tirzepatide 10 mg was also associated with higher direct costs than three comparators, but

was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg. ICERs for tirzepatide 10 mg ranged between £[redacted] and £[redacted] per QALY gained
- Tirzepatide 15 mg was also projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg and had ICERs ranging between £[redacted] and £[redacted] per QALY gained versus the other comparators
- The PSA indicated that there is a high probability ([redacted]% to [redacted]%, depending on the comparator) that tirzepatide would be cost-effective versus all comparators evaluated, assuming a willingness to pay of £20,000 per QALY gained
- The findings of the base case analysis remained robust under changes to a range of assumptions, including changes in risk factors associated with treatment, duration of therapy, quality of life benefits, and clinical assumptions used in the base case analysis
- Scenario analyses showed that changing the clinical input dataset and modifying assumptions around treatment intensification (specifically continuing GLP-1 RA or tirzepatide therapy after addition of basal insulin) produced comparable cost-effectiveness outcomes to the base case analysis

Conclusions

- Tirzepatide represents a cost-effective use of NHS resources versus commonly used GLP-1 RAs in England; sensitivity analyses showed that the ICERs calculated are robust when changing the modelling parameters

B.3.1 *Published cost-effectiveness studies*

A literature review targeting previously published health economic evaluations was performed via searches of four databases (PubMed, EconLit, EMBASE and Cochrane Library) and identified 804 articles. A total of 104 duplicates were removed, resulting in 700 unique hits. First-round screening of titles and abstracts identified 63 studies for full-text review. Additional hand searches identified 1,039 articles, six of which were duplicates, 994 excluded during first-round screening, and 39 unique hits identified for full-text review. During full-text review of hits identified through database and hand searches, 11 articles were excluded. The final review included 91 articles for data extraction. From these studies, focus was given to modern cost-effectiveness analyses published since 2018 in the UK. Table 72 summarises the included studies which were published since January 1, 2018 in a UK setting, published as full-text articles, and which were not comparing insulin-based treatment regimens.

The majority of summarised studies evaluated the cost-effectiveness of empagliflozin (eight studies), liraglutide (six studies), and once-weekly semaglutide (five studies). Utilised models were primarily UKPDS-based, with the most commonly used health economic model being the IQVIA CORE Diabetes Model (ten studies) followed by the UKPDS Outcomes Model 2 (two studies), two bespoke discrete-event simulation models (two studies) and the Cardiff Diabetes Model (one study). Approximately half of the studies (eight) incorporated data from cardiovascular outcomes trials (CVOTs), with five of these being calibration studies that adjusted model outputs to match CVOT findings.

Table 72: Summary list of published cost-effectiveness studies

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
AWMSG[124]	2018	CORE Diabetes Model (v9.0) 50-year time horizon No CVOT hazard ratios used	SUSTAIN 3, 5, and 7 trials Patients with inadequate glycaemic control on 1–2 OADs (SUSTAIN 3), basal insulin with or without metformin (SUSTAIN 5), or metformin monotherapy (SUSTAIN 7)	Once-weekly semaglutide 0.5 mg (9.00) and 1 mg (9.06) versus dulaglutide 1.5 mg (8.96) in dual therapy (differences: 0.04, 0.10) Once-weekly semaglutide 0.5 mg (9.32) and 1 mg (9.37) versus dulaglutide 1.5 mg (9.31) and liraglutide 1.8 mg (9.29) in triple therapy (differences: 0.01, 0.06, 0.03, 0.08) Once-weekly semaglutide 0.5 mg (9.27) versus liraglutide 1.2 mg (9.22) in triple therapy (difference: 0.06) Once-weekly semaglutide 1 mg (8.97) versus liraglutide 1.2 mg (8.86) in triple therapy (difference: 0.12)	Total costs for once-weekly semaglutide not reported (censored in commercial confidence) Dulaglutide 1.5 mg in dual therapy: 21,693 Dulaglutide 1.5 mg in triple therapy: 22,422 Liraglutide 1.8 mg in triple therapy: 23,799 Liraglutide 1.2 mg in triple therapy (versus once-weekly semaglutide 0.5 mg): 22,744 Liraglutide 1.2 mg in triple therapy (versus once-weekly semaglutide 1 mg): 22,127	ICERs not reported (censored in commercial confidence)
Bain *et al.*[125]	2020	CORE Diabetes Model (v9.0) 50-year time horizon No CVOT hazard ratios used	PIONEER 2, 3 and 4 trials Patients with inadequate glycaemic control on metformin	Oral semaglutide 14 mg (8.58, 8.20, 8.53) versus empagliflozin 25 mg (8.49), sitagliptin 100 mg (8.00), and liraglutide 1.8 mg (8.46)	Oral semaglutide 14 mg (25,856, 27,226, 27,868) versus empagliflozin 25 mg (24,885), sitagliptin 100 mg (26,263), and liraglutide 1.8 mg (29,418)	11,006 for oral semaglutide 14 mg versus empagliflozin 25 mg 4,930 for oral semaglutide 14 mg

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
			monotherapy (PIONEER 2), metformin with or without a sulfonylurea (PIONEER 3), or metformin with or without an SGLT-2 inhibitor (PIONEER 4)	Differences: 0.09, 0.20, 0.07	Differences: 971, 963, −1,551	versus sitagliptin 100 mg Oral semaglutide 14 mg dominant versus liraglutide 1.8 mg
Barnett *et al.*[126]	2020	CORE Diabetes Model (v8.5+) 50-year time horizon No CVOT hazard ratios used	LIRA-SWITCH trial Patients with inadequate glycaemic control on metformin plus sitagliptin	Liraglutide 1.8 mg (9.18) versus sitagliptin 100 mg (9.02) Difference: 0.15	Liraglutide 1.8 mg (24,737) versus sitagliptin 100 mg (22,362) Difference: 2,375	15,423 for liraglutide 1.8 mg versus sitagliptin 100 mg
Becker *et al.*[127]	2022	UKPDS OM2 Lifetime time horizon No CVOT hazard ratios used	EXSEL trial Patients with or without previous CVD	Once-weekly exenatide plus usual care (9.33) versus placebo plus usual care (9.18) Difference: 0.15	Once-weekly exenatide plus usual care (54,325) versus placebo plus usual care (47,968) Difference: 6,357	42,589 for once-weekly exenatide plus usual care versus placebo plus usual care
Capehorn *et al.*[128]	2021	CORE Diabetes Model (v9.0) 50-year time horizon No CVOT hazard ratios used	Meta-analysis of SUSTAIN 2, 3, and 8 and PIONEER 2 Patients with inadequate glycaemic control on metformin monotherapy	Once-weekly semaglutide 1 mg (7.28) versus empagliflozin 25 mg (7.05) Difference: 0.23	Once-weekly semaglutide 1 mg (27,144) versus empagliflozin 25 mg (26,127) Difference: 1,017	4,439 for once-weekly semaglutide 1 mg versus empagliflozin 25 mg
Johansen *et al.*[129]	2020	CORE Diabetes Model (v9.0) 50-year time horizon	SUSTAIN 10 trial Patients with inadequate	Once-weekly semaglutide 1 mg (6.58) versus liraglutide 1.2 mg (6.28)	Once-weekly semaglutide 1 mg (25,972) versus liraglutide 1.2 mg (26,112) Difference: −140	Once-weekly semaglutide 1 mg dominant versus liraglutide 1.2 mg

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
		No CVOT hazard ratios used	glycaemic control on 1–3 OADs	Difference: 0.30		
Kansal *et al.*[130]	2019	Discrete-event simulation model Specifically based on data from EMPA-REG OUTCOME trial for ten cardiovascular and renal events Lifetime time horizon Based on data from a CVOT	EMPA-REG OUTCOME trial Patients with established CVD	Standard of care plus empagliflozin (7.80) versus standard of care (6.80) Difference: 1.00	Standard of care plus empagliflozin (19,776) versus standard of care (16,040) Difference: 3,737	4,083 for standard of care plus empagliflozin versus standard of care
McEwan *et al.*[131]	2021	Cardiff Diabetes Model Lifetime time horizon Model calibrated to survival curves from the DECLARE-TIMI 58 trial	DECLARE-TIMI 58 trial Patients with or at risk of atherosclerotic CVD	Dapagliflozin (10.48) versus placebo (10.43) Difference: 0.06	Dapagliflozin (36,899) versus placebo (39,451) Difference: −10,730	Dapagliflozin dominant versus placebo
NICE NG-28[123]	2022	UKPDS OM2 Lifetime time horizon CVOT hazard ratios applied in base case analysis	THIN dataset and the National Diabetes Audit Adults with type 2 diabetes (interventions evaluated as additions or replacements, and as initial, first intensification, and second intensification therapies)	QALY outcomes are summarised as ranges across the evaluated populations (medication additions or replacements, initial therapy, first intensification and second intensification) Alogliptin: 7.81–9.51 Linagliptin: 7.90–9.58 Saxagliptin: 7.59–9.31 Sitagliptin: 7.89–9.61	Cost outcomes are summarised as ranges across the evaluated populations (medication additions or replacements, initial therapy, first intensification and second intensification) Alogliptin: 22,061–23,704 Linagliptin: 22,813–24,350 Saxagliptin: 23,806–25,203 Sitagliptin: 23,387–24,936 Dulaglutide: 30,154–31,056	SGLT-2 inhibitors and injectable semaglutide were the only treatments to have ICERs in the range of £20,000–30,000 across all populations, and dapagliflozin was the only intervention to have an ICER below £20,000 Both DPP-4 inhibitors and GLP-1 RAs other than injectable

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
			Sub-group analyses evaluated patients with: a BMI of greater than or equal to 30kg/m^2 high risk of a cardiovascular event who have not had a prior event a prior cardiovascular event	Dulaglutide: 8.05–9.71 Exenatide: 7.93–9.64 Liraglutide: 7.92–9.54 Lixisenatide: 7.67–9.22 Semaglutide (injection): 8.33–10.05 Semaglutide (oral): 7.49–9.31 Pioglitazone: 7.72–9.54 Canagliflozin: 8.05–9.83 Dapagliflozin: 8.24–9.90 Empagliflozin: 8.10–9.80 Ertugliflozin: 8.07–9.75	Exenatide: 30,446–31,203 Liraglutide: 35,927–37,441 Lixisenatide: 26,543–27,630 Semaglutide (injection): 30,130–31,067 Semaglutide (oral): 31,586–32,385 Pioglitazone: 19,212–21,665 Canagliflozin: 24,485–25,972 Dapagliflozin: 23,399–25,030 Empagliflozin: 23,785–25,329 Ertugliflozin: 22,316–23,967	semaglutide were either dominated or had very large ICERs compared to the standard of care arm
Pawaskar *et al.*[132]	2018	CORE Diabetes Model (v9.0) Lifetime time horizon CVOT hazard ratios applied in sensitivity analyses	GE Healthcare database Patients with inadequate glycaemic control on metformin monotherapy	SGLT-2 inhibitor plus metformin and sitagliptin (9.40) versus insulin NPH plus metformin and sitagliptin (9.22) Difference: 0.18	SGLT-2 inhibitor plus metformin and sitagliptin (25,747) versus insulin NPH plus metformin and sitagliptin (26,095) Difference: −348	SGLT-2 inhibitor plus metformin and sitagliptin dominant versus insulin NPH plus metformin and sitagliptin
Ramos *et al.*[133]	2019	CORE Diabetes Model (v9.0) 50-year time horizon Model calibrated to reproduce outcomes from EMPA-REG OUTCOME, TECOS, and SAVOR-TIMI 53 trials	EMPA-REG OUTCOME trial Patients with established CVD	Empagliflozin plus standard of care (6.41) versus sitagliptin plus standard of care (5.92), and saxagliptin plus standard of care (5.70) Differences: 0.49, 0.70	Empagliflozin plus standard of care (50,801) versus sitagliptin plus standard of care (47,627), and saxagliptin plus standard of care (48,071) Differences: 3,174, 2,730	6,464 for empagliflozin plus standard of care versus sitagliptin plus standard of care 3,878 for empagliflozin plus standard of care versus saxagliptin plus standard of care

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
Ramos *et al.*[134]	2020	CORE Diabetes Model (v9.5) 50-year time horizon Hazard ratio from real-world cardiovascular outcomes study used in empagliflozin arm of base case analysis	PIONEER 2 trial Patients with inadequate glycaemic control on metformin monotherapy	Empagliflozin 25 mg (9.27) versus oral semaglutide 14 mg (9.25) Difference: 0.02	Empagliflozin 25 mg (28,193) versus oral semaglutide 14 mg (34,441) Difference: −6,248	Empagliflozin 25 mg dominant versus oral semaglutide 14 mg
Ramos *et al.*[135]	2020	CORE Diabetes Model (v9.0) 50-year time horizon Model calibrated to reproduce outcomes from EMPA-REG OUTCOME and LEADER trials	EMPA-REG OUTCOME trial Patients with established CVD	Empagliflozin plus standard of care (6.41) versus liraglutide plus standard of care (6.19), and standard of care alone (5.84) Differences: 0.22, 0.57	Empagliflozin plus standard of care (50,801) versus liraglutide plus standard of care (54,185), and standard of care alone (47,137) Differences: −3,384, 3,664	Empagliflozin plus standard of care dominant versus liraglutide plus standard of care 6,428 for empagliflozin plus standard of care versus standard of care alone
Reifsnider *et al.*[136]	2020	Discrete-event simulation model Specifically based on data from EMPA-REG OUTCOME trial for ten cardiovascular and renal events (adapted for updated data for heart failure) Lifetime time horizon Based on data from a CVOT	EMPA-REG OUTCOME trial Patients with established CVD	Empagliflozin plus standard of care (6.27) versus standard of care (5.62) Difference: 0.65	Empagliflozin plus standard of care (18,197) versus standard of care (16,829) Difference: 1,367	2,093 for empagliflozin plus standard of care versus standard of care
Viljoen *et al.*[137]	2019	CORE Diabetes Model (v9.0)	SUSTAIN 7 trial	Once weekly semaglutide 0.5 mg (9.00) and 1 mg (9.06)	Once weekly semaglutide 0.5 mg (21,659) and 1 mg	Once-weekly semaglutide 0.5 mg and 1 mg dominant

Study	Year of publication	Summary of model	Patient population	Summary of QALY outcomes	Summary of cost outcomes (£)	ICER (£ per QALY gained)
		50-year time horizon No CVOT hazard ratios used	Patients with inadequate glycaemic control on metformin monotherapy	versus dulaglutide 1.5 mg (8.96) Differences: 0.04, 0.10	(21,588) versus dulaglutide 1.5 mg (21,693) Differences: −35, −106	versus dulaglutide 1.5 mg

Abbreviations: CVD: cardiovascular disease; CVOT: cardiovascular outcomes trial; OAD: oral antidiabetic medication; OM: Outcomes Model; QALY: quality-adjusted life year; SGLT-2: sodium-glucose cotransporter-2; UKPDS: United Kingdom Prospective Diabetes Study.

B.3.2 Economic Analysis

B.3.2.1 Developing a new health economic model of type 2 diabetes

The progress in diabetes management has been reflected in advances in computer simulation models of T2D, as evidenced in the data presented at the regular Mount Hood Diabetes Challenge Meetings.[138, 139] However, recent comparisons of health economic model projections with CVOT data has highlighted the need for ongoing development of T2D models to remain relevant.[140]

Models developed prior to 2016, including UKPDS OM1 and OM2 and the IQVIA CORE Diabetes Model, have been shown to under predict CV benefits from the GLP-1 RA class in certain situations. [140] One reason for this could be that models developed earlier than 2016 do not fully capture the benefits of reduced body weight as they tend to be based on cohorts using traditional therapies without any weight loss benefit. More recent risk equations have been shown to perform better in this regard.[140] Health technology assessment (HTA) agencies have also noted that diabetes models based on historic clinical data may be of limited relevance to modern care standards.

In 2019, in response to these criticisms, a literature review was performed to identify clinical data that could be used to support diabetes model development. The review focused on identifying data published from 2014 (when a previous review to support the PRIME Type 1 Diabetes Model had been conducted) through to December 2018 and included publications on RCTs, longitudinal studies, systematic reviews, meta-analyses and economic models in T2D. [141] With respect to existing T2D computer simulation models for health economic analysis, the literature review informed the following observations:

- Fifteen models were identified that could be used to perform health economic evaluations of different interventions in T2D. Most have participated in Mount Hood challenge meetings
- Many models rely on the same risk formulae (from the United Kingdom Prospective Diabetes Study [UKPDS]) to project the risk of macrovascular complications
- Outside of the Mount Hood Challenge meetings, external validation of existing models is limited
- None of the existing models use the novel approaches (e.g. model averaging) included in the PRIME Type 1 Diabetes Model[141]
- Many models rely on old data (pre-1990) and/or data from type 1 diabetes or mixed populations
- There may be scope for improvement in terms of the technologies used to run modelling simulations in type 2 diabetes
- Little has been done to date to validate existing models against recent data from CVOTs

In response to these observations, a decision was made to develop the PRIME Type 2 Diabetes Model (PRIME T2D Model), a new health economic model of T2D designed to meet the following criteria:[142]

- Based on the best available published data

- Developed exclusively from data on populations with type 2 diabetes
- Uses a model averaging approach to estimate the risk of macrovascular complications for a range of populations
- Meets International Society for Pharmacoeconomics and Outcomes Research [ISPOR] good modelling practice guidelines[143]
- Flexible enough to capture all relevant aspects of modern therapies for T2D
- Can be considered to have moved the field of type 2 diabetes modelling forward

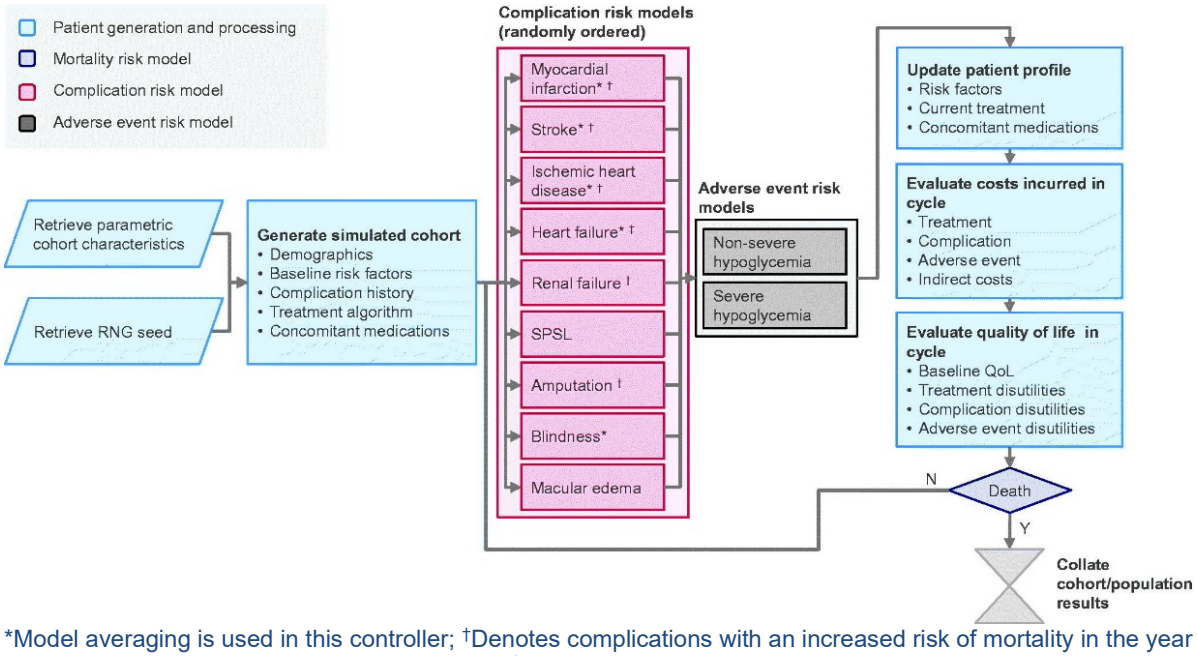
A manuscript describing the PRIME T2D Model and its validation has been published in the *Journal of Medical Economics*.[142] The model has been reviewed through the NICE PRIMA process in Q3 2021. During validation analyses, the PRIME T2D model has been shown to project long-term patient outcomes consistent with those reported for several long-term studies, including cardiovascular outcome trials.[142]

B.3.2.2 Model structure

The PRIME T2D Model runs as a patient-level simulation and is capable of simulating treatment algorithms, risk factor progression, and projecting the cumulative incidence of macrovascular and microvascular complications as well as hypoglycaemic events. Approaches novel to T2D modelling were utilised, including combining risk formulae using a weighted model averaging approach that takes into account patient characteristics to evaluate complication risk.[142]

A schematic diagram of the PRIME T2D is presented in Figure 69. A full description of the PRIME T2D Model is provided in the technical report accompanying this submission; the published model description and validation has been included in the reference pack.[142] A comparison of the PRIME T2D Model and the IQVIA CORE Diabetes Model (version 9.5), which has been used in prior NICE appraisals, is presented in Table 73.

Figure 69: Schematic diagram of the PRIME T2D Model



*Model averaging is used in this controller; †Denotes complications with an increased risk of mortality in the year of complication onset and in subsequent years; ‡Denotes complications with an increased risk of mortality associated with a history of this complication.
Abbreviations: QoL: quality of life; RNG: random number generator; SPSL: severe pressure sensation loss.
Source: Pollock et al. (2022)[142]

Table 73: Summary of similarities and differences between the PRIME T2D Model and the CORE Diabetes Model (version 9.5)

Feature	PRIME T2D Model	CORE Diabetes Model
Model type	Discrete time event, patient-level simulation	Discrete time event, patient-level simulation
Software / access	Calculation engine is coded in Java 10 with access to the model via an online interface	Calculation engine is coded in C++ with access to the model via an online interface
Source code audit	Source code was externally audited by a third-party to ensure accuracy/reliability	No reported source code audit
Data sources	Clinical data used to estimate complication risk is exclusively from T2D populations, based on a wide-ranging literature	Publications on the CORE Diabetes Model do not describe a literature review and suggest that data from T1D and mixed (T1D and T2D) populations is used to estimate the risk of microvascular complications
Validation	Recent validation published in 2022, including CVOT validation	Last validation published as part of Mount Hood Challenge in 2020 with calibration required to simulate outcomes from two CVOTs. Last broader validation of the model was published in 2014.
Macrovascular complication risk estimation	Model averaging approach to estimate risk in a range of simulation populations, combining risk equations from the UKPDS OM2, BRAVO Model and the Hong Kong Diabetes Registry. Model automatically weights risk equations for different populations.	Model provides different risk equations (selected by the user), grouped by setting. For Europe, these include: UKPDS OM1 UKPDS OM2 Swedish National Diabetes Registry ADVANCE PROCAM No information available on how different endpoint definitions were combined or how “missing” endpoints from the above data sources are handled
Individual patients modeled	Each simulated patient in the model is distinct (due to sampling at baseline of cohort characteristics)	All patients are identical in first order simulations (sampling is only performed in PSA)
Combined mortality approach to produce plausible outcomes	Combines complication-related mortality and country-specific data, using UKPDS OM2 complication mortality risk and cause-subtracted life tables	Similar combined approach is available, but usually with all-cause mortality life tables which may influence mortality estimates

Feature	PRIME T2D Model	CORE Diabetes Model
eGFR progression influences nephropathy outcomes	Progression to ESRD can be modelled based on declining eGFR	eGFR in the CORE Diabetes Model is distinct from renal outcomes (and possibly risk estimation for renal disease)
Full access and availability of model code to HTA authorities	Full access, including model source code, available on request (under the condition of confidentiality)	Limited to online access
Simulation time	Infrastructure to support high-end performance (short simulation times) means the model runs 1,000,000 unique patients through first order simulation in approximately 19 minutes	Standard simulation times for 1,000 patients through 1,000 iterations (all patients identical, first order simulation) appear to be around 90 minutes (the CORE Diabetes Model vendor claim shorter simulation times are possible)

Abbreviations: CVOT: cardiovascular outcomes trial; eGFR: estimated glomerular filtration rate; ESRD: end-stage renal disease; OM: Outcomes Model; PSA: probabilistic sensitivity analysis; T1D: type 1 diabetes; T2D: type 2 diabetes; UKPDS: United Kingdom Prospective Diabetes Study.

In the absence of any recent, publicly-available NICE technology appraisals of GLP-1 RA therapy, the key features of the present analysis were aligned with the evaluation described in NICE guideline NG28 (Table 74).[123]

Table 74: Key features of the economic analysis

	NG28 evaluation	Present analysis	Justification
Model	UKPDS OM2 based model	PRIME T2D Model	To evaluate the risk of diabetes-related complications based on combined risk estimates suited to a modern T2D population treated with GLP-1 RA
Time horizon	40 years	50 years	Long-term time horizon was employed to capture end-stage complications and in line with diabetes modelling guidance[144]
Cohort(s)	THIN initial cohort, THIN first intensification cohort and the THIN second intensification cohort	THIN second intensification cohort	This cohort was chosen to align to the anticipated positioning of tirzepatide within UK clinical practice
Treatment intensification	Patients intensified when HbA1c rose above 7.5%	Patients intensified to basal insulin therapy when HbA1c rose above 7.5%	To align with NG28 evaluation approach and NICE Guidance
Long-term HbA1c progression	UKPDS OM1 progression equation	UKPDS OM2 progression equation	To align with NG28 evaluation approach
Source of complication costs	UKPDS costs inflated to 2021 values	Literature review (see Section B.3.5.2), expressed as 2021 values	To provide up to date cost estimates
Source of utilities	UKPDS OM1	Literature review (see Section B.3.4.3)	To provide up to date utility estimates for T2D patients in the UK
Evaluation of complication risk	UKPDS OM2	UKPDS OM2 and BRAVO Model, plus other sources	Combined approaches were utilized to optimise risk evaluation for early and advanced populations (see PRIME T2D Model technical report)
Evaluation of mortality risk	UKPDS OM2	Hybrid approach using UKPDS OM2 and UK-specific life tables	To provide up to date country specific mortality estimates

Abbreviations: HbA1c: glycated haemoglobin; UKPDS OM1: United Kingdom Prospective Diabetes Study Outcomes Model 1; UKPDS OM2: United Kingdom Prospective Diabetes Study Outcomes Model 2.

B.3.2.3 Patient population

The patient population for the modelling analysis was intended to be representative of T2D patients in the UK who would currently be treated with GLP-1 RAs and was based on previously published information from NICE as part of evidence generation for the update of NICE Guideline NG28.[123] This stage in the T2D treatment algorithm is consistent with the anticipated positioning of tirzepatide; for patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. A simulated cohort of patients was defined with baseline demographics, complications and risk factors as summarised in Table 75. Where possible, baseline characteristics were taken from the THIN second intensification cohort to align with NG28; for missing inputs, corresponding values from the SURPASS-2 clinical trial cohort were used. The SURPASS-2 cohort had a comparable duration of diabetes to that of the THIN second intensification cohort (mean: 8.6 years and 8.5 years, respectively), making it a suitable proxy for risk factors and history of complications. The influence of cohort characteristics on cost-effectiveness outcomes was investigated in sensitivity analyses, where outcomes were assessed in a cohort with characteristics matching that in the SURPASS-2 trial (Section B.3.7).

The underlying risk equations in the model require differently specified inputs with respect to ethnicity. For baseline assignment of ethnic group, it was assumed that 4.5% of the population was classified as Black in line with the THIN second intensification cohort description,[123] which will adjust risk estimates from the BRAVO Model risk equations (but not UKPDS), and that 13.1% were classified as Indian, which will adjust risk estimates from UKPDS risk equations (but not BRAVO). The rationale for this approach is that including risk adjustment from both sets of equations may over-estimate the impact of ethnicity on outcomes. The impact of these assumptions was investigated in sensitivity analysis (Section B.3.7).

Table 75: Summary of cohort characteristics

	Mean	Standard deviation	Source
Demographics			
Percentage male (%)	57.0	Not applicable	THIN second intensification cohort (Table HE005)
Percentage with college education or higher (%)	25.97	Not applicable	PRIME Model index value[145]
Percentage smokers (%)	17.0%	Not applicable	THIN second intensification cohort (Table HE005)
Age (years)	63.95	10.4*	THIN second intensification cohort (Table HE005)
Duration of diabetes (years)	8.5	6.50*	THIN second intensification cohort (Section 2.3.1.1)
Ethnic group			
Percentage White (%)	82.4	Not applicable	THIN second intensification cohort (Table HE002)
Percentage Black (%)	4.5	Not applicable	THIN second intensification cohort (Table HE002)
Percentage Hispanic (%)	0.0	Not applicable	Assumed
Percentage Southeast Asian (%)	0.0	Not applicable	Assumed
Percentage Indian (%)	13.1	Not applicable	THIN second intensification cohort (Table HE002)

	Mean	Standard deviation	Source
Percentage Afro/Caribbean (%)	0.0	Not applicable	Assumed
Percentage Other (%)	0.0	Not applicable	Assumed
Baseline risk factors			
HbA1c (%)	7.50	1.03*	THIN second intensification cohort (Table HE005)
Systolic blood pressure (mmHg)	134.44	13.8*	THIN second intensification cohort (Table HE005)
Total cholesterol (mmol/L)	4.53	1.06	SURPASS-2 CSR, ITT population, Table GPGL.8.43
Low density lipoprotein cholesterol (mmol/L)	2.29	0.89*	THIN second intensification cohort (Table HE005)
High density lipoprotein cholesterol (mmol/L)	1.23	0.29*	THIN second intensification cohort (Table HE005)
Body mass index (kg/m^2)	30.7	6.90*	THIN second intensification cohort (2015 Report Table 20)[123]
Estimated glomerular filtration rate (mL/min/1.73 m^2)	71.37	17.10*	THIN second intensification cohort (Table HE005)
White blood cell count (10^6 cells/mL)	7.5	1.8**	THIN second intensification cohort (Table HE005)
Heart rate (beats per minute)	72.0	10.1*	THIN second intensification cohort (Table HE005)
Haemoglobin (g/dL)	14.5	1.42*	THIN second intensification cohort (Table HE005)
Complication history			
Percentage with atrial fibrillation at baseline (%)	1.2%	Not applicable	SURPASS-2 CSR, ITT population, Table GPGL.8.10
Percentage with urinary albumin ≥50mg/L at baseline (%)	22.6%	Not applicable	THIN second intensification cohort (Table HE004)
Percentage with peripheral vascular disease at baseline (%)	1.9%	Not applicable	SURPASS-2 CSR, ITT population, Table GPGL.8.10
Percentage with history of myocardial infarction at baseline (%)	2.0%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with history of stroke at baseline (%)	1.3%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with ischemic heart disease at baseline (%)	6.0%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with coronary revascularization at baseline (%)	3.0%	Not applicable	SURPASS-2 CSR, ITT population, Table GPGL.8.10
Percentage with heart failure at baseline (%)	1.9%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with foot ulcer at baseline (%)	0.8%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with amputation at baseline (%)	0.2%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with blindness at baseline (%)	1.3%	Not applicable	THIN second intensification cohort (Table HE006)

	Mean	Standard deviation	Source
Percentage with renal failure at baseline (%)	0.4%	Not applicable	THIN second intensification cohort (Table HE006)
Percentage with SPSL/neuropathy at baseline (%)	9.0%	Not applicable	SURPASS-2 CSR, ITT population, Table GPGL.8.11

* standard deviation value taken from the SURPASS-2 cohort as value was not reported in the source material. ** value assumed as not reported in source material.
Abbreviations: HbA1c: glycated haemoglobin; SPSL: severe pressure sensation loss.

B.3.2.4 Intervention technology and comparators

The intervention of interest is tirzepatide 5 mg, 10 mg or 15 mg, which is administered via injection every week (QW), using a single-dose pre-filled autoinjector pen device. Tirzepatide is initiated at 2.5 mg QW. After 4 weeks, the dose is increased to 5 mg QW. If needed, the dose can be increased in 2.5 mg increments every 4 weeks up to 15 mg. The recommended maintenance doses are 5 mg, 10 mg and 15 mg.

Tirzepatide was compared with the GLP-1 RAs currently in common use for the management of T2D in the UK, as tirzepatide is anticipated to be used for patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. Currently, triple therapy including a GLP-1 RA is considered for patients for whom triple therapy with metformin and two other oral drugs is not effective or tolerated, and who:

- have a BMI ≥35 kg/m^2 (adjusted accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity **or**
- who have a BMI <35 kg/m^2 and
  - for whom insulin therapy would have significant occupational implications or
  - when weight loss would benefit other significant obesity related comorbidities[36]

The following comparators were therefore considered for the base case analysis:

- **Dulaglutide 1.5, 3.0 and 4.5 mg**: Dulaglutide (Trulicity®) is a once-weekly injectable GLP-1 RA indicated for the treatment of adults with insufficiently controlled T2D as an adjunct to diet and exercise in combination with other therapies (or as monotherapy in cases where metformin is not tolerated/contraindicated)
- **Injectable Semaglutide 0.5 and 1.0 mg**: Semaglutide (Ozempic®) is a once-weekly injectable GLP-1 RA indicated for the treatment of adults with insufficiently controlled T2D as an adjunct to diet and exercise in combination with other therapies (or as monotherapy in cases where metformin is not tolerated/contraindicated)
- **Oral Semaglutide 7 and 14 mg**: Oral semaglutide (Rybelsus®) is a daily, oral administered, GLP-1 RA indicated for the treatment of adults with insufficiently controlled T2D to improve glycaemic control as an adjunct to diet and exercise in combination with other therapies (or as monotherapy in cases where metformin is not tolerated/contraindicated)
- **Liraglutide 1.2 and 1.8 mg**: Liraglutide (Victoza®) a daily injectable GLP-1 RA indicated for the treatment of adults, adolescents and children aged 10 years and above with

insufficiently T2D as an adjunct to diet and exercise in combination with other therapies (or as monotherapy in cases where metformin is not tolerated/contraindicated)

The following GLP-1 RAs were not included in the main analysis as recent estimates indicate that currently market share in the UK is less than 2%:

- **Exenatide 2.0 mg QW**: Exenatide (Bydureon®) is a once weekly, injectable GLP-1 RA indicated in adults 18 years and older with T2D to improve glycaemic control in combination with other glucose lowering medicinal products when the therapy in use, together with diet and exercise, does not provide adequate glycaemic control
- **Exenatide 10 µg BID**: Exenatide (Byetta®) is a twice daily, injectable GLP-1 RA indicated for the treatment of T2D in combination with other therapies in adults who have not achieved adequate glycaemic control on maximally tolerated doses of these oral therapies
- **Lixisenatide 20 µg QD**: Lixisenatide (Lyxumia®) is a once daily, injectable GLP-1 RA indicated for the treatment of adults with T2D to achieve glycaemic control in combination with oral glucose lowering medicinal products and/or basal insulin when these, together with diet and exercise, do not provide adequate glycaemic control

It is evident from the trial data that the GLP-1 RAs and tirzepatide exhibit a dose-response relationship, with higher doses being more efficacious in glycaemic control and weight loss outcomes whilst eliciting a greater level of gastrointestinal side-effects. It should therefore be considered that patients unable to tolerate higher doses of one GLP-1 RA would likely not be expected to tolerate higher doses of another GLP-1 RA or tirzepatide; therefore, for the base case analysis, comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps. Against a background of possible dose escalation, the comparison of a low maintenance dose of one agent with a high dose of another was considered to provide little meaningful information on the relative cost-effectiveness of interventions. The comparisons presented in the base case analysis are listed in Table 76.

Table 76: Overview of comparators and doses for the base case analysis

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide 5 mg	Dulaglutide 1.5 mg Semaglutide 0.5 mg Oral Semaglutide 7 mg Liraglutide 1.2 mg
Tirzepatide 10 mg	Dulaglutide 3.0 mg Semaglutide 1.0 mg Oral Semaglutide 14 mg Liraglutide 1.8 mg
Tirzepatide 15 mg	Dulaglutide 4.5 mg Semaglutide 1.0 mg Oral Semaglutide 14 mg Liraglutide 1.8 mg

B.3.3 *Clinical Parameters and Variables*

B.3.3.1 Treatment effects

Change from baseline in key risk factors including HbA1c, body weight, systolic blood pressure and serum lipid levels were informed by the NMA described in Section B.2.9. The NMA provided model inputs for all comparators outlined in Section B.3.2.4 and as such was used as input for the base case modelling analysis. Risk factor changes associated with therapy in the modelling analysis are summarized in Table 77,

Table 78 and

Table 79.

In the absence of missing inputs from the NMA, a conservative "nearest neighbour" approach was used to fill data gaps. Where inputs were missing, the corresponding input from the same compound was used as a proxy, wherever possible using higher (more efficacious) doses of comparator. For example, missing changes in serum lipid levels for semaglutide 0.5 mg were taken from the corresponding values for semaglutide 1.0 mg. This approach was also used for missing values for oral semaglutide.

It was assumed for the modelling analysis that tirzepatide and all comparators had an equivalent effect on renal function, with changes from baseline in estimated glomerular filtration rate (eGFR) set to zero for all treatments. Whilst this endpoint was included in the NMA, limitations with the available data meant that there was high uncertainty within the network: the eGFR network had a limited number of studies (seven studies in total) with variability between studies in terms of background therapies and change from baseline data for eGFR. Therefore, the eGFR NMA results were not considered robust enough to draw any conclusions. Of note, however, the measured eGFR decreases in patients treated with tirzepatide and injectable semaglutide 1 mg were similar in SURPASS-2.

Similarly, change from baseline in white blood cell count and haemoglobin levels were set to zero for tirzepatide and all comparators, as these endpoints were not included in the NMA.

For the model inputs, changes from baseline in BMI were calculated based on changes in body weight reported in the NMA. This is because whilst BMI was included in the NMA outputs, values were not available for all comparators whereas changes in body weight were available for all comparators. To avoid the use of proxy inputs from "nearest neighbour" comparators and as done in the NG28 economic analysis, BMI changes were calculated from the NMA-reported body weight changes and an assumed cohort height of 1.68 m, in line with the mean value reported for the THIN population.[123] Weight changes from the NMA and corresponding BMI changes used in the modelling analysis are summarised in Table 80.

Table 77: Treatment effects applied in the first year of the simulation for tirzepatide 5 mg and comparators

	Tirzepatide 5 mg mean (SD)	DULA 1.5 mg mean (SD)	SEMA 0.5 mg mean (SD)	ORAL SEMA 7 mg mean (SD)	LIRA 1.2 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with semaglutide 0.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA: liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: semaglutide.

Table 78: Treatment effects applied in the first year of the simulation for tirzepatide 10 mg and comparators

	Tirzepatide 10 mg mean (SD)	DULA 3.0 mg mean (SD)	SEMA 1.0 mg mean (SD)	ORAL SEMA 14 mg mean (SD)	LIRA 1.8 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with dulaglutide 1.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA, liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: semaglutide.

Table 79: Treatment effects applied in the first year of the simulation for tirzepatide 15 mg and comparators

	Tirzepatide 15 mg mean (SD)	DULA 4.5 mg mean (SD)	SEMA 1.0 mg mean (SD)	ORAL SEMA 14 mg mean (SD)	LIRA 1.8 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with dulaglutide 1.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA: liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: Semaglutide.

Table 80: Calculation of change from baseline in BMI based on changes in body weight from the NMA

	Mean value	Standard deviation
Change from baseline in body weight from the NMA (kg)		
Tirzepatide 5 mg	[redacted]	[redacted]
Tirzepatide 10 mg	[redacted]	[redacted]
Tirzepatide 15 mg	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]
Semaglutide 0.5 mg	[redacted]	[redacted]
Semaglutide 1.0 mg	[redacted]	[redacted]
Oral semaglutide 7 mg	[redacted]	[redacted]
Oral semaglutide 14 mg	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]
Calculated change from baseline in BMI (kg/m^2)		
Tirzepatide 5 mg	[redacted]	[redacted]
Tirzepatide 10 mg	[redacted]	[redacted]
Tirzepatide 15 mg	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]
Semaglutide 0.5 mg	[redacted]	[redacted]
Semaglutide 1.0 mg	[redacted]	[redacted]
Oral semaglutide 7 mg	[redacted]	[redacted]
Oral semaglutide 14 mg	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]

Calculations were based on an assumed height of 1.68, in line with the THIN population.[123]
Abbreviations: BMI: body mass index; NMA: network meta-analysis.

B.3.3.2 Treatment intensification

Simulated patients in the modelling analysis were assumed to intensify therapy when HbA1c levels rose above 7.5%, in line with NICE guidance for the management of T2D (NG28).[36] Section 1.6.8 of the recommendations state:

In adults with type 2 diabetes, if HbA1c levels are not adequately controlled by a single drug and rise to 58 mmol/mol (7.5%) or higher:

- *reinforce advice about diet, lifestyle and adherence to drug treatment and*
- *support the person to aim for an HbA1c level of 53 mmol/mol (7.0%) and*
- *intensify drug treatment.*

In the present modelling analysis, simulated patients were assumed to switch to basal insulin therapy on intensification and to remain on basal insulin therapy for the rest of the simulation. This assumption was based on NG28 guidance that states: [36]

For adults with type 2 diabetes, only offer combination therapy with a GLP‑1 mimetic and insulin along with specialist care advice and ongoing support from a consultant-led multidisciplinary team.

It was therefore assumed that, following treatment intensification, the majority of patients would not continue tirzepatide or GLP-1 RA therapy in the base case modelling analysis. To investigate the impact of this assumption on the modelled outcomes, a scenario analysis was performed where patients continued tirzepatide or comparator therapy after initiation of basal insulin, before continuing on to basal-bolus insulin therapy at which time tirzepatide or GLP-1 RA therapy was stopped (further details can be found in Section B.3.7.2).

On intensification to basal insulin in the base case analysis:

- HbA1c was assumed to decrease by a mean of 0.84% based on the formula for "all" input parameters published by Willis *et al.* in 2017.[146] This approach is aligned with the NICE modelling analysis for NG28, as well as recent GLP-1 RA submissions to other HTA bodies.[147] The standard deviation could not be estimated using Willis *et al.*, so was assumed to be 0.15% based on the HbA1c change from baseline on tirzepatide 5 mg in the NMA
- BMI was assumed to return to baseline levels in the first year of basal insulin therapy. This is a conservative assumption, given that initiation of basal insulin therapy was associated with a modest increase in bodyweight in the 4T trial.[4, 148] This approach was adopted due to the absence of data to differentiate between bodyweight responses upon initiation of insulin therapy following treatment with tirzepatide or GLP-1 RAs
- All other risk factors were assumed to return to baseline levels upon initiation of insulin therapy, as there was no evidence on durability of effect at the time of modelling analysis

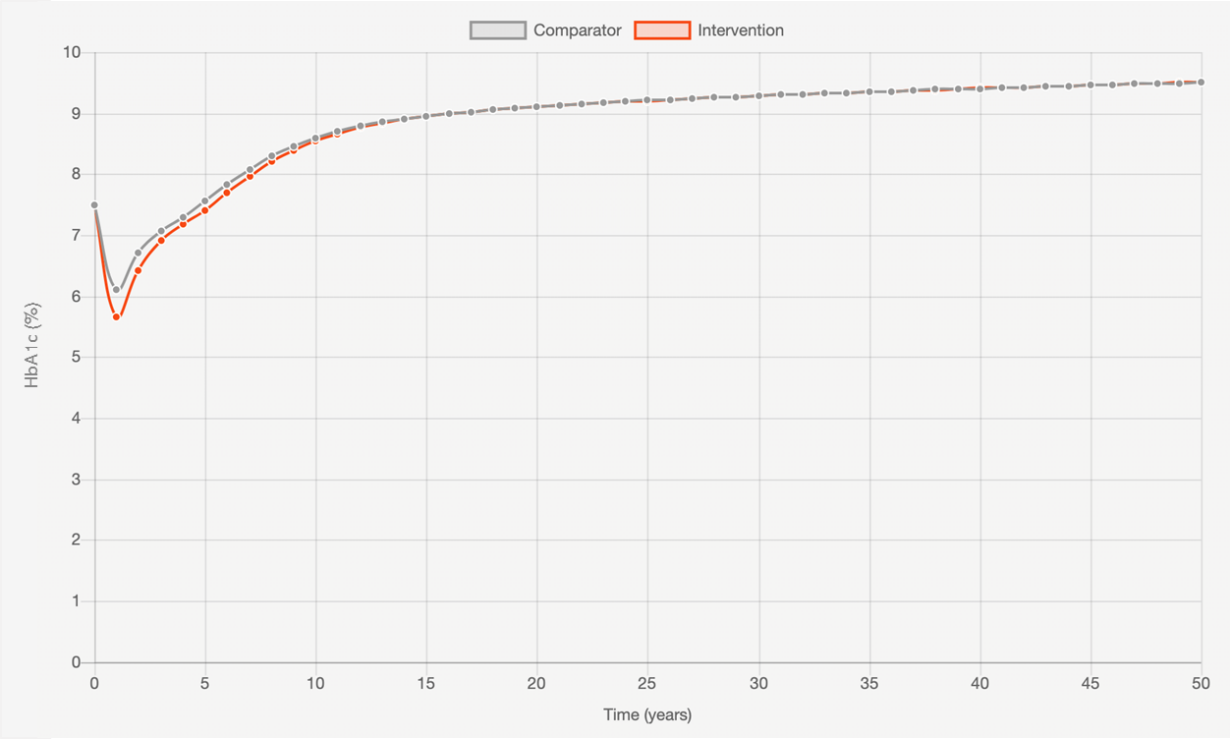
A simplifying assumption of only one intensification step was assumed in this evaluation. Complicating the analysis with subsequent intensification steps would have very little impact on cost-effectiveness, as the changes would be similar in both treatment arms. The impact of this assumption, as well as the duration of therapy and intensification threshold, were investigated in sensitivity analyses (Section B.3.7.1).

B.3.3.3 Long-term risk factor progression

B.3.3.3.1 HbA1c

In alignment with previous appraisals and the modelling analysis performed to support NG28, long-term HbA1c progression in the modelling analysis was based on the UKPDS Outcomes Model 2 equation published by Leal *et al.* in 2021.[149] Mean HbA1c progression curves from the PRIME T2D Model for the base case analysis are shown in Figure 70 for the comparison of tirzepatide 10 mg with semaglutide 1.0 mg over a 50-year time horizon. It should be noted that there is a flattening of the curves between years 3 and 8, this is due to simulated patients in the analysis intensifying therapy; intensification to basal insulin was associated with a mean HbA1c decrease of 0.84%, leading to a lower cohort average HbA1c level than would be expected from the UKPDS progression curve during this period.

Figure 70: Long-term HbA1c progression in the modelling analysis



Population mean values are shown in orange for tirzepatide 10 mg (intervention arm) and in grey for SEMA 1.0 mg (comparator arm)
Abbreviations: HbA1c: glycated haemoglobin.

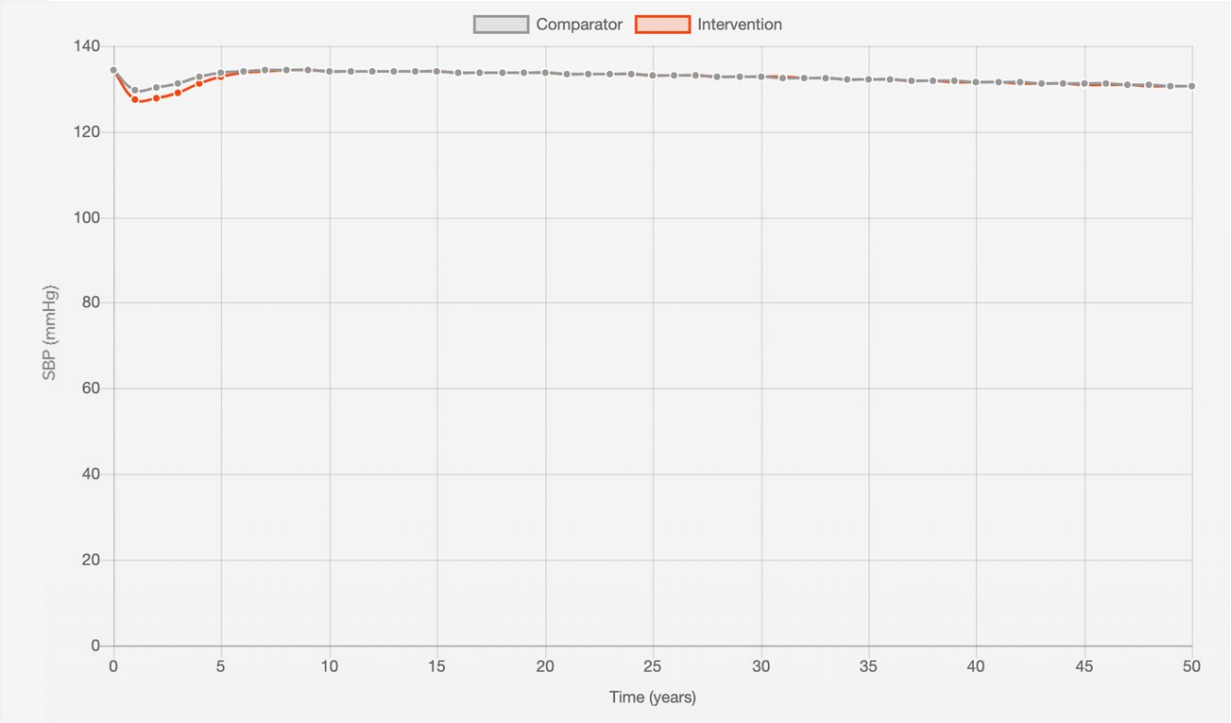
B.3.3.3.2 Other risk factors

For all other risk factors, it was assumed that no long-term changes would be applied in the modelling analysis. In effect, risk factor changes associated with treatment were applied in year 1 of the simulation and remained constant until treatment intensification when they returned to baseline levels. This simplifying, conservative assumption was used because the clinical benefits associated with therapy were applied for the duration of treatment (therefore balancing costs and effects for all comparators in the analysis) and there was little long-term data on the durability of treatment effects and effects of switching to basal insulin from GLP-1 RA on individual risk factors. A sensitivity analysis was performed using UKPDS OM2 risk factor progression equations to explore the impact of this approach on modelled outcomes (Section B.3.7.1).

Systolic blood pressure

The long-term SBP progression in the modelling analysis is presented in Figure 71. Of note, there is a difference between the population means for the tirzepatide and comparator arms over the first 6–7 years of the simulation, due to patients in the tirzepatide arm initiating insulin therapy later on average than those taking any of the comparators. Further, a slight decrease in mean SBP is evident beyond year 25 of the simulation due to patients with higher SBP levels experiencing a higher average mortality risk than those with lower SBP levels, leading to a slight decrease in the population mean in the advanced years of the simulation.

Figure 71: Long-term SBP progression in the modelling analysis

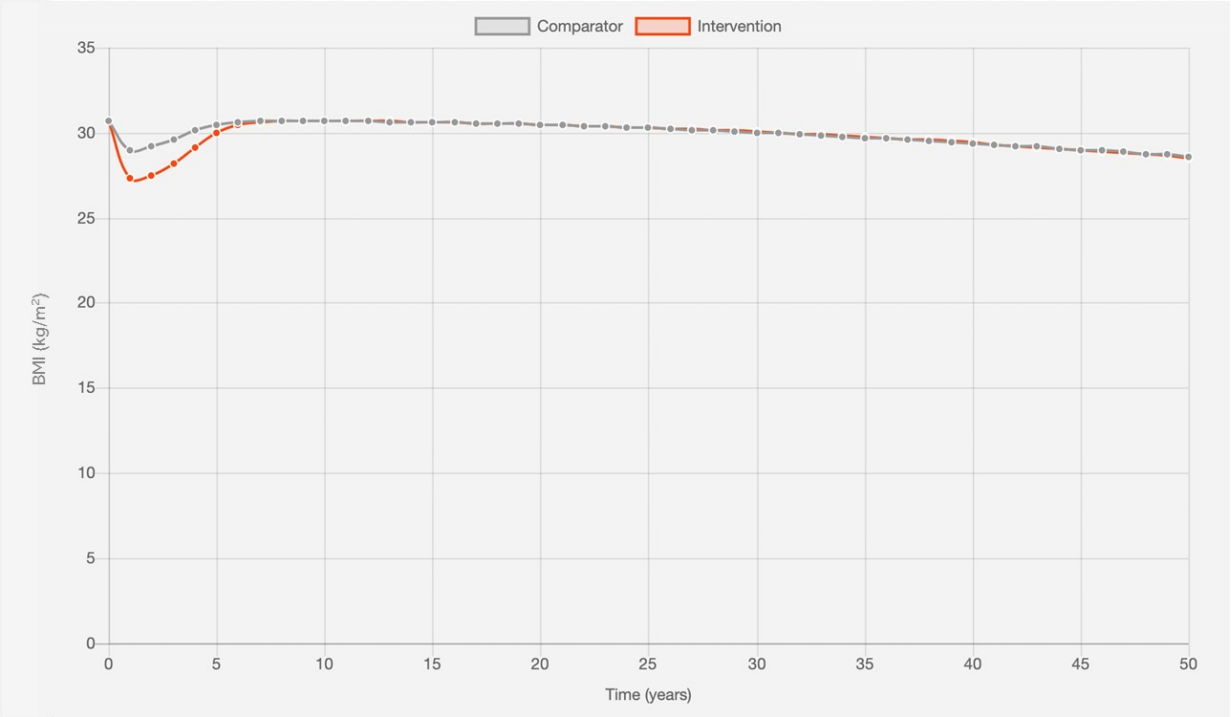


Population mean values are shown in orange for tirzepatide 10 mg (intervention arm) and in grey for semaglutide 1.0 mg (comparator arm)
Abbreviations: SBP: systolic blood pressure.

BMI

No long-term changes in BMI were applied in the modelling analysis (Figure 72). Differences between the tirzepatide and comparator population mean curves were evident in the initial years of the simulation due to the differential effects of BMI in year 1, followed by BMI returning to baseline levels when patients intensify to basal insulin therapy (primarily between years 3 and 7). In addition, a slight decrease in mean BMI is evident beyond year 25 of the simulation. Like SBP, this effect is due to patients with higher BMI levels experiencing a higher average mortality risk than those with lower BMI levels, leading to a slight decrease in the population mean in the advanced years of the simulation.

Figure 72: Long-term BMI progression in the modelling analysis

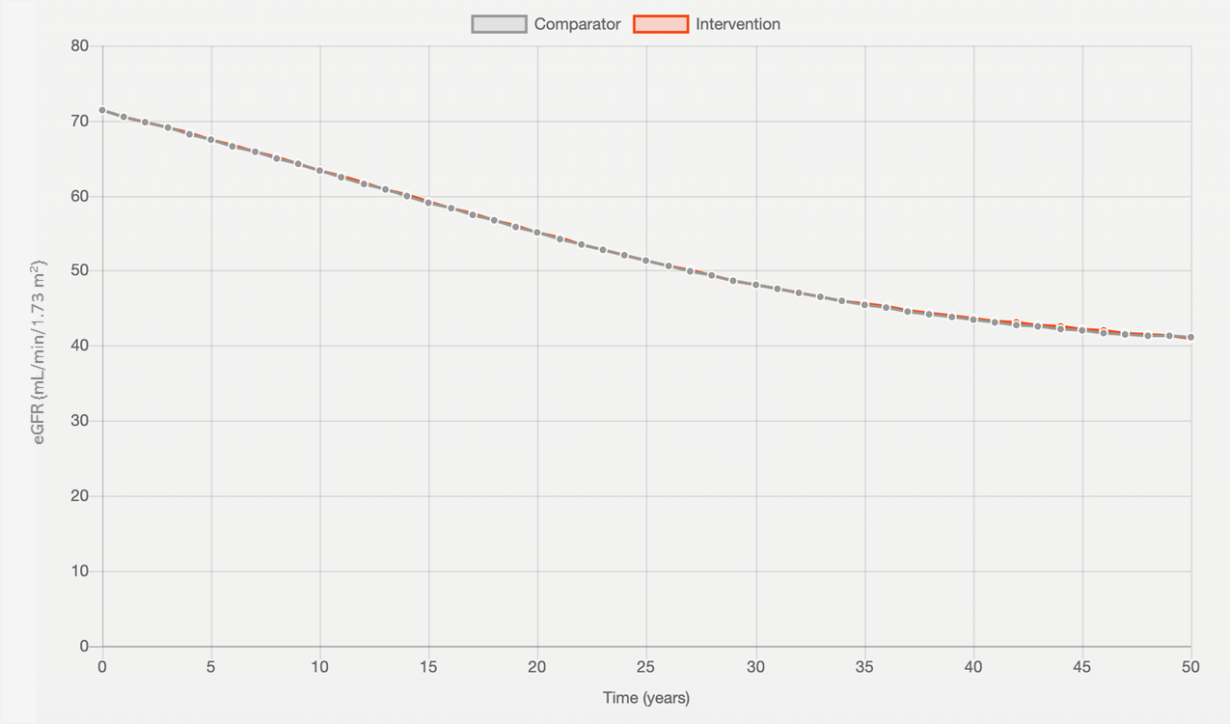


Population mean values are shown in orange for tirzepatide 10 mg (intervention arm) and in grey for semaglutide 1.0 mg (comparator arm).
Abbreviations: BMI: body mass index.

eGFR

Long-term eGFR progression in the modelling analysis was the same in both treatment arms and was based on data published by Zoppini *et al.* showing a progressive decrease in renal function over time (Figure 73).[150] This approach was preferred for the base case analysis over the UKPDS OM2 progression equation for eGFR as it represents a more clinically plausible decrease over time for a range of different baseline eGFR levels (whereas the UKPDS OM2 eGFR progression formula has all patients tending towards a mean value over time). Further details on eGFR progression are included in the PRIME T2D Model Technical Report. The impact of using the UKPDS OM2 progression equation for eGFR was investigated in sensitivity analysis (Section B.3.7.1).

Figure 73: Long-term eGFR progression in the modelling analysis

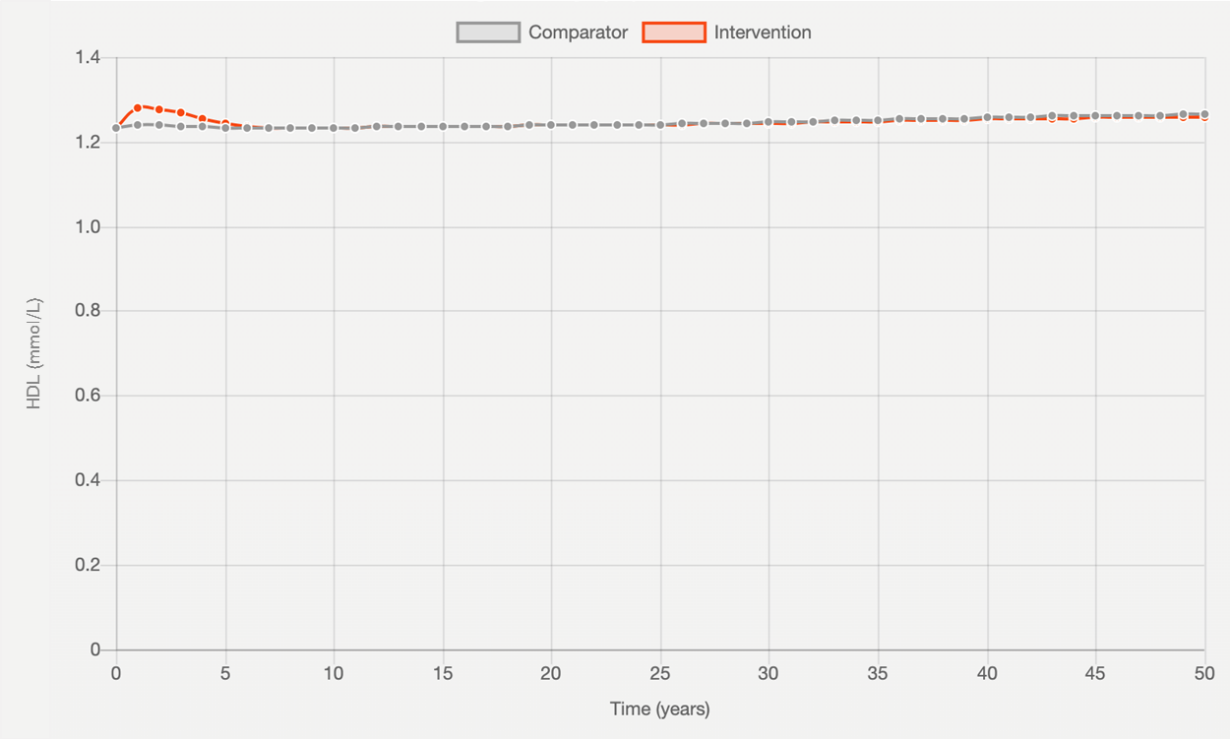


Abbreviations: eGFR: estimated glomerular filtration rate.

HDL and LDL cholesterol

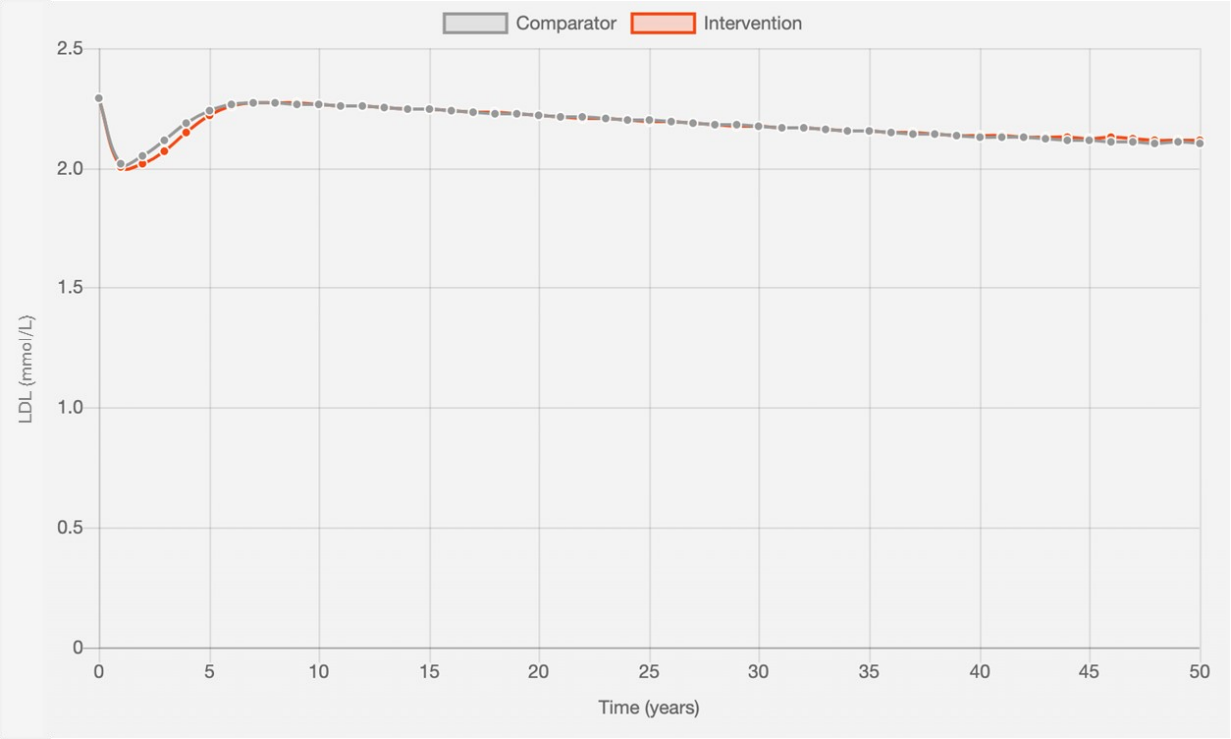
No long-term changes in HDL or LDL cholesterol were applied in the modelling analysis (Figure 74 and Figure 75). This assumption was based on the observation that risk factor progression formulae show only modest changes over time and was investigated in sensitivity analysis where UKPDS OM2 formulae were used to model progression of all risk factors over time in the simulation.[149] Changes from baseline in serum lipid levels were modest in the modelling analysis, therefore the effect of HDL or LDL levels returning to baseline on intensification to basal insulin therapy had little impact on population mean levels.

Figure 74: Long-term HDL cholesterol progression in the modelling analysis



Abbreviations: HDL: high density lipoprotein.

Figure 75: Long-term LDL cholesterol progression in the modelling analysis



Abbreviations: LDL: low density lipoprotein.

B.3.4 *Measurement and valuation of health effects*

B.3.4.1 **Health-related quality-of-life data from clinical trials**

No health-related quality of life data from clinical trials was included in the present analysis, as the clinical trials were not designed to measure health-related quality of life outcomes over long time periods.

B.3.4.2 **Mapping**

No mapping techniques were employed as part of the present analysis.

B.3.4.3 **Health-related quality-of-life studies**

A literature review was performed via searches of the PubMed, EMBASE, EconLit and Cochrane Library databases. Supplementary hand searches were also performed to identify abstracts published at major congresses of interest in 2020 through to 2022 (e.g. the virtual International Society for Pharmacoeconomics and Outcomes Research [ISPOR] meeting, Diabetes UK and the European Association for the Study of Diabetes (EASD) and American Diabetes Association (ADA) annual congresses). Hand searches of HTA agency websites and other relevant online resources were also searched (see Appendix H for details).

The literature searches across the four major databases (PubMed, EconLit, EMBASE and Cochrane Library) identified a total of 2,611 articles, including 323 duplicates, resulting in 2,288 unique hits. First round screening of titles and abstracts identified 73 hits for full text review. Additional hand searches HTA body websites identified 842 articles, 833 of which were excluded in first-round review, 1 hit was a duplicate, and 8 unique hits were identified for full text review. During second-round screening, 60 articles were excluded. The final review included 21 articles and data was extracted as described in Appendix H.

Utilities used in the modelling analysis are described in Section B.3.4.5.

B.3.4.4 **Adverse reactions**

As mentioned in Section B.2.9, GLP-1 RAs are known to be associated with GI AEs, including nausea and vomiting, in the early months of treatment.[151] Nausea rates for tirzepatide and all comparators were derived from the NMA and were assumed to negatively impact quality of life in year 1 of the simulation (see Section B.3.4.5). Nausea rates used in the modelling analysis are summarised in Table 81.

Table 81: Summary of nausea rates used in the modelling analysis

	Percentage of patients experiencing nausea	Source
Tirzepatide 5 mg	[redacted]	NMA
Tirzepatide 10 mg	[redacted]	NMA
Tirzepatide 15 mg	[redacted]	NMA
Dulaglutide 1.5 mg	[redacted]	NMA
Dulaglutide 3.0 mg	[redacted]	NMA*
Dulaglutide 4.5 mg	[redacted]	NMA*
Semaglutide 0.5 mg	[redacted]	NMA

	Percentage of patients experiencing nausea	Source
Semaglutide 1.0 mg	[redacted]	NMA
Oral semaglutide 7 mg	[redacted]	NMA*
Oral semaglutide 14 mg	[redacted]	NMA*
Liraglutide 1.2 mg	[redacted]	NMA
Liraglutide 1.8 mg	[redacted]	NMA
Basal insulin	0	Assumed

*Nausea rates based on nearest neighbour substitutions.
Abbreviations: NMA: network meta-analysis.

Rates of hypoglycaemia were not reported in the NMA due to many studies reporting zero events; therefore rates of hypoglycaemia were set to zero for tirzepatide and all comparators in the base case analysis. This assumption is likely to be a reasonable approximation for the interventions included in the present analysis based on the hypoglycaemia rates in the SURPASS trial programme and clinical studies of other T2D medications such as GLP-1 RAs.[152] For basal insulin therapy, hypoglycaemic event rates were aligned with those used in the NICE 2022 health economic report used to inform NG28. [123] The rates for severe and non-severe hypoglycaemia used in the modelling analysis were as follows (with standard deviations in parentheses, assumed to be approximately 10% of the mean value in the absence of reported variance):

- Mean annual severe hypoglycaemia rate 0.32 (0.03) events per patient year
- Mean annual non-severe hypoglycaemia rate 3.84 (0.38) events per patient year

B.3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis

Quality-adjusted life expectancy was evaluated in the modelling analysis using an additive approach using data sourced from the literature review described in Section B.3.4.3. Utilities from the review were prioritised for inclusion based on whether they were derived using the EQ-5D instrument, with preference given to UK-specific studies where available. For each simulated patient, a base utility score of 0.815 was assigned in each year they were alive in the simulation, in line with the recent NICE modelling analysis for NG28, and for each complication or AE experienced, disutilities were added to evaluate an annual quality of life utility score based on individual patient profiles. The disutilities used in the analysis are summarized in Table 82. No age-adjustment was used in the base case analysis; the inclusion of age-adjustment was explored in sensitivity analyses using the methodology of Ara and Brazier (2010) and was found to have little impact.[153]

Table 82: Utilities and disutilities used in the modelling analysis for diabetes-related complications and hypoglycaemic events

Baseline	Utility	Source
T2D with no complications	+0.815	NICE HE Report 2022 (Table HE027)[123]Error! Bookmark not defined.
Complication / adverse event	**Disutility**	**Source**
Macrovascular complications		

Baseline	Utility	Source
Myocardial infarction event	−0.055	NICE HE Report 2022 (Table HE027) [123]
History of myocardial infraction	−0.055	NICE HE Report 2022 (Table HE027) [123]Error! Bookmark not defined.
Stroke event	−0.164	NICE HE Report 2022 (Table HE027) [123]
History of stroke	−0.164	NICE HE Report 2022 (Table HE027) [123]
Ischemic heart disease (each year)	−0.090	NICE HE Report 2022 (Table HE027) [123]
Congestive heart failure (each year)	−0.108	NICE HE Report 2022 (Table HE027) [123]
Microvascular complications		
Foot ulcer (year of event)	−0.170	NICE HE Report 2022 (Table HE027) [123]
Lower extremity amputation (year of event)	−0.280	NICE HE Report 2022 (Table HE027) [123]
Lower extremity amputation (subsequent years)	−0.122	Nauck *et al.* (2019)[154]
Blindness (each year)	−0.074	NICE HE Report 2022 (Table HE027) [123]
Macular oedema (first year)	−0.047	Mitchell *et al.* (2012)[155] assumed, corresponding to best corrected visual acuity change from 76-85 to 66-75
Macular oedema (subsequent years)	0	Assumed
Neuropathy / SPSL (each years)	−0.066	Shao *et al.* (2019)[156]
Renal complications		
KDIGO CKD eGFR stage 1	0	Assumed
KDIGO CKD eGFR stage 2	0	Assumed
KDIGO CKD eGFR stage 3	−0.004	Assumed based on Nauck *et al.* (2019) [123, 154]
KDIGO CKD eGFR stage 4	−0.004	Assumed based on Nauck *et al.* (2019) [154]
KDIGO CKD eGFR stage 5	−0.164	NICE HE Report 2022 (Table HE027) [123]
Adverse events		
Severe hypoglycaemic event	−0.062	NICE HE Report 2022 (Section 2.3.5) [123]
Non-severe hypoglycaemic event	−0.005	Evans *et al.* (2013)[157]

Abbreviations: CKD: chronic kidney disease; eGFR: estimated glomerular filtration rate; KDIGO: Kidney Disease Improving Global Outcomes; SPSL: severe pressure sensation loss; T2D: type 2 diabetes.

Costs and utilities associated with KDIGO stages in the model can be assigned using the schema outlined in Table 83.[158]

Table 83: Summary of KDIGO stages in the PRIME Type 2 Diabetes Model

KDIGO stage	Corresponding AER level (assumed)

KDIGO stage 1 (eGFR ≥90 ml/min/1.73m^2)	Normalbuminuria (AER <20 µg/min)
KDIGO stages 2 and 3a (eGFR ≥45 ml/min/1.73m^2)	Microalbuminuria (AER 20-200 µg/min)
KDIGO stages 3b and 4 (eGFR ≥15 ml/min/1.73m^2)	Macroalbuminuria (AER 20-200 µg/min)
KDIGO stage 5 (eGFR <15 ml/min/1.73m^2)	ESRD (AER >200 µg/min)

Abbreviations: AER: albumin excretion rate; eGFR: estimated glomerular filtration rate; ESRD: end stage renal disease; KDIGO: Kidney Disease: Improving Global Outcomes.

In addition to disutilities relating to complications and adverse events, the effects of nausea, BMI and the device used to administer GLP-1 RAs were accounted for in the estimation of quality-adjusted life expectancy:

- For each patient experiencing nausea, a disutility of 0.04 was applied in the first year of the simulation, based on the data for nausea and vomiting reported by Matza *et al.* in 2007.[159] This was considered the most appropriate utility estimate identified by literature review for patients on GLP-1 RAs experiencing nausea adverse events
- For patients receiving tirzepatide and dulaglutide, a device utility of 0.007 was applied in the first year on treatment based on an analysis by Boye *et al.* (2019). This is based on the fact that tirzepatide will be administered using the same pen device as dulaglutide, which has shown a utility benefit over the semaglutide administration device[160]
- To capture the improvement in quality of life associated with bodyweight reductions in the first year of GLP-1 RA therapy, utilities from the Boye *et al.* (2022) study were used.[161] Linear interpolation of the utilities summarised in Table 84 were used to evaluate the impact of weight loss in year 1 of the simulation for tirzepatide and comparator treatments
- In each subsequent year (beyond year 1), the impact of bodyweight/BMI on quality of life was captured using a disutility of -0.0061 for each unit of BMI over 25. This is based on data reported by Bagust and Beale in 2005 and used in the NG28 modelling analysis.[123] Different utilities estimated the quality of life impact of *change* in BMI versus having a specific BMI health state, as literature review showed that the quality of life impact of BMI/bodyweight changes was generally much greater than the corresponding BMI/bodyweight state utilities (see Section B.3.4.3)
- A simplifying assumption of no injection-related disutility was used for all of the injectable formulations, including insulin, because this would be equal for all weekly injection formulations and potentially lead to a greater quality of life decrement with daily liraglutide injections. For oral semaglutide, the only non-injectable comparator in the modelling analysis, a utility of +0.004 was applied for each year on therapy (to improve quality of life versus injectable comparators). This value was estimated based on the single daily injection utility of 0.029, divided by 7 to compare with weekly injectables, derived from the NICE 2022 health economic report for NG28[123]

Table 84: Published utility estimates from the Boye et al. (2022) quality of life study

Pairs of Health States	Utility, mean (SD)	Difference score, mean (SD)	T-value (paired)	p-value
A. Current weight	0.797 (0.184)			
B. 2.5% less weight	0.808 (0.176)	0.011 (0.029)	5.3	<0.0001
C. 5% less weight	0.820 (0.171)	0.023 (0.050)	6.4	<0.0001
D. 10% less weight	0.839 (0.167)	0.042 (0.068)	8.7	<0.0001
E. 15% less weight	0.850 (0.158)	0.053 (0.087)	8.7	<0.0001
F. 20% less weight	0.857 (0.155)	0.060 (0.093)	9.1	<0.0001

Abbreviations: SD: standard deviation

B.3.5 *Cost and healthcare resource use identification, measurement and valuation*

All costs in the present analysis were accounted from a healthcare payer perspective. No indirect costs were included.

B.3.5.1 Intervention and comparators' costs and resource use

The annual costs associated with treatment were estimated based on June 2022 NHS Electronic Drug Tariff costs and resource use estimates aligned with the NICE modelling analysis for NG28.[123] For oral semaglutide, no cost was available from the June 2022 NHS Electronic Drug Tariff and costs were based on the values used in the NICE modelling analysis for NG28. This was assumed to be a reasonable proxy value as none of the other GLP-1 RA costs had changed from 2021 values. Annual cost estimates for each therapy are summarized in Table 86 and the unit costs used in these calculations are provided in Table 85.

Table 85: Summary of pharmacy costs used in the modelling analysis

	Weekly dose	Pack contents	Cost per pack (£)	Weekly cost (£)	Proportion	Annual cost (£)
Tirzepatide 5 mg						
Tirzepatide	1 injection	4 injections	[redacted]	[redacted]	[redacted]	[redacted]
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	0	1	0.05	0.00	0%	0.00
SMBG	Not included for tirzepatide					0.00
Initiation cost						40.33
Total in year 1						[redacted]
Total in years 2+						[redacted]
Tirzepatide 10 mg						
Tirzepatide	1 injection	4 injections	[redacted]	[redacted]	[redacted]	[redacted]
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	0	1	0.05	0.00	0%	0.00
SMBG	Not included for tirzepatide					0.00
Initiation cost						40.33
Total in year 1						[redacted]
Total in years 2+						[redacted]
Tirzepatide 15 mg						
Tirzepatide	1 injection	4 injections	[redacted]	[redacted]	[redacted]	[redacted]
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	0	1	0.05	0.00	0%	0.00
SMBG	Not included for tirzepatide					0.00
Initiation cost						40.33
Total in year 1						[redacted]
Total in years 2+						[redacted]
DULAGLUTIDE (all doses)						

	Weekly dose	Pack contents	Cost per pack (£)	Weekly cost (£)	Proportion	Annual cost (£)
Dulaglutide	1 injection	4 injections	73.25	18.31	100%	955.52
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	0	1	0.05	0.00	0%	0.00
SMBG	Not included for dulaglutide					0.00
Initiation cost						40.33
Total in year 1						1,036.03
Total in years 2+						995.70
SEMAGLUTIDE (all doses)						
Semaglutide	1 injection	4 injections	73.25	18.31	100%	955.52
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	0	1	0.05	0.00	0%	0.00
SMBG	Not included for semaglutide					0.00
Initiation cost						40.33
Total in year 1						1,036.03
Total in years 2+						995.70
ORAL SEMAGLUTIDE (all doses)						
Oral semaglutide						955.00
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	Not included for oral semaglutide					0.00
SMBG	Not included for oral semaglutide					0.00
Initiation cost						40.33
Total in year 1						1,035.51
Total in years 2+						995.18
LIRAGLUTIDE 1.2 mg						
Liraglutide	7 x 1.2 mg	6 x 3 x 2 mg	78.48	18.31	100%	955.49
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18

	Weekly dose	Pack contents	Cost per pack (£)	Weekly cost (£)	Proportion	Annual cost (£)
Needles	7	1	0.05	0.35	100%	18.26
SMBG	Not included for liraglutide 1.2 mg					0.00
Initiation cost						40.33
Total in year 1						1,054.27
Total in years 2+						1,013.93
LIRAGLUTIDE 1.8 mg						
Liraglutide	7 x 1.8 mg	6 x 3 x 2 mg	78.48	27.47	100%	1,433.24
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	7	1	0.05	0.35	100%	18.26
SMBG	Not included for liraglutide 1.8 mg					0.00
Initiation cost						40.33
Total in year 1						1,532.01
Total in years 2+						1,491.68
BASAL INSULIN						
NPH insulin	7 x 40 IU	100 x 3 x 5 IU	19.08	3.56	100%	185.84
Metformin	7 x 2000 mg	28 x 500 mg	0.77	0.77	100%	40.18
Needles	7	1	0.05	0.35	100%	18.26
SMBG	10.5 tests	1	0.26	2.73	100%	142.45
Initiation cost						141.17
Total in year 1						527.89
Total in years 2+						386.73

Abbreviations: IU: insulin units; SMBG: self-monitoring of blood glucose.

Table 86: Unit costs used in the modelling analysis

Treatment	Pack contents	Description	Cost (£)	Source
Tirzepatide 5 mg	4	pre-filled pens for weekly injection	[redacted]	
Tirzepatide 10 mg	4	pre-filled pens for weekly injection	[redacted]	
Tirzepatide 15 mg	4	pre-filled pens for weekly injection	[redacted]	
Dulaglutide 1.5mg/0.5ml solution for injection pre-filled disposable devices	4	pre-filled disposable injection	73.25	June 2022 NHS Electronic Drug Tariff
Dulaglutide 3mg/0.5ml solution for injection pre-filled disposable devices	4	pre-filled disposable injection	73.25	June 2022 NHS Electronic Drug Tariff
Dulaglutide 4.5mg/0.5ml solution for injection pre-filled disposable devices	4	pre-filled disposable injection	73.25	June 2022 NHS Electronic Drug Tariff
Semaglutide 0.5mg/0.37ml solution for injection 1.5ml pre-filled disposable device	1	pre-filled disposable injection	73.25	June 2022 NHS Electronic Drug Tariff
Semaglutide 1mg/0.74ml solution for injection 3ml pre-filled disposable device	1	pre-filled disposable injection	73.25	June 2022 NHS Electronic Drug Tariff
Oral semaglutide 7 mg	Annual cost	N/A	955	NICE 2022 HE Report, Table HE016: Unit costs of CVOT treatments
Oral semaglutide 14 mg	Annual cost	N/A	955	NICE 2022 HE Report, Table HE016: Unit costs of CVOT treatments
Liraglutide 6mg/ml solution for injection 3ml pre-filled disposable devices	2	pre-filled disposable injection	78.48	June 2022 NHS Electronic Drug Tariff
Liraglutide 6mg/ml solution for injection 3ml pre-filled disposable devices	2	pre-filled disposable injection	78.48	June 2022 NHS Electronic Drug Tariff
Metformin 500mg tablets	28	tablet	0.77	June 2022 NHS Electronic Drug Tariff
Insulin isophane human 100units/ml suspension for injection 3ml cartridges	5	vial	19.08	June 2022 NHS Electronic Drug Tariff

Abbreviations: CVOT: cardiovascular outcomes trial.

To estimate the annual costs of treatment, the following assumptions were applied:

- A simplifying assumption was made that background therapy was metformin only. As this was the same across all comparators, the costs associated with background therapy were expected to have little impact on cost-effectiveness. Sensitivity analyses were performed including sulfonylurea as background therapy (Section B.3.7.1)

- Simulated patients were assumed to take 2,000 mg metformin per day based on the World Health Organization defined daily dose (DDD) for metformin[162]
- NPH insulin was dosed at 40 IU per day in line with the World Health Organization defined daily dose estimate for insulin glargine[163]
- The costs of needles were included in the estimation of annual costs using the same assumptions applied in the health economic analysis for NG28 (Table HE022) [123]
- No self-monitoring of blood glucose (SMBG) costs were applied for tirzepatide or GLP-1 RA treatments. SMBG costs were accounted for basal insulin therapy in line with the approach used in the health economic analysis for NG28 (Table HE022), with each SMBG test assumed to cost £0.26 and patients using 10.5 tests per week[123]
- Assumptions on training/administration resource use for the initiation of tirzepatide/GLP-1 RA and basal insulin therapy were also aligned with the health economic analysis for NG28 (Tables HE021 and HE022).[123] These assumptions are also assumed to cover training/administration resource use associated with dose escalation
    - Tirzepatide or GLP-1 RA therapy was initiated with 2 x 20 minute appointments with a band 6 or band 7 nurse (weighted average cost of £55 and £66 per hour). Costs were sourced from PSSRU Unit Costs Database of Health and Social Care Professionals 2020/21[164]
    - Insulin therapy was initiated with 1 x 40 minute and 5 x 20 minute appointments with a band 6 or band 7 nurse (weighted average cost of £55 and £66 per hour). Costs were sourced from PSSRU Unit Costs Database of Health and Social Care Professionals 2020/21[164]

B.3.5.2 Health-state unit costs and resource use

A literature review was performed to identify unit costs and resource use via searches of four databases (PubMed, EconLit, EMBASE, and Cochrane Library) from 2015-2022, with supplementary hand searches to identify abstracts published at major congresses of interest in 2020 through to 2022 (e.g. the virtual International Society for Pharmacoeconomics and Outcomes Research [ISPOR] meeting, Diabetes UK and the European Association for the Study of Diabetes (EASD) and American Diabetes Association (ADA) annual congresses). Hand searches of HTA agency website and other relevant online resources were also performed (Appendix I).

The literature searches across all four databases identified a total of 1,652 articles, including 339 duplicates, resulting in 1,313 unique hits. First round screening of titles and abstracts identified 71 hits for full text review. Additional hand searches of HTA body websites identified 775 articles, 766 of which were excluded in first-round review, 5 hits were duplicates, and 4 unique hits were identified for full text review. During second-round screening, 70 articles were excluded. Data extraction was performed on five studies (Appendix I).

The cost set used in the modelling analysis is summarised in Table 87.

Table 87: Summary of direct costs associated with diabetes-related complications used in the modelling analysis

	Source cost (£)	Source year	2021 value (£)	Source
Macrovascular complications				
Myocardial infarction, year 1	8,419	2020	8,678	NICE HE Report 2022 (Table HE018) [123]
Myocardial infarction, year 2	2,093	2020	2,157	NICE HE Report 2022 (Table HE018) [123]
Stroke, year 1	9,054	2020	9,333	NICE HE Report 2022 (Table HE018) [123]
Stroke, year 2	2,157	2020	2,223	NICE HE Report 2022 (Table HE018) [123]
Ischemic heart disease, year 1	12,190	2020	12,565	NICE HE Report 2022 (Table HE018) [123]
Ischemic heart disease, year 2	2,143	2020	2,209	NICE HE Report 2022 (Table HE018) [123]
Congestive heart failure, year 1	4,782	2020	4,929	NICE HE Report 2022 (Table HE018) [123]
Congestive heart failure, year 2	2,805	2020	2,891	NICE HE Report 2022 (Table HE018) [123]
Microvascular complications				
Foot ulcer, year 1	3,620	2020	3,731	NICE HE Report 2022 (Table HE018) [123]
Foot ulcer, year 2	0	2021	0	Assumed
Amputation, year 1	14,041	2020	14,473	NICE HE Report 2022 (Table HE018) [123]
Amputation, year 2	3,902	2020	4,022	NICE HE Report 2022 (Table HE018) [123]
Blindness, year 1	3,606	2020	3,717	NICE HE Report 2022 (Table HE018) [123]
Blindness, year 2	1,366	2020	1,408	NICE HE Report 2022 (Table HE018) [123]
Macular oedema	661	2020	681	NHS reference costs 2019/2020*
Neuropathy/SPSL, all years	1,082	2016	293	Hunt et al. (2017)[165]
Renal complications				

	Source cost (£)	Source year	2021 value (£)	Source
KDIGO CKD eGFR stage	0	2021	0	Assumed
KDIGO CKD eGFR stage 2	0	2021	0	Assumed
KDIGO CKD eGFR stage 3	0	2021	0	Assumed
KDIGO CKD eGFR stage 4	393	2011	465	Kent et al. (2015)[166]
KDIGO CKD eGFR stage 5	20,897	2020	21,541	NICE HE Report 2022 (Table HE018) [123]

Abbreviations: CKD: chronic kidney disease; eGFR: estimated glomerular filtration rate; SPSL: severe pressure sensation loss; KDIGO: Kidney Disease Improving Global Outcomes.
*Day Case, BZ87A, Minor vitreous retinal procedures, 19 years and over.[167]

B.3.5.3 Adverse reaction unit costs and resource use

It was assumed that there were no direct costs associated with nausea for the modelling analysis. Severe hypoglycaemia costs were based on the recent NICE modelling analysis for NG28 (Table 88). [123] Hypoglycaemia costs were applied per event.

Table 88: Summary of direct costs associated with adverse events

Adverse event	Source cost (£)	Source year	2021 value (£)	Source
Nausea	0	2021	0	Assumed
Severe hypoglycaemic event	373	2020	384	NICE HE Report 2022 (Table HE023) [123]
Non-severe hypoglycaemic event	0	2020	0	NICE HE Report 2022 (Table HE023) [123]

Abbreviations: HE: health economics; NICE: National Institute of Health and Care Excellence.

B.3.6 *Severity*

No severity weights were used in the evaluation of quality-adjusted life expectancy in the present analysis.

B.3.7 *Uncertainty*

B.3.7.1 Sensitivity analysis

In addition to PSA (see the PRIME T2D Model Technical Report for details), one-way and multi-way sensitivity analyses were performed using the same first order Monte Carlo simulation approach as in the base case analysis (Table 89). After running the base case simulations (Section B.3.10), it was decided to perform sensitivity analyses focused on the comparisons of tirzepatide with semaglutide, as semaglutide was the most cost-effective of the comparators examined in the base case. The rationale for this approach was that key drivers of outcomes in the modelling analysis would be similar across all comparators, but were most likely to affect the cost-effectiveness of tirzepatide in comparison with semaglutide.

PSA was performed for all 12 base case simulations. The model generated results based on a nonparametric bootstrapping approach, in which samples from 1% of the simulated population (in this case comprising 3,000 patients) were drawn 1,000 times from the full patient data set at the end of the simulation. Sampling was performed with replacement. The population mean and confidence intervals were then calculated by rerunning the cost and quality-of-life estimators on each sampled population and generating a set of descriptive statistics.

Table 89: Summary of base case, one-way and multi-way sensitivity analyses

Number	Intervention	Comparator	Description
Base case			
1	Tirzepatide 5 mg	Dulaglutide 1.5 mg	Base case
2	Tirzepatide 5 mg	Semaglutide 0.5 mg	Base case
3	Tirzepatide 5 mg	Oral semaglutide 7 mg	Base case
4	Tirzepatide 5 mg	Liraglutide 1.2 mg	Base case
5	Tirzepatide 10 mg	Dulaglutide 3.0 mg	Base case
6	Tirzepatide 10 mg	Semaglutide 1.0 mg	Base case
7	Tirzepatide 10 mg	Oral semaglutide 14 mg	Base case
8	Tirzepatide 10 mg	Liraglutide 1.8 mg	Base case
9	Tirzepatide 15 mg	Dulaglutide 4.5 mg	Base case
10	Tirzepatide 15 mg	Semaglutide 1.0 mg	Base case
11	Tirzepatide 15 mg	Oral semaglutide 14 mg	Base case
12	Tirzepatide 15 mg	Liraglutide 1.8 mg	Base case
Probabilistic sensitivity analysis (PSA)			
13	Tirzepatide 5 mg	Dulaglutide 1.5 mg	Base case
14	Tirzepatide 5 mg	Semaglutide 0.5 mg	Base case
15	Tirzepatide 5 mg	Oral semaglutide 7 mg	Base case
16	Tirzepatide 5 mg	Liraglutide 1.2 mg	Base case
17	Tirzepatide 10 mg	Dulaglutide 3.0 mg	Base case
18	Tirzepatide 10 mg	Semaglutide 1.0 mg	Base case
19	Tirzepatide 10 mg	Oral semaglutide 14 mg	Base case
20	Tirzepatide 10 mg	Liraglutide 1.8 mg	Base case
21	Tirzepatide 15 mg	Dulaglutide 4.5 mg	Base case
22	Tirzepatide 15 mg	Semaglutide 1.0 mg	Base case
23	Tirzepatide 15 mg	Oral semaglutide 14 mg	Base case
24	Tirzepatide 15 mg	Liraglutide 1.8 mg	Base case
Housekeeping (only run for comparisons of tirzepatide versus semaglutide)			
25	Tirzepatide 5 mg	Semaglutide 0.5 mg	5-year time horizon
26	Tirzepatide 5 mg	Semaglutide 0.5 mg	10-year time horizon
27	Tirzepatide 5 mg	Semaglutide 0.5 mg	15-year time horizon
28	Tirzepatide 5 mg	Semaglutide 0.5 mg	20-year time horizon
29	Tirzepatide 10 mg	Semaglutide 1.0 mg	5-year time horizon
30	Tirzepatide 10 mg	Semaglutide 1.0 mg	10-year time horizon
31	Tirzepatide 10 mg	Semaglutide 1.0 mg	15-year time horizon
32	Tirzepatide 10 mg	Semaglutide 1.0 mg	20-year time horizon

Number	Intervention	Comparator	Description
33	Tirzepatide 15 mg	Semaglutide 1.0 mg	5-year time horizon
34	Tirzepatide 15 mg	Semaglutide 1.0 mg	10-year time horizon
35	Tirzepatide 15 mg	Semaglutide 1.0 mg	15-year time horizon
36	Tirzepatide 15 mg	Semaglutide 1.0 mg	20-year time horizon
37	Tirzepatide 5 mg	Semaglutide 0.5 mg	6% discount rates on future clinical and cost benefits
38	Tirzepatide 5 mg	Semaglutide 0.5 mg	No discounting on future clinical and cost benefits
39	Tirzepatide 10 mg	Semaglutide 1.0 mg	6% discount rates on future clinical and cost benefits
40	Tirzepatide 10 mg	Semaglutide 1.0 mg	No discounting on future clinical and cost benefits
41	Tirzepatide 15 mg	Semaglutide 1.0 mg	6% discount rates on future clinical and cost benefits
42	Tirzepatide 15 mg	Semaglutide 1.0 mg	No discounting on future clinical and cost benefits
43	Tirzepatide 5 mg	Semaglutide 0.5 mg	SURPASS-2 cohort
44	Tirzepatide 10 mg	Semaglutide 1.0 mg	SURPASS-2 cohort
45	Tirzepatide 15 mg	Semaglutide 1.0 mg	SURPASS-2 cohort
Clinical drivers (only run for tirzepatide 10 mg versus semaglutide 1.0 mg)			
46	Tirzepatide 10 mg	Semaglutide 1.0 mg	No HbA1c difference between treatments (tirzepatide HbA1c changed matched to SEMA)
47	Tirzepatide 10 mg	Semaglutide 1.0 mg	No SBP difference treatments (tirzepatide SBP changed matched to SEMA)

Number	Intervention	Comparator	Description
48	Tirzepatide 10 mg	Semaglutide 1.0 mg	No serum lipids difference between treatments (tirzepatide serum lipid levels changed matched to SEMA value)
49	Tirzepatide 10 mg	Semaglutide 1.0 mg	No BMI difference treatments (tirzepatide BMI changed matched to SEMA)
50	Tirzepatide 10 mg	Semaglutide 1.0 mg	Only HbA1c difference between treatments (all other risk factor changes matched to SEMA values)
51	Tirzepatide 10 mg	Semaglutide 1.0 mg	Only BMI difference between treatments (all other risk factor changes matched to SEMA values)
52	Tirzepatide 10 mg	Semaglutide 1.0 mg	Only HbA1c and BMI differences between treatments (all other risk factor changes matched to SEMA values)
Duration of therapy			
53	Tirzepatide 10 mg	Semaglutide 1.0 mg	Treatment intensification (switch to insulin) after 3 years in both treatment arms
54	Tirzepatide 10 mg	Semaglutide 1.0 mg	Treatment intensification (switch to insulin) after 5 years in both treatment arms
55	Tirzepatide 10 mg	Semaglutide 1.0 mg	Include second intensification step from basal insulin to basal-bolus therapy when HbA1c reaches 7.5% for a second time
56	Tirzepatide 10 mg	Semaglutide 1.0 mg	Treatment intensification when HbA1c is higher than 8.5%
57	Tirzepatide 10 mg	Semaglutide 1.0 mg	Treatment intensification when HbA1c is higher than 9.5%
Quality of life utilities			
58	Tirzepatide 10 mg	Semaglutide 1.0 mg	No weight change utility (only BMI state utilities applied)
59	Tirzepatide 10 mg	Semaglutide 1.0 mg	No weight/BMI utilities
60	Tirzepatide 10 mg	Semaglutide 1.0 mg	No device utilities
61	Tirzepatide 10 mg	Semaglutide 1.0 mg	No nausea utilities
62	Tirzepatide 10 mg	Semaglutide 1.0 mg	No hypoglycaemia utilities

Number	Intervention	Comparator	Description
63	Tirzepatide 10 mg	Semaglutide 1.0 mg	QALY age-adjustment based on Ara and Brazier
64	Tirzepatide 10 mg	Semaglutide 1.0 mg	Multiplicative approach to combining utilities
Other base case assumptions			
65	Tirzepatide 10 mg	Semaglutide 1.0 mg	Cohort ethnic groups settings changed from Black to Afro-Caribbean (to use UKPDS OM2 risk adjustment)
66	Tirzepatide 10 mg	Semaglutide 1.0 mg	Include sulfonylurea as background therapy (with metformin) in both treatment arms
67	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use change in BMI values directly from the NMA (as opposed to BMI calculated based on weight change)
68	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use UKPDS OM2 risk factor progression functions for all risk factors
69	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use complication costs identified by literature review (alternative cost set)
70	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use UKPDS OM2 risk equation to estimate renal failure
71	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use UKPDS OM2 risk equation and UKPDS OM2 eGFR progression to estimate renal failure
72	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use only UKPDS OM2 equations to estimate mortality risk
73	Tirzepatide 10 mg	Semaglutide 1.0 mg	Use only cause-subtracted life tables to estimate mortality risk

Abbreviations: BMI: body mass index; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; SBP: systolic blood pressure; SEMA: semaglutide; UKPDS OM2: United Kingdom Prospective Diabetes Study Outcomes Model 2.

B.3.7.2 Scenario analysis

Two scenario analysis simulations were performed. The first was designed to understand the effect of using model input data from the SURPASS-2 head-to-head trial of tirzepatide versus semaglutide 1.0 mg. The second scenario analysis was performed to evaluate the influence of a decision to continue tirzepatide or semaglutide therapy after the initiation of basal insulin on cost-effectiveness.

B.3.7.2.1 SURPASS-2 model inputs

For the SURPASS-2 based analysis, cohort characteristics and treatment effects for tirzepatide 10 mg and semaglutide 1.0 mg were derived from the SURPASS-2 CSR. All other settings and

assumptions in this scenario analysis matched the base case analysis. Simulation cohort characteristics based on SURPASS-2 are summarized in Table 90. Changes in risk factors associated with tirzepatide 10 mg and semaglutide 1.0 mg are summarized in Table 91. Severe hypoglycaemia rates for both treatments were set to zero (as no severe hypoglycaemic events were reported in the trial), but semaglutide 1.0 mg was associated with a non-severe hypoglycaemia rate of 0.057 events per patient per year based on incidence of events with blood glucose <54 mg/dL (3.0 mmol/L) reported in the CSR. The equivalent non-severe hypoglycaemia rate for tirzepatide 10 mg was zero. The proportion of patients experiencing nausea in the first year on therapy of the modelling analysis was 19.2% on tirzepatide 10 mg and 17.9% on semaglutide 1.0 mg based on values reported in the CSR.

Table 90: Summary of cohort characteristics in the SURPASS-2 scenario analysis

	Mean	Standard deviation	Source
Demographics			
Percentage male (%)	47.0	Not applicable	SURPASS-2 CSR
Percentage with college education or higher (%)	25.97	Not applicable	PRIME Model index value[145]
Percentage smokers (%)	17.0%	Not applicable	THIN second intensification cohort (Table HE005)
Age (years)	56.6	10.4	SURPASS-2 CSR
Duration of diabetes (years)	8.6	6.5	SURPASS-2 CSR
Ethnic group			
Percentage White (%)	82.6	Not applicable	SURPASS-2 CSR
Percentage Black (%)	4.2	Not applicable	SURPASS-2 CSR
Percentage Hispanic (%)	0.0	Not applicable	Assumed
Percentage Southeast Asian (%)	1.3	Not applicable	SURPASS-2 CSR
Percentage Indian (%)	0	Not applicable	Assumed
Percentage Afro/Caribbean (%)	0	Not applicable	Assumed
Percentage Other (%)	11.9	Not applicable	Assumed
Baseline risk factors			
HbA1c (%)	8.28	1.03	SURPASS-2 CSR
Systolic blood pressure (mmHg)	130.6	13.8	SURPASS-2 CSR
Total cholesterol (mmol/L)	4.41	1.10	SURPASS-2 CSR
Low density lipoprotein cholesterol (mmol/L)	2.27	0.57	SURPASS-2 CSR
High density lipoprotein cholesterol (mmol/L)	1.11	0.28	SURPASS-2 CSR
Body mass index (kg/m2)	34.2	6.9	SURPASS-2 CSR
Estimated glomerular filtration rate (mL/min/1.73 m2)	96.0	17.10	SURPASS-2 CSR
White blood cell count (106 cells/mL)	6.8	1.8	PRIME Model index value[168]
Heart rate (beats per minute)	74.8	10.1	SURPASS-2 CSR
Haemoglobin (g/dL)	14.5	13.0	PRIME Model index value[168]

	Mean	Standard deviation	Source
Complication history			
Percentage with atrial fibrillation at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with urinary albumin ≥50mg/L at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with peripheral vascular disease at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with history of myocardial infarction at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with history of stroke at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with ischemic heart disease at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with coronary revascularization at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with heart failure at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with foot ulcer at baseline (%)	0%	Not applicable	Assumed (not reported in the SURPASS-2 CSR)
Percentage with amputation at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with blindness at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with renal failure at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR
Percentage with SPSL/neuropathy at baseline (%)	[redacted]	Not applicable	SURPASS-2 CSR

Abbreviations: SPSL: severe pressure sensation loss.

Table 91: Treatment effects applied in the first year of the scenario analysis based on SURPASS-2

	Tirzepatide 10 mg mean (SD)	Semaglutide 1.0 mg mean (SD)	Source
HbA1c change from baseline (%)	−2.37 (0.97)	−1.86 (0.98)	SURPASS-2 CSR, ITT population, efficacy estimand
SBP change from baseline (mmHg)	−5.3 (12.2)	−3.60 (12.2)	SURPASS-2 CSR, ITT population
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	SURPASS-2 CSR, ITT population, efficacy estimand
HDL change from baseline (mmol/L)	0.09 (0.18)	0.05 (0.18)	SURPASS-2 CSR, ITT population
LDL change from baseline (mmol/L)	−0.13 (0.68)	−0.14 (0.69)*	SURPASS-2 CSR, ITT population
eGFR change from baseline ($ml/min/1.73m^2$)	[redacted]	[redacted]	SURPASS-2 CSR, ITT population

Abbreviations: BMI: body mass index; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; SBP: systolic blood pressure; SD: standard deviation.

B.3.7.2.2 Intensification of therapy by adding basal insulin

To investigate the scenario where patients continued tirzepatide or GLP-1 RA therapy after initiation of basal insulin, three step treatment algorithms were created in the model for tirzepatide 10 mg and semaglutide 1 mg as follows:

- **Step 1:** Tirzepatide or semaglutide therapy with metformin background as per the base case analysis
- **Step 2:** Tirzepatide or semaglutide therapy with metformin background in combination with basal insulin. HbA1c was decreased by 0.84% on basal insulin initiation as per the base case analysis. All other risk factors remained unchanged on addition of basal insulin. Costs were adjusted to include the costs of NPH insulin, needles, SMBG and training (first year of basal insulin treatment only)
- **Step 3:** Basal bolus insulin therapy was initiated with tirzepatide or semaglutide and metformin stopped. HbA1c was decreased 0.24% based on the Willis et al. (2017) "all" formula for insulin experienced patients and all other risk factors returned to base line levels. Basal bolus insulin costs were evaluated assuming daily doses of 40 IU NPH insulin and 40 IU insulin in line with World Health Organization DDD estimates, as outlined in Table 92

Table 92: Costs associated with basal bolus insulin therapy in the scenario analysis of continuing tirzepatide/semaglutide after intensification

	Weekly dose	Pack contents	Cost per pack (£)	Weekly cost (£)	Proportion	Annual cost (£)
NPH insulin	7 x 40 IU	100 x 3 x 5 IU	19.08	3.56	100%	185.84
Insulin aspart	7 x 40 IU	100 x 10 x 1 IU	14.08	3.94	100%	205.50
Needles	14	1	0.05	0.70	100%	36.53
SMBG	10.5 tests	1	0.26	2.73	100%	142.45
Initiation cost*						40.33
Total in year 1						**610.65**
Total in years 2+						**570.31**

*Assumed to be 2 x 20 minute training sessions with a band 6/7 nurse as basal bolus was a progression from basal insulin therapy. Unit costs were taken from June, 2022 NHS Electronic Drug Tariff costs and estimates aligned with the NICE modelling analysis for NG28.[123]
Abbreviations: IU: insulin units; SMBG: self-monitoring of blood glucose.

B.3.7.3 Stability analysis

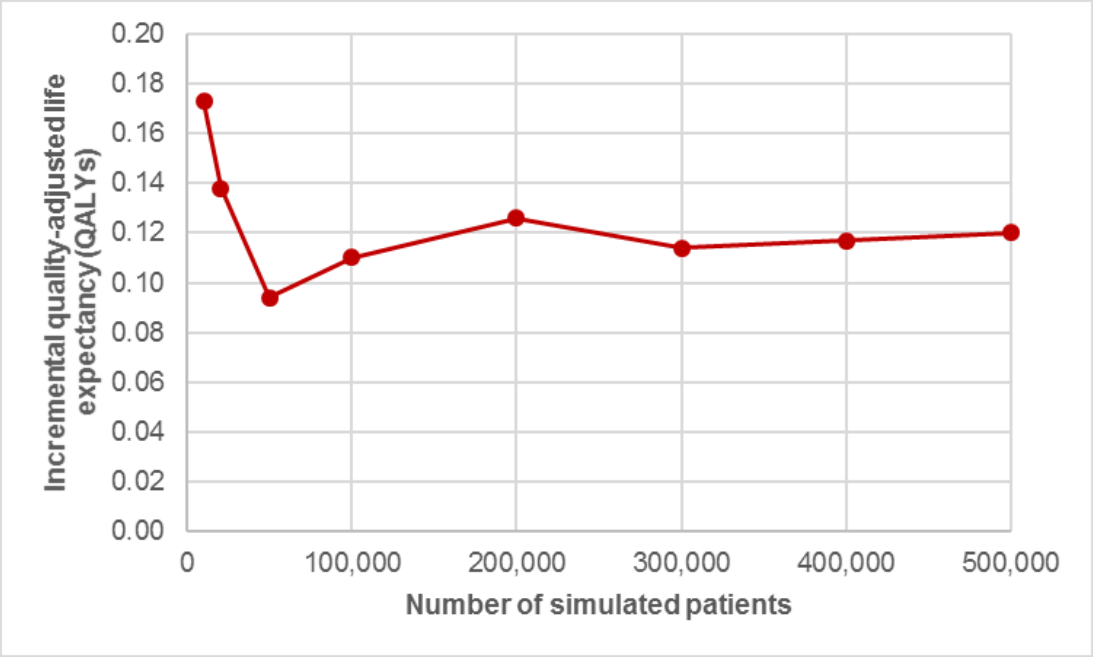
A stability analyses was performed to identify the number of iterations (i.e. the number of simulated patients) required for long-term outcomes to be stable (i.e. where the effect of random statistical variation is not a key driver of outcomes) in the base case analysis. Simulations were performed using the base case settings for the comparison of tirzepatide 10 mg with semaglutide 1.0 mg, with the number of simulated patients increased from 10,000 through to 500,000 in each treatment arm. Summary cost-effectiveness outcomes from these simulations are shown in Figure 76, Figure 77 and Figure 78.

Graphical analysis showed that, with patient numbers of 200,000 and above, the outcomes of the simulations appeared to stabilise. On more detailed analysis of the simulation outputs from 200,000 to 500,000 patients, it was decided to use a patient number of 300,000 for cost-effectiveness simulations because total costs and quality-adjusted life expectancy in both treatment arms were within 0.15% of the median estimates across these four simulations. This ensured a balance between computational time and the stability of model outputs (Table 93).

Figure 76: Incremental direct costs by number of simulated patients for tirzepatide 10 mg versus semaglutide 1.0 mg (stability analysis)

Abbreviations: GBP: Great British Pounds.

Figure 77: Incremental quality-adjusted life expectancy by number of simulated patients for tirzepatide 10 mg versus semaglutide 1.0 mg (stability analysis)



Abbreviations: QALY, quality-adjusted life year.

Figure 78: Incremental cost-effectiveness ratio by number of simulated patients for tirzepatide 10 mg versus semaglutide 1.0 mg (stability analysis)

[redacted]

Abbreviations: QALY, quality-adjusted life year; TZP: tirzepatide.

Table 93: Summary of stability analysis outcomes (200,000 to 500,000 simulated patients)

Number of simulated patients / Statistic	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	SEMA 1.0 mg	Tirzepatide 10 mg	Difference	SEMA 1.0 mg	Tirzepatide 10 mg	Difference	
200,000	[redacted]	[redacted]	[redacted]	9.419	9.544	0.126	[redacted]
300,000	[redacted]	[redacted]	[redacted]	9.421	9.535	0.114	[redacted]
400,000	[redacted]	[redacted]	[redacted]	9.429	9.546	0.117	[redacted]
500,000	[redacted]	[redacted]	[redacted]	9.419	9.539	0.120	[redacted]
Minimum	[redacted]	[redacted]	[redacted]	9.419	9.523	0.094	[redacted]
Maximum	[redacted]	[redacted]	[redacted]	9.448	9.621	0.173	[redacted]
Median	[redacted]	[redacted]	[redacted]	9.430	9.545	0.119	[redacted]
Mean	[redacted]	[redacted]	[redacted]	9.430	9.554	0.124	[redacted]

Abbreviations: ICER; incremental cost-effectiveness ratio; SEMA: semaglutide; QALY, quality-adjusted life year.

B.3.8 *Managed access proposal*

Not applicable.

B.3.9 *Summary of base-case analysis inputs and assumptions*

B.3.9.1 Summary of base-case analysis inputs

For the base case analysis, the model simulated 300,000 patients in each treatment arm with the inputs described in Table 94. The simulation was a first order Monte Carlo simulation (see the PRIME T2D Model Technical Report for details) and PSA was performed separately.

Table 94: Summary of base case analysis inputs

Model input	Setting	Section / Report
Cohort	THIN second intensification cohort (supplemented by SURPASS-2 data where required)	Section B.3.2.2
Model	PRIME T2D Model	PRIME T2D Model Technical Report
Treatment effects	NMA	Section B.3.3.1
Treatment intensification	Intensify to basal insulin when HbA1c is above 7.5%	Section B.3.3.2
Risk factor progression	UKPDS progression for HbA1c, other risk factors remain constant, except eGFR which declines over time	Section B.3.3.3
Utilities	Various sources based on literature review	Section B.3.4.5
Costs	Various sources based on literature review	Section B.3.5.2
Discount rates (clinical / costs)	3.5% / 3.5%	NICE Health Technology Evaluations Manual[169]
Mortality estimation	Hybrid approach using UKPDS OM2 mortality equations for mortality risk following complications and cause-subtracted UK life tables for mortality from other causes	PRIME T2D Model Technical Report

Abbreviations: eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; NMA: network meta-analysis; UKPDS OM2, United Kingdom Prospective Diabetes Study Outcomes Model 2.

B.3.9.2 Assumptions

Key assumptions used in the base case analysis are summarised below:

- It was assumed that the treatment effects derived from the NMA would be of a similar magnitude in a UK population close to the THIN second intensification cohort. In the absence of UK-specific data for tirzepatide versus multiple comparators, this approach uses the best available data and is aligned with the approach used in previous technology appraisals

- It was assumed that HbA1c progression would follow a curve similar to that described by UKPDS researchers. This assumption may not hold, particularly following intensification to insulin therapy, where doses can be titrated to maintain glycaemic control. The UKPDS progression assumption was used to align with a previous health economic analysis by NICE (NG28). It is noteworthy that, following intensification, the impact of HbA1c progression assumptions on cost-effectiveness is minimal in the present analysis as it is similar in both treatment arms
- It was assumed that patients would intensify therapy when HbA1c rose above 7.5%. Whilst this threshold is aligned with published guidance, it may be higher in routine clinical practice as individualised HbA1c targets are recommended. This base case assumption is in line with previous NICE analyses (NG28) and has been explored in sensitivity analyses (Section B.3.7.1)
- Upon intensification, it was conservatively assumed that BMI and other risk factors would return to baseline levels in both treatment arms. In the absence of data on the durability of treatment effects following intensification, this may be the most appropriate assumption available
- It was assumed that patients would stop tirzepatide or comparator therapy when intensifying to basal insulin based on NG28 recommendations. However, there is a clear clinical rationale for continuing tirzepatide or GLP-1 RA therapy with basal insulin. This assumption was explored in scenario analyses (Section B.3.7.1)
- A simplifying assumption of only one treatment intensification step was assumed in the base case analysis. This assumption minimised model complexity in an aspect of the analysis that had little impact on cost-effectiveness. This assumption was explored in sensitivity analyses (Section B.3.7.1)

B.3.10 *Base case results*

B.3.10.1 Base-case incremental cost-effectiveness analysis results

Long-term projections with the PRIME T2D Model indicated that use of all three doses of tirzepatide was associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 95, Table 96 and Table 97). Tirzepatide 5 mg was associated with greater lifetime direct costs than the four comparators, with incremental costs ranging between £[redacted] and £[redacted] and incremental cost-effectiveness ratios (ICERs) ranging between £[redacted] and £[redacted] per QALY gained (Table 95). Tirzepatide 10 mg was also associated with higher direct costs than three comparators, but was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg. ICERs for tirzepatide 10 mg ranged between £[redacted] and £[redacted] per QALY gained (Table 96). A similar pattern of results was projected for tirzepatide 15 mg, which was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg and had ICERs ranging between £[redacted] and £[redacted] per QALY gained versus the other comparators (Table 97).

Table 95: Summary of base case results for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg	[redacted]	13.132	9.488	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.053	9.347	[redacted]	+0.078	+0.140	[redacted]
Semaglutide 0.5 mg	[redacted]	13.074	9.374	[redacted]	+0.057	+0.114	[redacted]
Oral semaglutide 7 mg	[redacted]	13.030	9.319	[redacted]	+0.101	+0.169	[redacted]
Liraglutide 1.2 mg	[redacted]	13.022	9.310	[redacted]	+0.110	+0.178	[redacted]

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 96: Summary of base case results for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 10 mg	[redacted]	13.138	9.535	--	--	--	--
Dulaglutide 3.0 mg	[redacted]	13.063	9.377	[redacted]	+0.075	+0.157	[redacted]
Semaglutide 1.0 mg	[redacted]	13.092	9.421	[redacted]	+0.046	+0.114	[redacted]
Oral semaglutide 14 mg	[redacted]	13.078	9.388	[redacted]	+0.060	+0.147	[redacted]
Liraglutide 1.8 mg	[redacted]	13.025	9.320	[redacted]	+0.113	+0.214	[redacted]

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 97: Summary of base case results for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 15 mg	[redacted]	13.165	9.581	--	--	--	--
Dulaglutide 4.5 mg	[redacted]	13.087	9.406	[redacted]	+0.078	+0.174	[redacted]
Semaglutide 1.0 mg	[redacted]	13.092	9.421	[redacted]	+0.073	+0.160	[redacted]
Oral semaglutide 14 mg	[redacted]	13.078	9.388	[redacted]	+0.087	+0.193	[redacted]
Liraglutide 1.8 mg	[redacted]	13.025	9.320	[redacted]	+0.141	+0.260	[redacted]

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator.

The base case cost-effectiveness analysis data was also used to generate cost-effectiveness frontiers for each dose of tirzepatide (Figure 79, Figure 80 and Figure 81). In all three cases, the frontier was found between tirzepatide and semaglutide, with all other comparators above (to the North West of) the frontier represented by the ICERs of £[redacted] per QALY gained for tirzepatide 5 mg versus semaglutide 0.5 mg, £[redacted] per QALY gained for tirzepatide 10 mg versus semaglutide 1.0 mg, and £[redacted] per QALY gained for tirzepatide 15 mg versus semaglutide 1 mg.

PSA was used to investigate the statistical uncertainty around the base case ICERs and is presented in Section B.3.11.2. Based on the finding of the base case analysis, basic sensitivity analysis was performed on the three comparisons of tirzepatide with semaglutide (Section B.3.11.3):

- Tirzepatide 5 mg versus semaglutide 0.5 mg
- Tirzepatide 10 mg versus semaglutide 1.0 mg
- Tirzepatide 15 mg versus semaglutide 1.0 mg

Additional sensitivity analyses simulations performed to identify key drivers of cost-effectiveness were performed for the tirzepatide 10 mg versus semaglutide 1.0 mg comparison, as this had the highest pairwise ICER and was anticipated to be most sensitive to changes in base case assumptions.

Figure 79: Cost-effectiveness frontier for tirzepatide 5 mg versus comparators

The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: GBP: Great British Pounds; QALY: quality-adjusted life year; TZP: tirzepatide.

Figure 80: Cost-effectiveness frontier for tirzepatide 10 mg versus comparators

The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: GBP: Great British Pounds; QALY: quality-adjusted life year; TZP: tirzepatide.

Figure 81: Cost-effectiveness frontier for tirzepatide 15 mg versus comparators

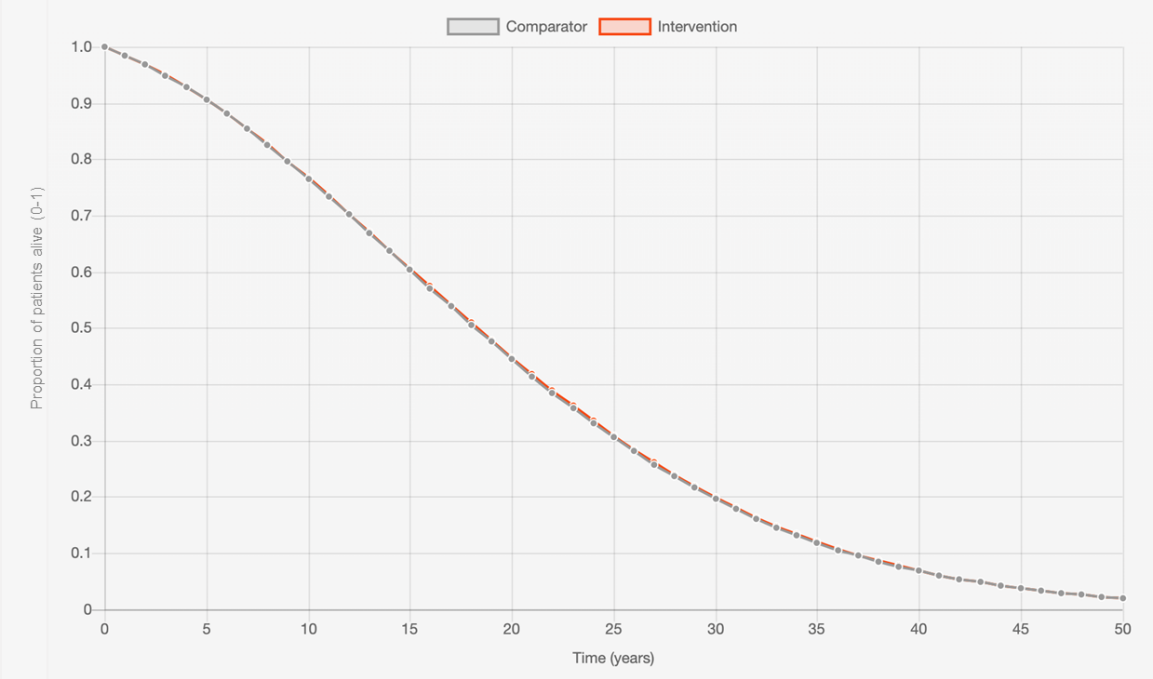
The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: QALY: quality-adjusted life year; TZP: tirzepatide.

B.3.10.1.1 Clinical outcomes from the base case analysis

Clinical outcomes from the base case analysis are described here only for the comparison of tirzepatide 10 mg with semaglutide 1.0 mg; results for other comparisons are presented in Appendix J.

Survival curves demonstrated only modest separation between tirzepatide 10 mg and semaglutide 1.0 mg during years 15–30 of the modelling analysis, consistent with the modest difference in life expectancy (0.046 years) in the base case (Figure 82).

Figure 82: Population mean survival curves from the base case analysis



Population mean values are shown in orange for tirzepatide 10 mg (intervention arm) and in grey for semaglutide 1.0 mg (comparator arm)

Reductions in diabetes-related complications associated with reductions in HbA1c and BMI on tirzepatide 10 mg therapy are summarised in Table 98. Modest reductions in cumulative incidence were observed for most diabetes-related complications, with the exceptions of renal failure (where risk was based on eGFR progression, which was the same in both treatment arms) and ischaemic heart disease, blindness, and amputation (where survival paradox and the modest impact of HbA1c as a risk factor meant the incidences of these complications were similar in both treatment arms. The development of diabetes-related complications over time in the simulation is summarised in Figure 83 through to Figure 87.

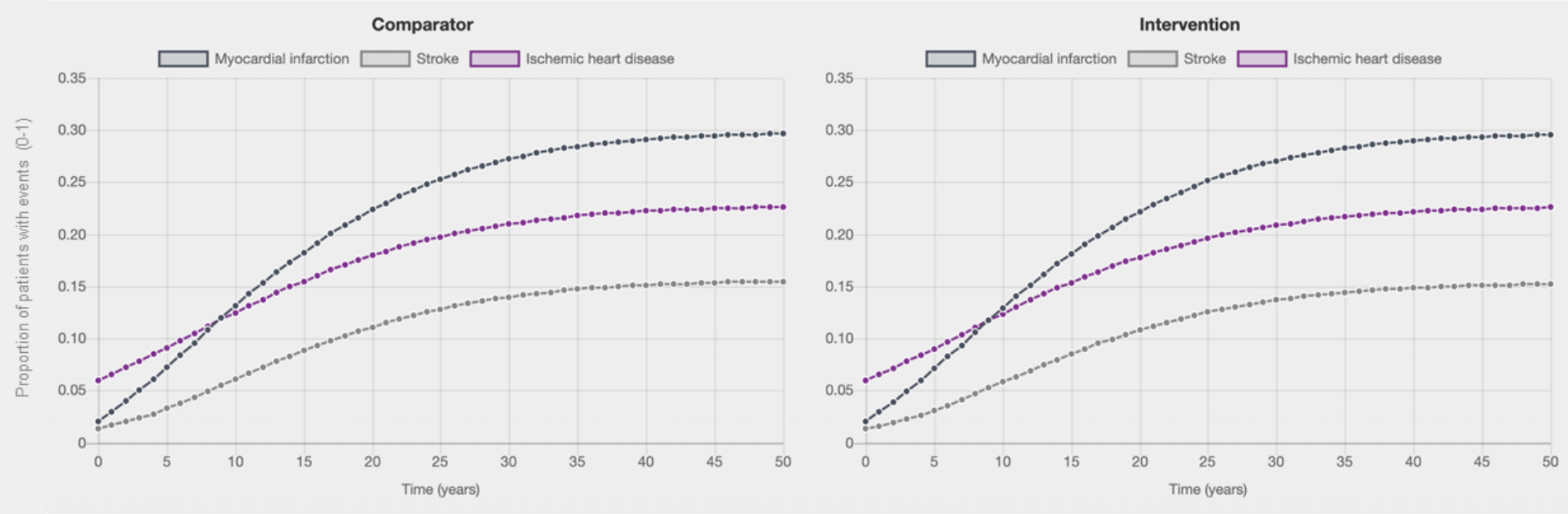
Table 98: Cumulative incidence of diabetes-related complications for tirzepatide 10 mg versus semaglutide 1.0 mg

	Tirzepatide 10 mg	SEMA 1.0 mg	Incremental value
Myocardial infarction (%)	29.6	29.7	−0.1
Stroke (%)	15.2	15.5	−0.3
IHD (%)	22.6	22.6	0
Congestive heart failure (%)	21.4	21.7	−0.3
Revascularization (%)	47.9	48.6	−0.7

Renal failure (%)	8.6	8.6	0
SPSL/neuropathy (%)	61.6	62.0	−0.4
Ulcer (%)	4.4	4.5	−0.1
Amputation (%)	6.4	6.3	+0.1
Blindness (%)	8.3	8.3	0
Macular oedema (%)	26.3	26.6	−0.3

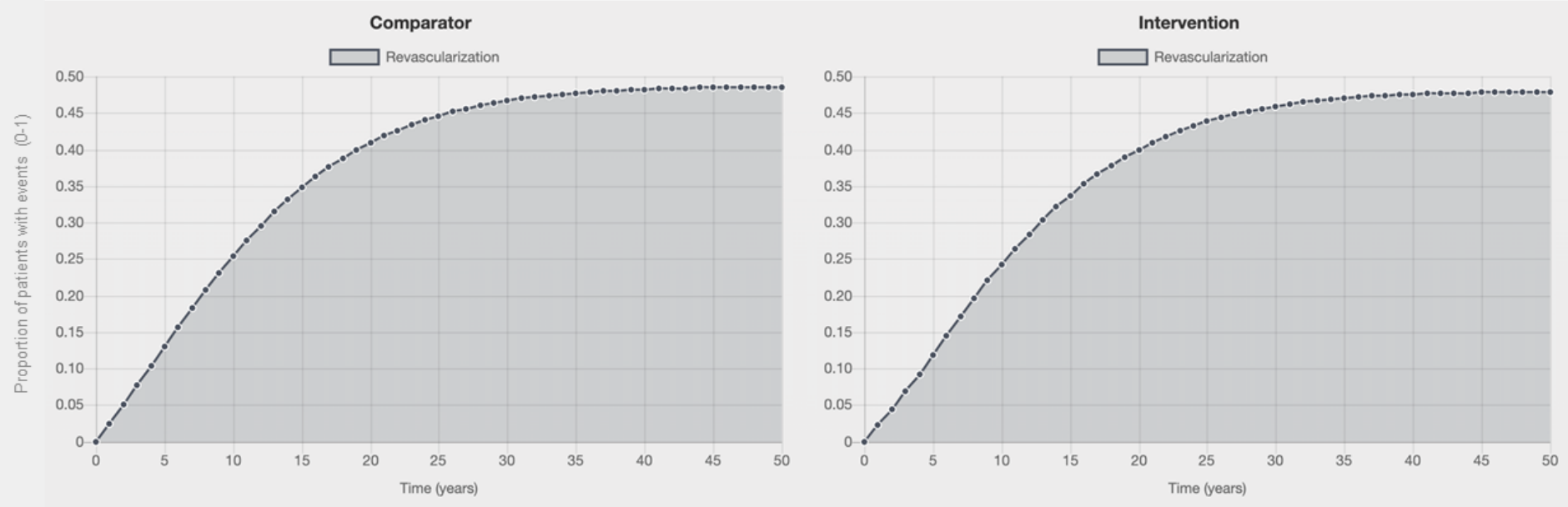
Abbreviations: IHD: ischaemic heart disease; SEMA: semaglutide; SPSL: severe pressure sensation loss.

Figure 83: Cumulative incidence of macrovascular complications in the base case analysis



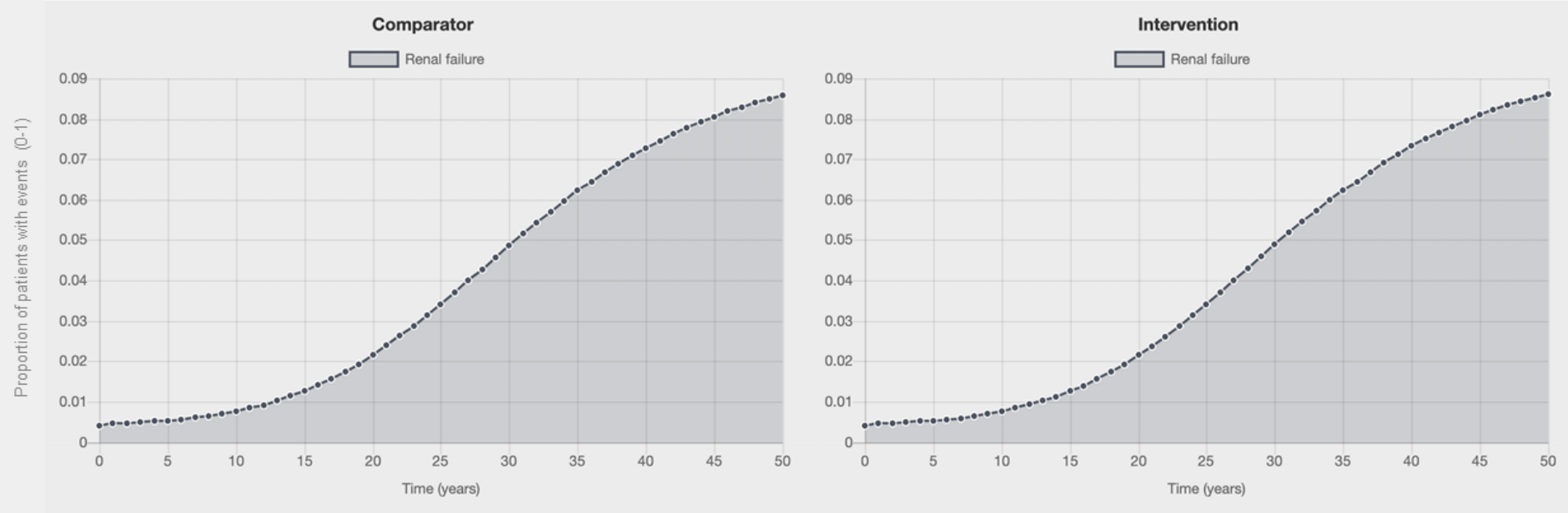
Cumulative incidences, expressed as the mean number of events per patient, are shown in the right hand figure for tirzepatide 10 mg (intervention arm) and in the left hand figure for semaglutide 1.0 mg (comparator arm).

Figure 84: Cumulative incidence of revascularization in the base case analysis



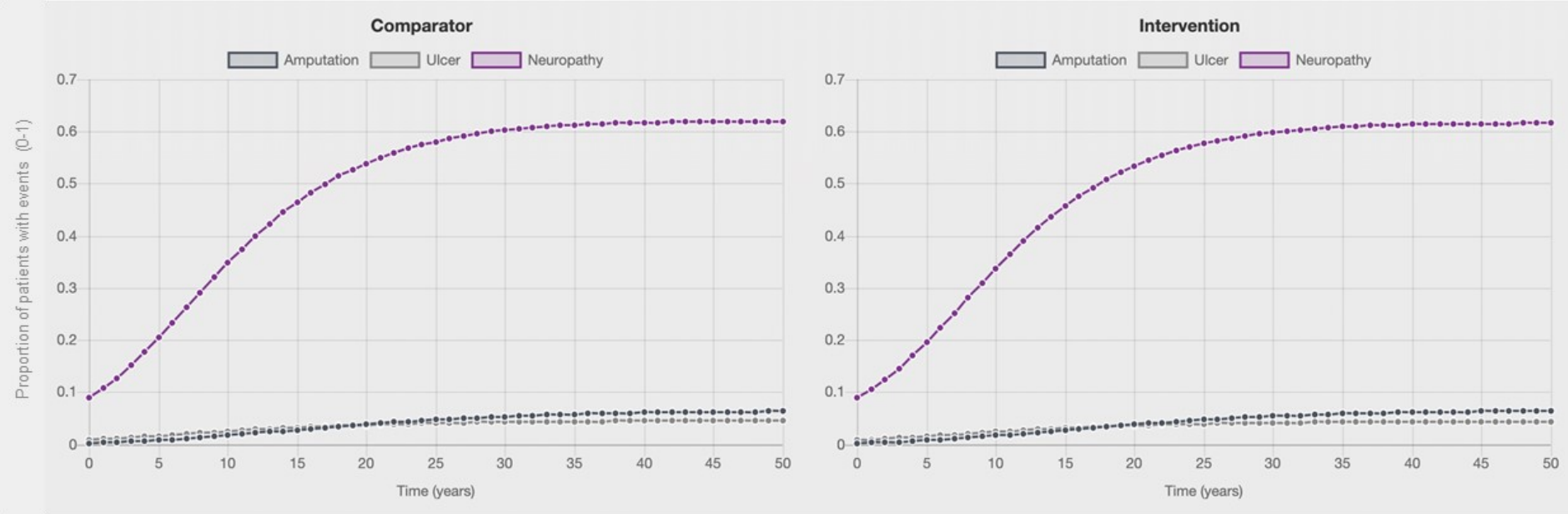
Cumulative incidences, expressed as the mean number of events per patient, are shown in the right hand figure for tirzepatide 10 mg (intervention arm) and in the left hand figure for semaglutide 1.0 mg (comparator arm).

Figure 85: Cumulative incidence of renal failure in the base case analysis



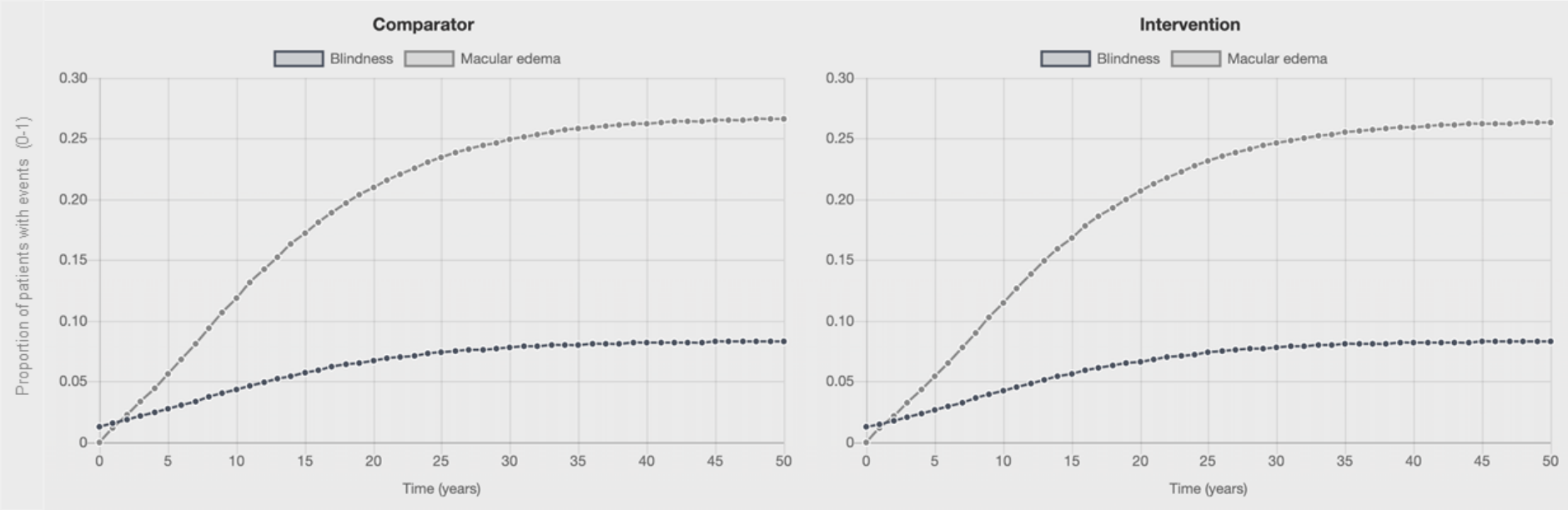
Cumulative incidences, expressed as the mean number of events per patient, are shown in the right hand figure for tirzepatide 10 mg (intervention arm) and in the left hand figure for semaglutide 1.0 mg (comparator arm).

Figure 86: Cumulative incidence of foot ulcer, amputation and neuropathy in the base case analysis



Cumulative incidences, expressed as the mean number of events per patient, are shown in the right hand figure for tirzepatide 10 mg (intervention arm) and in the left hand figure for semaglutide 1.0 mg (comparator arm).

Figure 87: Cumulative incidence of blindness and macular oedema in the base case analysis



Cumulative incidences, expressed as the mean number of events per patient, are shown in the right hand figure for tirzepatide 10 mg (intervention arm) and in the left hand figure for semaglutide 1.0 mg (comparator arm).

B.3.10.1.2 Cost outcomes from the base case analysis

The breakdown of discounted costs from the base case analysis showed that the additional treatment costs associated with tirzepatide (due to the higher pharmacy costs and a longer time to intensification) were offset by reductions in diabetes-related complication costs (Table 99). The greatest cost savings were associated with cardiovascular events avoided (approximately £[redacted] per patient), driven by the improvements in HbA1c and BMI associated with tirzepatide over semaglutide. Tracking the accumulation of costs over time showed that the costs of therapy were the biggest cost driver through to year 15 of the simulation, at which point the continuing accumulation of macrovascular complication costs became the main contributor to overall costs (Figure 88). This remained true through to the end of the simulation.

Table 99: Breakdown of discounted costs for the base case analysis

	TZP 10 mg (£)	SEMA 1.0 mg (£)	Incremental (£)
Myocardial infarction	[redacted]	[redacted]	[redacted]
Stroke	[redacted]	[redacted]	[redacted]
Ischaemic heart disease	[redacted]	[redacted]	[redacted]
Heart failure	[redacted]	[redacted]	[redacted]
Revascularization	[redacted]	[redacted]	[redacted]
Blindness	[redacted]	[redacted]	[redacted]
Macular oedema	[redacted]	[redacted]	[redacted]
Renal disease (pre-ESRD)	[redacted]	[redacted]	[redacted]
Renal disease (ESRD)	[redacted]	[redacted]	[redacted]
Neuropathy/SPSL	[redacted]	[redacted]	[redacted]
Amputation	[redacted]	[redacted]	[redacted]
Ulcer	[redacted]	[redacted]	[redacted]
Hypoglycaemia (all)	[redacted]	[redacted]	[redacted]
Treatment (all)	[redacted]	[redacted]	[redacted]
Total (£)	[redacted]	[redacted]	[redacted]

Abbreviations: SEMA: semaglutide; SPSL: severe pressure sensation loss.

Figure 88: Mean cumulative direct medical costs over time (per patient) in the base case analysis

Population mean values are shown in top figure for tirzepatide 10 mg (intervention arm, right side) and in the bottom figure for semaglutide 1.0 mg (comparator arm, left side).

B.3.11 *Exploring uncertainty*

B.3.11.1 Overall assessment of uncertainty

Uncertainty in the modelling analysis was investigated by running PSA for all base case comparisons to help quantify statistical uncertainty, by performing one-way and multi-way sensitivity analysis on the comparison of tirzepatide with semaglutide to identify key drivers of projected outcomes, and by performing scenario analysis where changes to broader base case assumptions were explored (Sections B.3.11.2 to B.3.11.4).

- PSA indicated that there is a high probability ([redacted]% to [redacted]%, depending on the comparator) that tirzepatide would be cost-effective versus all comparators evaluated, assuming a willingness to pay of £20,000 per QALY gained
- Tirzepatide remained cost-effective even at a 5-year time horizon, with ICERs below £[redacted] per QALY gained for comparisons of all three doses of tirzepatide with semaglutide. These scenarios do not fully capture the clinical benefits associated with improvements in risk factors on tirzepatide therapy due to their short duration
- The findings of the base case analysis remained robust under changes to a range of assumptions, including changes in risk factors associated with treatment, duration of therapy, quality of life benefits, and clinical assumptions used in the base case analysis
- Scenario analyses showed that changing the clinical input dataset and modifying assumptions around treatment intensification (specifically continuing GLP-1 RA or tirzepatide therapy after addition of basal insulin) produced comparable cost-effectiveness outcomes to the base case analysis

B.3.11.2 Probabilistic sensitivity analysis

For PSA, the model reported results based on a nonparametric bootstrapping approach, in which samples from 1% of the simulated population (in this case comprising 3,000 patients) were drawn 1,000 times from the full patient data set at the end of the simulation. Sampling was performed with replacement. The population mean and confidence intervals were then calculated by rerunning the cost and quality-of-life estimators on each sampled population and generating a set of descriptive statistics.

PSA indicated that there was a [redacted]% to [redacted]% probability that tirzepatide 5 mg would be cost-effective, assuming a willingness to pay threshold of £20,000 per QALY gained versus the four comparators evaluated (Table 100). The 95% credible intervals around the improvement in quality-adjusted life expectancy associated with tirzepatide did not cross zero for any of the comparisons. Scatterplots and acceptability curves for the comparisons of tirzepatide 5 mg with dulaglutide 1.5 mg, semaglutide 0.5 mg, oral semaglutide 7 mg and liraglutide 1.2 mg are presented in Appendix J.

Similarly, PSA for tirzepatide 10 mg versus comparators suggested that there was a [redacted]% to [redacted]% probability that tirzepatide 10 mg would be cost-effective against all four comparators, assuming a willingness to pay threshold of £20,000 per QALY gained (

Table 101). The lowest probability (with [redacted]% of 1,000 iterations producing an ICER below £20,000 per QALY gained) was observed for the comparison with [redacted]. The highest ([redacted]%) was versus [redacted]. As with the tirzepatide 5 mg analysis, none of the 95% credible intervals around the improvement in quality-adjusted life expectancy associated with tirzepatide 10 mg crossed zero for any of the comparisons evaluated. The incremental cost-effectiveness scatter plot and acceptability curve for the tirzepatide 10 mg versus semaglutide 1.0 mg are shown in Figure 89 and Figure 90, respectively. The proportion of PSA iterations with ICERs below £10,000 per QALY gained was [redacted]%, increasing to [redacted]% at £30,000 per QALY gained for tirzepatide 10 mg versus semaglutide 1.0 mg.

Scatterplots and acceptability curves for the comparisons of tirzepatide 10 mg with dulaglutide 3.0 mg, oral semaglutide 14 mg and liraglutide 1.8 mg are presented in Appendix J.

For tirzepatide 15 mg, PSA indicated that there was a [redacted]% to [redacted]% probability that tirzepatide would be cost-effective against the four comparators evaluated, assuming a willingness to pay threshold of £20,000 per QALY gained (

Table 102). As with the tirzepatide 5 mg and 10 mg analyses, none of the 95% credible intervals around the improvement in quality-adjusted life expectancy associated with tirzepatide 15 mg crossed zero for any of the comparisons evaluated. Scatterplots and acceptability curves for the comparisons of tirzepatide 15 mg with dulaglutide 4.5 mg, semaglutide 1.0 mg, oral semaglutide 14 mg and liraglutide 1.8 mg are presented in Appendix J.

Table 100: Summary of probabilistic sensitivity analysis results for tirzepatide 5 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)	Probability of tirzepatide being cost-effective**
Tirzepatide 5 mg	[redacted]	7.827 (7.751 – 7.902)	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	7.684 (7.608 – 7.760)	[redacted]	0.144 (0.036 – 0.249)	[redacted]	81.9%
Semaglutide 0.5 mg	[redacted]	7.711 (7.639 – 7.790)	[redacted]	0.116 (0.002 – 0.221)	[redacted]	73.9%
Oral semaglutide 7 mg	[redacted]	7.657 (7.583 – 7.733)	[redacted]	0.171 (0.059 – 0.278)	[redacted]	92.4%
Liraglutide 1.2 mg	[redacted]	7.647 (7.571 – 7.722)	[redacted]	0.180 (0.074 – 0.286)	[redacted]	90.5%

Values shown are means with 95% credible intervals in parentheses. * for tirzepatide versus comparator; **assuming a willingness to pay threshold of £20,000 per QALY again.
Abbreviations: QALY: quality-adjusted life year.

Table 101: Summary of probabilistic sensitivity analysis results for tirzepatide 10 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)	Probability of tirzepatide being cost-effective**
Tirzepatide 10 mg	[redacted]	7.904 (7.830 – 7.979)	--	--	--	--
Dulaglutide 3.0 mg	[redacted]	7.711 (7.636 – 7.786)	[redacted]	0.193 (0.089 – 0.298)	[redacted]	91.7%
Semaglutide 1.0 mg	[redacted]	7.777 (7.704 – 7.853)	[redacted]	0.126 (0.023 – 0.229)	[redacted]	76.8%
Oral semaglutide 14 mg	[redacted]	7.721 (7.648 – 7.794)	[redacted]	0.182 (0.077 – 0.290)	[redacted]	89.2%
Liraglutide 1.8 mg	[redacted]	7.657 (7.586 – 7.736)	[redacted]	0.246 (0.135 – 0.352)	[redacted]	99.4%

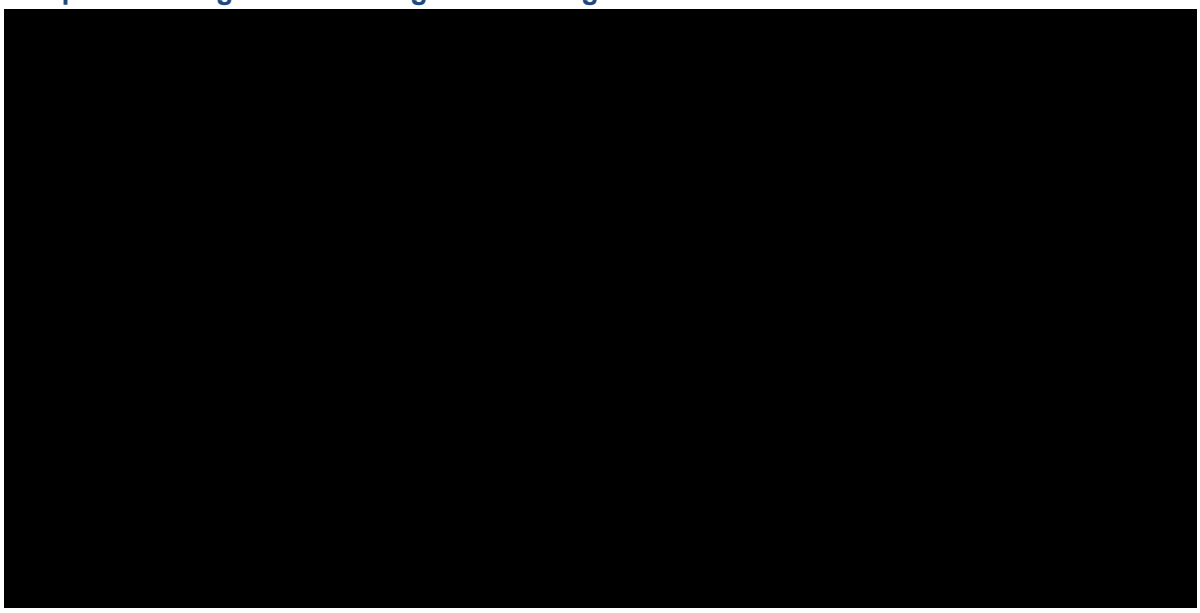
Values shown are means with 95% credible intervals in parentheses. * for tirzepatide versus comparator; **assuming a willingness to pay threshold of £20,000 per QALY again.
Abbreviations: QALY: quality-adjusted life year.

Table 102: Summary of probabilistic sensitivity analysis results for tirzepatide 15 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)	Probability of tirzepatide being cost-effective**
Tirzepatide 15 mg	[redacted]	7.959 (7.885 – 8.034)	--	--	--	--
Dulaglutide 4.5 mg	[redacted]	7.747 (7.676 – 7.824)	[redacted]	0.212 (0.103 – 0.316)	[redacted]	94.0%
Semaglutide 1.0 mg	[redacted]	7.777 (7.704 – 7.853)	[redacted]	0.182 (0.079 – 0.284)	[redacted]	90.5%
Oral semaglutide 14 mg	[redacted]	7.721 (7.648 – 7.794)	[redacted]	0.238 (0.134 – 0.343)	[redacted]	96.2%
Liraglutide 1.8 mg	[redacted]	7.657 (7.586 – 7.736)	[redacted]	0.302 (0.193 – 0.405)	[redacted]	99.7%

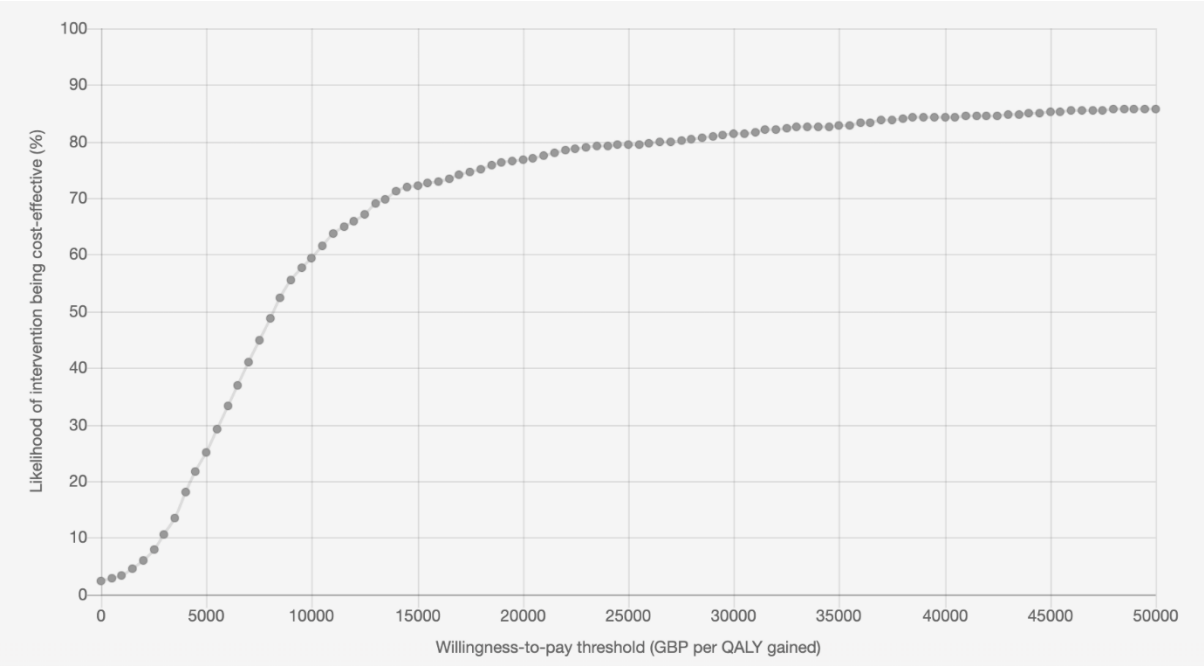
Values shown are means with 95% credible intervals in parentheses. * for tirzepatide versus comparator; **assuming a willingness to pay threshold of £20,000 per QALY again.
Abbreviations: QALY: quality-adjusted life year.

Figure 89: Cost-effectiveness scatterplot from probabilistic sensitivity analysis of tirzepatide 10 mg versus semaglutide 1.0 mg



Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 90: Cost-effectiveness acceptability curve based on the probabilistic sensitivity analysis of tirzepatide 10 mg versus semaglutide 1.0 mg



Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

B.3.11.3 One-way and multi-way sensitivity analysis

One-way and multi-way sensitivity analysis showed that tirzepatide remained cost-effective by commonly quoted standards versus comparators despite variation in a range of modelling input assumptions (Table 103, Table 104, Table 105, and Table 106).

B.3.11.3.1 Sensitivity analysis with all three doses of tirzepatide versus semaglutide

[redacted] 103 [redacted] 104 [redacted] 105 [redacted] Overall, the conclusions drawn from the model are robust to changes in model time horizon, with tirzepatide remaining cost-effective in all scenarios.

Changing the model discount rates affected the incremental costs and QALYs in all three comparisons but did not meaningfully impact the ICER relative to the base case analysis. Changing the cohort characteristics to match those from the SURPASS-2 cohort, which was approximately 7 years younger than the base case cohort, did not notably change the cost-effectiveness profile of tirzepatide versus semaglutide in relation to the base case analysis.

B.3.11.3.2 Sensitivity analysis with tirzepatide 10 mg versus semaglutide 1.0 mg

Clinical drivers

Assuming the HbA1c benefit with tirzepatide 10 mg was equivalent to that with semaglutide 1.0 mg increased the ICER to approximately £[redacted] per QALY gained (Table 106). This change had the effect of reducing the clinical benefits associated with tirzepatide, leading to more diabetes-related complications and a lower incremental QALY benefit. However, it also led to patients intensifying to basal insulin therapy earlier (relative to the base case) on tirzepatide as the HbA1c threshold of 7.5% was reached sooner; this reduced pharmacy costs, leading to lower incremental costs overall, despite slightly higher complication costs for tirzepatide relative to the base case. Assuming equivalent treatment effects on tirzepatide and semaglutide for SBP, serum lipid levels, eGFR and hypoglycaemia rates had little impact on the ICER compared with the base case analysis. However, assuming equivalent BMI improvements in both treatment arms lead to a higher ICER than in the base case analysis. In this scenario, there were smaller QALY benefits associated with weight loss and subsequent BMI state for tirzepatide versus the base case, which produced an incremental benefit of 0.070 QALYs and an ICER of £[redacted] per QALY gained for tirzepatide 10 mg versus semaglutide 1.0 mg. Modelling only the HbA1c benefit and BMI benefit for tirzepatide over semaglutide showed that these two risk factors were the main drivers of the improvements in quality-adjusted life expectancy reported in the base case analysis. Sensitivity analyses which modelled only the HbA1c and BMI differences between tirzepatide and semaglutide produced outcomes comparable to the base case analysis.

Duration of therapy

Sensitivity analyses demonstrated that the duration of therapy and intensification thresholds had little impact on the incremental cost-effectiveness ratio for tirzepatide versus semaglutide (Table 106). Limiting the duration of therapy with tirzepatide and semaglutide to 3 or 5 years reduced the incremental QALY benefit with tirzepatide but also the incremental costs in relation to the

base case, leaving the ICER largely unchanged. Similarly, increasing the duration of therapy by having higher intensification thresholds of HbA1c 8.5% or 9.5% caused an increase in the incremental QALY benefits with tirzepatide but also the incremental costs (due to longer duration of therapy) relative to the base case.

Quality of life utilities

Omitting quality-of-life utilities associated with treatment effects (e.g. device utilities or year 1 weight loss utilities) or adverse events (e.g. nausea or hypoglycaemic events) had little impact on the ICER for tirzepatide versus semaglutide compared to the base case analysis (Table 106). The exception was the sensitivity analysis where no utilities associated with weight loss or BMI were included in the simulation. In this scenario, tirzepatide 10 mg was associated with a quality-adjusted life expectancy benefit of 0.068 QALYs over semaglutide 1.0 mg (relative to 0.114 QALYs in the base case), which resulted in an ICER of £[redacted] per QALY gained for tirzepatide versus semaglutide. Including age adjustment for utility scores or using a multiplicative approach to combine utilities had little impact on the ICER.

Other base case assumptions

Risk adjustments were made to several risk equations in the model based on ethnic group. For example, the risk of myocardial infarction is adjusted for ethnicity in the UKPDS OM2 risk equations, as is the risk of myocardial infarction, revascularization, blindness and neuropathy/SPSL in the BRAVO risk equations (see Appendix N for details). In the base case analysis, 4.5% of the cohort were Black, triggering risk adjustment based on BRAVO but not UKPDS OM2 equations. A sensitivity analysis was performed by changing this label to Afro-Caribbean, triggering risk adjustment with the UKPDS OM2 equations and not with BRAVO. This change had very little impact on the outcomes of the analysis relative to the base case (Table 106).

Including sulfonylurea as a background therapy alongside metformin similarly had little impact on the relative cost-effectiveness of tirzepatide 10 mg versus semaglutide 1.0 mg. For this analysis, it was assumed that treatment with tirzepatide or semaglutide in addition to metformin and sulfonylurea would be associated with annual rates of severe and non-severe hypoglycaemia of 0.09 and 1.94 events per patient year, respectively.[123] The additional costs of sulfonylurea were based on gliclazide 60 mg modified release tablets, once daily, from the June 2022, NHS Electronic Drug Tariff estimates (annual cost £62.22). This had the effect of increasing incremental costs (due to patients being on tirzepatide, metformin and sulfonylurea for a longer duration than in the comparator arm) relative to the base case, and produced an ICER of £[redacted] per QALY gained.

Further changes which were tested and had little impact on incremental QALYs, costs, and the cost-effectiveness of tirzepatide versus semaglutide compared to the base case analysis included: using changes from baseline in BMI taken directly from the NMA (as opposed to estimating BMI change based on weight loss), using UKPDS OM2 formulae to model the progression of risk factors, taking complication costs from alternative published sources based on the literature review, using UKPDS renal failure risk estimation with or without UKPDS OM2 eGFR estimation, and using UKPDS OM2 mortality estimation in the modelling analysis (Table 106).

Mortality estimation in the PRIME T2D Model is calculated by combining risk estimates from diabetes-related complications (largely, CVD and end-stage renal disease) and cause-subtracted

life tables (all other causes of mortality, with CVD and ESRD removed to avoid double counting). In the scenario where only cause-subtracted life tables were used to evaluate mortality risk, this did not fully capture the increased risk of mortality associated with diabetes-related complications and led to patients living much longer (on average, surviving until the end of the simulation at 100 years of age) and accumulating substantial diabetes-related complication costs. This demonstrates that mortality in the base case is driven by diabetes-related complications. This scenario will have underestimated the risk of mortality and, as a result, provides little insight into the cost-effectiveness evaluation.

Table 103: Summary of one-way sensitivity analysis results with tirzepatide 5 mg versus semaglutide 0.5 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 5 mg	Semaglutide 0.5 mg	Incremental value	Tirzepatide 5 mg	Semaglutide 0.5 mg	Incremental value	
Base case	[redacted]	[redacted]	[redacted]	*9.488*	*9.374*	*+0.114*	[redacted]
Sensitivity analysis							
5-year time horizon	[redacted]	[redacted]	[redacted]	3.412	3.345	+0.067	[redacted]
10-year time horizon	[redacted]	[redacted]	[redacted]	5.723	5.638	+0.085	[redacted]
15-year time horizon	[redacted]	[redacted]	[redacted]	7.275	7.178	+0.098	[redacted]
20-year time horizon	[redacted]	[redacted]	[redacted]	8.257	8.152	+0.105	[redacted]
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	13.800	13.651	+0.149	[redacted]
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	7.719	7.620	+0.099	[redacted]
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	9.844	9.739	+0.105	[redacted]

Abbreviations: QALY: quality-adjusted life years.

Table 104: Summary of one-way sensitivity analysis results with tirzepatide 10 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Base case	[redacted]	[redacted]	[redacted]	*9.535*	*9.421*	*+0.114*	[redacted]
Sensitivity analysis							

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
5-year time horizon	[redacted]	[redacted]	[redacted]	3.446	3.376	+0.070	[redacted]
10-year time horizon	[redacted]	[redacted]	[redacted]	5.763	5.679	+0.084	[redacted]
15-year time horizon	[redacted]	[redacted]	[redacted]	7.315	7.224	+0.092	[redacted]
20-year time horizon	[redacted]	[redacted]	[redacted]	8.299	8.200	+0.099	[redacted]
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	13.859	13.706	+0.152	[redacted]
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	7.761	7.663	+0.098	[redacted]
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	9.882	9.784	+0.098	[redacted]

Abbreviations: QALY: quality-adjusted life years.

Table 105: Summary of one-way sensitivity analysis results with tirzepatide 15 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	
Base case	[redacted]	[redacted]	[redacted]	*9.581*	*9.421*	*+0.160*	[redacted]
Sensitivity analysis							
5-year time horizon	[redacted]	[redacted]	[redacted]	3.468	3.376	+0.092	[redacted]
10-year time horizon	[redacted]	[redacted]	[redacted]	5.791	5.679	+0.112	[redacted]
15-year time horizon	[redacted]	[redacted]	[redacted]	7.350	7.224	+0.127	[redacted]
20-year time horizon	[redacted]	[redacted]	[redacted]	8.338	8.200	+0.138	[redacted]
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	13.928	13.706	+0.221	[redacted]

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	7.799	7.663	+0.136	[redacted]
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	9.934	9.784	+0.150	[redacted]

Abbreviations: QALY: quality-adjusted life years.

Table 106: Summary of additional one-way and multi-way sensitivity analysis results for tirzepatide 10 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Base case	[redacted]	[redacted]	[redacted]	*9.535*	*9.421*	*+0.114*	[redacted]
Clinical drivers							
No HbA1c difference	[redacted]	[redacted]	[redacted]	9.485	9.421	+0.064	[redacted]
No SBP difference	[redacted]	[redacted]	[redacted]	9.530	9.421	+0.109	[redacted]
No serum lipids difference	[redacted]	[redacted]	[redacted]	9.539	9.421	+0.118	[redacted]
No BMI difference	[redacted]	[redacted]	[redacted]	9.491	9.421	+0.070	[redacted]
Only HbA1c difference between treatments	[redacted]	[redacted]	[redacted]	9.475	9.421	+0.054	[redacted]
Only BMI difference between treatments	[redacted]	[redacted]	[redacted]	9.467	9.421	+0.046	[redacted]
Only HbA1c and BMI differences between treatments	[redacted]	[redacted]	[redacted]	9.534	9.421	+0.113	[redacted]
Duration of therapy							

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Intensification to insulin after 3 years	[redacted]	[redacted]	[redacted]	9.517	9.438	+0.079	[redacted]
Intensification to insulin after 5 years	[redacted]	[redacted]	[redacted]	9.658	9.548	+0.110	[redacted]
Second intensification to basal-bolus therapy	[redacted]	[redacted]	[redacted]	9.555	9.428	+0.127	[redacted]
Intensification at HbA1c 8.5% threshold	[redacted]	[redacted]	[redacted]	9.829	9.679	+0.150	[redacted]
Intensification at HbA1c 9.5% threshold	[redacted]	[redacted]	[redacted]	10.392	10.122	+0.270	[redacted]
Quality of life utilities							
No weight change utility	[redacted]	[redacted]	[redacted]	9.478	9.374	+0.104	[redacted]
No weight/BMI utilities	[redacted]	[redacted]	[redacted]	9.888	9.820	+0.068	[redacted]
No device utility	[redacted]	[redacted]	[redacted]	9.528	9.421	+0.107	[redacted]
No nausea utilities	[redacted]	[redacted]	[redacted]	9.548	9.432	+0.116	[redacted]
No hypoglycaemia utilities	[redacted]	[redacted]	[redacted]	9.885	9.785	+0.099	[redacted]
QALY age-adjustment based on Ara and Brazier	[redacted]	[redacted]	[redacted]	8.832	8.724	+0.109	[redacted]
Multiplicative approach to combining utilities	[redacted]	[redacted]	[redacted]	9.354	9.271	+0.083	[redacted]
Other base case assumptions							
Cohort ethnic groups changed from Black to Afro-Caribbean	[redacted]	[redacted]	[redacted]	9,559	9.451	+0.108	[redacted]
Sulfonylurea added to background therapy	[redacted]	[redacted]	[redacted]	9.490	9.369	+0.122	[redacted]

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Change in BMI values taken directly from NMA	[redacted]	[redacted]	[redacted]	9.544	9.436	+0.108	[redacted]
UKPDS OM2 risk factor progression for all risk factors	[redacted]	[redacted]	[redacted]	9,487	9.385	+0.101	[redacted]
Complication costs taken from alternative sources (lit. review)	[redacted]	[redacted]	[redacted]	9.535	9.421	+0.114	[redacted]
UKPDS OM2 renal failure estimation	[redacted]	[redacted]	[redacted]	9.474	9.368	+0.107	[redacted]
UKPDS OM2 eGFR progression and renal failure estimation	[redacted]	[redacted]	[redacted]	9.816	9.694	+0.122	[redacted]
UKPDS OM2 mortality risk estimation	[redacted]	[redacted]	[redacted]	8.240	8.156	+0.084	[redacted]
Cause-subtracted life tables for mortality risk estimation	[redacted]	[redacted]	[redacted]	12.151	12.063	+0.089	[redacted]

Abbreviations: BMI: body mass index; eGFR: estimated glomerular filtration rate; NMA: network meta-analysis; QALY: quality-adjusted life years; UKPDS OM2: United Kingdom Prospective Diabetes Study Outcomes Model 2.

B.3.11.4 Scenario analysis

B.3.11.4.1 SURPASS-2 model inputs

For the SURPASS-2 based analysis, cohort characteristics and treatment effects for tirzepatide 10 mg and semaglutide 1.0 mg were derived from the SURPASS-2 CSR. All other settings and assumptions in this scenario analysis matched the base case analysis. In this scenario, life expectancy was projected to be approximately 1 year longer in both treatment arms and direct costs were around £3,000 higher per patient over a lifetime (Table 107). However, incremental costs and QALYs were comparable to the base case analysis, resulting in a ICER of £[redacted] per QALY gained for tirzepatide 10 mg versus semaglutide 1.0 mg.

B.3.11.4.2 Intensification of therapy by adding basal insulin

In the scenario where tirzepatide and semaglutide therapy were continued after initiation of basal insulin, projected clinical outcomes were broadly similar to the base case analysis, with only slight improvements in quality-adjusted life expectancy observed (Table 107). This was due to more durable improvements in risk factors leading to slightly lower rates of diabetes-related complications in both treatment arms. Lifetime costs were higher in both treatment arms, by approximately £[redacted] in the tirzepatide arm and £[redacted] in the semaglutide arm. In this scenario, the ICER for tirzepatide 10 mg versus semaglutide 1.0 mg was approximately £[redacted] per QALY gained.

Table 107: Summary of scenario analysis results

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER (£ per QALY gained)
Base case (based on NMA results and switch to basal insulin on intensification)							
Tirzepatide 10 mg	[redacted]	13.138	9.535	--	--	--	--
Semaglutide 1.0 mg	[redacted]	13.092	9.421	[redacted]	+0.046	+0.114	[redacted]
SURPASS-2 based analysis							
Tirzepatide 10 mg	[redacted]	14.080	9.784	--	--	--	--
Semaglutide 1.0 mg	[redacted]	13.963	9.636	[redacted]	+0.116	+0.148	[redacted]
Intensification of therapy by adding basal insulin							
Tirzepatide 10 mg	[redacted]	13.195	9.594	--	--	--	--
Semaglutide 1.0 mg	[redacted]	13.153	9.486	[redacted]	+0.042	+0.108	[redacted]

* for tirzepatide versus comparator.
Abbreviations: QALY: quality-adjusted life year.

B.3.12 *Subgroup analysis*

No sub-group analyses were performed.

B.3.13 *Benefits not captured in the QALY calculation*

Whilst every effort was made to capture the quality of life implications of therapy in the present analysis, the benefits associated with weight loss may be underestimated. The substantial weight loss reported with tirzepatide may produce quality of life benefits greater than those reported by Bagust and Beale (2005) in their analysis of CODE-2 data, which was used in the present analysis; this may be due to the fact substantial weight loss has only been possible with newer interventions for type 2 diabetes (which have become available after the Bagust and Beale analysis was conducted).[161] Whilst studies of obesity suggest that greater improvements in quality of life than those included in the base case may be associated with substantial weight loss, there remains considerable heterogeneity in the published data.[170, 171] However, it is notable that there may be several other benefits of weight loss (e.g. reduced cancer risk) not captured in the present analysis that could have significant bearing on quality of life.[172]

Most of the comparators in the present evaluation had a similar mode of administration (once weekly injectables). However, evidence published by Boye *et al.* in 2019 found that there was a patient preference for the type of administration device used, namely a preference for the device used for tirzepatide and dulaglutide administration.[160] The quality of life benefit for using this type of administration device was, conservatively, only applied in the first year of the modelling analysis (and not in all years of treatment) as the durability of this preference is not yet established. Given the similar mode of administration for most comparators, no disutilities associated with injection were included in the base case analysis (simplifying assumption). This assumption is likely to have favoured liraglutide (daily injection) over formulations designed for weekly injection. Moreover, the earlier initiation of insulin with most comparators (relative to tirzepatide) was not modelled to be associated with a negative impact on quality of life due to more frequent injections. The exception, in terms of mode of administration, was oral semaglutide. An adjustment was made in the base case analysis by increasing the annual utility score for simulated patients on oral semaglutide, in line with the injection disutility previously used by NICE.[123]

B.3.14 *Validation*

B.3.14.1 Validation of cost-effectiveness analysis

The PRIME T2D Model has been the subject of internal and external validation analyses: face validity of the model via review by clinical and diabetes modelling experts, internal validation of the model code to ensure the model was coded correctly and could accurately reproduce the results of the studies used to develop the model, and external validation where the PRIME T2D Model was used to reproduce the outcomes of published studies in type 2 diabetes, including long-term outcomes and outcomes from CVOT studies. Full details and results of the PRIME T2D Model validation are provided in the PRIME T2D Model Technical Report.

B.3.15 *Interpretation and conclusions of economic evidence*

Long-term cost-effectiveness analyses using the PRIME T2D Model to compare tirzepatide with GLP-1 RAs in common use in the UK setting, based on NMA data, have shown that:

- All three doses of tirzepatide (5, 10 and 15 mg) were associated with improvements in life expectancy and quality-adjusted life expectancy over the comparators evaluated. These benefits were driven by reductions in HbA1c and BMI associated with tirzepatide in the modelling analysis
- Overall, direct costs were generally higher for tirzepatide than for comparators. [redacted] Higher lifetime costs versus other comparators were driven by higher pharmacy costs in the tirzepatide arms due to higher drug acquisition costs and a longer time on therapy (prior to intensification to basal insulin). The longer time on therapy was driven by greater improvements in HbA1c with tirzepatide, resulting in a longer time to reach the intensification threshold of 7.5%. Higher pharmacy costs with tirzepatide were partially offset by reduced complication costs, in particular the reduced costs associated with macrovascular complications on tirzepatide versus comparators
- Cost-effectiveness analysis of tirzepatide 5 mg showed that ICERs ranged between £[redacted] per QALY gained versus oral semaglutide 7 mg and £[redacted] per QALY gained versus injectable semaglutide 0.5 mg
- In the analysis of tirzepatide 10 mg, ICERs ranged between £[redacted] per QALY gained versus dulaglutide 3.0 mg and £[redacted] per QALY gained versus semaglutide 1.0 mg. [redacted]
- For tirzepatide 15 mg, ICERs were in the range of £[redacted] per QALY gained versus dulaglutide 4.5 mg and £[redacted] per QALY gained versus semaglutide 1.0 mg. As with the 10 mg dose, [redacted]

Exploration of uncertainty around the base case modelling analysis led to the following observations:

- PSA indicated that there is a high probability ([redacted]% to [redacted] depending on the comparator) that tirzepatide would be cost-effective versus all comparators evaluated, assuming a willingness to pay of £20,000 per QALY gained or more
- Changing cohort characteristics or assumptions around treatment intensification (including duration of therapy) or background therapy did not notably impact the cost-effectiveness profile of tirzepatide. It could be assumed tirzepatide would have a comparable cost-effectiveness profile earlier or later in the T2D treatment algorithm based on these findings. Scenario analysis using data from SURPASS-2, with tirzepatide or semaglutide as an add-on to metformin, showed tirzepatide 10 mg is likely to be cost-effective versus semaglutide 1.0 mg, with an ICER of £[redacted] per QALY gained

- Tirzepatide remained cost-effective even at a 5-year time horizon, with ICERs below £[redacted] per QALY gained for comparisons of all three doses with semaglutide. These scenarios can be considered conservative as they do not fully capture the clinical benefits associated with improvements in risk factors on tirzepatide therapy due to their short duration
- Scenario analyses showed that changing the clinical input dataset and modifying assumptions around treatment intensification (specifically continuing GLP-1 RA or tirzepatide therapy after addition of basal insulin) produced comparable cost-effectiveness outcomes to the base case analysis

A literature review did not identify any published evaluations of tirzepatide to date, making comparison with the published literature impossible (beyond the model validation analysis already conducted). The economic evaluation was designed to be generalisable to clinical practice in England by using a cohort representative of patients who would currently be treated with GLP-1 RAs for type 2 diabetes and following methodological approaches consistent with those used previously by NICE. Whilst using NMA data to model treatment effects represents the best approach currently available, it is not known how closely the treatment effects from the NMA would match those patients who would currently be treated with GLP-1 RAs in routine clinical practice in England.

The PRIME T2D Model represents a novel approach for cost-effectiveness analyses in England but utilises risk equations from UKPDS in a model averaging approach designed to adjust for high risk and low risk patients as well as local costs, utilities and mortality risk estimation. The PRIME T2D Model has been shown to validate well against the UK-based Lipids in Diabetes Study using model averaging (see PRIME T2D Model Technical Report for details). It is not currently known how well the BRAVO Model risk equations alone would validate for risk estimation in a UK population, as they were derived from a North American dataset. However, it is known that the UKPDS OM2 risk equations may not perform particularly well in older UK patients with T2D or in populations with a higher cardiovascular risk profile than that in the UKPDS study.[139, 173] Validation evidence suggests that integration of the BRAVO Model risk equations into the PRIME T2D Model may well represent a viable solution to the challenge of modelling outcomes for patients with high risk profiles and/or advanced disease (as is important when modelling long-term time horizons).

Hazard ratios from CVOTs were not used to calibrate the present modelling analysis. This type of calibration approach has been used several times in recent years, since the 2018 Mount Hood Challenge meeting showed that several diabetes models at that time needed calibration to reproduce outcomes from the EMPA-REG OUTCOME trial and the CANVAS Program.[174] Integration of hazard ratios from CVOTs into a modelling analysis designed to evaluate complication risk based on patient characteristics and risk factors is problematic; the published hazard ratios from CVOTs are, generally, derived from placebo-controlled trials and are not adjusted for conventional risk factors, such as HbA1c or SBP, or the overall risk profile of the population.[174] This means that the use of any such hazard ratios runs the risk of double-counting benefits associated with specific interventions. Moreover, for several CVOTs, hazard ratios are only presented for composite endpoints (most commonly 3-point MACE) which makes their application to models that report individual endpoints (e.g. myocardial infarction, stroke, etc.)

challenging. Currently, CVOT data for tirzepatide are not available (expected in 2025) and it was decided not to include hazard ratios in the present analysis for three main reasons:

- The PRIME T2D Model has validated well against the EMPA-REG OUTCOME trial (see the PRIME T2D Model Technical Report for details) suggesting that any such calibration is not required in the present modelling analysis
- Given the limitations of a hazard ratio calibration approach, the potential for the inclusion of unadjusted hazard ratios from different sources to bias the analysis and obfuscate the outcomes was considered very high
- Conversion of composite endpoint hazard ratios for application to individual endpoints was not possible without the risk of biasing the analysis

CVOT data that are generalisable to a UK-based population, offer hazard ratios for individual endpoints for all relevant comparators, are adjusted for changes in conventional risk factors associated with each comparator, and are adjusted for population characteristics, would enhance not only the present modelling analysis but all T2D health economic modelling studies for the UK setting. Given the purported mechanism of action of the GLP-1 RA cardioprotective effects, and the magnitudes of these benefits reported in CVOTs, it is highly likely that including hazard ratios for all of the interventions in the present cost-effectiveness evaluation would not have altered the relative outcomes, i.e. tirzepatide would remain cost-effective or dominant versus commonly used GLP-1 RAs in England.[175]

B.3.15.1 Conclusions

In summary, tirzepatide represents a new treatment option that can improve the glycaemic control and weight loss of patients with T2D who have an unmet need in these areas on currently-available treatments. Tirzepatide represents a cost-effective use of NHS resources versus commonly used GLP-1 RAs in England; sensitivity analyses showed that the ICERs calculated are robust to changes in the modelling parameters. Tirzepatide is therefore a valuable new addition to the clinical pathway of care for T2D, providing patients with an effective, tolerable therapy for T2D that addresses the unmet needs identified in Section B.1.3.2.

References

1. clinicaltrials.gov. A Study of Tirzepatide (LY3298176) Compared With Dulaglutide on Major Cardiovascular Events in Participants With Type 2 Diabetes (SURPASS-CVOT). Available at: https://clinicaltrials.gov/ct2/show/NCT04255433/ [Accessed 25 April 2022].
2. Coskun T, Sloop KW, Loghin C, et al. LY3298176, a novel dual GIP and GLP-1 receptor agonist for the treatment of type 2 diabetes mellitus: From discovery to clinical proof of concept. Mol Metab 2018;18:3-14.
3. clinicaltrials.gov. A Study of Tirzepatide in Participants With Type 2 Diabetes Mellitus (T2DM). Available at: https://clinicaltrials.gov/ct2/show/NCT03951753/ [Accessed 27 April 2022].
4. Samms RJ, Coghlan MP, Sloop KW. How May GIP Enhance the Therapeutic Efficacy of GLP-1? Trends Endocrinol Metab 2020;31:410-421.
5. Eli Lilly Data on File. Tirzepatide Draft Summary of Product Characteristics (SmPC).
6. DeFronzo RA, Ferrannini E, Groop L, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers 2015;1:15019.
7. Diabetes UK. Number of people with diabetes reaches 4.7 million. 2019. Available at: https://www.diabetes.org.uk/about_us/news/new-stats-people-living-with-diabetes/ [Last accessed: 10 February 2022].
8. Holman N, Wild SH, Khunti K, et al. Incidence and Characteristics of Remission of Type 2 Diabetes in England: A Cohort Study Using the National Diabetes Audit. Diabetes Care 2022.
9. International Diabetes Federation (IDF). IDF Diabetes Atlas. 10th Edition. Available at: https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf/ [Accessed 25 April 2022]. 2021.
10. National Health Service (NHS). National Diabetes Audit 2019/20. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/ [Last accessed 10 February 2022].
11. Sprenger L, Mader A, Larcher B, et al. Type 2 diabetes and the risk of cardiovascular events in peripheral artery disease versus coronary artery disease. BMJ Open Diabetes Res Care 2021;9.
12. Taylor R. Type 2 diabetes: etiology and reversibility. Diabetes Care 2013;36:1047-55.
13. Zurita-Cruz JN, Manuel-Apolinar L, Arellano-Flores ML, et al. Health and quality of life outcomes impairment of quality of life in type 2 diabetes mellitus: a cross-sectional study. Health Qual Life Outcomes 2018;16:94.
14. Carey VJ, Walters EE, Colditz GA, et al. Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. Am J Epidemiol 1997;145:614-9.
15. Wu Y, Ding Y, Tanaka Y, et al. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci 2014;11:1185-200.
16. Bellou V, Belbasis L, Tzoulaki I, et al. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. PLoS One 2018;13:e0194127.
17. Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. Nature 2006;444:881-7.
18. Skyler JS, Bakris GL, Bonifacio E, et al. Differentiation of Diabetes by Pathophysiology, Natural History, and Prognosis. Diabetes 2017;66:241-255.
19. Abate N, Garg A, Peshock RM, et al. Relationships of generalized and regional adiposity to insulin sensitivity in men. J Clin Invest 1995;96:88-98.
20. Cerf ME. Beta cell dysfunction and insulin resistance. Front Endocrinol (Lausanne) 2013;4:37.
21. Van Gaal L, Scheen A. Weight management in type 2 diabetes: current and emerging approaches to treatment. Diabetes Care 2015;38:1161-72.
22. Hunger M, Schunk M, Meisinger C, et al. Estimation of the relationship between body mass index and EQ-5D health utilities in individuals with type 2 diabetes: evidence from the population-based KORA studies. J Diabetes Complications 2012;26:413-8.
23. Redekop WK, Koopmanschap MA, Stolk RP, et al. Health-related quality of life and treatment satisfaction in Dutch patients with type 2 diabetes. Diabetes Care 2002;25:458-63.
24. Hänninen J, Takala J, Keinänen-Kiukaanniemi S. Quality of life in NIDDM patients assessed with the SF-20 questionnaire. Diabetes Res Clin Pract 1998;42:17-27.

25. Kolotkin RL, Crosby RD, Williams GR. Assessing weight-related quality of life in obese persons with type 2 diabetes. Diabetes Res Clin Pract 2003;61:125-32.
26. National Health Service (NHS). Long Term Plan. Diabetes. Available at: https://www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/better-care-for-major-health-conditions/diabetes/ [Last accessed: 10 February 2022].
27. Blonde L, Rosenstock J, Del Prato S, et al. Switching to iGlarLixi Versus Continuing Daily or Weekly GLP-1 RA in Type 2 Diabetes Inadequately Controlled by GLP-1 RA and Oral Antihyperglycemic Therapy: The LixiLan-G Randomized Clinical Trial. Diabetes Care 2019;42:2108-2116.
28. Fang M, Wang D, Coresh J, et al. Trends in Diabetes Treatment and Control in U.S. Adults, 1999-2018. N Engl J Med 2021;384:2219-2228.
29. Stark Casagrande S, Fradkin JE, Saydah SH, et al. The prevalence of meeting A1C, blood pressure, and LDL goals among people with diabetes, 1988-2010. Diabetes Care 2013;36:2271-9.
30. Public Health England. Adult obesity and type 2 diabetes. 2014. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/338934/Adult_obesity_and_type_2_diabetes_.pdf/ [Last accessed: 10 February 2022].
31. Boye KS, Lage MJ, Shinde S, et al. Trends in HbA1c and Body Mass Index Among Individuals with Type 2 Diabetes: Evidence from a US Database 2012-2019. Diabetes Ther 2021;12:2077-2087.
32. Cheng YH, Tsao YC, Tzeng IS, et al. Body mass index and waist circumference are better predictors of insulin resistance than total body fat percentage in middle-aged and elderly Taiwanese. Medicine (Baltimore) 2017;96:e8126.
33. Racette SB, Evans EM, Weiss EP, et al. Abdominal adiposity is a stronger predictor of insulin resistance than fitness among 50-95 year olds. Diabetes Care 2006;29:673-8.
34. Wing RR, Bond DS, Gendrano IN, 3rd, et al. Effect of intensive lifestyle intervention on sexual dysfunction in women with type 2 diabetes: results from an ancillary Look AHEAD study. Diabetes Care 2013;36:2937-44.
35. Rubin RR, Wadden TA, Bahnson JL, et al. Impact of intensive lifestyle intervention on depression and health-related quality of life in type 2 diabetes: the Look AHEAD Trial. Diabetes Care 2014;37:1544-53.
36. National Institute of Health and Care Excellence (NICE). Guideline: type 2 diabetes management. Available at: https://www.nice.org.uk/guidance/ng28/ [Last accessed 18 February 2022].
37. Landgraf R, Aberle J, Birkenfeld AL, et al. Therapy of Type 2 Diabetes. Exp Clin Endocrinol Diabetes 2019;127:S73-s92.
38. Reed J, Bain S, Kanamarlapudi V. A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives. Diabetes Metab Syndr Obes 2021;14:3567-3602.
39. Diabetes UK. Number of people living with diabetes doubles in twenty years. Available at: https://www.diabetes.org.uk/about_us/news/diabetes-prevalence-statistics. Accessed on: 13 Jul 22.
40. National Health Service (NHS). Young People with Type 2 Diabetes 2019–20. National Diabetes Audit. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/young-people-with-type-2-diabetes-2019--20/ [Accessed: 25 April 2022].
41. Royal College of Paediatrics and Child Health. National Paediatric Diabetes Audit Report 2016-17. Care processes and outcomes. Available at: https://www.rcpch.ac.uk/sites/default/files/2018-07/npda_annual_report_2016_-_2017_april_2018_final_updated_3.pdf/ [Accessed 25 April 2022].
42. Lawrence JM, Divers J, Isom S, et al. Trends in Prevalence of Type 1 and Type 2 Diabetes in Children and Adolescents in the US, 2001-2017. Jama 2021;326:717-727.
43. Zoungas S, Woodward M, Li Q, et al. Impact of age, age at diagnosis and duration of diabetes on the risk of macrovascular and microvascular complications and death in type 2 diabetes. Diabetologia 2014;57:2465-74.

44. Lin X, Xu Y, Pan X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 2020;10:14790.
45. Department of Health. National Service Framework for Diabetes: Standards. 2001. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198836/National_Service_Framework_for_Diabetes.pdf/ [Accessed 22 February 2022].
46. Gorst C, Kwok CS, Aslam S, et al. Long-term Glycemic Variability and Risk of Adverse Outcomes: A Systematic Review and Meta-analysis. Diabetes Care 2015;38:2354-69.
47. Whyte MB, Joy M, Hoang U, et al. Glycaemic trajectories in type 2 diabetes and risk of cardiovascular disease: Exploring the link. Oral presentation at the 55th Annual Meeting of the European Association for the Study of Diabetes; 16-20 September 2019; Barcelona, Spain. Abstract 122. Available at: https://ora.ox.ac.uk/objects/uuid:2807559f-fd6e-442e-8013-53bf1aa7c1d1/ [Accessed 3 May 2022]. 2019.
48. Zoungas S, Arima H, Gerstein HC, et al. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. Lancet Diabetes Endocrinol 2017;5:431-437.
49. Zoungas S, Chalmers J, Ninomiya T, et al. Association of HbA1c levels with vascular complications and death in patients with type 2 diabetes: evidence of glycaemic thresholds. Diabetologia 2012;55:636-43.
50. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008;359:1577-89.
51. Laiteerapong N, Ham SA, Gao Y, et al. The Legacy Effect in Type 2 Diabetes: Impact of Early Glycemic Control on Future Complications (The Diabetes & Aging Study). Diabetes Care 2019;42:416-426.
52. Peng XV, Ayyagari R, Lubwama R, et al. Impact of Simultaneous Versus Sequential Initiation of Basal Insulin and Glucagon-like Peptide-1 Receptor Agonists on HbA1c in Type 2 Diabetes: A Retrospective Observational Study. Diabetes Ther 2020;11:995-1005.
53. Mulnier HE, Seaman HE, Raleigh VS, et al. Mortality in people with type 2 diabetes in the UK. Diabet Med 2006;23:516-21.
54. Solli O, Stavem K, Kristiansen IS. Health-related quality of life in diabetes: The associations of complications with EQ-5D scores. Health Qual Life Outcomes 2010;8:18.
55. Grandy S, Fox KM. Change in health status (EQ-5D) over 5 years among individuals with and without type 2 diabetes mellitus in the SHIELD longitudinal study. Health Qual Life Outcomes 2012;10:99.
56. Grandy S, Fox KM. EQ-5D visual analog scale and utility index values in individuals with diabetes and at risk for diabetes: Findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Health Qual Life Outcomes 2008;6:18.
57. Holmes J, McGill S, Kind P, et al. Health-related quality of life in type 2 diabetes (TARDIS-2). Value Health 2000;3 Suppl 1:47-51.
58. Clouet F, Excler-Cavailher G, Christophe B, et al. [Type 2 Diabetes and Short Form 36-items Health Survey]. Diabetes Metab 2001;27:711-7.
59. Kuznetsov L, Long GH, Griffin SJ, et al. Are changes in glycaemic control associated with diabetes-specific quality of life and health status in screen-detected type 2 diabetes patients? Four-year follow up of the ADDITION-Cambridge cohort. Diabetes Metab Res Rev 2015;31:69-75.
60. Svedbo Engström M, Leksell J, Johansson UB, et al. Health-related quality of life and glycaemic control among adults with type 1 and type 2 diabetes - a nationwide cross-sectional study. Health Qual Life Outcomes 2019;17:141.
61. Rodríguez-Almagro J, García-Manzanares Á, Lucendo AJ, et al. Health-related quality of life in diabetes mellitus and its social, demographic and clinical determinants: A nationwide cross-sectional survey. J Clin Nurs 2018;27:4212-4223.
62. Beaudet A, Clegg J, Thuresson PO, et al. Review of utility values for economic modeling in type 2 diabetes. Value Health 2014;17:462-70.
63. Polonsky WH, Fisher L, Hessler D, et al. Toward a more comprehensive understanding of the emotional side of type 2 diabetes: A re-envisioning of the assessment of diabetes distress. J Diabetes Complications 2022;36:108103.

64. Bradley C, Eschwège E, de Pablos-Velasco P, et al. Predictors of Quality of Life and Other Patient-Reported Outcomes in the PANORAMA Multinational Study of People With Type 2 Diabetes. Diabetes Care 2018;41:267-276.
65. Lean ME, Leslie WS, Barnes AC, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018;391:541-551.
66. Timar R, Velea I, Timar B, et al. Factors influencing the quality of life perception in patients with type 2 diabetes mellitus. Patient Prefer Adherence 2016;10:2471-2477.
67. National Health Service (NHS). National Diabetes Audit 2020/21. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/e4-data/ [Last accessed 1 August 2022].
68. Blak BT, Rigney U, Sternhufvud C, et al. Weight change and healthcare resource use in English patients with type 2 diabetes mellitus initiating a new diabetes medication class. Int J Clin Pract 2016;70:45-55.
69. Weiss T, Iglay K, Carr R, et al. 987-P: Real-World Weight Change, Adherence, and Discontinuation among Type 2 Diabetes (T2D) Patients Initiating GLP-1 Receptor Agonists (GLP-1RAs) in the UK. Diabetes 68 (Supplement 1). doi: 10.2337/db19-987-P. 2019.
70. Valabhji J. NHS Blog. Type 2 diabetes and the importance of prevention. 2018. Available at: https://www.england.nhs.uk/blog/type-2-diabetes-and-the-importance-of-prevention/#:~:text=Around%20nine%20out%20of%2010,and%20through%20successful%20weight%20loss. [Last accessed 22 February 2022].
71. Diabetes UK. Diabetes Statistics. 2019. Available at: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics/ [Last accessed 22 February 2022].
72. Hex N, Bartlett C, Wright D, et al. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabet Med 2012;29:855-62.
73. Lopes R, Ethgen O, Alzaid A. The cost of complications of type 2 diabetes (T2D) in established and emerging markets – estimates from the G15 review (PDB66). Presented at ISPOR Europe 2019, Copenhagen. Value Health 22(S3):S584–5. 2019.
74. Dilla T, Valladares A, Nicolay C, et al. Healthcare costs associated with change in body mass index in patients with type 2 diabetes mellitus in Spain: the ECOBIM study. Appl Health Econ Health Policy 2012;10:417-30.
75. Johnston SS, Ammann EM, Kashyap SR, et al. Body mass index and insulin use as identifiers of high-cost patients with type 2 diabetes: A retrospective analysis of electronic health records linked to insurance claims data. Diabetes Obes Metab 2019;21:1419-1428.
76. Li Q, Blume SW, Huang JC, et al. The Economic Burden of Obesity by Glycemic Stage in the United States. Pharmacoeconomics 2015;33:735-48.
77. Meyers J, Hoog M, Mody R, et al. Burden of cardiovascular disease and cardiovascular (CV) risk in patients with type 2 diabetes mellitus (T2DM): an analysis of a US health insurance database. Poster presented at the 2020 ISPOR Virtual Conference; May 2020. [abstract] Value Health 23 (Suppl 1):S115. doi: 10.1016/j.jval.2020.04.371. . 2020.
78. Lage MJ, Boye KS. The relationship between HbA1c reduction and healthcare costs among patients with type 2 diabetes: evidence from a U.S. claims database. Curr Med Res Opin 2020;36:1441-1447.
79. McBrien KA, Manns BJ, Chui B, et al. Health care costs in people with diabetes and their association with glycemic control and kidney function. Diabetes Care 2013;36:1172-80.
80. Bell K, Parasuraman S, Shah M, et al. Economic implications of weight change in patients with type 2 diabetes mellitus. Am J Manag Care 2014;20:e320-9.
81. Davis WA, Bruce DG, Davis TM. Economic impact of moderate weight loss in patients with Type 2 diabetes: the Fremantle Diabetes Study. Diabet Med 2011;28:1131-5.
82. Fridman M, Lucas ME, Paprocki Y, et al. Impact of Weight Change in Adults with Type 2 Diabetes Mellitus: A Literature Review and Critical Analysis. Clinicoecon Outcomes Res 2020;12:555-566.
83. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018;41:2669-2701.

84. Dahl D, Onishi Y, Norwood P, et al. Effect of Subcutaneous Tirzepatide vs Placebo Added to Titrated Insulin Glargine on Glycemic Control in Patients With Type 2 Diabetes: The SURPASS-5 Randomized Clinical Trial. JAMA 2022;327:534-545.
85. Del Prato S, Kahn SE, Pavo I, et al. Tirzepatide versus insulin glargine in type 2 diabetes and increased cardiovascular risk (SURPASS-4): a randomised, open-label, parallel-group, multicentre, phase 3 trial. Lancet 2021;398:1811-1824.
86. Frías JP, Davies MJ, Rosenstock J, et al. Tirzepatide versus Semaglutide Once Weekly in Patients with Type 2 Diabetes. N Engl J Med 2021;385:503-515.
87. Ludvik B, Giorgino F, Jódar E, et al. Once-weekly tirzepatide versus once-daily insulin degludec as add-on to metformin with or without SGLT2 inhibitors in patients with type 2 diabetes (SURPASS-3): a randomised, open-label, parallel-group, phase 3 trial. Lancet 2021;398:583-598.
88. Rosenstock J, Wysham C, Frías JP, et al. Efficacy and safety of a novel dual GIP and GLP-1 receptor agonist tirzepatide in patients with type 2 diabetes (SURPASS-1): a double-blind, randomised, phase 3 trial. Lancet 2021;398:143-155.
89. Eli Lilly Data on File. SURPASS-5 CSR.
90. Eli Lilly Data on File. SURPASS-2 CSR.
91. Eli Lilly Data on File. SURPASS-3 CSR.
92. Eli Lilly Data on File. SURPASS-4 CSR.
93. Centre for Reviews and Dissemination (CRD). Systematic reviews: CRD's guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination). Available at: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf/ [Accessed 5 August 2022].
94. Pan C, Gross JL, Yang W, et al. A Multinational, Randomized, Open-label, Treat-to-Target Trial Comparing Insulin Degludec and Insulin Glargine in Insulin-Naïve Patients with Type 2 Diabetes Mellitus. Drugs R D 2016;16:239-49.
95. Zinman B, Philis-Tsimikas A, Cariou B, et al. Insulin degludec versus insulin glargine in insulin-naive patients with type 2 diabetes: a 1-year, randomized, treat-to-target trial (BEGIN Once Long). Diabetes Care 2012;35:2464-71.
96. Orme ME, Nguyen H, Lu JY, et al. Comparative effectiveness of glycemic control in patients with type 2 diabetes treated with GLP-1 receptor agonists: a network meta-analysis of placebo-controlled and active-comparator trials. Diabetes Metab Syndr Obes 2017;10:111-122.
97. Pedder H, Dias S, Bennetts M, et al. Modelling time-course relationships with multiple treatments: Model-based network meta-analysis for continuous summary outcomes. Res Synth Methods 2019;10:267-286.
98. Rucker G. Network meta-analysis, electrical networks and graph theory. Res Synth Methods 2012;3:312-24.
99. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making 2013;33:657-70.
100. Pfeffer MA, Claggett B, Diaz R, et al. Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. N Engl J Med 2015;373:2247-57.
101. Holman RR, Bethel MA, Mentz RJ, et al. Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2017;377:1228-1239.
102. Hernandez AF, Green JB, Janmohamed S, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. Lancet 2018;392:1519-1529.
103. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2016;375:311-22.
104. Gerstein HC, Colhoun HM, Dagenais GR, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. Lancet 2019;394:121-130.
105. Marso SP, Bain SC, Consoli A, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med 2016;375:1834-1844.
106. Husain M, Birkenfeld AL, Donsmark M, et al. Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med 2019;381:841-851.

107. North EJ, Newman JD. Review of cardiovascular outcomes trials of sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists. Curr Opin Cardiol 2019;34:687-692.
108. Hiramatsu T, Asano Y, Mabuchi M, et al. Liraglutide relieves cardiac dilated function than DPP-4 inhibitors. Eur J Clin Invest 2018;48:e13007.
109. Wang X, Zhang H, Zhang Q, et al. Exenatide and Renal Outcomes in Patients with Type 2 Diabetes and Diabetic Kidney Disease. Am J Nephrol 2020;51:806-814.
110. Tuttle KR, Lakshmanan MC, Rayner B, et al. Dulaglutide versus insulin glargine in patients with type 2 diabetes and moderate-to-severe chronic kidney disease (AWARD-7): a multicentre, open-label, randomised trial. Lancet Diabetes Endocrinol 2018;6:605-617.
111. Davies MJ, Bain SC, Atkin SL, et al. Efficacy and Safety of Liraglutide Versus Placebo as Add-on to Glucose-Lowering Therapy in Patients With Type 2 Diabetes and Moderate Renal Impairment (LIRA-RENAL): A Randomized Clinical Trial. Diabetes Care 2016;39:222-30.
112. Mosenzon O, Blicher TM, Rosenlund S, et al. Efficacy and safety of oral semaglutide in patients with type 2 diabetes and moderate renal impairment (PIONEER 5): a placebo-controlled, randomised, phase 3a trial. Lancet Diabetes Endocrinol 2019;7:515-527.
113. Charbonnel B, Steinberg H, Eymard E, et al. Efficacy and safety over 26 weeks of an oral treatment strategy including sitagliptin compared with an injectable treatment strategy with liraglutide in patients with type 2 diabetes mellitus inadequately controlled on metformin: a randomised clinical trial. Diabetologia 2013;56:1503-11.
114. Yang J, Xiao W, Guo L, et al. Efficacy and safety of generic exenatide injection in Chinese patients with type 2 diabetes: a multicenter, randomized, controlled, non-inferiority trial. Acta Diabetol 2020;57:991-1000.
115. Pieber TR, Bode B, Mertens A, et al. Efficacy and safety of oral semaglutide with flexible dose adjustment versus sitagliptin in type 2 diabetes (PIONEER 7): a multicentre, open-label, randomised, phase 3a trial. Lancet Diabetes Endocrinol 2019;7:528-539.
116. Harreiter J, Just I, Leutner M, et al. Combined exenatide and dapagliflozin has no additive effects on reduction of hepatocellular lipids despite better glycaemic control in patients with type 2 diabetes mellitus treated with metformin: EXENDA, a 24-week, prospective, randomized, placebo-controlled pilot trial. Diabetes Obes Metab 2021;23:1129-1139.
117. Azar ST, Echtay A, Wan Bebakar WM, et al. Efficacy and safety of liraglutide compared to sulphonylurea during Ramadan in patients with type 2 diabetes (LIRA-Ramadan): a randomized trial. Diabetes Obes Metab 2016;18:1025-33.
118. Distiller LA, Nortje H, Wellmann H, et al. A 24-week, prospective, randomized, open-label, treat-to-target pilot study of obese type 2 diabetes patients with severe insulin resistance to assess the addition of exenatide on the efficacy of u-500 regular insulin plus metformin. Endocr Pract 2014;20:1143-50.
119. Bolli GB, Munteanu M, Dotsenko S, et al. Efficacy and safety of lixisenatide once daily vs. placebo in people with Type 2 diabetes insufficiently controlled on metformin (GetGoal-F1). Diabet Med 2014;31:176-84.
120. Ahren B, Leguizamo Dimas A, Miossec P, et al. Efficacy and safety of lixisenatide once-daily morning or evening injections in type 2 diabetes inadequately controlled on metformin (GetGoal-M). Diabetes Care 2013;36:2543-50.
121. clinicaltrials.gov. A Study of Tirzepatide (LY3298176) Versus Insulin Lispro (U100) in Participants With Type 2 Diabetes Inadequately Controlled on Insulin Glargine (U100) With or Without Metformin (SURPASS-6). Available at: https://clinicaltrials.gov/ct2/show/NCT04537923. [Accessed 5 May 2022].
122. clinicaltrials.gov. A Study of Tirzepatide (LY3298176) in Participants With Type 2 Diabetes on Metformin With or Without Sulfonylurea (SURPASS-AP-Combo). Available at: https://clinicaltrials.gov/ct2/show/NCT04093752 [Accessed 5 May 2022].
123. National Institute of Health and Care Excellence (NICE). NG28. Type 2 diabetes in adults: management. Health Economic Model Report. Available at: https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845/ [Accessed 10 April 2022].
124. All Wales Therapeutics and Toxicology Centre. AWMSG SECRETARIAT ASSESSMENT REPORT Semaglutide (Ozempic®) 1.34 mg/ml solution for injection in pre-filled pen Reference number: 1842 FULL SUBMISSION. 2018. Accessed on February 17, 2022.

Available at: https://awmsg.nhs.wales/medicines-appraisals-and-guidance/medicines-appraisals/semaglutide-ozempic/.
125. Bain SC, Hansen BB, Malkin SJP, et al. Oral Semaglutide Versus Empagliflozin, Sitagliptin and Liraglutide in the UK: Long-Term Cost-Effectiveness Analyses Based on the PIONEER Clinical Trial Programme. Diabetes Ther 2020;11:259-277.
126. Barnett AH, Arnoldini S, Hunt B, et al. Switching from sitagliptin to liraglutide to manage patients with type 2 diabetes in the UK: A long-term cost-effectiveness analysis. Diabetes Obes Metab 2018;20:1921-1927.
127. Becker F, Dakin HA, Reed SD, et al. Lifetime cost-effectiveness simulation of once-weekly exenatide in type 2 diabetes: A cost-utility analysis based on the EXSCEL trial. Diabetes Res Clin Pract 2022;183:109152.
128. Capehorn M, Hallén N, Baker-Knight J, et al. Evaluating the Cost-Effectiveness of Once-Weekly Semaglutide 1 mg Versus Empagliflozin 25 mg for Treatment of Patients with Type 2 Diabetes in the UK Setting. Diabetes Ther 2021;12:537-555.
129. Johansen P, Chubb B, Hunt B, et al. Evaluating the Long-Term Cost-Effectiveness of Once-Weekly Semaglutide Versus Once-Daily Liraglutide for the Treatment of Type 2 Diabetes in the UK. Adv Ther 2020;37:2427-2441.
130. Kansal A, Reifsnider OS, Proskorovsky I, et al. Cost-effectiveness analysis of empagliflozin treatment in people with Type 2 diabetes and established cardiovascular disease in the EMPA-REG OUTCOME trial. Diabet Med 2019;36:1494-1502.
131. McEwan P, Morgan AR, Boyce R, et al. The cost-effectiveness of dapagliflozin in treating high-risk patients with type 2 diabetes mellitus: An economic evaluation using data from the DECLARE-TIMI 58 trial. Diabetes Obes Metab 2021;23:1020-1029.
132. Pawaskar M, Bilir SP, Kowal S, et al. Cost-effectiveness of intensification with sodium-glucose co-transporter-2 inhibitors in patients with type 2 diabetes on metformin and sitagliptin vs direct intensification with insulin in the United Kingdom. Diabetes Obes Metab 2019;21:1010-1017.
133. Ramos M, Foos V, Ustyugova A, et al. Cost-Effectiveness Analysis of Empagliflozin in Comparison to Sitagliptin and Saxagliptin Based on Cardiovascular Outcome Trials in Patients with Type 2 Diabetes and Established Cardiovascular Disease. Diabetes Ther 2019;10:2153-2167.
134. Ramos M, Cummings MH, Ustyugova A, et al. Long-Term Cost-Effectiveness Analyses of Empagliflozin Versus Oral Semaglutide, in Addition to Metformin, for the Treatment of Type 2 Diabetes in the UK. Diabetes Ther 2020;11:2041-2055.
135. Ramos M, Ustyugova A, Hau N, et al. Cost-effectiveness of empagliflozin compared with liraglutide based on cardiovascular outcome trials in Type II diabetes. J Comp Eff Res 2020;9:781-794.
136. Reifsnider OS, Kansal AR, Franke J, et al. Cost-effectiveness of empagliflozin in the UK in an EMPA-REG OUTCOME subgroup with type 2 diabetes and heart failure. ESC Heart Fail 2020;7:3910-8.
137. Viljoen A, Hoxer CS, Johansen P, et al. Evaluation of the long-term cost-effectiveness of once-weekly semaglutide versus dulaglutide for treatment of type 2 diabetes mellitus in the UK. Diabetes Obes Metab 2019;21:611-621.
138. Palmer AJ, Si L, Tew M, et al. Computer Modeling of Diabetes and Its Transparency: A Report on the Eighth Mount Hood Challenge. Value Health 2018;21:724-731.
139. Si L, Willis MS, Asseburg C, et al. Evaluating the Ability of Economic Models of Diabetes to Simulate New Cardiovascular Outcomes Trials: A Report on the Ninth Mount Hood Diabetes Challenge. Value Health 2020;23:1163-1170.
140. Shao H, Shi L, Fonseca VA. Using the BRAVO Risk Engine to Predict Cardiovascular Outcomes in Clinical Trials With Sodium-Glucose Transporter 2 Inhibitors. Diabetes Care 2020;43:1530-1536.
141. Valentine WJ, Pollock RF, Saunders R, et al. The Prime Diabetes Model: Novel Methods for Estimating Long-Term Clinical and Cost Outcomes in Type 1 Diabetes Mellitus. Value Health 2017;20:985-991.
142. Pollock RF, Norrbacka K, Boye KS, et al. The PRIME Type 2 Diabetes Model: a novel, patient-level model for estimating long-term clinical and cost outcomes in patients with type 2 diabetes mellitus. J Med Econ 2022;25:393-402.

143. Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value Health 2012;15:796-803.
144. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. Diabetes Care 2004;27:2262-5.
145. Shao H, Fonseca V, Stoecker C, et al. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). Pharmacoeconomics 2018;36:1125-1134.
146. Willis M, Asseburg C, Nilsson A, et al. Multivariate Prediction Equations for HbA(1c) Lowering, Weight Change, and Hypoglycemic Events Associated with Insulin Rescue Medication in Type 2 Diabetes Mellitus: Informing Economic Modeling. Value Health 2017;20:357-371.
147. National Centre for Pharmacoeconomics (NCPE). The cost-effectiveness of oral semaglutide (Rybelsus®) for the treatment of adult patients with type II diabetes mellitus. Available at: https://www.ncpe.ie/wp-content/uploads/2020/03/Technical-Summary-Oral-Semaglutide-Rybelsus-HTA-ID-20011.pdf. Accessed on: 5 Aug 2022, 2021.
148. Holman RR, Thorne KI, Farmer AJ, et al. Addition of biphasic, prandial, or basal insulin to oral therapy in type 2 diabetes. N Engl J Med 2007;357:1716-30.
149. Leal J, Alva M, Gregory V, et al. Estimating risk factor progression equations for the UKPDS Outcomes Model 2 (UKPDS 90). Diabet Med 2021;38:e14656.
150. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. Clin J Am Soc Nephrol 2012;7:401-8.
151. Witkowski M, Wilkinson L, Webb N, et al. A Systematic Literature Review and Network Meta-Analysis Comparing Once-Weekly Semaglutide with Other GLP-1 Receptor Agonists in Patients with Type 2 Diabetes Previously Receiving Basal Insulin. Diabetes Ther 2018;9:1233-1251.
152. Nauck MA, Quast DR, Wefers J, et al. GLP-1 receptor agonists in the treatment of type 2 diabetes - state-of-the-art. Mol Metab 2021;46:101102.
153. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health 2010;13:509-18.
154. Nauck MA, Buse JB, Mann JFE, et al. Health-related quality of life in people with type 2 diabetes participating in the LEADER trial. Diabetes Obes Metab 2019;21:525-532.
155. Mitchell P, Annemans L, Gallagher M, et al. Cost-effectiveness of ranibizumab in treatment of diabetic macular oedema (DME) causing visual impairment: evidence from the RESTORE trial. Br J Ophthalmol 2012;96:688-93.
156. Shao H, Yang S, Fonseca V, et al. Estimating Quality of Life Decrements Due to Diabetes Complications in the United States: The Health Utility Index (HUI) Diabetes Complication Equation. Pharmacoeconomics 2019;37:921-929.
157. Evans M, Khunti K, Mamdani M, et al. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. Health Qual Life Outcomes 2013;11:90.
158. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Available at: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf/ [Accessed 31 May 2021].
159. Matza LS, Boye KS, Yurgin N, et al. Utilities and disutilities for type 2 diabetes treatment-related attributes. Qual Life Res 2007;16:1251-65.
160. Boye KS, Matza LS, Stewart KD, et al. Patient preferences and health state utilities associated with dulaglutide and semaglutide injection devices among patients with type 2 diabetes in Italy. J Med Econ 2019;22:806-813.
161. Boye KS, Matza LS, Stewart KD, et al. Health state utilities associated with weight loss in type 2 diabetes and obesity. J Med Econ 2022;25:14-25.
162. WHO Collaborating Centre for Drug Statistics Methodology. A10B Drugs Used in Diabetes: Blood Glucose Lowering Drugs, Excluding Insulins. Bigudanides. Available at: https://www.whocc.no/atc_ddd_index/?code=A10BA02/ [Accessed 11 July 2022].
163. WHO Collaborating Centre for Drug Statistics Methodology. A10A Drugs Used in Diabetes: Insulins and Analogues. Available at: https://www.whocc.no/atc_ddd_index/?code=A10AE04/ [Accessed 31 May 2022].

164. Personal Social Services Research Unit (PSSRU). Unit Costs Database of Health and Social Care Professionals 2020/2021. Available at: https://www.pssru.ac.uk/project-pages/unit-costs/ [Accessed 11 July 2022].
165. Hunt B, Ye Q, Valentine WJ, et al. Evaluating the Long-Term Cost-Effectiveness of Daily Administered GLP-1 Receptor Agonists for the Treatment of Type 2 Diabetes in the United Kingdom. Diabetes Ther 2017;8:129-147.
166. Kent S, Schlackow I, Lozano-Kühne J, et al. What is the impact of chronic kidney disease stage and cardiovascular disease on the annual cost of hospital care in moderate-to-severe kidney disease? BMC Nephrol 2015;16:65.
167. National Health Service (NHS). 2019/20 National Cost Collection Data Publication. Available at: https://www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/ [Accessed 11 July 2022].
168. Hayes AJ, Leal J, Gray AM, et al. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia 2013;56:1925-33.
169. National Institute for Health and Care Excellence (NICE). Resource impact template (type 2 diabetes and cardiovascular risk update). 2022 [NG28]. Available at: https://www.nice.org.uk/guidance/ng28/resources/ [Accessed 02 May 2022]. 2022.
170. Andersen JR, Aasprang A, Karlsen TI, et al. Health-related quality of life after bariatric surgery: a systematic review of prospective long-term studies. Surg Obes Relat Dis 2015;11:466-73.
171. Kolotkin RL, Andersen JR. A systematic review of reviews: exploring the relationship between obesity, weight loss and health-related quality of life. Clin Obes 2017;7:273-289.
172. Ma C, Avenell A, Bolland M, et al. Effects of weight loss interventions for adults who are obese on mortality, cardiovascular disease, and cancer: systematic review and meta-analysis. Bmj 2017;359:j4849.
173. Keng MJ, Leal J, Mafham M, et al. Performance of the UK Prospective Diabetes Study Outcomes Model 2 in a Contemporary UK Type 2 Diabetes Trial Cohort. Value Health 2022;25:435-442.
174. Mannucci E, Monami M. Cardiovascular Safety of Incretin-Based Therapies in Type 2 Diabetes: Systematic Review of Integrated Analyses and Randomized Controlled Trials. Adv Ther 2017;34:1-40.
175. Ma X, Liu Z, Ilyas I, et al. GLP-1 receptor agonists (GLP-1RAs): cardiovascular actions and therapeutic potential. Int J Biol Sci 2021;17:2050-2068.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Tirzepatide for the treatment of patients with type 2 diabetes

[ID3938]

Summary of Information for Patients (SIP)

9th August 2022

File name	Version	Contains confidential information	Date
ID3938_Eli Lilly_Tirzepatide for T2D_SIP [NoACIC]_08Aug		No	9th August 2022

Summary of Information for Patients (SIP):
The pharmaceutical company perspective

What is the SIP?

The Summary of Information for Patients (SIP) is written by the company who is seeking approval from NICE for their treatment to be sold to the NHS for use in England. It is a plain English summary of their submission written for patients participating in the evaluation. It is not independently checked, although members of the public involvement team at NICE will have read it to double-check for marketing and promotional content before it is sent to you.

The **Summary of Information for Patients** template has been adapted for use at NICE from the Health Technology Assessment International – Patient & Citizens Involvement Group (HTAi PCIG). Information about the development is available in an open-access IJTAHC journal article

SECTION 1: Submission summary

1a) Name of the medicine (generic and brand name):

Generic name: Tirzepatide

Brand name: Mounjaro®

1b) Population this treatment will be used by. **Please outline the main patient population that is being appraised by NICE:**

The population that this treatment will be used for is people living with type 2 diabetes (T2D) who would otherwise be treated with a type of medicine known as a **glucagon-like polypeptide-1 receptor agonist** (**GLP-1 RA**).

1c) Authorisation: Please provide marketing authorisation information, date of approval and link to the regulatory agency approval. If the marketing authorisation is pending, please state this, and reference the section of the company submission with the anticipated dates for approval.

The **Medicines and Healthcare products Regulatory Agency** (**MHRA**) is reviewing whether tirzepatide should be approved and granted **marketing authorisation** as a treatment for adults with T2D. More information on this can be found in **Document B**, **Section B.1.2**.[a]

[a]Please note that further explanations for the phrases highlighted in **orange** are provided in the glossary. Cross-references to other sections of this document or other appraisal documents are highlighted in **green**.

1d) Disclosures. Please be transparent about any existing collaborations (or broader conflicts of interest) between the pharmaceutical company and patient groups relevant to the medicine. Please outline the reason and purpose for the engagement/activity and any financial support provided:

In 2022, Lilly provided sponsorship funding to the following patient groups:

Table 1: Summary of patient group sponsorship funding

Patient organisation	Project	Financial Support
Diabetes UK	Sponsorship of DUK Professional Conference 2022	£55,000
Diabetes UK	International Scholarship supporting participation in EASD 2022 congress	£20,000
Diabetes Africa	Sponsorship of 'Diabetes Health Matters' patient webinar series	£46,000

SECTION 2: Current landscape

2a) The condition – clinical presentation and impact

Please provide a few sentences to describe the condition that is being assessed by NICE and the number of people who are currently living with this condition in England.

Please outline in general terms how the condition affects the quality of life of patients and their families/caregivers. Please highlight any mortality/morbidity data relating to the condition if available. If the company is making a case for the impact of the treatment on carers this should be clearly stated and explained.

Tirzepatide is being considered for the treatment of type 2 diabetes.

What is T2D?

T2D is a condition where **glucose** (sugar) levels in the blood are uncontrolled. This is caused by the **pancreas** not making enough **insulin**, or where the **insulin** made by the pancreas does not work properly. **Insulin** is a **hormone** which allows **glucose** to move from the blood into cells where it is used to provide energy.[1]

How many people have T2D?

It is estimated that one in ten adults over 40 years of age in the UK have T2D. As of 2019, over 3.9 million people in the UK were living with diabetes, and 90% of those people had T2D. In addition, it is estimated that almost a million people had **undiagnosed** T2D in the UK in 2019.[2]

What is the impact of T2D?

Life expectancy is reduced by up to 10 years on average in people with T2D.[3] Uncontrolled blood sugar levels in T2D can increase the risk of other serious conditions, such as damage to the eyes, heart and feet.[1] People with T2D should keep an eye on their health and have regular check-ups, because T2D can lead to:

- Heart disease and stroke
- Loss of feeling and pain (nerve damage)
- Foot problems, like sores and infections
- Vision loss and blindness
- Miscarriage and stillbirth
- Problems with the kidneys
- Sexual problems – like problems getting or keeping an erection

T2D can also impact everyday life, with side effects such as increased tiredness, a reduced ability to take part in daily activities and can affect mental health. But with the right treatment and care, people can live well with T2D.

T2D and weight

Weight is a key **risk factor** for developing T2D and can also affect the person's ability to control their blood sugar and blood pressure.[1] For adults with T2D who are overweight, NICE guidelines recommend discussing and agreeing an initial target of 5% to 10% body weight loss.[4] Losing weight can support blood sugar management and may be beneficial for the overall health of individuals living with T2D.[5]

2b) Diagnosis of the condition (in relation to the medicine being evaluated)

Please briefly explain how the condition is currently diagnosed and how this impacts patients. Are there any additional diagnostic tests required with the new treatment?

How is T2D diagnosed?

T2D is diagnosed following blood tests which can be arranged through a GP.

A blood test that detects the level of sugar (**glucose**) is used to confirm a T2D diagnosis. Many people with T2D do not get any symptoms or symptoms develop gradually and people do not notice them.

Figure 1 summarises the main T2D symptoms.

Figure 1: Symptoms of T2D

There are different types of blood test that can be used to diagnose T2D:

- **Glycated haemoglobin test** (also known as an **HbA1c** test) measures an average of blood sugar over 2–3 months
- **Non-fasting blood test** to measure blood sugar
- **Fasting blood test** to measure blood sugar, where food and drink (apart from water) cannot be consumed for at least eight hours before the test
- **Glucose tolerance test**, which has two parts:
    - First, a fasting blood test is done to measure blood sugar levels
    - Then a sugary drink is consumed, and a second blood test is done two hours later to measure blood sugar levels and see if a normal blood sugar level was maintained

Other types of test, such as urine tests, at-home diabetes testing kits, and eye tests, cannot diagnose diabetes. However, these tests can identify symptoms of diabetes and suggest that the person should get tested for diabetes.

Introducing tirzepatide for the treatment of T2D would not require any additional tests.

2c) Current treatment options:

The purpose of this section is to set the scene on how the condition is currently managed:

- What is the treatment pathway for this condition and where in this pathway the medicine is likely to be used? Please use diagrams to accompany text where possible. Please give emphasis to the specific setting and condition being considered by NICE in this review. For example, by referencing current treatment guidelines. It may be relevant to show the treatments people may have before and after the treatment under consideration in this SIP.
- Please also consider:
  - if there are multiple treatment options, and data suggest that some are more commonly used than others in the setting and condition being considered in this SIP, please report these data.
  - are there any drug–drug interactions and/or contraindications that commonly cause challenges for patient populations? If so, please explain what these are.

What current treatment options are available for T2D?

Some people can manage their blood sugar levels through changes to diet and lifestyle. Many people also need medicine to control their T2D, as this helps to keep blood sugar levels as normal as possible and prevent health problems.[6]

People with T2D are usually first prescribed **metformin**, which can be taken in combination with a **sodium glucose co-transporter inhibitor** (**SGLT2i**) in people who have heart disease, or a high risk of developing heart disease. If blood sugar levels continue to be uncontrolled with this medicine, changing treatments or adding a second and third medicine can be considered.

One option for a medicine that can be added is a **glucagon-like polypeptide-1 receptor agonist** (**GLP-1 RA**), which is considered for people with T2D who continue to have uncontrolled blood sugar levels when receiving **metformin** with two other drugs and who either:

- have a high **BMI** and **psychological** or medical problems associated with **obesity**
- do not have a very high **BMI**, but **insulin** therapy would have an impact on their health, or weight loss would benefit their health

The anticipated positioning of tirzepatide is as a new medicine that can be prescribed for patients with T2D that is not adequately controlled with three or more medicines, whenever **GLP-1 RAs** would otherwise be considered.

2d) Patient-based evidence (PBE) about living with the condition

Context:

- **Patient-based evidence (PBE)** is when patients input into scientific research, specifically to provide experiences of their symptoms, needs, perceptions, quality of life issues or experiences of the medicine they are currently taking. PBE might also include carer burden and outputs from patient preference studies, when conducted in order to show what matters most to patients and carers and where their greatest needs are. Such research can inform the selection of patient-relevant endpoints in clinical trials.

In this section, please provide a summary of any PBE that has been collected or published to demonstrate what is understood about **patient needs and disease experiences**. Please include the methods used for collecting this evidence. Any such evidence included in the SIP should be formally referenced wherever possible and references included.

T2D from the patient perspective

T2D is a serious condition and living with T2D can be very challenging. In order to maintain control of blood sugar levels patients have to make many changes to their daily lives, and this can have a big impact on the individuals. When patients were interviewed as part of a research study about how they feel about managing T2D, all of them reported that T2D influenced their daily life but the amount of impact it had on them varied from person to person. Lifestyle changes, medication and knowledge/control were all commonly reported as having an influence on patients' daily lives.[7]

In another study of how T2D affects the lives of patients currently being treated for T2D, the most common responses described by patients were concerns related to diet (82%), health complications related to T2D (74%), and weight changes or control (68%).[8] The majority of these patients also reported that achieving lower blood sugar levels would change their lives through both physical and psychological improvements.

SECTION 3: The treatment

3a) How does the new treatment work?

What are the important features of this treatment?

Please outline as clearly as possible important details that you consider relevant to patients relating to the mechanism of action and how the medicine interacts with the body

Where possible, please describe how you feel the medicine is innovative or novel, and how this might be important to patients and their communities.

If there are relevant documents which have been produced to support your regulatory submission such as a summary of product characteristics or patient information leaflet, please provide a link to these.

About tirzepatide and how it works

After food has been eaten, the amount of sugar in the blood increases. At the same time, **cells** in the gut produce **hormones** called **incretins**. The two main incretins are **GIP** and **GLP-1**, which control the amount of **insulin** that is produced. **Insulin** is a hormone that tells the cells to remove sugar from the blood; the cells can then use this sugar to generate energy.

In people with T2D, as more food is eaten the amount of sugar in the blood increases, but this is not controlled. This is because there is not enough **insulin** produced, or the **insulin** that is produced does not work.

Tirzepatide works in the same way as the incretin hormones **GIP** and **GLP-1**:

- It tells your **pancreas** how much **insulin** to release after you eat
- It helps improve how sensitive the body is to the effects of **insulin**, and it helps to lower body weight which may also contribute to improving **insulin sensitivity**
- It helps slow down how quickly food leaves your stomach, this may help **glucose** enter the blood more slowly

3b) Combinations with other medicines

Is the medicine intended to be used in combination with any other medicines?

- Yes / No

If yes, please explain why and how the medicines work together. Please outline the mechanism of action of those other medicines so it is clear to patients why they are used together.

If yes, please also provide information on the availability of the other medicine(s) as well as the main side

effects.

If this submission is for a combination treatment, please ensure the sections on efficacy (3e), quality of life (3f) and safety/side effects (3g) focus on data that relate to the combination, rather than the individual treatments.

Tirzepatide can be used either on its own, or with other T2D medicines if they are not controlling blood sugar levels well enough. Tirzepatide does not have a specific combination of medicines it must be taken with, and the combination of medicines should be taken in line with advice from a healthcare professional (HCP).

In the NHS, tirzepatide is expected to be used whenever **GLP-1 RAs** would otherwise be considered.

3c) Administration and dosing

How and where is the treatment given or taken? Please include the dose, how often the treatment should be given/taken, and how long the treatment should be given/taken for.

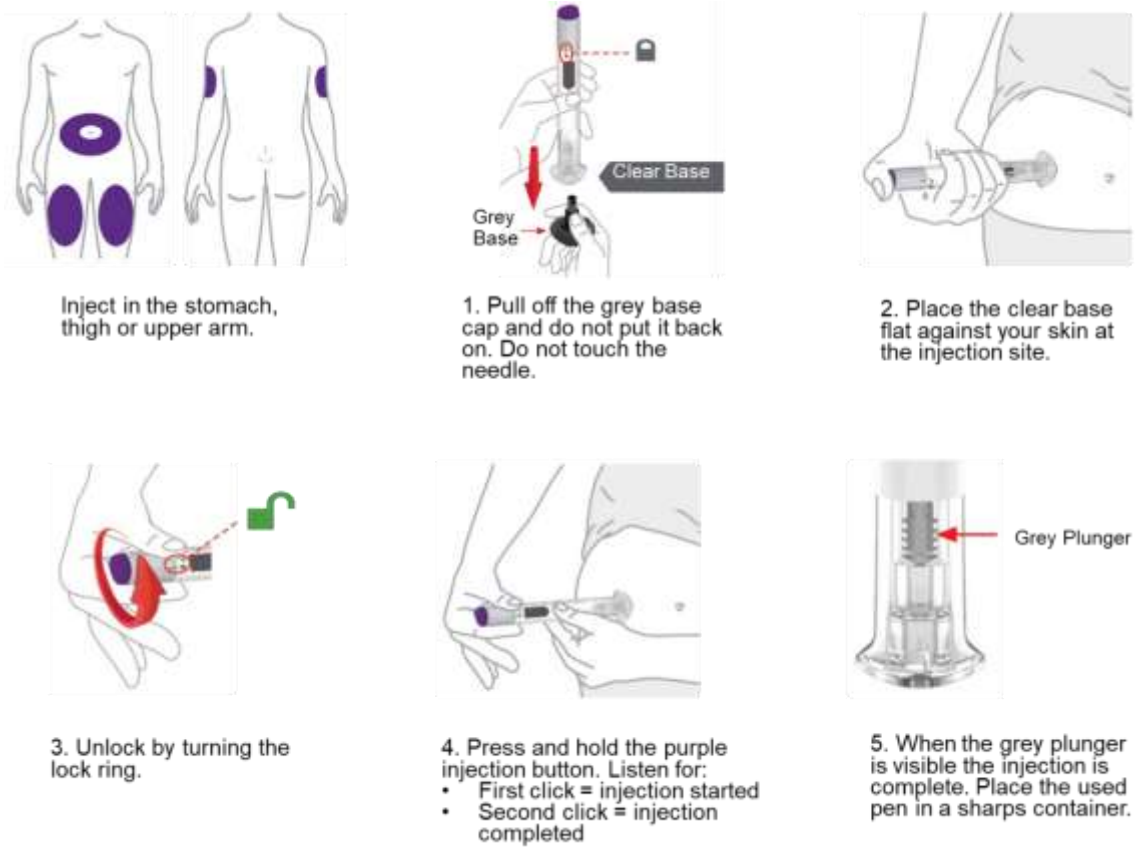
How will this administration method or dosing potentially affect patients and caregivers? How does this differ to existing treatments?

How is tirzepatide taken?

Tirzepatide should be used exactly as the doctor or pharmacist has instructed. Tirzepatide is injected under the skin of the stomach area or upper leg by the patient, using a **pre-filled pen**. The patient may require help from someone else if injecting into the upper arm. The area of the body that is used should be rotated each week. Tirzepatide can be injected at any time of day, with or without meals.[9]

Instructions for injecting tirzepatide are shown in **Figure 2**.

Figure 2. How to inject tirzepatide



Inject in the stomach, thigh or upper arm.

1. Pull off the grey base cap and do not put it back on. Do not touch the needle.

2. Place the clear base flat against your skin at the injection site.

3. Unlock by turning the lock ring.

4. Press and hold the purple injection button. Listen for:
- First click = injection started
- Second click = injection completed

5. When the grey plunger is visible the injection is complete. Place the used pen in a sharps container.

How much medicine do people with T2D take and when?

Tirzepatide is injected once weekly and the **dose** will be determined by a HCP. Tirzepatide dosing starts at 2.5 mg every week to help the patient adjust to the treatment and is increased to 5 mg every week after 4 weeks. The HCP may increase the **dose** further if needed. Recommended **doses** are 5, 10 or 15 mg every week; **doses** of 7.5 and 12.5 mg every week may be given for 4 weeks when changing between the recommended **doses** to help the patient adjust to the new **dose**; in each case the HCP will provide instructions on how long each **dose** should be taken for.

Treatment is continued for as long as the person's diabetes is controlled, unless it causes unmanageable **side effects** or the HCP and patient decide to stop the treatment.

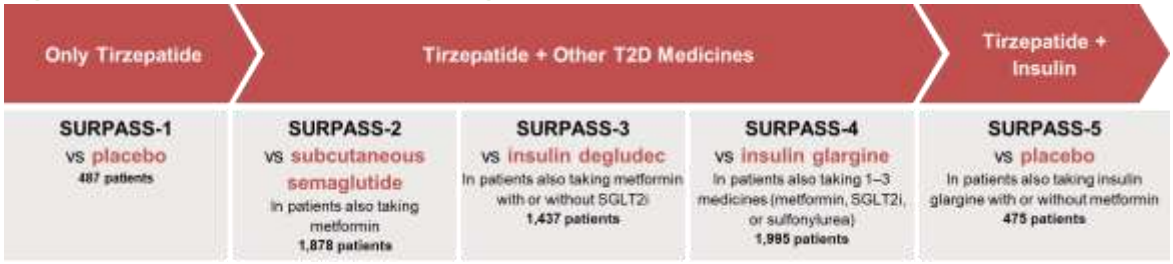
3d) Current clinical trials

Please provide a list of completed or ongoing clinical trials for the treatment. Please provide a brief top-level summary for each trial, such as title/name, location, population, patient group size, comparators, key inclusion and exclusion criteria and completion dates etc. Please provide references to further information about the trials or publications from the trials.

Studies of tirzepatide in T2D

The SURPASS **clinical trial** programme studied tirzepatide for the treatment of T2D. This programme consisted of several **phase 3** clinical trials, which means it tested the **efficacy** and safety of tirzepatide compared to other T2D medicines or **placebo**. A summary of the SURPASS clinical trials is shown in **Figure 3**Error! Reference source not found..

Figure 3: SURPASS clinical trial programme



How were the trials carried out?

People with T2D in the SURPASS clinical trials received either tirzepatide 5 mg, 10 mg or 15 mg, or received a **comparator** (either another T2D medicine or **placebo**). The **efficacy** of tirzepatide was tested by assessing participants' blood sugar levels before they started treatment, then at different timepoints during treatment, to see if blood sugar levels improved.

Lower **BMI** is related to better blood sugar control and maintaining a healthy weight can improve blood sugar levels and blood pressure. Therefore, body weight was also measured before and during treatment as a further **efficacy** measure. As well as these key changes, other measures were taken before and during treatment to more fully understand the effect of tirzepatide on people with T2D.

Ongoing studies of tirzepatide in T2D

As of August 2022, two further SURPASS clinical trials are currently. SURPASS-6 is similar to the trials described above, testing the **efficacy** and safety of tirzepatide compared to other T2D medicines and SURPASS-CVOT is designed to test the safety of tirzepatide treatment in people with T2D and heart disease. The results of these trials are expected in 2023 for SURPASS-6 and 2025 for SURPASS-CVOT. A further trial, SURPASS-AP-Combo, has recently been completed and also tested the **efficacy** and safety of tirzepatide compared to other T2D medicines.

3e) Efficacy

Efficacy is the measure of how well a treatment works in treating a specific condition.

In this section, please summarise all data that demonstrate how effective the treatment is compared with current treatments at treating the condition outlined in section 2a. Are any of the outcomes more important to patients than others and why? Are there any limitations to the data which may affect how to interpret the results? Please do not include academic or commercial in confidence information but where necessary reference the section of the company submission where this can be found.

Trial results

The SURPASS **clinical trials** measured the effect of tirzepatide on blood sugar control and weight loss. In summary, patients treated with all **doses** of tirzepatide had greater reductions in blood sugar levels and body weight compared with other T2D treatments or **placebo**. Higher proportions of patients treated with all **doses** of tirzepatide achieved blood sugar levels of **HbA1c** <7.0% (53 mmol/mol), ≤6.5% (≤48 mmol/mol) and <5.7% (<39 mmol/mol) compared with other T2D treatments or **placebo** and higher proportions of patients treated with tirzepatide achieved body weight reductions of ≥5%, ≥10%, or ≥15% compared with other T2D treatments or **placebo**. The key **efficacy** results are shown in **Table 2** and **Table 3**. More

efficacy results, such as the percentage of patients achieving weight loss targets and additional HbA1c targets, can be found in **Document B**, **Section B.2.6**.

Table 2: Blood sugar efficacy results from the SURPASS trials

Efficacy measure	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 vs injectable semaglutide 1 mg (40 weeks)				
Blood sugar change from baseline, % (mmol/mol)	−2.09 (−22.8)	−2.37 (−25.9)	−2.46 (−26.9)	−1.86 (−20.3)
Percentage of patients who achieved a target blood sugar level of HbA1c <7.0% (%)	85.5	88.9	92.2	81.1
SURPASS-3 vs insulin degludec (52 weeks)				
Blood sugar change from baseline, % (mmol/mol)	−1.93 (−21.1)	−2.20 (−24.0)	−2.37 (−26.0)	−1.34 (−14.6)
Percentage of patients who achieved a target blood sugar level of HbA1c <7.0% (%)	82.4	89.7	92.6	61.3
SURPASS-4 vs insulin glargine (52 weeks)				
Blood sugar change from baseline, % (mmol/mol)	−2.24 (−24.5)	−2.43 (−26.6)	−2.58 (−28.2)	−1.44 (−15.7)
Percentage of patients who achieved a target blood sugar level of HbA1c <7.0% (%)	81.0	88.2	90.7	50.7
SURPASS-5 vs placebo (40 weeks)				
Blood sugar change from baseline, % (mmol/mol)	−2.23 (−24.4)	−2.59 (−28.3)	−2.59 (−28.3)	−0.93 (−10.2)
Percentage of patients who achieved a target blood sugar level of HbA1c <7.0% (%)	93.0	97.4	94.0	33.9

Table 3: Weight loss efficacy results from the SURPASS trials

Efficacy measure	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 vs injectable semaglutide 1 mg (40 weeks)				
Body weight change from baseline (kg)	−7.8	−10.3	−12.4	−6.2
SURPASS-3 vs insulin degludec (52 weeks)				
Body weight change from baseline (kg)	−7.5	−10.7	−12.9	2.3
SURPASS-4 vs insulin glargine (52 weeks)				
Body weight change from baseline (kg)	−7.1	−9.5	−11.7	1.9
SURPASS-5 vs placebo (40 weeks)				

Body weight change from baseline (kg)	−6.2	−8.2	−10.9	1.7

Indirect treatment comparison

For practical and ethical reasons, clinical trials usually only directly compare a small number of medicines. To compare tirzepatide with all other treatments that people with T2D might receive, **indirect comparisons** are used. This is a common approach in evaluations of new medicines. An indirect comparison was done in this submission to compare tirzepatide with other **GLP-1 RAs**. This indirect comparison is explained in further detail in **Document B**, **Section B.2.9**.

3f) Quality of life impact of the medicine and patient preference information

What is the clinical evidence for a potential impact of this medicine on the quality of life of patients and their families/caregivers? What quality of life instrument was used? If the EuroQol-5D (EQ-5D) was used does it sufficiently capture quality of life for this condition? Are there other disease specific quality of life measures that should also be considered as supplementary information?

Please outline in plain language any quality of life related data such as **patient reported outcomes (PROs).**

Please include any **patient preference information (PPI)** relating to the drug profile, for instance research to understand willingness to accept the risk of side effects given the added benefit of treatment. Please include all references as required.

How was quality of life measured?

The SURPASS trials assessed the **quality of life** of people with T2D through several measures:

- The **APPADL questionnaire** asked people how difficult they found it to take part in activities of normal daily life, such as walking, standing, and climbing stairs
- The **IWQOL-Lite-CT questionnaire** specifically tested the impact of weight on physical and mental health

Tirzepatide impact on quality of life

At the start and end of the SURPASS **clinical trials**, patients completed questionnaires which asked them how difficult it was to do normal activities such as walking and climbing stairs. They were also asked how their weight impacts their **quality of life**. Comparing the questionnaire scores at the start and the end of the trials showed whether patients thought their **quality of life** had improved.

In all trials, there was an improvement in activities of normal daily life scores for the groups who took the two highest **doses** of tirzepatide, and there was an improvement in the lowest **dose** group in all trials except SURPASS-5. There was also an improvement in daily living scores in the semaglutide group (an alternative drug that can be prescribed), but the improvement was greater for the highest **dose** tirzepatide group. There was no improvement in daily living scores in the **insulin** or **placebo** groups.

All three tirzepatide groups and the semaglutide group showed improvements in the weight-related **quality of life** scores, and the improvements for the higher **dose** tirzepatide groups were greater than for the semaglutide group.

3g) Safety of the medicine and side effects

When NICE appraises a treatment, it will pay close attention to the balance of the benefits of the treatment in relation to its potential risks and any side effects. Therefore, please outline the main side effects (as opposed to a complete list) of this treatment and include details of a benefit/risk assessment where possible. This will support patient reviewers to consider the potential overall benefits and side effects that the medicine can offer.

Based on available data, please outline the most common side effects, how frequently they happen compared with standard treatment, how they could potentially be managed and how many people had treatment adjustments or stopped treatment. Where it will add value or context for patient readers, please include references to the Summary of Product Characteristics from regulatory agencies etc.

Every medicine has **side effects,** and the same medicine can produce different reactions in different people. In the SURPASS **clinical trials**, tirzepatide was generally a well-tolerated treatment option for people with T2D. There were a higher number of side effects in people taking higher **doses** of tirzepatide.

Table 4 compares the percentage of patients reporting side effects in two of the clinical trials who were either taking tirzepatide or a **placebo**. The most common side effects of tirzepatide were **gastrointestinal events**, including nausea (feeling sick), diarrhoea, vomiting and constipation. To reduce these side effects, tirzepatide is started on a lower **dose**, then increased after every 4 weeks until the patient and HCP agree the **dose** is appropriate. The most common side effects seen with **GLP-1 RA** treatment were also **gastrointestinal events** and had a similar frequency to tirzepatide.

Some patients in the clinical trials stopped taking the treatment because of **gastrointestinal side effects**. The percentage of these patients is shown in **Table 5**. Overall, stopping rates with tirzepatide were low.

Results from an assessment of the risk of heart problems showed that tirzepatide was not linked with excess risk of events caused by heart problems.

The risk of significant **hypoglycaemia** (blood **glucose** < 3mmol/L [54mg/dL]) was higher when tirzepatide was used in combination with other therapies known to make hypoglycaemia more likely, than when tirzepatide was studied as the only treatment. The risk of severe **hypoglycaemia** (needing assistance) was low.

The side effects experienced by people treated with tirzepatide can be managed by following advice from the HCP.

Table 4: Percentage of patients taking tirzepatide who reported gastrointestinal side effects compared to those taking a placebo (data from SURPASS-1 and SURPASS-5) [10, 11]

	Tirzepatide			Placebo
	5 mg (n=237)	10 mg (n=240)	15 mg (n=241)	(n=235)
Nausea	12.2	15.4	18.3	4.3
Diarrhoea	11.8	13.3	16.6	8.9
Decreased appetite	5.5	9.6	11.2	1.3
Dyspepsia	8.0	7.5	5.4	2.6
Vomiting	5.1	5.0	9.1	2.1

Constipation	5.9	5.8	6.6	1.3
Abdominal pain	5.9	4.6	5.4	4.3

Table 5: Percentage of patients in the tirzepatide and placebo groups who stopped taking the treatment because of gastrointestinal side effects (data from SURPASS-1 and SURPASS-5) [10, 11]

	Tirzepatide			Placebo
	5 mg (n=237)	10 mg (n=240)	15 mg (n=241)	(n=235)
% discontinuation rates	3	5.4	6.6	0.4

Side effects associated with tirzepatide[9]

Very Common (may affect more than 1 in 10 people):

- Feeling sick (nausea)
- Diarrhoea
- Low blood sugar – this is very common when tirzepatide is used with medicines that contain a **sulphonylurea** and/or **insulin**
    - Symptoms of low blood sugar include headaches, drowsiness, weakness, dizziness, feeling hungry, confusion, irritability, fast heartbeat and sweating

Common (may affect up to 1 in 10 people):

- Low blood sugar – this is common when tirzepatide is used with both **metformin** and an **SGLT2i** (a type of diabetes medicine)
- Feeling less hungry
- Stomach pain
- Being sick (vomiting)
- Indigestion
- Constipation
- Bloating of the stomach
- Burping
- Gas (flatulence)
- Reflux or heartburn
- Feeling tired
- Injection site reactions (e.g. itching or redness)
- Increase of **enzymes** in the **pancreas** (such as lipase and amylase)

Uncommon (may affect up to 1 in 100 people):

- Low blood sugar – this is uncommon when tirzepatide is used with **metformin**
- Gallstones
- Fast pulse

Serious side effects (uncommon; may affect up to 1 in 100 people):

- **Inflamed pancreas** – causes severe pain in the stomach and back which does not go away. Patients should see a doctor immediately if they experience these symptoms

For further information on side effects and their frequency, see **Document B**, **Section B.2.10**.

3h) Summary of key benefits of treatment for patients

Issues to consider in your response:

- Please outline what you feel are the key benefits of the treatment for patients, caregivers and their communities when compared with current treatments.
- Please include benefits related to the mode of action, effectiveness, safety and mode of administration

Improved control of blood sugar levels

Tirzepatide treatment allows people with T2D to better control their blood sugar levels. In the SURPASS **clinical trials**, people receiving tirzepatide had a bigger decrease in blood sugar, getting closer to normal levels, compared with other T2D medicines. Better control of blood sugar levels means people are at lower risk of experiencing serious conditions associated with T2D, such as damage to the heart, eyes and feet.[12]

Weight loss

Tirzepatide treatment also helps people with T2D achieve weight loss. In the SURPASS trials, participants receiving all **doses** of tirzepatide achieved greater weight loss compared with other T2D medicines.

Achieving weight loss is important for people with T2D who are overweight or **obese**, as weight loss is associated with better control of blood sugar levels.[5]

3i) Summary of key disadvantages of treatment for patients

Issues to consider in your response:

- Please outline what you feel are the key disadvantages of the treatment for patients, caregivers and their communities when compared with current treatments. Which disadvantages are most important to patients and carers?
- Please include disadvantages related to the mode of action, effectiveness, side effects and mode of administration
- What is the impact of any disadvantages highlighted compared with current treatments

Tirzepatide treatment can cause some **side effects**. The SURPASS clinical trials showed that **gastrointestinal events** were most frequently reported by people with T2D treated with tirzepatide. These side effects can limit the use of higher **doses** of tirzepatide in some people with T2D. However, the side effects of tirzepatide treatment are similar to currently available T2D medicines and can mostly be managed by following advice from a HCP.

Tirzepatide is a medicine which is taken by injection, which some people with T2D may consider to be a disadvantage.

3i) Value and economic considerations

Introduction for patients:

Health services want to get the most value from their budget and therefore need to decide whether a new treatment provides good value compared with other treatments. To do this they consider the costs of treating patients and how patients' health will improve, from feeling better and/or living longer, compared with the treatments already in use. The drug manufacturer provides this information, often presented using a health economic model.

In completing your input to the NICE appraisal process for the medicine, you may wish to reflect on:

- The extent to which you agree/disagree with the value arguments presented below (e.g., whether you feel these are the relevant health outcomes, addressing the unmet needs and issues faced by patients; were any improvements that would be important to you missed out, not tested or not proven?)
- If you feel the benefits or side effects of the medicine, including how and when it is given or taken, would have positive or negative financial implications for patients or their families (e.g., travel costs, time-off work)?
- How the condition, taking the new treatment compared with current treatments affects your quality of life.

Introduction for patient groups

Healthcare administrators need to get the most value from limited budgets. To do this, they need to check whether a new medicine provides good value for money compared to other medicines. They will look at the costs of the new medicine and how the health of people is likely to improve if they take it. The pharmaceutical company that makes the medicines provides this information to healthcare administrators using a **health economic model**.

A **budget impact model** was also created to assess the costs of introducing tirzepatide as a new medicine for T2D. This model showed that the cost of tirzepatide does not go over the cost threshold specified by the NHS.

How the model reflects the condition

The health economic model simulates people with T2D with characteristics similar to those of people who would receive tirzepatide treatment in the NHS.

The effect of treatment with tirzepatide on T2D was modelled using the changes in blood sugar and weight seen in the SURPASS **clinical trials** and the **indirect comparison**. Other effects of treatment were included in the model to better represent the overall impact of treatment on people with T2D, such as changes in **blood pressure** and **cholesterol**.

Modelling how much a treatment improves quality of life

As well as direct changes to patient health, the model measured the impact of treatment on patient **quality of life**; this can include improvements in **quality of life** due to reduced symptoms or decreases in **quality of life** due to side effects of treatment.

Tirzepatide treatment helps people lose weight, which can improve **quality of life** by allowing them to more easily participate in daily activities. This was considered in the model by including an increase in **quality of life** if weight decreased.

The model also included reductions in **quality of life** whenever a patient with T2D had a serious T2D-related condition that would affect their health, such as problems with the heart, eyes, feet and kidney. Further reductions in **quality of life** were included when people experienced **side effects** of tirzepatide treatment, such as nausea and vomiting.

Uncertainty

All model results are to some extent uncertain. Analyses were conducted to test the uncertainty around the model inputs which found that there was a 73.9% to 99.7% probability that tirzepatide is cost-effective.

Results of the economic analysis

All of these considerations affect whether tirzepatide represents good value for money and a good use of NHS resources. Based on the evidence that is available and the results of the economic analysis, tirzepatide is considered to be a good use of NHS resources for the treatment of people with T2D who have uncontrolled blood sugar levels when taking three T2D medicines.

Tirzepatide is injected by the patient, so there are no additional costs for health services to give treatment.

The cost effectiveness results of the economic analysis are presented in **Document B**, **Section B.3**.

3j) Innovation

NICE considers how innovative a new treatment is when making its recommendations.
If the company considers the new treatment to be innovative please explain how it represents a 'step change' in treatment and/ or effectiveness compared with current treatments. Are there any QALY benefits that have not been captured in the economic model that also need to be considered (see section 3f)

Tirzepatide works differently to currently available T2D medicines. Tirzepatide can **bind** to two different proteins found on the surface of **cells**, called **GIP** and **GLP-1 receptors**, and is the first medicine to be able to do this.

This binding causes greater **insulin** production from the cells and greater sugar removal from the bloodstream into the cells, reducing the amount of sugar in the blood to lower levels compared with other T2D medicines it has been tested against.[13] More patients taking tirzepatide met blood sugar targets compared to patients taking the other medicines or **placebo**.

Tirzepatide also brings benefits in weight loss and has shown better weight reduction compared to **placebo** or other medicines. **Obesity** is linked to poorer blood sugar control. Across the trials, more patients taking tirzepatide met the initial target of 5% to 10% body weight loss set by NICE guidelines[4] compared to patients taking the other medicines or **placebo**. Losing weight can support blood sugar management and may be beneficial for the overall health of individuals living with T2D. [5]

Tirzepatide represents a new and effective treatment option for people with T2D that may address some of the unmet needs that still exist with this very common and serious disease. It is anticipated that clinicians will use tirzepatide as a treatment option in people living with T2D who would otherwise be treated with a **GLP-1 RA**.

3k) Equalities

Are there any potential equality issues that should be taken into account when considering this condition and this treatment? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics More information on how NICE deals with equalities issues can be found in the NICE equality scheme Find more general information about the Equality Act and equalities issues here
There are no equality issues associated with T2D and tirzepatide treatment.

SECTION 4: Further information, glossary and references

4a) Further information

Feedback suggests that patients would appreciate links to other information sources and tools that can help them easily locate relevant background information and facilitate their effective contribution to the NICE assessment process. Therefore, please provide links to any relevant online information that would be useful, for example, published clinical trial data, factual web content, educational materials etc. Where possible, please provide open access materials or provide copies that patients can access.
Information on T2D for people with T2D: • What is Type 2 Diabetes? https://www.nhs.uk/conditions/type-2-diabetes/ • Type 2 Diabetes: https://www.diabetes.org.uk/diabetes-the-basics/types-of-diabetes/type-2 • Type 2 Diabetes Risk Factors: https://www.diabetes.org.uk/preventing-type-2-diabetes/diabetes-risk-factors • Getting Tested for Diabetes: https://www.diabetes.org.uk/diabetes-the-basics/test-for-diabetes • Symptoms. Type 2 Diabetes: https://www.nhs.uk/conditions/type-2-diabetes/symptoms/ • What are the Signs and Symptoms of Diabetes? https://www.diabetes.org.uk/diabetes-the-basics/diabetes-symptoms • Understanding Medicine. Type 2 Diabetes: https://www.nhs.uk/conditions/type-2-diabetes/understanding-medication/ **Further information on NICE and the role of patients:** • Public Involvement at NICE Public involvement | NICE and the public | NICE Communities | About | NICE • NICE's guides and templates for patient involvement in HTAs Guides to developing our guidance | Help us develop guidance | Support for voluntary and community sector (VCS) organisations | Public involvement | NICE and the public | NICE Communities | About | NICE • EUPATI guidance on patient involvement in NICE: https://www.eupati.eu/guidance-patient-involvement/ • EFPIA – Working together with patient groups: https://www.efpia.eu/media/288492/working-together-with-patient-groups-23102017.pdf • National Health Council Value Initiative. https://nationalhealthcouncil.org/issue/value/ • INAHTA: http://www.inahta.org/ • European Observatory on Health Systems and Policies. Health technology assessment - an introduction to objectives, role of evidence, and structure in Europe: http://www.inahta.org/wp-

content/themes/inahta/img/AboutHTA_Policy_brief_on_HTA_Introduction_to_Objectives_Role_of_Evidence_Structure_in_Europe.pdf

4b) Glossary of terms

This glossary explains terms highlighted in **orange** in this document. At times, an explanation for a term might mean you need to read other terms to understand the original terms.

APPADL questionnaire	This is a questionnaire that assesses the 'Ability to Perform Physical Activities of Daily Living' of patients with type 2 diabetes.
Binding	This occurs when a molecule (such as **insulin**) attaches to a **receptor** and causes a response in the **cell.**
Blood pressure	The pressure of blood in your arteries. Usually you see it as two numbers e.g. 120/80. The top number is your systolic pressure. This is measured when your heart contracts (pumps). The lower number is your diastolic pressure. This is the pressure when you heart relaxes.
BMI	A calculation often used to work out your weight compared to your height. You can calculate this by dividing your weight (in kg) by your height (in metres squared).
Budget impact model	A way of estimating the extra cost that the NHS would have to pay once a new treatment is approved and used by patients.
Cholesterol	A type of fat that is found in the blood that helps the brain, skin, and other organs do their jobs.
Clinical trial	A type of research study that tests how well new medical approaches work in people. These studies test new methods of screening, prevention, diagnosis or treatment of a disease. Also called a clinical study.
Comparator	The standard (for example, another medicine or usual care) against which a medicine is compared in a study. The comparator can be no intervention (for example, best supportive care).
Dose	The measured amount of a medicine that is taken at a particular time.
Efficacy	The ability of a medicine to produce a desired positive effect on your disease or illness in a **clinical trial**.
Enzymes	These proteins help speed up chemical reactions in the human body.
Fasting blood test	A blood test that is taken after several hours of fasting (not eating) and is used to help diagnose diabetes.
Gastrointestinal events	**Side effects** related to the organs that food and liquids travel through when they are swallowed, digested, absorbed and leave the body (such as the stomach and intestines).

GIP	An **incretin hormone** found in the intestine.
GLP-1	An **incretin hormone** found in the intestine.
GLP-1 RA	A GLP-1 receptor agonist is a type of medicine that is used to treat diabetes.
Glucose	The main type of sugar found in the blood. Glucose is the main source of energy for the body's cells.
Glucose tolerance test	This test measures the body's response to **glucose** and can be used to screen for type 2 diabetes. In the test a patient drinks a sugary drink and their blood **glucose** level is measured before and at intervals after the sugary drink is taken.
Glycated haemoglobin test	Also known as an HbA1c test, this test measures the average blood sugar (**glucose**) levels for the last 2–3 months.
HbA1c (glycated haemoglobin)	Glycated haemoglobin or HbA1c is made when **glucose** in the blood sticks to the red blood cells. High levels of HbA1c mean there is too much sugar in the blood.
Health economic model	A way to predict the costs and effects of a technology over time or in patient groups not covered in a **clinical trial**.
Hormones	Chemical substances that carry messages within the body to help coordinate different bodily functions.
Hypoglycaemia	Low blood sugar level.
Incretins	**Hormones** that are released after eating and cause blood sugar levels to decrease.
Indirect comparisons	An analysis that compares medicines that have not been compared directly in a head-to-head, randomised trial.
Inflamed pancreas	A swollen **pancreas** that causes tenderness and pain. This condition also known as pancreatitis.
Insulin	Insulin is a **hormone** created by your **pancreas** that controls the amount of **glucose** in your blood.
Insulin sensitivity	How sensitive the body is to the effects of **insulin**. People with high insulin sensitivity need smaller amounts of **insulin** to lower blood sugar and having low insulin sensitivity can lead to health problems.
IWQOL-Lite-CT questionnaire	The 'Impact of Weight on **Quality of Life**' questionnaire measures **quality of life** in patients with **obesity**.
Marketing authorisation	The legal approval by a **regulatory body** that allows a medicine to be given to patients in a particular country.
Metformin	A type of medicine, usually the first to be prescribed for type 2 diabetes.

MHRA	The **regulatory body** that evaluates, approves and supervises medicines throughout the European Union.
Non-fasting blood test	A blood test that is used to help diagnose diabetes. Unlike a **fasting blood test**, there is no need to not eat before this test.
Obese	This term describes a person who is very overweight, with a lot of body fat.
Pancreas	An organ the lies behind the stomach and helps produce **enzymes** that are released into the small intestine to help with digestion.
Phase 3 clinical trial	This type of **clinical trial** that tests the safety and how well a new treatment works compared with a standard treatment. For example, it evaluates which group of patients has better survival rates or fewer **side effects**.
Placebo	A treatment that appears real but has no therapeutic benefit. It is used in **clinical trials** to compare treatments.
Pre-filled pen	Device used to inject tirzepatide under the skin. The pen has a hidden needle which will automatically insert into the skin and inject tirzepatide when the injection button is pressed. The pen will automatically retract the needle when the injection is completed.
Psychological	Psychological problems relate to a person's mental or emotional state, rather than physical.
Quality of life	The overall enjoyment of life. Many **clinical trials** assess the effects of a disease and its treatment on the **quality of life** of patients. These studies measure aspects of a patient's sense of well-being and their ability to carry out activities of daily living.
Receptors	A structure on the surface of a **cell** that detects stimuli.
Regulatory body	These are legal bodies that review the quality, safety and **efficacy** of medicines and medical technologies.
Risk factor	Any aspect of a person's lifestyle, environment or pre-existing health condition that may increase their risk of developing a specific disease or condition.
SGLT2i	Sodium glucose co-transporter inhibitors are a type of medicine that is used to treat type 2 diabetes. They are taken as a tablet.
Side effects	An unexpected medical problem that arises during treatment. Side effects may be mild, moderate or severe.
Sulphonylurea	A type of medicine used to treat type 2 diabetes, they are taken as a tablet.
Undiagnosed	A disease that has not yet been identified.

References

1. Diabetes UK. Diabetes: The Basics. Available at: https://www.diabetes.org.uk/diabetes-the-basics. Accessed on: 28 Jun 2022.
2. Diabetes UK. Number of people with diabetes reaches 4.7 million. 2019. Available at: https://www.diabetes.org.uk/about_us/news/new-stats-people-living-with-diabetes/ [Last accessed: 10 February 2022].
3. Department of Health. National Service Framework for Diabetes: Standards. 2001. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198836/National_Service_Framework_for_Diabetes.pdf/ [Accessed 22 February 2022].
4. National Institute of Health and Care Excellence (NICE). Guideline: type 2 diabetes management. Available at: https://www.nice.org.uk/guidance/ng28/ [Last accessed 18 February 2022].
5. Diabetes UK. Weight loss and diabetes. Available at: https://www.diabetes.org.uk/guide-to-diabetes/enjoy-food/eating-with-diabetes/whats-your-healthy-weight/lose-weight. Accessed on: 29 Jun 2022.
6. NHS.uk. Type 2 Diabetes. Available at: https://www.nhs.uk/conditions/type-2-diabetes/understanding-medication/. Accessed on: 29 Jun 22.
7. van Smoorenburg AN, Hertroijs DFL, Dekkers T, et al. Patients' perspective on self-management: type 2 diabetes in daily life. BMC Health Serv Res 2019;19:605.
8. Gelhorn H, Balantac Z, Shinde S, et al. The Burden of Type 2 Diabetes and the Value of Achieving Near Normoglycemia from the Patient Perspective. Diabetes Ther 2021;12:1821-1837.
9. Eli Lilly Data on File. Tirzepatide Draft Summary of Product Characteristics (SmPC).
10. Rosenstock J, Wysham C, Frías JP, et al. Efficacy and safety of a novel dual GIP and GLP-1 receptor agonist tirzepatide in patients with type 2 diabetes (SURPASS-1): a double-blind, randomised, phase 3 trial. Lancet 2021;398:143-155.
11. Dahl D, Onishi Y, Norwood P, et al. Effect of Subcutaneous Tirzepatide vs Placebo Added to Titrated Insulin Glargine on Glycemic Control in Patients With Type 2 Diabetes: The SURPASS-5 Randomized Clinical Trial. JAMA 2022;327:534-545.
12. Diabetes UK. Complications of Diabetes. Available at: https://www.diabetes.org.uk/guide-to-diabetes/complications. Accessed on: 29 Jun 2022.
13. Thomas MK, Nikooienejad A, Bray R, et al. Dual GIP and GLP-1 Receptor Agonist Tirzepatide Improves Beta-cell Function and Insulin Sensitivity in Type 2 Diabetes. J Clin Endocrinol Metab 2021;106:388-396.

Sall NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Tirzepatide for the treatment of patients with type 2 diabetes

[ID3938]

Clarification questions

February 2023

File name	Version	Contains confidential information	Date
Tirzepatide EAG clarification letter to PM_Redacted		**Yes**	**20.02.2023**

Notes for company

Highlighting in the template

Square brackets and [redacted] highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [redacted] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

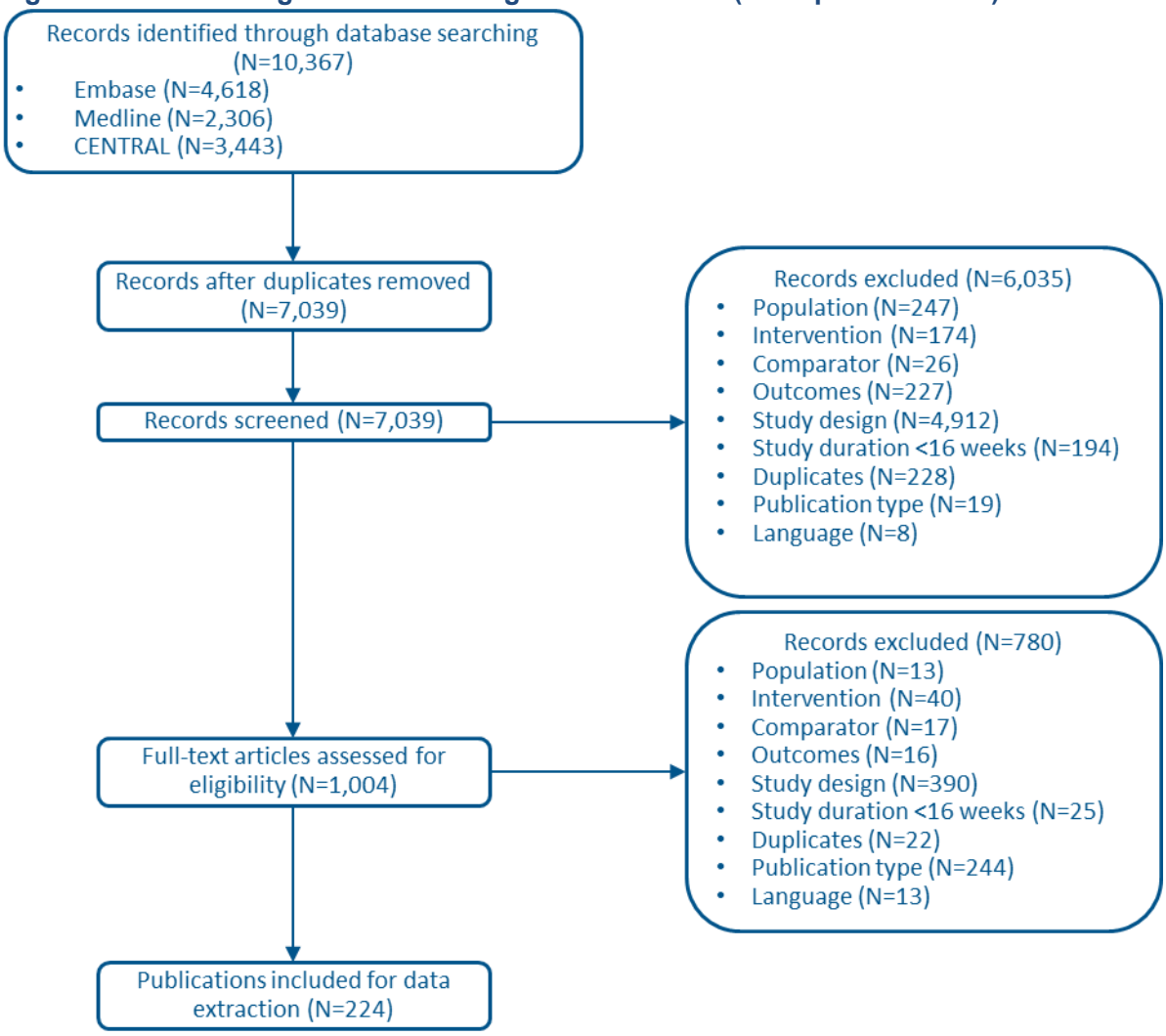
Section A: Clarification on effectiveness data (Heading 1)

Literature searches (Heading 2)

A 1. Please provide the correct PRISMA diagram for the original SLR. The diagram presented in on p.39 of Appendix D is for the October 2021 update.

The correct PRISMA diagram for the original systematic literature review (SLR) search (22 September 2021) is provided in Figure 1.

Figure 1: PRISMA diagram from the original SLR search (22 September 2021)



Abbreviations: SLR: systematic literature review.

A 2. Please provide the exact dates on which all of the clinical effectiveness SLR searches were conducted, and for each database please state the date range that was searched.

The dates that the searches were conducted (including the date range for each search) in electronic databases for the clinical effectiveness SLR are presented in Table 1 below.

Table 1: Clinical effectiveness SLR search dates

Search	Date Conducted	Date Range
Original Search	22nd September 2021	• Embase: 1st January 1990 to 22nd September 2021 • MEDLINE: 1st January 1990 to 22nd September 2021 • CENTRAL: 1st January 1990 to 22nd September 2021
1st Update Search	18th October 2021	• Embase: 22nd September 2021 to 18th October 2021 • MEDLINE: 22nd September 2021 to 18th October 2021 • CENTRAL: 22nd September 2021 to 18th October 2021

Search	Date Conducted	Date Range
2nd Update Search	21st June 2022	• Embase: 18th October 2021 to 21st June 2022 • MEDLINE: 18th October 2021 to 21st June 2022 • CENTRAL: 18th October 2021 to 21st June 2022

A 3. Please provide correct details of the database host(s) used to search CENTRAL, MEDLINE and Embase for the original September 2021 SLR searches. These appear to be searched via the Ovid host, rather than via ProQuest and the Cochrane Library as stated on p.18 in Appendix D.

For the original search (22nd September 2021) Embase, MEDLINE and CENTRAL were all searched via the Ovid search platform. For the update searches (18th October 2021 and 21st June 2022) Embase and MEDLINE were searched via the ProQuest search platform and CENTRAL was searched via the Cochrane Library search platform.

A 4. Please provide details of the database host(s) used to search Embase, the Cochrane Library and EconLit for the cost-effectiveness, HRQoL and cost/resource use studies documented in Appendices G-I.

All the databases searched in the literature review were accessed directly (via the respective online interface) as follows:

- EMBASE: https://www.embase.com/search/quick
- The Cochrane Library: https://www.cochranelibrary.com/advanced-search
- EconLit: https://search.ebscohost.com/

Decision problem

A 5. Priority question. The decision problem addressed by the company in the submission is much narrower that NICE scope and the MHRA marketing authorization. Table 1 (pg 16) of the CS also states that the rationale of the change from the scope includes that the intended position of tirzepatide in the care pathway as *'whenever GLP-1 RAs would otherwise be considered.'* and that 'This is the anticipated positioning of tirzepatide in UK clinical practice.' No population is specified for tirzepatide monotherapy. In addition, the patients' populations addressed in the company's clinical trials (SURPASS-2-5) appear to be misaligned with the population of the decision problem in the CS regarding specific treatment experience and treatment line (See Table 1). In fact, two trials exclude triple therapy experience within the 3 months prior to Visit 1: SURPASS-2

excludes any antihyperglycemic medication except metformin, SURPASS-3 excludes any other than metformin or and SGLT-2i, SURPASS-4 does permit triple therapy of metformin, an SGLT2i and an SU, which applied to only about [redacted] of the trial population (Table 14).

Table 2: comparison of various population definitions (NICE scope, license, trials, NICE guideline, decision problem)

NICE scope	**MHRA therapeutic indications**	**SURPASS trials populations**	**SURPASS trials treatment positioning**	**Treatment positioning of tirzepatide according to CS and NICE NG28**	**Decision problem addressed in the company submission**
Tirzepatide monotherapy: • Adults with type 2 diabetes (T2D) that is inadequately controlled with diet and exercise alone and in whom the use of metformin is considered inappropriate Tirzepatide with other antidiabetic agents: • Adults with type 2 diabetes that is inadequately controlled with one or more antidiabetic agents	Mounjaro is indicated for the treatment of adults with insufficiently controlled type 2 diabetes mellitus as an adjunct to diet and exercise • as monotherapy when metformin is considered inappropriate due to intolerance or contraindications • in addition to other medicinal products for the treatment of diabetes.	**SURPASS-2** Patients with T2D, who had inadequate glycaemic control with metformin monotherapy (≥1500 mg/day) and had not been treated with any other OADs during the 3 months prior to the start of the study	Second-line treatment	Third line of therapy: When triple therapy with metformin and two other oral drugs, one of which is a GLP-1 RA, is not effective, tolerated or contraindicated **and** patients: • have a BMI ≥35 kg/m^2 (adjust accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity, **or** • have a BMI <35 kg/m^2 **and**: ○ for whom insulin therapy would have significant occupational implications **or** ○ when weight loss would benefit other significant obesity related comorbidities, then change the GLP-1 RA to tirzepatide	Tirzepatide with other antidiabetic agents: • Adults with T2D that is inadequately controlled with three or more antidiabetic agents
		SURPASS-3 Patients with T2D, who had inadequate glycaemic control on stable doses of metformin with or without an SGLT2i	Second- or third-line treatment		
		SURPASS-4 Patients with T2D with high CVD risk, who had inadequate glycaemic control on stable doses of at least 1 and no more than 3 oral antidiabetic drugs (OADs), including metformin, an SGLT2i and/or an SU	Second-, third-or later line treatment		
		SURPASS-5 Patients with T2D, with background therapy of insulin glargine with or without metformin	Second-line treatment		

a) **Should the decision problem population be narrowed further according to the NG28 restrictions for GLP-1 RAs in terms of BMI, obesity related problems and potential weight loss benefit?**

Lilly consider that tirzepatide would be an option whenever glucagon-like peptide-1 receptor agonists (GLP-1 RAs) would otherwise be considered. While the current NG28 applies specific additional criteria to GLP-1 RA use, as mentioned in the question, Lilly consider that the likely position of tirzepatide in NHS practice will be driven by GLP-1 RA use rather than driven by the specific criteria in NG28 themselves and as such have defined the decision problem addressed in terms of GLP-1 RA use rather than the criteria listed in NG28. Nonetheless, given the NG28 restrictions apply in current practice there is no difference in the population between Lilly's definition and the NG28 GLP-1 RA population described in the question.

b) **Please confirm that the population in the decision problem is the one described by the company or that which includes the NG28 restrictions for GLP -1 RAs i.e., that it is the intention of the submission to address the clinical and cost effectiveness of tirzepatide as a combination therapy only and in the restricted population described.**

Lilly can confirm that the decision problem is intended to address the clinical and cost effectiveness of tirzepatide as a combination therapy only and in the restricted population of adults with T2D that is inadequately controlled with three or more antidiabetic agents as a more efficacious option whenever GLP-1 RAs would otherwise be considered, as this is the anticipated positioning of tirzepatide in UK clinical practice. This anticipated position aligns with current NHS clinical practice in England and reflects the highest unmet need for a more effective treatment option for patients for whom the alternative is a GLP-1 RA, which may not sufficiently control their HbA1c level and/or provide sufficient weight loss.

c) **Please confirm that the company has no intention of positioning tirzepatide beyond this restricted population.**

In alignment with current National Health Service (NHS) clinical practice, we believe that the data available at this time supports positioning as an alternative to GLP-1 RA therapies with the restrictions as described above. Within the framework of the current NG28 algorithm, there is no intention to position tirzepatide beyond the restricted population with the current data available.

d) **Given this population misalignment, please justify the decision to focus the submission on an alternative population, for which no direct evidence is presented, and to restrict the list of relevant comparators accordingly i.e. not to include a comparison with sulfonylureas, DDP-4 inhibitors, pioglitazone, GLP-1 mimetics, SGLT-2 inhibitors, or insulin, as monotherapy or in a combination regimens.**

The selected population for the submission aligns with the expected position of tirzepatide in current NHS clinical practice in England where GLP-1 RAs represent the most relevant comparators for the submission. Although there is no direct trial evidence in this population, an NMA was conducted to establish comparative efficacy for tirzepatide versus GLP-1 RAs.

Due to the anticipated positioning of tirzepatide, sulfonylureas, DDP-4 inhibitors, pioglitazone, SGLT-2 and insulin do not represent relevant comparators as they are prescribed at a different position within the treatment pathway; it was therefore not relevant to provide comparative efficacy for them.

The results of the SURPASS-4 subgroup analysis of baseline oral antidiabetic medication also provide reassurance as to the generalisability of the tirzepatide results irrespective of baseline therapy. The results were in line with those of the main analysis, demonstrating their generalisability: the subgroups in this analysis were defined as metformin alone, metformin + SU, metformin + sodium-glucose transport protein 2 inhibitor (SGLT2i), metformin + SU + SGLT2i, and Other (which comprised patients with either SU alone [3.7% of patients], SGLT2i alone [0.6% of the patients] and SU plus SGLT2i [0.8% of patients]). For body weight change from baseline to week 52 there was no significant treatment-by-subgroup interaction. For HbA1c change from baseline to week 52, the treatment-by-subgroup interaction was statistically significant, but this was likely due to the small sample size in the 'other' category leading to high variability. The results for body weight and HbA1c are presented in Figure 2 and Figure 3.

Figure 2: Forest plot for HbA1c (%) change from baseline to Week 52 baseline antidiabetic medication subgroup analysis

Abbreviations: HbA1c: haemoglobin A1c; IG: Insulin Glargine; LSMean: least square mean; SGLT2i: sodium-glucose co-transporter-2 inhibitor; SU: Sulfonylureas; TZP: tirzepatide.

Figure 3: Forest plot for body weight change from baseline to Week 52 baseline antidiabetic medication subgroup analysis

Abbreviations: IG: Insulin Glargine; LSMean: least square mean; SGLT2i: sodium-glucose co-transporter-2 inhibitor; SU: Sulfonylureas; TZP: tirzepatide.

e) Please provide the rationale for the design of the SURPASS trials given that the intended placement of tirzepatide in the care pathway is so different to the trial populations.

The evidence base for tirzepatide as a treatment for T2D is provided by the SURPASS trial programme. The SURPASS trials were designed to meet regulatory requirements of different authorities around the globe and to provide clinically meaningful data on the use of tirzepatide at different stages of T2D and its treatment continuum from monotherapy to the failure of basal insulin treatment. It was not feasible to assess all interim treatment stage scenarios.

As the positioning of GLP-1 RAs for the treatment of T2D varies globally, the trial designs do not completely align with UK clinical practice or the decision problem addressed within this submission. Such a misalignment is not uncommon in NICE appraisals and in this context the EAG may wish briefly to consider the Phase 3 trial designs listed in e.g. the semaglutide SmPC which are likewise not aligned to how this GLP-1 RA is used in NHS practice.[1]

Nonetheless, the trials do provide robust evidence for tirzepatide in T2D and their relevance to the decision problem is further discussed in Section B.2.2 of the submission. To account for the lack of direct comparative evidence provided by the trials, an NMA was conducted to establish comparative efficacy for tirzepatide generalisable to the relevant population.

Whilst none of the SURPASS trials exactly match the population as proposed for positioning in UK clinical practice, through the NMA comparative efficacy is demonstrated across multiple relevant efficacy outcomes, including the most critical clinical endpoints in the management of diabetes, such as change from baseline in glycated haemoglobin (HbA1c), weight and body mass index (BMI). These endpoints are the key drivers of clinical- and cost-effectiveness analyses of a treatment in T2D.

f) Please confirm that there are no studies (completed or ongoing) which evaluate tirzepatide, as a combination treatment, in the population/line of therapy specified (adults with T2D that was inadequately controlled with three or more antidiabetic agents).

SURPASS-4 is a completed study included in the submission which included patients inadequately controlled on up to three oral antidiabetic agents. There are no additional studies ongoing with this combination.

It may be noted that a cardiovascular outcome trial to evaluate the effect of tirzepatide versus dulaglutide on major adverse cardiovascular events (MACE) in patients with T2D is ongoing (SURPASS-CVOT). Participants in this study are permitted to use a variety of background diabetes therapies (except other GLP-1 RAs, DPP4is and pramlintide) as required per standard of care, and it is expected that the final population enrolled will include patients in whom tirzepatide will be used in combination with triple oral antidiabetic medicines. This trial is ongoing and does not adjust the position in the current framework.

g) Given the discrepancy between the SURPASS populations and the decision problem, please construct a decision problem that is within scope and more consistent with the trial evidence i.e., at an earlier line of therapy or in addition to insulin (as in SURPASS-5) and with the comparators appropriate to such a line of therapy.

The SURPASS trial programme was designed to inform licensing around the globe; as noted, the current NHS use of GLP-1 RAs may be less typical and at variance with use in other countries, nonetheless GLP-1 RAs remain the key comparator. The SURPASS trial programme is broadly similar to other trial programmes assessing the latest GLP-1 RAs and, alongside the NMA, provides a robust evidence base for tirzepatide in T2D and its use in comparison to GLP-1 RAs.[2-12] Clinical data available indicates significant efficacy measured by change from baseline in HbA1c regardless of number and type of combined oral agents. Whilst significant efficacy has been demonstrated across these trials, GLP1-RAs are not used at an earlier position in the NICE pathway, although earlier use of injectable therapies is accepted globally depending on patient characteristics.[13]

In conclusion, the company maintain that the decision problem addressed within the submission is appropriately aligned with UK clinical practice and have not presented an updated decision problem.

A 6. Priority question. The cost-effectiveness analysis compares tirzepatide to comparators defined by dose according to dose i.e., the comparators for 5 mg,10 mg and 15 mg are not identical. Table 46 shows the same for the NMA. However, Table 2 in the CS suggests that all patients are treated the same and will move from a maintenance dose of 5 mg to 10 mg or 15 mg as required.

a) **Please provide a justification for these three subgroups defined by tirzepatide dose in terms of UK clinical practice and the clinical criteria by which these subgroups would be identified for prescription of the appropriate dose.**

This question implies that the CS has been misunderstood: the CS does *not* specify three subgroups. The CS states that in seeking to *meaningfully interpret* the comparative efficacy results for tirzepatide 5, 10 and 15 mg with the comparators in Table 46 of CS it is more meaningful to compare within dose steps rather than between, as in any given individual patient the dose will have been titrated according to the balance of patient tolerability and observed treatment effect.

In clinical practice, patients are expected to be titrated up the recommended maintenance doses as required, and the most appropriate dose will be determined by the clinician based on clinical characteristics and patient tolerability, aligned to the SmPC which states for tirzepatide:

- The starting dose of tirzepatide is 2.5 mg once weekly. After 4 weeks, the dose should be increased to 5 mg once weekly. If needed, dose increases can be made in 2.5 mg increments after a minimum of 4 weeks on the current dose
- The recommended maintenance doses are 5, 10 and 15 mg
- The maximum dose is 15 mg once weekly

It is not anticipated that clinical practice will divide patients into subgroups with target doses upon initiating treatment. In the SURPASS trials, even among patients treated with the 5 mg tirzepatide dose, a number of patients achieved HbA1c as low as <5.7%. It is not possible to identify *a priori* patients who will respond sufficiently to the lower doses without a trial period at that maintenance dose. Given type 2 diabetes (T2D) is a chronic and progressive disease we anticipate that over time patients may require escalation through the doses, the timing and extent of escalation will be tailored to the individual.

b) **Given that in SURPASS-2 only one dose of semaglutide (1 mg) is considered as the comparator for all three dosed of tirzepatide and that three further doses of 2.0, 7.0 and 14 mg for semaglutide are reported to be licensed in section B.2.9.2, please justify the choice of doses for semaglutide in the NMA, as well as all other comparators by reference to UK clinical practice.**

The NMA was conducted on a global level and therefore includes a wide range of comparators and doses to account for various global markets. However, the discussion within the submission focuses in on relevant comparators to align with the treatments and doses available in UK.

To clarify, injectable semaglutide (branded Ozempic) is currently available in the UK at three doses; 0.25 mg (titration dose), 0.5 mg and 1.0 mg. Semaglutide has an additional licenced dose of 2.0 mg but, as stated in Section B.2.9.2 of the CS, this was not available at the time of the clinical trials and remains unavailable at this time in the UK. Given this, 1.0 mg injectable semaglutide is the highest available dose and is therefore the most appropriate comparator for the highest doses of tirzepatide (10 mg and 15 mg). However, as described in Table 76 of Document B in the submission, the lowest dose of tirzepatide (5 mg) was compared with the lower dose of 0.5 mg injectable semaglutide. As discussed in Section B.3.2.4, comparisons were made within each recommended maintenance dose step, rather than between.

Semaglutide is additionally available in 3 doses as an oral formulation (branded Rybelsus); 3.0 mg (titration dose), 7.0 mg and 14.0 mg. The oral formulation has a low absolute bioavailability and variable absorption. The exposure after 14.0 mg oral semaglutide is equivalent to injectable 0.5 mg semaglutide.[14, 15] There is no evidence available to suggest that oral semaglutide (7.0 mg or 14.0 mg) has greater efficacy than 1.0 mg injectable semaglutide. These doses were therefore considered separately from the available doses of injectable semaglutide in the NMA and cost-effectiveness analysis.

c) **If these are not subgroups and all patients will move to higher doses as required in clinical practice, then please conduct the NMA and cost-effectiveness analysis without subgroups and allowing a comparison with all comparators at all doses.**

The NMA presented in the submission does compare all comparators at all doses. The submission suggests *interpreting* it within maintenance dose steps, but the analysis itself is not constrained in this way. As discussed in the CS, the overall aim of the NMA was to provide robust results on the comparative efficacy and safety of tirzepatide 5 mg, 10 mg and 15 mg versus GLP-1 RAs available in NHS practice.

A 7. Priority question. According to the NICE scope, NICE NG28 only recommends insulin-based treatment if metformin is contraindicated or not tolerated or with metformin only if HbA1c not controlled on dual therapy. However, about 32% and 68% of patients in SURPASS-3 and SURPASS-4, respectively, where the comparator included insulin in some form, were on metformin alone. Please provide a rationale for the combination of insulin and only metformin in SURPASS-3 and SURPASS-4.

NICE NG28 represents current NHS practice however the clinical trial program was designed to meet regulatory requirements of different authorities around the globe and to provide clinically meaningful data on use of tirzepatide at different stages of T2D and its treatment continuum from monotherapy to the failure of basal insulin treatment relevant to all countries rather than to the unique situation in the NHS. The background treatments reflect the global nature of these trials.

Participants in the SURPASS-3 study were required to be on metformin with or without SGLT-2i, 68% were on metformin monotherapy. Participants in SURPASS-4 were required to be on 1–3 oral agents, 32% were on metformin monotherapy. The proportions on metformin alone reflected the treatment of the population enrolled into the trial and were not a design specification of the protocol.

A 8. Priority question. Eye complications and mortality are outcomes defined in the NICE scope but not included in the decision problem addressed by the company. On the other hand, change in body weight was not defined as an outcome in the NICE scope but has been added in the decision problem.

a) **Please provide a rationale for these differences.**

Outcome measures were chosen to align where possible with the scope and provide clinically relevant outcomes. Mortality data were provided in the appendices of the submission (Appendix F.1) and, as requested, eye complication data have been provided below.

Data on change in body weight from baseline have been included because body weight is clinically relevant and important to patients. Body weight is closely related to BMI and is an easily communicated and monitored target for patients. Additionally, body weight was more widely reported than BMI in comparator trials informing the NMA.

Substantial numbers of patients with T2D do not meet adequate weight loss goals on current T2D treatments.[16] Additionally, as obesity is the greatest risk factor for T2D and high BMI is linked to higher mortality and morbidity in T2D patients,[17-20] there is a clear unmet need for more efficacious treatment options to help more patients achieve body weight reduction alongside glycaemic control.[16] Whilst a target of 5–10% weight loss has been included in NG28 for a number of years, the importance of weight reduction as a targeted intervention for people with T2D is increasingly recognised and has been a key element in the recently updated ADA/EASD consensus report (September 2022).[13] The report comments on potential benefits that may extend beyond glycaemic management to improved risk factors for cardiometabolic disease and quality of life particularly with higher magnitudes of weight loss. Therefore, change in body weight from baseline was considered an important and relevant outcome and included in the submission.

b) Please include all outcomes listed in the scope.

Eye complications

Eye complication data at this time is limited due to trial exclusion criteria, duration of studies and monitoring performed within these studies. As noted in Section B.2.3 of the submission, patients with a history of proliferative diabetic retinopathy, diabetic maculopathy, or non-proliferative diabetic retinopathy that required acute treatment were excluded based on a dilated fundoscopic examination performed by a qualified eye care professional during screening, and therefore those most at risk of eye complications were excluded. This was however still included as a safety endpoint.

Across the Phase 3 clinical trials, a worsening of fundoscopic examination result was observed in [redacted] tirzepatide-treated patients ([redacted]). No serious adverse events (SAEs) from the SOC of eye disorders were reported in any of these [redacted] tirzepatide-treated patients. [redacted] of tirzepatide-treated patients in the Phase 3 clinical trials reported a TEAE of potential diabetic retinopathy complication. No tirzepatide-treated patients in the placebo-controlled analysis set experienced a serious or severe TEAE of potential diabetic retinopathy complication. [redacted] tirzepatide-treated patients in the dose effect analysis set experienced a serious or severe TEAE of potential diabetic retinopathy complication.[21]

These results did not show increased risk of worsening of retinopathy with tirzepatide treatment in the studied population. A dedicated addendum study to SURPASS-CVOT is ongoing to further investigate the impact of tirzepatide treatment on diabetic retinopathy progression. Whilst the detailed outcome data of the study are being investigated, the label includes a specific caution aligned with semaglutide reflecting those patients most at risk of eye complications.[1, 22]

Mortality

Mortality data at this time are limited by the size, duration and population enrolled in the studies. The SURPASS-4 study included patients at high CV risk and a proportion of patients were

followed up for to 2 years to accumulate adequate major cardiovascular events. The data from this study are included in the CV metanalysis demonstrating CV safety. Mortality outcomes will be evaluated in the ongoing cardiovascular outcome trial.

Appendix F.1 of the company submission presents available mortality data across the studies. Following adjudication by an external clinical endpoint committee, none on the patient deaths reported during the trials were ruled as related to treatment with tirzepatide.

A 9. Priority question. Table 1 (pg 16) of the CS describes the intended position of tirzepatide in the care pathway as *'whenever GLP-1 RAs would otherwise be considered.'*

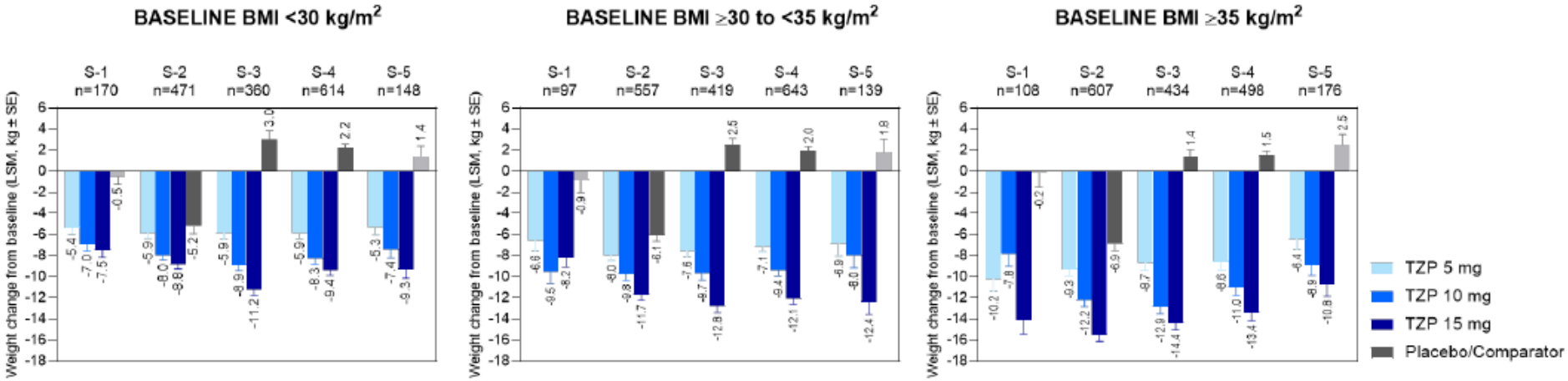
GLP-1 RAs are recommended in NG28 according to the following criteria:
'If triple therapy with metformin and 2 other oral drugs is not effective, not tolerated or contraindicated, consider triple therapy by switching one drug for a GLP-1 mimetic for adults with type 2 diabetes who:

- ***have a body mass index (BMI) of 35 kg/m^2 or higher (adjust accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity or***
- ***have a BMI lower than 35 kg/m^2 and for whom insulin therapy would have significant occupational implications or weight loss would benefit other significant obesity-related comorbidities.'***

However, the inclusion criteria for the SURPASS trials specify BMI of at least 25 kg/m2 (SURPASS-2 to 4) or at least 23 kg/m2 (SURPASS-5). Please discuss the implications of this lower BMI threshold in terms of clinical effectiveness.

Significant HbA1c and weight improvements have been seen regardless of baseline BMI (Figure 4 and Figure 5). Generally speaking, those with higher baseline BMI and on the higher dose of tirzepatide see greater reductions in HbA1c and body weight. For each study, subgroup analyses by baseline BMI (BMI <30 kg/m²; ≥30 to <35 kg/m²; ≥35 kg/m²) were performed for both change in HbA1c from baseline and change in weight from baseline to primary endpoint (40 or 52 weeks depending on the study), with p = 0.1 as the significance level.

Figure 4: LSM change from baseline in body weight (kg) at primary endpoint in BMI subgroups

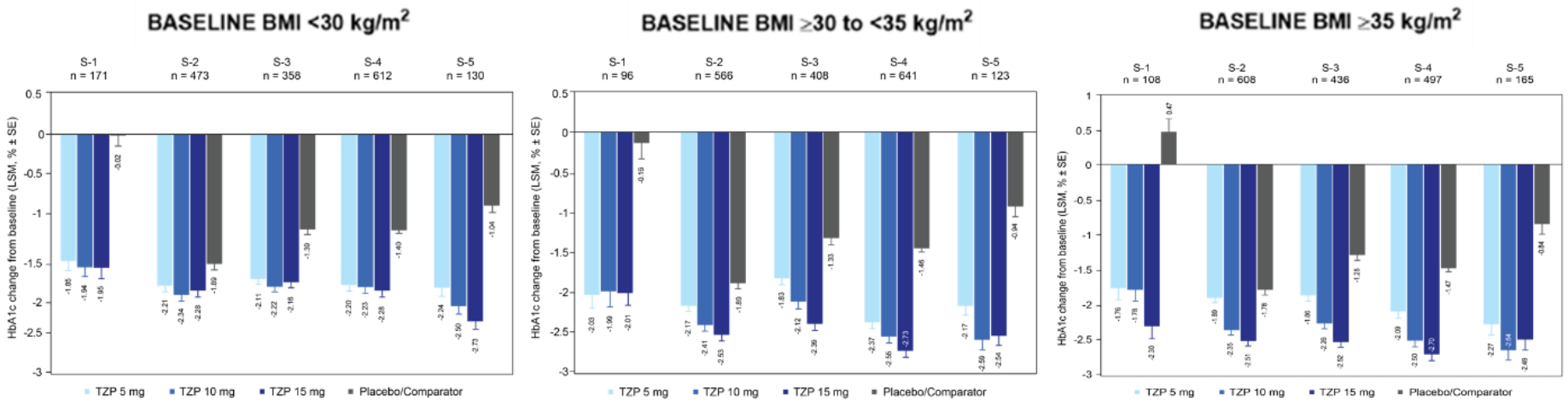


n = total number of patients overall in each SURPASS subgroup at primary endpoint. LSM change from baseline body weight to endpoint [Week 40 (S-1, -2 and -5) and Week 52 (S-3 and -4)] evaluated TZP (5, 10 & 15 mg) versus: Placebo (S-1); Semaglutide (S-2); Insulin degludec (S-3); Insulin glargine (S-4); Placebo(S-5). Note: LSM change was estimated using MMRM with treatment, visit, treatment-by-visit interaction, pooled country, baseline HbA1c group, baseline oral antihyperglycemic medication (when appropriate), and baseline weight as fixed effects, and patient as random effect.

Abbreviations: MMRM: mixed model for repeated measures; SLGT2i: sodium-glucose co-transporter 2 inhibitor; SU: sulfonylurea; LSM: least-squares mean; SE: standard error; TZP: tirzepatide.

Source: Kwan *et al,* 2022[23]

Figure 5: LSM change from baseline in HbA1c (%) at primary endpoint in BMI subgroups



n = total number of patients overall in each SURPASS subgroup at primary endpoint. LSM change from baseline body weight to endpoint [Week 40 (S-1, -2 and -5) and Week 52 (S-3 and -4)] evaluated TZP (5, 10 & 15 mg) versus: Placebo (S-1); Semaglutide (S-2); Insulin degludec (S-3); Insulin glargine (S-4); Placebo(S-5). Note: LSM change was estimated using MMRM with treatment, visit, treatment-by-visit interaction, pooled country, baseline HbA1c group, baseline oral antihyperglycemic medication (when appropriate), and baseline weight as fixed effects, and patient as random effect.
Abbreviations: MMRM: mixed model for repeated measures; SLGT2i: sodium-glucose co-transporter 2 inhibitor; SU: sulfonylurea; LSM: least-squares mean; SE: standard error; TZP: tirzepatide.
Source: Kwan *et al,* 2022[23]

A 10. According to the CS “studies in a specific population of patients with renal impairment (stage 3 or 4 CKD or macroalbuminuria) were excluded from the NMA.”. Nevertheless, the SURPASS trials included patients with varying degrees of renal impairment and the company states that it had “no significant effect on overall efficacy or safety results for TZP”. Please clarify that the decision problem population should exclude patients with renal impairment.

In the SURPASS trials the primary endpoint was the change in HbA1c level from baseline. In SURPASS 1–5 studies the majority of the patients had eGFR >60 mL/min per 1·73 m^2 (83–97%), so only a limited number of patients had decreased kidney function in these studies. In SURPASS 2–4, patients with eGFR <45 mL/min per 1.73 m^2 were excluded. The Chronic Kidney Disease stage 3 (eGFR 30-44) was included in SURPASS-1, but only [redacted] patients had eGFR <60 mL/min per 1·73 m^2. In SURPASS-5 only [redacted] of the patients had eGFR < 60 mL/min per 1.73 m^2.

A clinical pharmacology study has been performed to assess the impact of renal impairment on tirzepatide.[24] There have been no clinical efficacy studies exclusively in patients with CKD and T2D. Renal impairment does not affect pharmacokinetics (PK) of tirzepatide but patients with renal impairment are expected to have different efficacy and safety results than patients without renal disease. Hence studies conducted solely in a renal impairment population were excluded to align with SURPASS-2 and -3 as much as possible (exclusion criteria: patients with baseline eGFR <45). A total of 5 studies were excluded from the NMA.

Since the submission, a post hoc analysis has been performed on SURPASS-4 data to compare the effects of tirzepatide and insulin glargine on kidney parameters and outcomes in people with type 2 diabetes. Results of the analysis suggest that in people with type 2 diabetes and high cardiovascular risk, tirzepatide may slow the rate of eGFR decline and reduce the urine albumin-creatinine ratio compared with insulin glargine, warranting further research.[25]

The draft summary of product characteristics provides the following information to prescribers in relation to renal impairment:

- *Renal impairment does not impact the PK of tirzepatide. The PK of tirzepatide after a single 5 mg dose was evaluated in patients with different degrees of renal impairment (mild, moderate, severe, ESRD) compared with subjects with normal renal function and no clinically relevant differences were observed. This was also shown for patients with both type 2 diabetes mellitus and renal impairment based on data from clinical studies.*

- *No dose adjustment is required for patients with renal impairment including end stage renal disease (ESRD). Experience with the use of tirzepatide in patients with severe renal impairment and ESRD is limited. Caution should be exercised when treating these patients with Tirzepatide.*

Therefore, patients with renal impairment do not require exclusion from the decision problem population.

Systematic review

A 11. Section B.2.1 (pg 30) of the CS states that 'A clinical systematic literature review (SLR) was conducted in September 2021 to identify further relevant clinical evidence on the efficacy and safety of treatment of T2D, including tirzepatide, in patients with T2D who match the patient population of interest for this appraisal' and that this SLR is described in detail in Appendix D. Table 7 in Appendix D reports the eligibility criteria for this SLR, however, the reported criteria for population 'Adult patients (≥18 years of age) with T2D' do not appear to match the population for the decision problem, either as defined in the NICE scope or by the company (Table 1, pg 16 of the CS).

a) Please clarify whether (as indicated in in Appendix D, Table 7) the SLR included studies conducted in any adult population with T2D, irrespective of prior/background treatment or diabetes control.

Yes, the SLR included studies conducted in any adult population with T2D, irrespective of prior/background treatment or diabetes control.

b) If Table 7 in Appendix D accurately describes the inclusion criteria for the SLR, please explain why an SLR has been conducted using eligibility criteria which do not match the definition of the decision problem.

The SLR was designed with a broader scope than the company decision problem with the intention of meeting the needs of multiple HTA agencies globally. Although the SLR eligibility criteria were broader than the company decision problem, there was no risk that studies relevant to the company decision problem would be missed. At the stage of the NMA feasibility assessment, only studies from the SLR that were relevant to the company decision problem were considered.

c) Please specify how the list of interventions and comparators was populated.

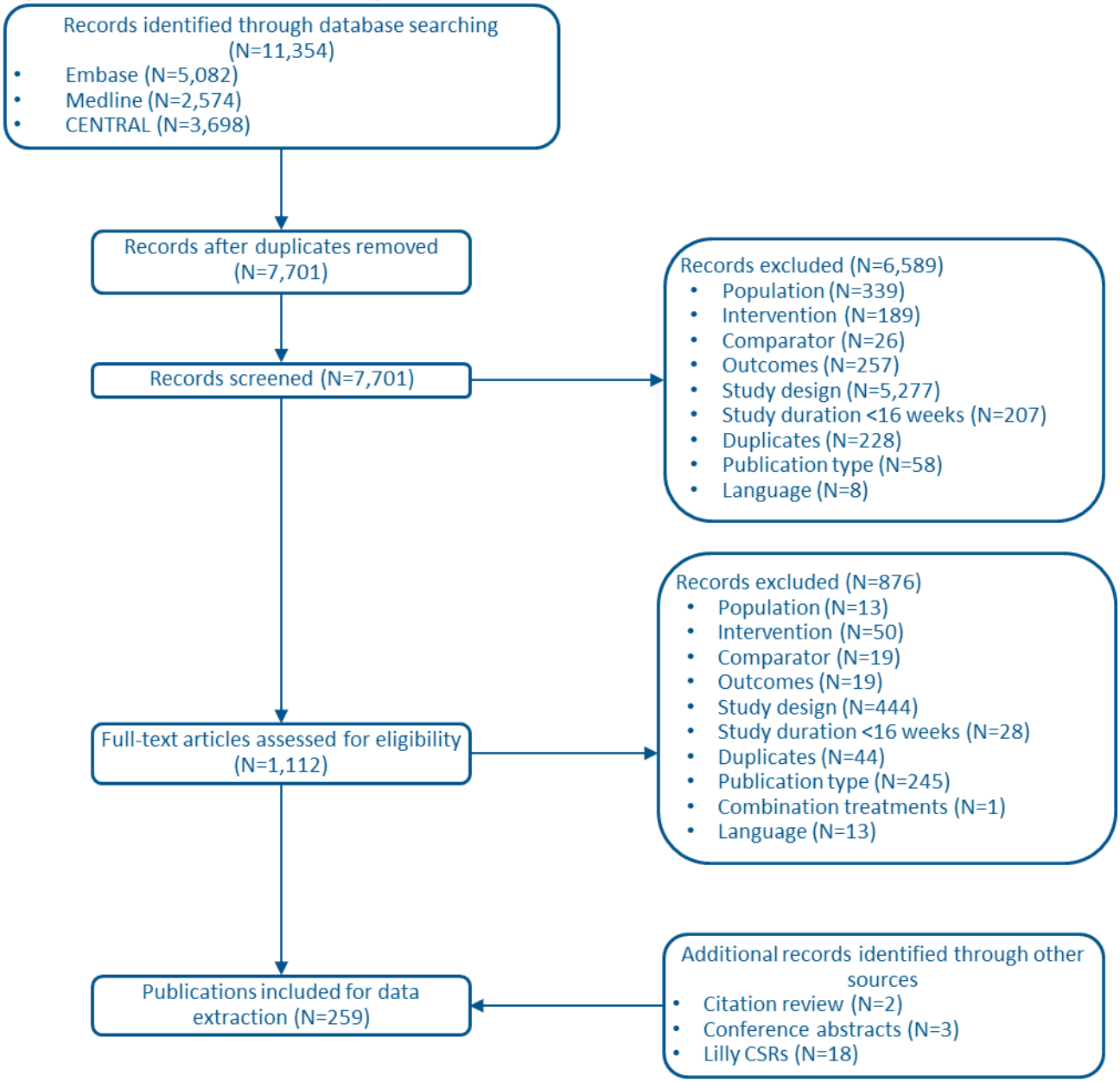
It is anticipated that HTA agencies around the globe will be interested in seeing evidence of indirect comparisons of tirzepatide to a range of GLP-1 RA therapies for the treatment of T2D. The list of interventions and comparators was populated to meet multiple HTA requirements.

A 12. Priority question. The PRISMA flow diagrams, presented in Figures 1 and 2 in Appendix D, are not consistent with the summary of search results provided in the text (section D.6.1). Figures 1 and 2 are described as flow charts for the first and up-date searches, respectively, however, the Figure 1 appears to be a duplicate of Figure 2. Please provide a single, complete PRISMA flow chart, illustrating the flow of studies through the review process, from search results to

inclusion/exclusion and including all searches and deduplication steps and providing the final number of included studies and publications.

A single, complete PRISMA diagram that combines the study flow from the original search (22nd September 2021), the first update search (18th October 2021) and the second update search (June 2022) has been provided in Figure 6 below.

Figure 6: Combined PRISMA diagram for all three searches



Abbreviations: CSR: clinical study report.

A 13. Section B.2.2 of the CS states that: "SURPASS-J-Mono and SURPASS-J-Combo were conducted in a Japanese population and are therefore not considered generalisable to the UK population; they are not presented as part of the clinical evidence in this appraisal." Please clarify whether similar, generalisability-based,

exclusion criteria were applied in the SLR and when selecting studies of comparators for inclusion in the NMA.

There were no exclusions in the SLR based on generalisability criteria. SURPASS-J-Mono and SURPASS-J-Combo were both included in the SLR. However, for the NMA, SURPASS-J-Mono and SURPASS-J-Combo did not meet the inclusion criteria of oral background treatment (SURPASS-J-Mono) or include comparator of interest (SURPASS-J-Combo) and were therefore excluded from the NMA

A 14. Please clarify if one or more reviewers were involved in the critical appraisal of the studies included in the CS as reported in section B.2.5 and Appendix D.10.

A single reviewer conducted the critical appraisal of any given study, and the findings of the critical appraisal were confirmed by a second reviewer. Any discrepancies between the findings were resolved either by agreement between the two reviewers or with referral to a third reviewer.

Tirzepatide trial evidence

A 15. Priority question. Please confirm that, as indicated in the CS, no data are available regarding the effects of tirzepatide on the complications of diabetes, or on mortality. Are you aware of any ongoing studies which will collect these data?

The SURPASS studies to date do not provide sufficient data to fully assess the impact on long term complications of diabetes or mortality. However, a pre-specified meta-analysis to compare time to first occurrence of confirmed four-component major adverse cardiovascular events (MACE-4; cardiovascular death, myocardial infarction, stroke and hospitalisation for unstable angina) has been conducted between pooled tirzepatide groups and control groups showing no increased risk. A Cardiovascular Outcome trial SURPASS-CVOT is currently ongoing (with expected completion in 2025) to further explore this important topic, this study includes UK clinical trial sites.

A 16. Priority question. Given the dose related subgroups in the cost-effectiveness analysis (see question A6), please explain why the tirzepatide trials were not stratified by eligibility for each tirzepatide dose with comparator dose chosen accordingly? In SURPASS-2 and SURPASS-3 randomization was stratified based on baseline HbA1c (≤8.5% or >8.5% [69 mmol/mol]).” Might HbA1c level be a proxy for eligibility for each tirzepatide dose?

HbA1c stratification in the clinical trials was to ensure balance in patient characteristics across the doses rather than to determine which dose the participants should have been assigned to. Imbalance between subgroups could result in bias in the analyses.

As highlighted in response to question A6, the most appropriate dose for an individual should be determined by the clinician in collaboration with the patient, based on clinical characteristics and patient tolerability and cannot be determined *a priori*; as further noted in the response to question A6, the maintenance dose steps are not defined subgroups but aid in meaningful interpretation of the NMA and CEM results.

A 17. Priority question. According to the CSRs of SURPASS-2, 3 and 4, a proportion of the patients was previously treated with a GLP-1 RA, <15%, <7% and <10% in every intervention arm, respectively.

a) Please discuss how previous use of GLP-1 RA has possibly affected the results of the trials.

For these studies, use of GLP-1 RAs in the three months before screening was an exclusion criterion. This would represent a suitable washout period based on the half-life of these therapies and is a standard practice in clinical trial design. Furthermore, GLP-1 RA therapy to date has no known disease-modifying effects and therefore the impact of history of GLP-1 RAs before the washout period is likely to be minimal if any.

b) Please provide details on the previous use of GLP-1 in the populations in SURPASS-5.

In the SURPASS-5 study, approximately [redacted] of patients had a history of GLP-1 RA use:

- Tirzepatide 5 mg: [redacted] patients ([redacted]%)
- Tirzepatide 10 mg: [redacted] patients ([redacted]%)
- Tirzepatide 15 mg: [redacted] patients ([redacted]%)
- Placebo: [redacted] patients ([redacted]%)

Please see Table 3 for more information.

Table 3: Summary of prior GLP-1 RA use three months prior to screening, by decreasing frequency (SURPASS-5)

Preferred Term n(%)	TZP 5 mg (N=116)	TZP 10 mg (N=119)	TZP 15 mg (N=120)	Placebo (N=120)	Total (N=475)	p-values*
Subjects with >= 1 Prior GLP-1 Therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ATC Level 4 GLP-1 RA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Liraglutide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dulaglutide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lixisenatide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Exenatide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Glucagon-like peptide-1 (GLP-1) analogues	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Semaglutide	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*p value for overall treatment effect was computed using Fisher's exact test.
WHODrug Version SEP20B3.
Abbreviations: ATC: Anatomic Therapeutic Chemical; N: number of subjects in population; n: number of subjects who received medication; TZP: tirzepatide.

c) Please provide evidence on whether previous use of GLP-1 has been explored in subgroup analysis and discuss possible implications.

Further information such as dose and duration of use are not available, and as such, previous use of GLP-1 RAs have not been explored within subgroup analyses.

A 18. Priority question. In UK clinical practise GLP-1 RAs appear to be recommended to be used at fourth line (see Table 1 above) by switching one of the drugs as part of triple therapy (NG28).

a) Please provide a breakdown of all the SURPASS trials in terms of line of therapy and treatment combination history including any class of antidiabetic drugs.

The SURPASS trials did not collect detailed data on treatment combination history, only on the treatments which were allowed within 3 months prior to the study, which were the background therapies in each trial. Beyond this the only data available are that GLP-1 RAs must have been stopped at least 3 months before study entry. Given this, it is not possible to answer this question as the data were not recorded. Furthermore, it should be noted that were such data relevant to the efficacy of tirzepatide it would remain the case that equivalent data are unlikely to be available for comparator therapies, thus precluding any analyses of comparative effectiveness by prior treatment history.

b) Have treatment combination history and line of therapy been considered as a potential treatment effect modifier?

As noted above, no data on treatment history are available within the SURPASS programme to directly analyse this question. In clinical practice individualised up-titration is undertaken for GLP-1 RAs, and will be for tirzepatide, to achieve the desired clinical outcome in terms of target HbA1c and BMI whilst remaining within the bounds of individual patient tolerability.

With respect to any treatment-by-subgroup interaction of current background therapy during the SURPASS-4 trial, please see the answer to question A5d above.

With respect to line of therapy, a proxy for this would be to examine the absolute outcomes for tirzepatide doses across the SURPASS trial programme, presented in the CS pages 27–28 and reproduced below. As seen in these results, each trial has demonstrated efficacy despite differences in line of therapy and background therapy.

Table 4: Summary of change in HbA1c from baseline across SURPASS trials

HbA1c change from baseline, % (mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	−2.1% (−22.8)	2.4% (−25.9)	−2.5% (−26.9)	−1.9% (−20.3)
SURPASS-3 (vs insulin degludec)	−1.9% (−21.1)	−2.5% (−24.0)	−2.4% (−26.0)	−1.3% (−14.6)
SURPASS-4 (vs insulin glargine)	−2.2% (−24.5)	−2.4% (26.6)	−2.6% (−28.2)	−1.4% (−15.7)
SURPASS-5 (vs placebo)	−2.2% [redacted]	−2.6% [redacted]	−2.6% [redacted]	−0.9% [redacted]

Source: Frías *et al*, 2021;[26] Ludvik *et al,* 2021;[27] Del Prato *et al,* 2021;[28] SURPASS-5 CSR.[29]

Table 5: Summary of proportion of patients achieving HbA1c <7.0% (<53 mmol/mol)across SURPASS trials

Patients achieving HbA1c <7.0% (<53 mmol/mol)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	85.5%	88.9%	92.2%	81.1%
SURPASS-3 (vs insulin degludec)	82.4%	89.7%	92.6%	61.3%
SURPASS-4 (vs insulin glargine)	81.0%	88.2%	90.7%	50.7%
SURPASS-5 (vs placebo)	93.0%	97.4%	94.0%	33.9%

Source: Frias *et al*, 2021;[26] Ludvik *et al,* 2021;[27] Del Prato *et al,* 2021;[28] Dahl *et al*, 2021.[30]

Table 6: Summary of change in body weight from baseline across SURPASS trials

Body weight change from baseline (kg)	Tirzepatide 5 mg	Tirzepatide 10 mg	Tirzepatide 15 mg	Comparator
SURPASS-2 (vs semaglutide 1 mg)	[redacted] (−7.8)	[redacted] (−10.3)	[redacted] (−12.4)	[redacted] (−6.2)
SURPASS-3 (vs insulin degludec)	[redacted] (−7.5)	[redacted] (−10.7)	[redacted] (−12.9)	[redacted] (2.3)
SURPASS-4 (vs insulin glargine)	[redacted] (−7.1)	[redacted] (−9.5)	[redacted] (−11.7)	[redacted] (1.9)
SURPASS-5 (vs placebo)	[redacted] (−6.2)	[redacted] (−8.2)	[redacted] (−10.9)	[redacted] (1.7)

Source: SURPASS-2 CSR;[31] Frias *et al*, 2021;[26] SURPASS-3 CSR;[32] Ludvik *et al,* 2021;[27] SURPASS-4 CSR;[33] Del Prato *et al,* 2021;[28] SURPASS-5 CSR;[29] Dahl *et al*, 2021.[30]

c) Please provide a subgroup analysis by treatment combination history and line of treatment for all the SURPASS trials included in the CS.

It is not possible to identify all previous treatments that a patient may have been on since their diagnosis of T2D in the trials as this is not routinely captured. Diabetes treatments are typically additive, as per NG28, due to differing and complementary mechanisms of action. Patients may discontinue treatments for a variety of reasons and clinician preference and experience has also historically been a factor in choice. So far, GLP-1 RA therapy has not been shown to have disease-modifying effects, and therefore the impact of the history of GLP-1 RAs before the washout period seems to be irrelevant.

SURPASS-5 is a useful indicator of potential prior treatment modifiers. Patients in this study have had a mean duration of type 2 diabetes of more than 13 years and are therefore likely to have received multiple treatments over the years. Significant HbA1c lowering efficacy in this study is therefore a reassuring indicator that past treatments have limited effect on efficacy of tirzepatide.

A 19. Priority question. The company states that a series of subgroup analyses was conducted in all the SURPASS trials for some patient characteristics (Table 48 of the CS). According to the CS "*Overall, analyses of change from baseline in both*

***HbA1c and body weight were generally consistent with the primary results in all of the SURPASS 2–5 trials, with the treatment difference favouring all three doses of tirzepatide compared with the comparator in the majority of subgroups.*"**
Nevertheless, according to the subgroup analysis results there were a few statistically significant differences between groups across the available CSRs.

In SURPASS-2:

"*…the treatment-by-subgroup interactions for*

- *race, baseline BMI group (<27 kg/m2 vs ≥27 kg/m2) were statistically significant using the treatment-regimen estimand.*
- *baseline BMI group (<30 kg/m2 vs ≥30 to <35 kg/m2 vs ≥35 kg/m2) was statistically significant using the efficacy estimand*."

In SURPASS-3:

"*…the treatment-by-subgroup interactions for*

- *age Group 1 (<65 years vs. ≥65 years), baseline BMI Group 1 (<27 kg/m2 vs. ≥27 kg/m2), baseline BMI Group 2 (<30 kg/m2 vs. ≥30 to <35 kg/m2, and ≥35 kg/m2), and ethnicity were statistically significant using the treatment-regimen estimand.*
- *baseline BMI Group 1 (<27 kg/m2 vs. ≥27 kg/m2), baseline BMI Group 2 (<30 kg/m2 vs. ≥30 to <35 kg/m2, and ≥35 kg/m2), and ethnicity were statistically significant using the efficacy estimand*."

In SURPASS-4:

"…the treatment-by-subgroup interactions for

- *Age Group 1 (<65 years vs. ≥65 years), Age Group 2 (<75 years vs. ≥75 years), BMI Group 1 (≤27 kg/m2 vs. >27 kg/m2), BMI Group 2 (<30 kg/m2 vs. ≥30 to <35 kg/m2, and I8F-MC-GPGM ≥35 kg/m2), baseline OAM use, and sex were statistically significant for the treatment regimen estimand, and*
- *BMI Group 1 (≤27 kg/m2 vs. >27 kg/m2), baseline BMI Group 2 (<30 kg/m2 vs. ≥30 to <35 kg/m2, and ≥35 kg/m2), baseline OAM use, and sex were statistically significant for the efficacy estimand."*

Please provide a discussion regarding the results of the subgroup analyses.

Lilly have seen consistent and robust improvement in glycaemia and bodyweight across the clinical trials. As noted, a number of subgroup analyses have been performed and for some of these significant treatment-by-subgroup interactions have been reported. Given the large number of analyses on smaller subgroups, differences are expected to be identified, particularly with small population sizes. It should be noted that in most analyses, despite the significant interaction, all three doses of TZP were favoured vs the comparator. For example, Figure 7 demonstrates this in the BMI subgroup analysis for SURPASS-4. Significant interactions were most probably driven either by a small sample size or by some greater effects identified in some subgroups.

Figure 7: Forest plot for HbA1c (%) change from baseline to Week 52 by BMI Group 2 subgroups

Abbreviations: BMI: body mass index; HbA1c: haemoglobin A1c; IG: Insulin Glargine; LSMean: least square mean; TZP: tirzepatide.

There are no consistent significant subgroup analyses driven by a lack of efficacy across the trial program and therefore no specific subgroups identified that would impact the decision problem.

A 20. Priority question. The CS states: "This submission will present the efficacy estimand to align with the results used within the cost-effectiveness model (CEM) and NMA." This estimand is then defined in relation to rescue therapy for severe persistent hyperglycaemia and differentiated from the 'treatment-regimen estimand'. However, no justification is made for the estimand in either the NMA or the cost-effectiveness section.

a) Please provide the definition of severe persistent hyperglycaemia.

Table 7 summarises the prespecified criteria for severe, persistent hyperglycaemia by study that investigators were to use when considering the addition of rescue medication in patients who did not reach glycaemic targets. Investigators were to first confirm the patient was fully compliant with the assigned therapeutic regimen and there was not an acute condition causing severe, persistent hyperglycaemia.

If new antihyperglycemic medication was initiated after completion/discontinuation of the study treatment, it was not considered as rescue therapy.

Table 7: Summary of prespecified criteria for severe, persistent hyperglycaemia

Study	>270 mg/dL (>15.0 mmol/L)	>240 mg/dL (>13.3 mmol/L)	>200 mg/dL (>11.1 mmol/L)	HbA1c ≥8.5% (≥69 mmol/mol)
GPGB[a]	Weeks 0–6	Weeks 7–26	N/A	N/A
GPGF[b]	Weeks 0–6	Weeks 7–12	N/A	N/A
SURPASS-1[c]	Weeks 0–6	Weeks 7–12	Week 13 to end of study	By and after Week 24
SURPASS-2[d]	Weeks 0–8	Weeks 9–16	Beyond Week 16	Week 24[e]
SURPASS-3[d,f]	Weeks 0–8	Weeks 9–16	Beyond Week 16	Week 24[e]
SURPASS-4[d,g]	Weeks 0–8	Weeks 9–16	Beyond Week 16	Beyond Week 24[h]
SURPASS-5[d]	Weeks 16–24	Weeks 25–32	Beyond Week 32	Week 24[e]
GPGO[i]	Weeks 0–8	Weeks 9–16	Beyond Week 16	Week 24[e]
GPGP[i]	Weeks 0–8	Weeks 9–16	Beyond Week 16	Week 24[e]

[a]Average FBG over at least 2 weeks (≥4 values/week) exceeds threshold. [b]Average FBG over at least 2 weeks (≥3 to 4 values/week) exceeds threshold. [c]Any FBG in 1 week for 2 consecutive weeks exceeds threshold (4 values/week). [d]Average daily BG from once-weekly 4-point SMBG over 2 consecutive weeks exceeds threshold. [e]With improvement in HbA1c from Week 12 to Week 24 that was <0.3%. [f]For the insulin degludec group, the above criteria were only applicable after Week 16.[g]For the insulin glargine group, the above criteria were only applicable after Week 12. [h]Two consecutive measurements separated by ≥8 weeks any time beyond Week 24.[i]BG by weekly 1-point SMBG before breakfast over at least 2 consecutive weeks exceeds threshold.
Abbreviations: BG: blood glucose; FBG: fasting blood glucose; HbA1c: haemoglobin A1c; N/A: not applicable; SMBG: self-monitored blood glucose.

b) Please explain why results based on the efficacy estimand were chosen for either the presentation of clinical effectiveness or for use in the cost-effectiveness analysis.

The CS presented the efficacy estimand throughout to align with the regulatory submission. Both estimands were provided during regulatory submission, but the efficacy estimand data were considered the primary source within the submission to the European Medicines Agency (EMA) and therefore the MHRA. The treatment-regimen estimand data were requested by the FDA, and this is therefore a U.S. targeted estimand. The treatment-regimen estimand are also presented in the CSRs for reference, but not included in the CS for brevity.

In the studies, analyses were conducted relative to two estimands of interest in evaluating primary and key secondary efficacy objectives:

- Efficacy estimand: assessed on-treatment efficacy using data up to the time of discontinuation of study drug or initiating rescue therapy for persistent hyperglycaemia. Analysis relative to the efficacy estimand was conducted using the efficacy analysis set. The efficacy analysis set is defined as data obtained from all randomly assigned patients who took at least 1 dose of study drug (in the event of a treatment error, patients were analysed according to the treatment they were randomized), excluding patients who discontinued study drug due to inadvertent enrolment and data after initiating rescue antihyperglycemic medication or prematurely stopping study drug. For this analysis continuous endpoints are analysed with the aid of a mixed model for repeated measures (MMRM) (please refer to individual CSRs and the SAP for details)
- Treatment regimen estimand assessed efficacy using all data irrespective of adherence to investigational product or introduction of rescue therapy for persistent hyperglycaemia. Analysis relative to treatment regimen estimand was conducted using the full analysis set defined as data obtained from all randomly assigned patients who took at least 1 dose of study drug (in the event of a treatment error, patients were analysed according to the treatment they were randomized), excluding patients who discontinued study drug due to inadvertent enrolment, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication. This analysis consisted of an ANCOVA, with multiple imputation of missing measures.

The NMA was designed to align with the EMA/MHRA primary data package. The notion of estimands is relatively new (the E9(R1) addendum having been adopted by ICH in November 2019) and in recent trials the main comparators present the efficacy estimand for the primary analysis (for example the semaglutide trial programme)[1]. Older trials do not describe their analyses in terms of the estimand framework, but the efficacy estimand was found to be the one corresponding in most cases to the primary analysis (for example the dulaglutide trial programme)[34] approach. As such, choosing the efficacy estimand strategy aligned with both the regulatory submission and the available data for the most relevant GLP-1 RA comparators.

c) Please discuss the pros and cons of the two estimands in relation to any difference in clinical effectiveness or cost effectiveness and applicability to clinical practice.

As described in Section B.2.4.2 of the submission, the two estimands assess treatment efficacy from different perspectives and account for intercurrent events differently.[27]

The efficacy estimand is the ‘ideal scenario’ where efficacy is measured while patients are on the drug, thus aligning to the NICE decision problem and the cost-effectiveness analysis, wherein costs and QALYs are associated with current treatment received and treatment switching results in changes in both costs and QALYs. The treatment-regimen estimand is a more ‘real life’ type of

estimand where patients can add or change drug – it does not capture the efficacy from only the one drug and does not therefore align with the decision problem to be addressed by NICE nor the cost-effectiveness analysis. Furthermore, this estimand is not always reported, especially in older trials in which the notion of estimand did not exist.

In terms of endpoints, the primary and key secondary endpoints were analysed using both estimands:

- SURPASS 2: a total of 32 patients (1.7%) required ≥1 rescue therapy as add-on treatment to study drug for severe, persistent hyperglycaemia. This was the case for 7 patients (1.5%) in tirzepatide 5 mg, 6 patients (1.3%) in tirzepatide 10 mg, 6 patients (1.3%) in tirzepatide 15 mg and 13 patients (2.8%) in semaglutide 1 mg. The main efficacy results are consistent between the 2 analyses
- SURPASS 3: a total of 16 patients (1.1%) patients received ≥1 rescue therapy as an add-on treatment to study drug for severe persistent hyperglycaemia. This was the case for 4 patients (1.1%) in TZP 5 mg, 4 patients (1.1%) in TZP 10 mg, 6 patients (1.7%) in TZP 15 mg (however one patient actually started rescue medication during the safety follow up so this does not qualify for rescue medication) and 2 patients (0.6%) in insulin glargine. The main efficacy results are consistent between the 2 analyses
- SURPASS 4: a total of 10 (0.5%) of patients were prescribed ≥1 rescue therapy as add-on treatment to study drug for severe persistent hyperglycaemia. This was the case for 1 patient (0.3%) for tirzepatide 5 mg, 1 patient (0.3%) for tirzepatide 10 mg, 3 patients (0.9%) for tirzepatide 15 mg and 5 patients (0.5%) for insulin glargine. The main efficacy results are consistent between the 2 analyses
- SURPASS 5: a total of 7 patients received antihyperglycemic rescue therapy during the planned treatment period. This was the case for 1 patient in TZP 5 mg, 0 patients in TZP 10 mg, 1 patient in TZP 15 mg, 5 patients in placebo. The main efficacy results are consistent between the 2 analyses

Among the 72 studies included in the main analysis of the NMA, only 8 studies reported results with both estimands. It should also be noted that among the 11 endpoints included in the economic model (and analysed in the NMA) only 2 would be impacted by this notion of estimand.

Overall, across the tirzepatide trial programme, considering the low number of patients reporting rescue therapy, the number of endpoints reported with both estimands and the consistent results between estimands, Lilly consider the impact of estimand choice to be low.

Network meta-analysis (NMA)

A 21. Priority question. Figure 3, section B.2.9.6 illustrates reasons for excluding studies identified in the SLR from the NMA as part of the feasibility assessment. Please provide details of which inclusion criteria were applied, when selecting studies for inclusion in the NMA, in order to ensure comparability between studies.

Overall, the aim of the NMA was to provide robust results on the comparative efficacy and safety of tirzepatide 5 mg, 10 mg and 15 mg versus relevant GLP-1 RAs at the second and third line of

treatment for T2D that can be considered generalisable to the use of tirzepatide as a more efficacious option whenever GLP-1 RAs would otherwise be considered.

Following exclusion of the studies detailed in Figure 3 of Document B, studies were included according to the following criteria:

- **Network**: The network was defined to align with SURPASS-2 and -3 trials and included studies conducted in patients with one to two OADs; whilst acknowledging some differences as discussed in earlier answers, these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice. The population also included studies with patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin. Studies with an unclear proportion of patients receiving metformin, as well as trials including patients on a background therapy of three OADs were included in a sensitivity analysis.
- **Treatments**: Studies were included if relevant comparator treatments (as detailed in Section B.2.9.2 of Document B) were evaluated
- **Endpoints:** Studies that presented data on the endpoints of interest for the NMA highlighted in Section B.2.9.4 were included

A 22. Priority question. There are dose related subgroups in the cost-effectiveness analysis (see question A6). Also, according to the CS "*As GLP-1 Ras and tirzepatide exhibit a dose-response relationship in terms of efficacy and gastrointestinal side-effects, when interpreting the NMA comparisons were made within each recommended maintenance dose step, rather than between recommended maintenance dose steps. Therefore, comparisons were made as per Table 46.*".

a) Please explain why only a single network per outcome was constructed i.e., the NMA was not stratified (separate networks constructed) by eligibility for each tirzepatide dose with comparator trials chosen according to dose as set out in Table 46?

As noted in the answer to question A6 above, up-titration is individualised to each patient based on observed response and the different maintenance dose steps are not defined subgroups. As noted in the text quoted in the question, the comparison within, rather than between, maintenance dose steps was intended as an aid to meaningful interpretation of the NMA and economic analyses, given the exhibited dose–response relationship of tirzepatide and GLP-1 RAs, not a constraint on how the NMA was analysed.

b) Please conduct the NMA with separate networks for sets of studies chosen by eligibility for each tirzepatide dose.

Given the studies in the main analysis were similar in terms of study design and patient characteristics and therefore met the inclusion and exclusion criteria, as described in Document B (Section B.2.9), it was decided that all eligible studies should be included in the analysis.

Moreover, interpretation of dose-related treatment comparisons can still be made from the NMA presented. It should be noted that comparisons made within each dose-escalation step rather than between, refers only to how the NMA results should be *interpreted* based on expected use in UK clinical practice, rather than how the NMA was conducted.

A 23. Priority question. Table 49 of the CS presents a summary of the of type of background therapy received across included studies in the NMA. However, lack of presentation by study precludes assessment of comparability.

a) Please clarify why these treatments are called 'background'. Please confirm that they were used in both the intervention and comparator arms of all the trials in the NMA?

Treatments allowed within studies other than the "investigational drug" and the "comparator" in the studies are considered as background therapy. These background therapies were used in both arms of all the trials and were used for the full duration of the study, with the exception of individuals requiring rescue therapy. The network was defined to ensure that background therapies were used in similar ways within all the trial arms of the included in the NMA.

b) Please report all the treatments/comparators of all the studies included in the NMA by study, including a full description of all treatment combinations i.e., all the antidiabetic treatments used in all the arms of the studies, including any 'background' treatments.

The available by-individual-study data are provided in the NMA input data file supplied alongside this response[35] but it should be noted that in any one trial patients may be on a mix of background therapies suitable to their personalised treatment at the time of the trial.

c) In the NMA is it appears to have been assumed that the treatment effect in each trial is independent of treatments common to all arms, which might have been labelled as 'background' treatments? If that is the case, then please provide a justification for this additive independence.

This assumption does underpin the NMA analysis, as it does the individual trials for both tirzepatide and comparators, and is necessary given that all treatment combinations are personalised and individual to each patient given their exhibited response at any given time. As such, all trials, other than monotherapy trials, will necessarily comprise patients on a mix of background therapies. The answers to questions A5b and A19b above do not reveal any evidence to suggest that this assumption is unreasonable but nonetheless the NMA inclusion criteria were set to reduce the heterogeneity of background therapy within the network and sensitivity analyses were undertaken around background therapy.

d) If it is the case that there is variation in 'background' treatments in type and possibly doses and durations between those comparators listed in Table 46 and explicitly identified as nodes in the NMA, then:

i. If it is the case that there is variation in 'background' treatments in number, type and possibly doses and durations between those comparators listed in Table 46 and explicitly identified as nodes in the NMA, then please describe whether and how the potential impact of interaction effects on estimates of the treatment effect have been considered.

ii. Please provide an updated NMA for all outcomes where each node is identified by the full combination and additive independence is not assumed.

A feasibility assessment was conducted to determine any potential heterogeneity between the included trials. All NMAs were random effects (RE) models to account for the heterogeneity between studies. These RE models converged and there were no signs of skewness in the between study SD, sigma. Tests of heterogeneity, inconsistency was conducted as well as leverage for each NMA according to NICE DSU TSD 3,4.[36, 37] Overall, limited concerns with regards to inconsistency and heterogeneity were identified. No concerns regarding inconsistency were identified for continuous or binary endpoints. Although heterogeneity was identified for some outcomes (as summarised in Table 52 in Document B), only a minority of studies contributed to the heterogeneity.

Furthermore, a meta-regression was conducted to adjust for the number of OADs, presented in Appendix D.8.1.8, where the median effects for all treatments were shown to be similar to the unadjusted analysis. Hence, the covariate effect was observed to be non-significant.

In addition, given the heterogeneity in OADs received by patients at second line (as demonstrated in the accompanying NMA input data file for this response[35]) having a separate node for each treatment by background therapy would greatly reduce the number of studies that could be included and would likely restrict the connection of network and could cause diagnostic issues with convergence and autocorrelation of the models.

Overall, given that limited inconsistency was identified between the included studies, and as the meta-regression adjusting for number of OADs produced similar results, no further investigation into separating treatment node by background therapy was considered necessary.

A 24. Priority question. A set of figures has been provided in Section B.2.9.6. of various baseline characteristics. However, an assessment of feasibility/comparability is hindered by not being able to see all relevant characteristics by study

a) Please provide a table with the study characteristics presented by study and arm for all the studies included in any part of the NMA (efficacy and safety).

This information has now been provided within the accompanying reference pack for these responses.[38]

b) Please provide a feasibility assessment, which explicitly considers all potential treatment effect modifiers, including at least: concomitant therapy, HbA1c,

comorbidities e.g., CVD/CV high risk, obesity, non-alcoholic fatty liver disease, sex, age, weight, BMI, and duration of diabetes.

Following the completion of the original SLR, the most important treatment effect modifiers were identified, and these were subsequently adjusted for using meta-regression, as described in Section B.2.9.6 of Document B of the CS. This approach was replicated for the SLR update. Other treatment effect modifiers that were not explored in the meta-regression analysis were excluded for the following reasons:

- **Background therapies**: Background therapies were reviewed and assessed to ensure that studies included were comparable to the SURPASS 2 and 3 trials and to EU guidelines. Background therapies were therefore not considered to be a key treatment-effect modifier. Information about background therapies are provided in the accompanying NMA input data file for this response.[35]
- **Comorbidities**: As shown in Figure 25, Document B of the CS, comorbidities were not systematically reported in most studies. In addition, the comorbidities listed below were not considered as key treatment effect modifiers for the following additional reasons:
    - **CVD and high CV risk**: CVD and high CV risk are specifically reported in CVOTs, given the specific requirements for CVOT trial design, population, objectives, and glycaemic control. As such, CVD and high CV risk were not considered to be a key treatment-effect modifier to be explored in this NMA which did not include CVOTs in the network.
    - **HbA1c**: Most studies reported consistent mean baseline HbA1c values between 8–8.5%, as shown in Figure 32 of the CS. Nevertheless, HbA1c was considered as a potential factor for consideration within the meta-regression to adjust for heterogeneity, as described in Appendix D.1.1.8 of the CS.
    - **Obesity**: Across all studies, patients were consistently either overweight or obese, as shown in Figure 31 of the CS. As such, all studies included in the NMA included patients with a BMI between 30–35 kg/m². Nevertheless, similar to HbA1c, weight and/or BMI were considered as a potential factor for consideration in the meta-regression.
    - **Baseline diabetes duration:** Baseline diabetes duration was generally similar between studies with most reporting a baseline mean between 6 and 8 years, as shown in Figure 33 of the CS; a very small number of outlier studies had lower durations.
    - **Patients age**: As shown in Figure 29 of the CS, mean age at baseline was between 50 and 60 years in almost all studies. For the studies not excluded for other reasons, age was not considered to be a key treatment effect modifier, given feasibility and clinical judgement.
    - **Sex:** The proportion female in each study did exhibit some variation, as shown in Figure 28 of the CS, but this was not considered as a reason to exclude studies; in principal this parameter could be added in the meta-regressions if it were considered a treatment effect modifier.
    - **NASH**: 6 included studies were among T2D patients with NASH – all other baseline characteristics from these studies were considered comparable to other studies, and these also had comparable study designs. Therefore, it was considered that these

studies could be included in the analysis and that NASH was not a treatment effect modifier.

A 25. Priority question. Subgroup analysis of the SURPASS trials revealed a some statistically significant effects of race and ethnicity.

a) Please provide the baseline characteristics for race and ethnicity for the studies included in the NMA.

Baseline characteristics for race and ethnicity for the studies included in the NMA are summarised in Figure 8–Figure 12 and in Table 8.

Figure 8: Summary of the proportion of Caucasian patients in each treatment arm for each study included in the main analyses

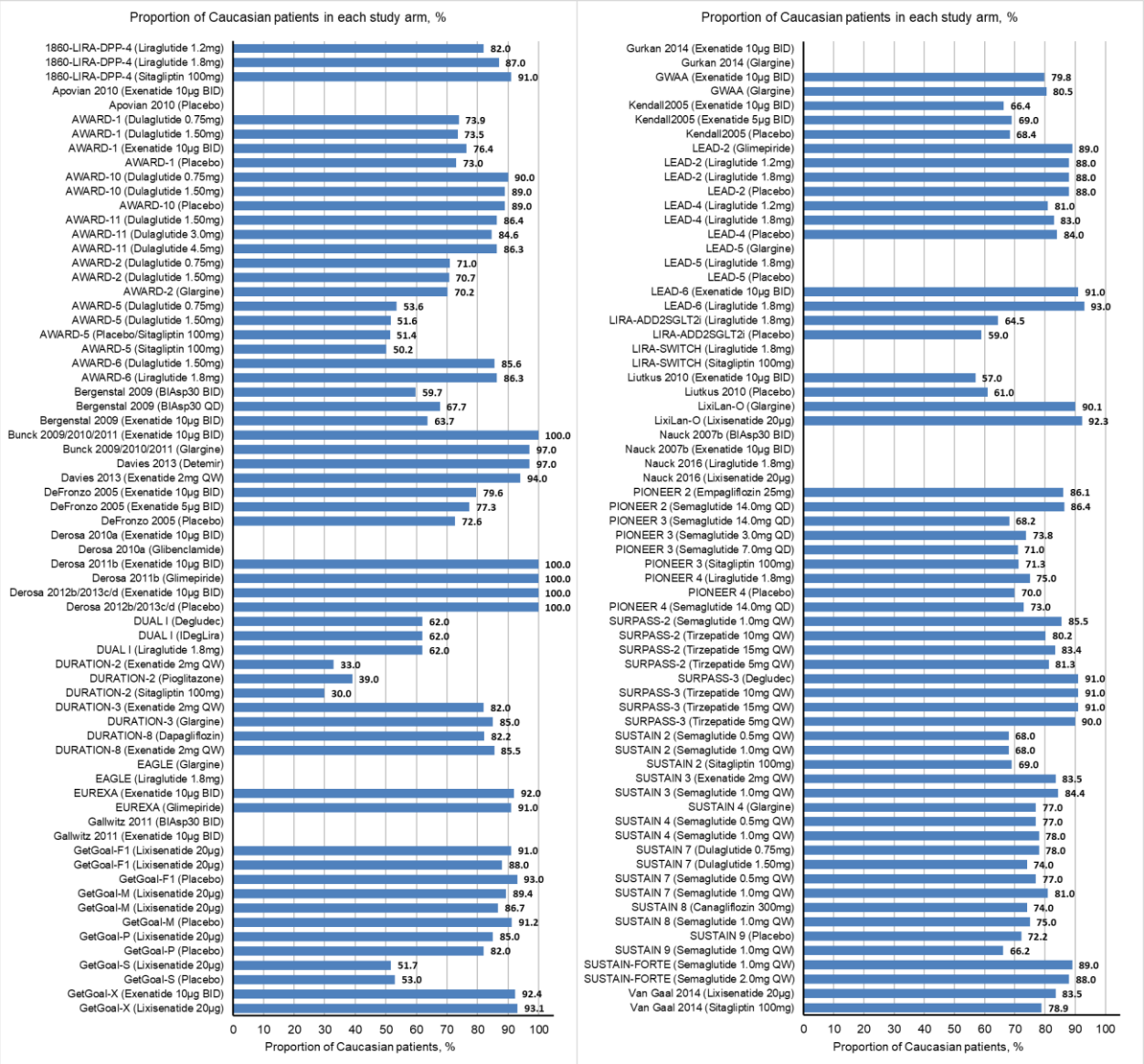


Figure 9. Summary of the proportion of Black patients in each treatment arm for each study included in the main analyses

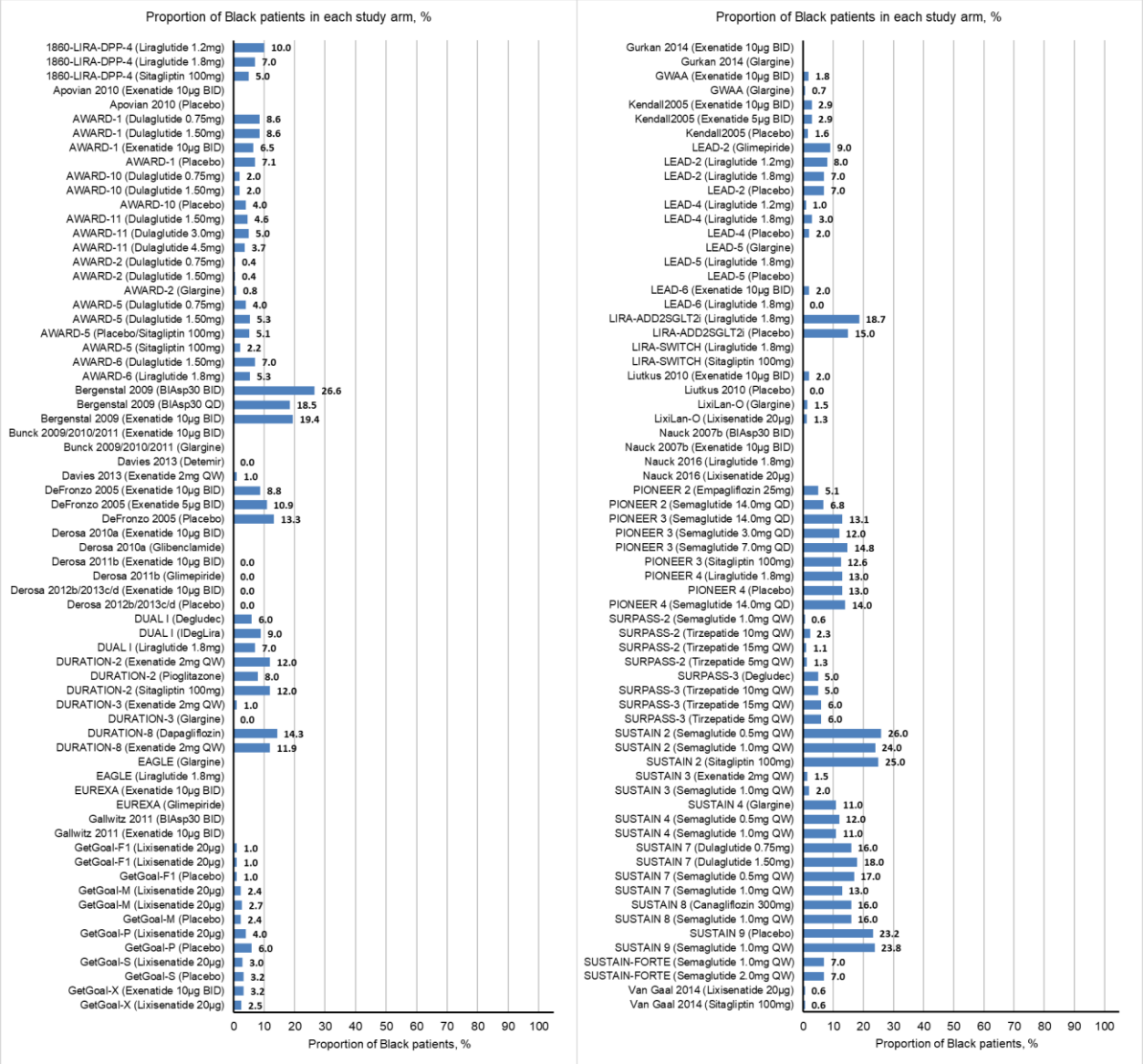


Figure 10. Summary of the proportion of Asian patients in each treatment arm for each study included in the main analyses

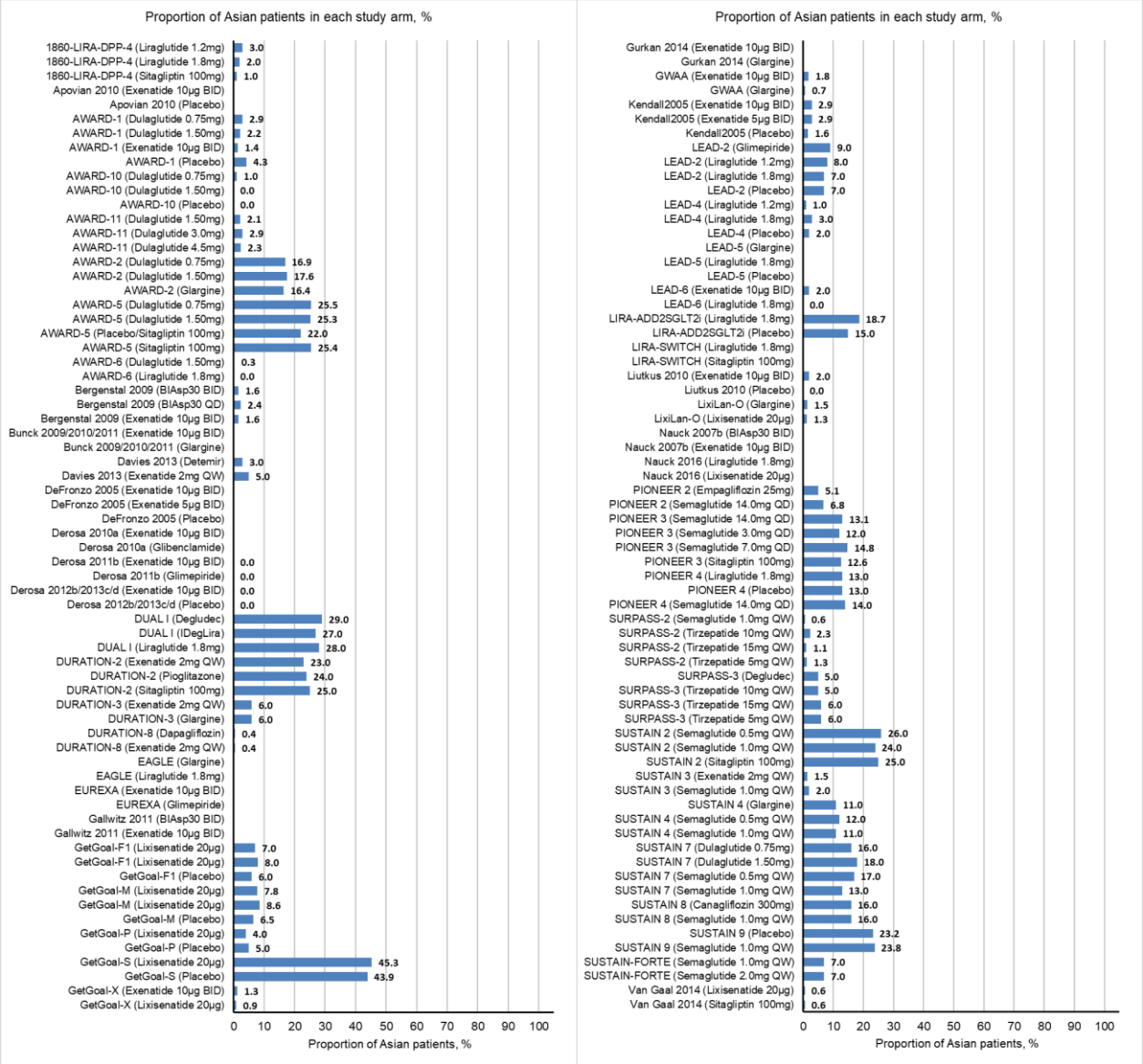


Figure 11. Summary of the proportion of Other patients in each treatment arm for each study included in the main analyses

Figure 12. Summary of the proportion of patients by ethnicity in each treatment arm for each study included in the main analyses

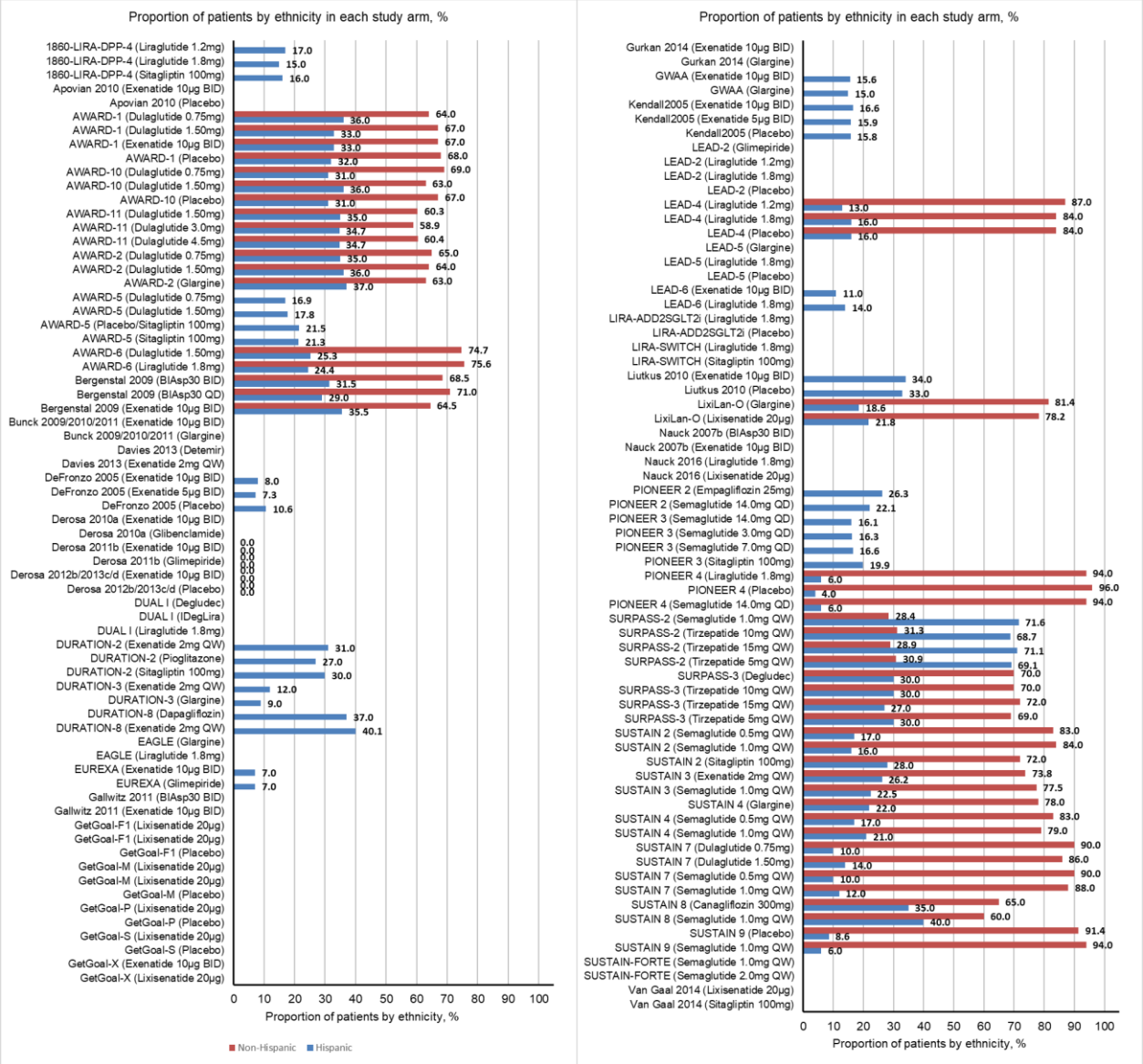


Table 8. Summary of baseline characteristics for race and ethnicity for the studies included in the NMA

Baseline characteristics	Mean value	Minimum value	Maximum value
Number of patients	264.7	17.0	834.0
Proportion of female patients, %	47.4	31.0	70.0
Mean age, years	55.9	42.7	59.8
Mean baseline weight, kg	91.9	80.2	101.9
Mean baseline BMI, kg/m^2	32.75	28.4	36.8
Mean baseline HbA1c, %	8.3	7.4	10.3
Mean baseline duration of diabetes, years	7.6	0.6	10.1
Mean treatment duration, weeks	46.5	24.0	156.0

Baseline characteristics	Mean value	Minimum value	Maximum value
Race			
Proportion of Caucasian patients, %	78.3	30.0	100.0
Proportion of Black patients, %	5.8	0.0	26.6
Proportion of Asian patients, %	9.0	0.0	45.3
Proportion of Other patients, %	4.6	0.0	21.0
Ethnicity			
Proportion of Hispanic patients, %	23.1	0.0	71.6
Proportion of non-Hispanic patients, %	66.4	0.0	96.0

b) Please include race and ethnicity in the feasibility assessment of the NMA.

Trials define and collect racial and ethnicity baseline characteristics in different ways, depending on both trial design and the requirements of the different countries and locations where the studies were undertaken. As such, a formal feasibility assessment relating to these protected characteristics has not been undertaken, but the baseline data as reported by each trial according to its own definitions has been reported above; it is apparent that there is some degree of heterogeneity in the trial populations which is most likely reflective of the location of the centres recruiting for each trial.

c) Please discuss why these characteristics were not identified as potential treatment effect modifying variables in the meta-regression analysis.

Race and ethnicity were not identified as potential treatment effect modifying variables in the meta-regression analysis on account of the following subgroup analysis results from the SURPASS-2 and -3 trials:

Race: Overall, analyses of change from baseline in both HbA1c and body weight based on race were consistent with the primary results in the SURPASS 2 and 3 trials (efficacy estimand). Nevertheless, Asian/non-Asian race was considered as a potential treatment effect modifier as baseline characteristics can differ between Asian and non-Asian participants.

Overall, approximately one-third of studies identified in the SLR were Asian-specific, defined as studies only recruiting in Asia or with ≥80% of Asian participants. In the main analyses, Asian-specific studies were excluded, and the effect of including them was instead explored in one sensitivity analysis. Results from this sensitivity analyses were reported in the CS Appendix and did not show any significant difference versus the main analysis.

Ethnicity: In SURPASS 2 and 3, subgroup analyses were only performed for Hispanic/Latino vs non-Hispanic/Latino ethnicities. However, these ethnic groups were not commonly reported in the comparator studies (see Figure 12); therefore, it was not possible to explore these as potential treatment effect modifiers in the meta-regression analysis.

d) Please provide a sensitivity analysis including these variables in the meta-regression analysis.

As discussed above, this analysis has not been undertaken.

A 26. Priority question. Section B.2.9.5.1 of the CS specifies, for the analysis population in the NMA, "studies including patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin."

a) Considering baseline comparability across studies, please provide a justification for the choice of these proportions, given that all patients in both of the SURPASS trials (-2 and -3) included in the NMA were receiving background therapy regimens that included metformin.

These proportions were chosen to ensure comparability to SURPASS-2, in which 100% of patients were on only metformin, and SURPASS-3, in which 68% of patients were on metformin only and 32% were on metformin and SGLT2i. Additionally, metformin is commonly used in first-line therapy, this is in line with clinical guidance and also reflects recommendations published by the ADA and European Association for the Study of Diabetes, and also aligns with metformin use as per NICE NG28.[39]

In reality, the threshold for metformin use within the included studies was higher than the proportions specified for inclusion. In the base case analysis, 100% of patients treated with an add-on to one OAD in studies were on metformin and ≥70% of patients treated with add-on to one to two OADs were on metformin.

b) Please conduct a sensitivity analysis including only studies where all patients received metformin. If inclusion of all relevant comparators is not feasible then consider a threshold that is higher than 50%.

Among the 53 studies included in the main analysis of the NMA, 25 included only patients on metformin, in which 100% patients receive metformin. In 16 of the 28 studies that include patients on combination therapy, 100% of patients are on metformin. In the remaining 12 studies, the proportion of patients on metformin varies between 72% and 97%. Thus, the vast majority of patients in the studies included in the NMA were receiving metformin. Given this information, a sensitivity analysis has not been conducted, as the suggested 50% threshold of patients receiving metformin has already been exceeded by a large margin.

A 27. Priority question. The duration of the studies included in the NMA is very variable (SURPASS-2: 40 weeks and SURPASS-3: 52 weeks, rest of the studies: 24-156 weeks). In addition, the duration to reach the target dose of tirzepatide at 15 mg (week 21) is also different. To address this mismatch the company decided to use a follow-up of 26 ± 4 (22–30) weeks for comparator data, compared to tirzepatide data at Week 40. To support this choice the company ran two sensitivity analyses, first, using the change from baseline at 24 weeks for tirzepatide and 26 ± 4 weeks for the comparators and second, the change from baseline at 40 weeks for tirzepatide and 32 ± 8 weeks for the comparators (the closest timepoint to 40 weeks was selected per study where available). According

to the outcomes of the sensitivity analyses the results were similar for the three outcomes of HbA1c (%), weight (kg) and BMI (kg/m2) change from baseline.

a) **Please provide a rationale why the much longer 40-week timepoint was used for the SURPASS trials instead of a shorter one to match the comparators (26 ± 4 weeks) or instead a longer timepoint of the comparator (32 ± 8 weeks) to match the 40-week timepoint of the SURPASS trials.**

A 40-week timepoint was used for the SURPASS trials to enable patients to have at least 16 weeks exposure following titration for each dose of tirzepatide, including the 15 mg dose which had the longest dose-escalation period. This exposure duration was considered appropriate to assess the effects and the benefit-risk profile of each maintenance dose of tirzepatide on both glycaemic control and body weight. As presented in part b of this question, there is a plateau in change in HbA1c from baseline in all of tirzepatide doses. Additionally Figure 13 shows that this plateau also occurs with semaglutide.

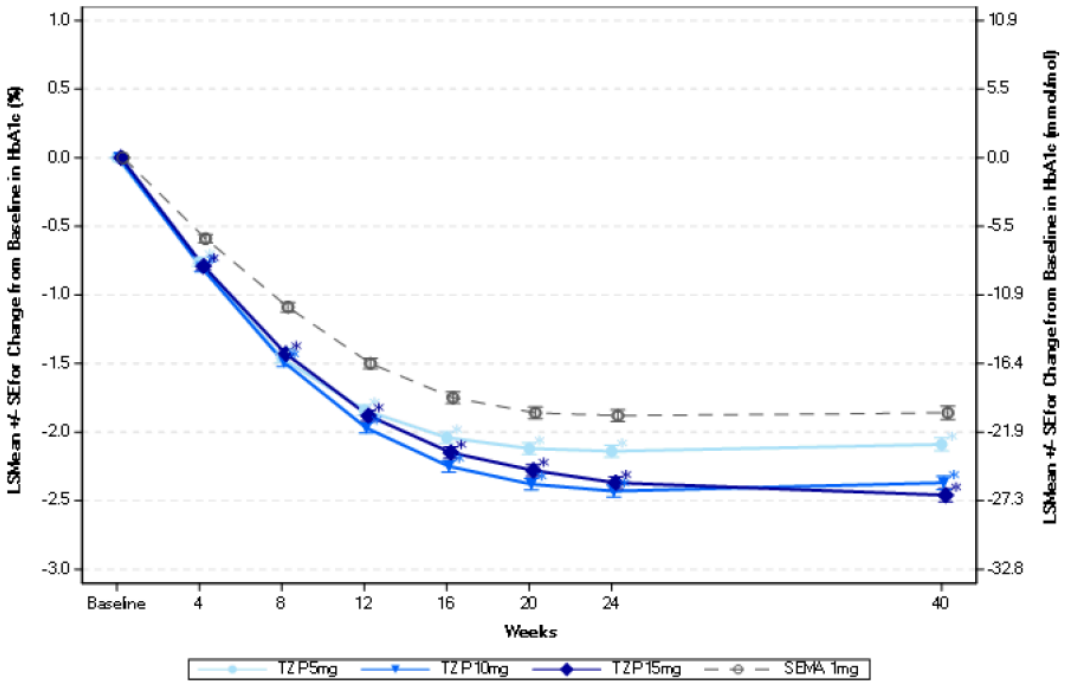
The duration of dose escalation employed to reach the tirzepatide target dose in the SURPASS trials was longer (0–20 weeks) than the corresponding durations used for the comparators in the comparator studies, which ranged from 0–12 weeks. Given this difference in dose-escalation, the duration of exposure in the comparator trials was anticipated to be similar to that of the SURPASS trials despite the shorter timepoint. It was therefore considered unlikely that including studies with different timepoints would bias against the comparators.

However, to investigate the impact of these assumptions, a number of sensitivity analyses were conducted and are described in Appendix D.8.1. These included one analysis including 24-week tirzepatide data, one allowing for a broader analysis interval for comparator trials (32 ± 8 weeks) and a model-based NMA, allowing for inclusion of multiple timepoints per outcome. For all three sensitivity analyses, results were consistent both in terms of significant difference and magnitude of the difference, indicating that the difference in timepoints did not have a significant difference on the results.

b) **The rationale is based on the dose of tirzepatide 15 mg which takes 20 weeks to reach the dose and then it is evaluated for 20 weeks further. How does this cover the inconsistency created for the doses of 5 mg, where the dose is reached on week 5 and evaluated for apr. 35 weeks; and 10 mg, where the dose in reached on 13 and is evaluated for apr. 27 weeks?**

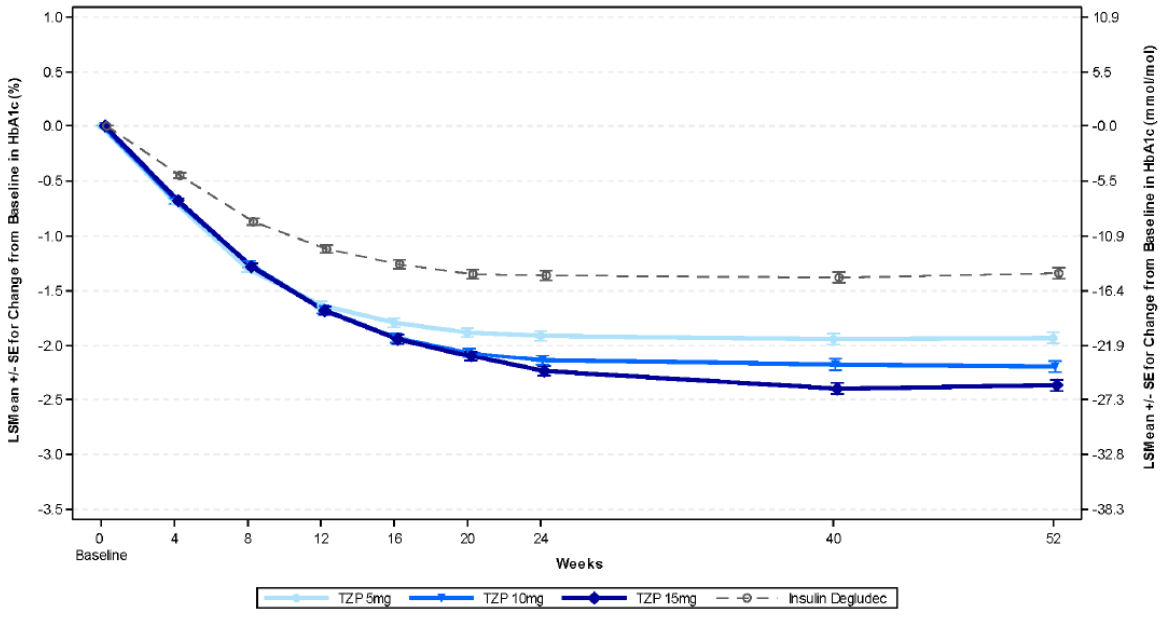
As demonstrated in Figure 13: Change in HbA1c from baseline across timepoints in SURPASS-2Figure 13 to Figure 16, there is a plateau for all three doses of tirzepatide from Week 24 onwards. There is therefore limited inconsistency expected using different timepoints for evaluation for each dose.

Figure 13: Change in HbA1c from baseline across timepoints in SURPASS-2



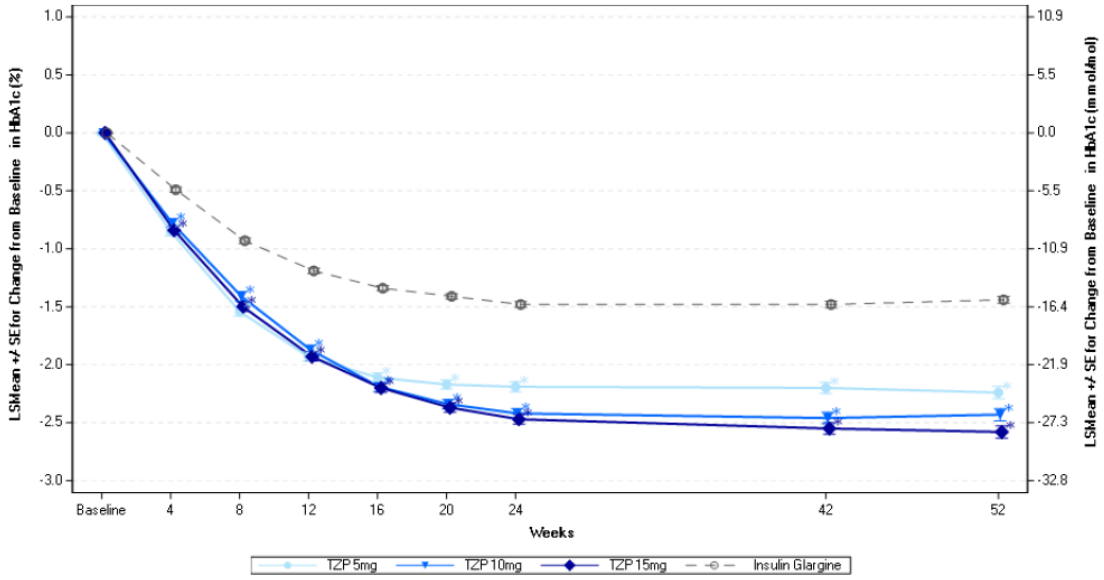
Abbreviations: HbA1c: glycated haemoglobin; SEMA: semaglutide; TZP: tirzepatide.
Source: SURPASS 2 CSR. [31]

Figure 14: Change in HbA1c from baseline across timepoints in SURPASS-3



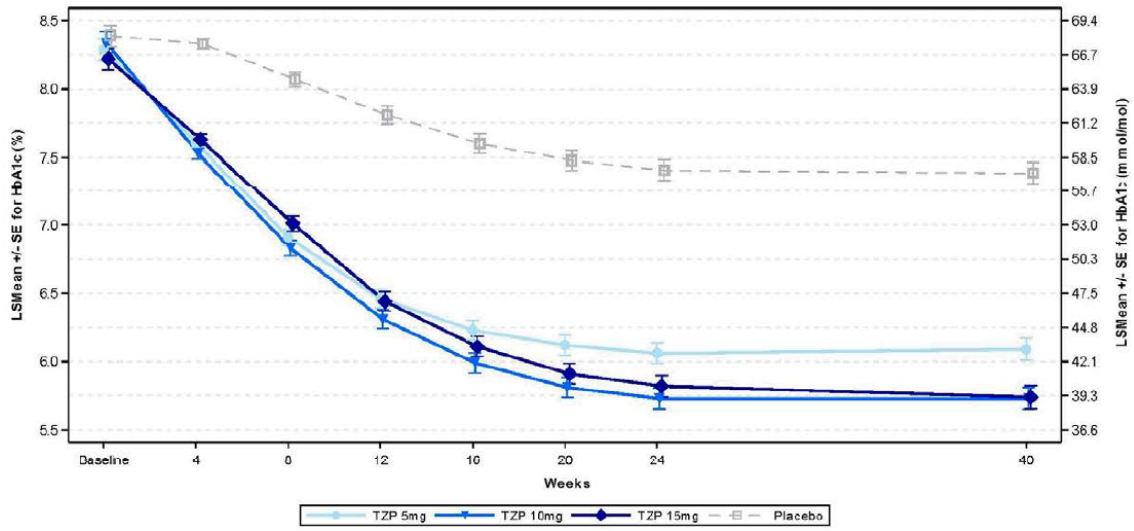
Abbreviations: HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: SURPASS-3 CSR.[32]

Figure 15: Change in HbA1c from baseline across timepoints in SURPASS-4



Abbreviations: HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: SURPASS-4 CSR.[33]

Figure 16: HbA1c levels across timepoints in SURPASS-5



Abbreviations: HbA1c: glycated haemoglobin; TZP: tirzepatide.
Source: SURPASS-5 CSR.[29]

c) Please assess the feasibility of including only studies and SURPASS data where the timepoint is more comparable i.e. identical, ± 2 weeks and ± 4 weeks

As highlighted in Figure 34 of Document B, there is variability in the treatment duration of each study arm included in the main analysis. Restricting the analysis to only include data from week 26 (the most commonly reported timepoint) would have produced results that were unfairly detrimental to tirzepatide, due to the titration schedule of tirzepatide. There is also a lack of data around the 40 week timepoint (which is closer aligned to the primary endpoint of the SURPASS trials), with 25 of the 53 studies included in the main analysis having a maximum duration of 24–26 weeks. A summary of the data availability at different timepoints is provided in the reference

pack.[35] As such limiting the NMA to include only studies where the timepoint is more comparable to that of the SURPASS trials would have substantially decreased the evidence based included.

Two sensitivity analysis were conducted to investigate the effect of changing the analysis time point on the results (Appendix D.8.1). The results of these sensitivity analysis demonstrated that neither inclusion of tirzepatide data from the 24-week timepoint nor the inclusion of comparator data at 32 ± 8 weeks timepoint had a considerable impact on the results of the NMA and as such, the results of the main analysis can be considered robust.

A 28. Priority question. Question A5 describes the inconsistency between the decision problem and the SURPASS trials in terms of treatment experience. Section B.2.9.1 states: "*The analysis population was defined to align with SURPASS-2 and 3 trials, and included studies conducted in patients with one to two OADs…*". Therefore, the population of the NMA is also inconstant with the decision problem i.e., inadequately controlled with three or more antidiabetic agents. This is despite the comparator trials being of GLP-1 RAs, which are recommended in the NICE guideline NG28 at the line of therapy specified in the decision problem Please discuss this inconsistency between the population addressed in the decision problem and the population addressed in the NMA.

As with the design of the trial programmes for tirzepatide and for comparators, the NMA was conducted on a global level to meet the needs of multiple countries, so does not exactly match the population in the decision problem that is specific to the relatively less common position to which the NHS restricts GLP-1 RAs. However, the NMA population is aligned with the SURPASS-2 and -3 trials as well as other GLP-1 RA comparator trials and considered generalisable to UK clinical practice as described above in the answer to question A5d and A5e.

Results adjusted for the number of background OADs (in the meta-regression analysis) were similar to unadjusted results for all tirzepatide doses compared to all GLP-1 RAs at the same recommended maintenance dose step for HbA1c change from baseline and weight change from baseline. In addition, as described in Appendix D.9.1.3, sensitivity analysis that included studies with patients on a background therapy of three OADs (e.g., SURPASS-4) did not significantly impact the NMA results. This supports the contention that results of the NMA are generalisable to patients in the target population.

A 29. Priority question. As stated in question A21, the CS states: "This submission will present the efficacy estimand from the SURPASS trials to align with the results used within the cost-effectiveness model (CEM) and NMA."

a) **Please confirm that the results based on the efficacy estimand were chosen for the NMA.**

Results in the NMA were based on efficacy estimands when reported; further discussion of these results are provided in response to Question A20.

b) **Please provide the estimand used for each of other trials in the NMA i.e. in relation to rescue therapy and a justification for the efficacy estimand in terms of the comparability of the trials in the NMA.**

The efficacy estimand was used for all SURPASS trials included in the NMA and was the preferred estimand for all comparator trials included, when the estimand was reported. In AWARD-10, AWARD-11, LIRA-ADD2SGLT2i, PIONEER 3, PIONEER 4, SUSTAIN-FORTE both the efficacy estimand and the treatment-regimen estimand were reported, and for these trials the efficacy estimand was used for the NMA. However, as discussed in A.20, some trials do not report the estimand used and in older trials, undertaken before the concept of estimands was adopted in trial design, often only one analysis is reported. In these situations, the only results that were reported were included within the NMA.

A 30. The variables that were identified as potential treatment effect modifiers and used in the meta-regression analyses were: assessment timepoint (weeks), number of OADs (1 vs. 2), baseline HbA1c and baseline weight. Please discuss how these variables were selected.

The variables identified as potential treatment effect modifiers were pre-selected during feasibility assessment based on clinical review of the included studies; refinement of the choice of selected variables was considering during heterogeneity checks undertaken when conducting the analysis.

During a series of in depth internal discussions at Lilly between medical and statistical experts, potential treatment effect modifiers were considered, including baseline characteristics (such as baseline HbA1c, weight and background therapies) and study design features (such as study durations, timepoints for reporting endpoints). Following these discussions, assessment timepoint (weeks), number of OADs (1 vs. 2), baseline HbA1c and baseline weight were all selected to be potentially significant sources of heterogeneity and, as such, were included in the meta-regression analysis.

A 31. Priority question. The simulation study seems to implement arm-specific treatment effects.

a) **Please clarify how these were estimated and if they were from the NMA reported in the CS.**

Assuming that the term “simulation study” refers to the economic model, treatment effects were taken from the NMA reported in B.2.9 of Document B.

b) **Please clarify if the NMA pooled relative treatment effect estimates e.g. mean difference between treatments, or whether a so-called arm-based NMA has been conducted (to obtain summary estimates of arm-specific treatment effects).**

Absolute treatment effects from baseline for the simulation study were taken from the relative effects NMA reported in B.2.9 of Document B. Absolute treatment effect estimate were created by adding the placebo effect from a separate baseline model to the relative effect estimate from the NMA model.

A 32. How was multi-collinearity assessed in the meta-regression models?

Meta-regression models were adjusted for one covariate at a time (independently) so there would be no multi-collinearity. This approach was taken as it was considered that there may be a risk of lack of data when adjusting for two or more covariates simultaneously in the meta-regression.

A 33. Priority question. Please provide full details of the network meta-regression, including model specification, how the network meta-regression models accounted for variation within studies and between studies, and tests of model fit.

The model specification for the meta-regression is the same as the corresponding unadjusted endpoint of the main analysis, as shown in Table 9 below.

Table 9: Goodness of fit statistics for all endpoints in meta-regression

Endpoint	Residual deviance	DIC
Change from baseline in HbA1c (%) - Adjusted for Number of Background OADs	106.28	198.44
Change from baseline in weight (kg) - Adjusted for Number of Background OADs	97.58	180.6

Abbreviations: OAD: Oral antidiabetic; HbA1c: haemoglobin A1C.

A 34. The NMA was conducted using "Just Another Gibbs Sampler (JAGS) version 4.2.0 software via R".

a) Please provide the details of the libraries used in R along with the data file with the point values used in the model so it can be validated. This refers to both the efficacy and the safety analysis data sets.

The R library used was RJAGS 4.10. The data file with the point values used in the model has been provided within the reference pack.[40]

b) Please provide the code used in JAGS for obtaining of the Markov chain Monte Carlo (MCMC) input.

Apologies for the error here, the CS should have stated that the NMA was conducted in JAGS version 4.3.0 rather than version 4.2.0 as was originally stated.

Example code in JAGS for obtaining of the Markov chain Monte Carlo (MCMC) input was provided in Appendix D8.5 for both continuous and binary outcomes and fixed and random effects models as well as the baseline models. The FitModel.R files which contains the JAGS code and input NMA data have been provided in the reference pack to allow replication for each endpoint.[41]

A 35. As well as in terms of baseline characteristics, there appear to be other notable differences between the studies included in the NMA.

- "*The majority of studies were single, double or triple-blind (29/53; 55%), although a large minority were open-label (22/53, 42%).*"

- The company reports that at least one study measured two combined interventions "…*Apovian (2010) all treatments were combined with starting a lifestyle modification programme which would be expected to have a positive effect on T2D alone*…".
- In the study by DeFronzo et al. 2015 on 5 or 10 µg exenatide "*Subjects were instructed to fast overnight during the study*."

a) Please discuss the implications of these variations.

In the main analysis, 22/53 of studies were open label. In diabetes it is common to design open label studies given differences between injection devices of various comparators, as well as the distinct tolerability profile associated with GLP-1 RAs. The risk of bias thus introduced is mitigated by the objective outcomes measures used for primary and key secondary outcomes, such as HbA1c and weight, although it is acknowledged that specific safety outcomes may be more subjective and thus open to bias. In SURPASS studies, even if studies were open-label, every effort was taken to minimise the potential for biases in the study design: the study team remained blinded to the treatment assignment, within tirzepatide arms, the dose was blinded to patients, investigator and sponsor.

Regarding the studies instructing the patients about lifestyle modification: the quoted discussion relates to the comparison of placebo arm outcomes. As noted in the CS, the lifestyle instructions given to patients are not likely to change the results of the NMA because they would affect both arms similarly, thus being unlikely to affect the *relative* treatment effects which are the basis for the NMA.

b) Were there other studies that included non-pharmacological treatments or instruction that might be treatment-effect modifying?

The feasibility assessment reported all but one trial as low risk in terms of blinding and as such, there is a low likelihood of blinding having an effect on the NMA results. Lifestyle modification programmes and fasting were not considered as background therapy within the NMA feasibility assessment, but given the nature of the disease it is standard for diabetes trials to include lifestyle advice for all trial participants in addition to the pharmacological background therapies, intervention and comparators.

A 36. Please provide details of the model fit statistics for each of the Model-based NMA time-course models.

The model fit statistics for the model-based NMA are provided in Table 10.

Table 10: Goodness of fit statistics for all endpoints in model-based NMA

Endpoint	Residual deviance	DIC
Change from baseline in HbA1c (%)	62.01	124.01
Change from baseline in weight (kg)	71.23	136.39

Abbreviations: DIC: Deviance Information Criterion; HbA1c: glycated haemoglobin; NMA: network meta-analysis.

Adverse events

A 37. Priority question. The misalignment between the NICE scope, the decision problem addressed by the company and the evidence coming from the SURPASS trials, that is described in detail in Q. A5 and Table 1, also relates to the safety analysis presented in section B.2.10 of the CS. Please justify how the AEs experienced by the participants of the SURPASS trials relate to the patients that would get tirzepatide in the UK clinical practice.

The safety analysis is an integrated analysis reflecting data from a broader population than would receive tirzepatide in UK clinical practice. The most common side effects experienced by participants are GI related and mostly mild to moderate in severity and in general occurred more often during the dose escalation and decreased over time. This finding was consistent across the studies.

Considering the proposed population, combination therapies are an important consideration. Frequency of hypoglycaemia in combination and steps to mitigate effects are well described in the SmPC and summarised here:

- Clinically significant hypoglycaemia is defined as blood glucose <3.0 mmol/L (<54 mg/dL) or severe hypoglycaemia (requiring the assistance of another person)
- Clinically significant hypoglycaemia occurred in 10 to 14 % (0.14 to 0.16 events/patient year) of patients when tirzepatide was added to sulphonylurea and in 14 to 19 % (0.43 to 0.64 events/patient year) of patients when tirzepatide was added to basal insulin (very common)
- When tirzepatide is added to existing therapy of a sulphonylurea and/or insulin, a reduction in the dose of sulphonylurea or insulin may be considered to reduce the risk of hypoglycaemia. Blood glucose self-monitoring is necessary to adjust the dose of sulphonylurea and insulin. A stepwise approach to insulin reduction is recommended
- Clinically significant hypoglycaemia is classified as common (≥ 1/100 to < 1/10) when tirzepatide is added to existing metformin and/or sodium-glucose co-transporter 2 inhibitor (SGLT2i) therapy, the current dose of metformin and/or SGLT2i can be continued.
- Hypoglycaemia is uncommon when used with metformin.

This language is consistent with the GLP-1 RA class of therapies. The safety profile of tirzepatide will be familiar to the healthcare community due to its similarities to this well-established class of therapies and can be readily managed by following the guidance in the SmPC.

A 38. According to the CS “SURPASS-J-Mono and SURPASS-J-Combo were conducted in a Japanese population and are therefore not considered generalisable to the UK population; they are not presented as part of the clinical evidence in this appraisal. Data from SURPASS-J-Mono and SURPASS-J-Combo are included in the safety analysis in Section B.2.9.”.

a) Please provide a rationale why the two trials were included in the safety analysis when they were excluded from the efficacy analysis.

The SURPASS-J-Mono and SURPASS-J-Combo were included in the safety analysis to provide maximum safety data for tirzepatide in T2D and ensure that important safety signals were not overlooked. However, the trials were conducted in a Japanese population so were considered less relevant from an efficacy standpoint so were not included in the efficacy analysis. Whilst regional differences may determine the most suitable studies for efficacy analyses, it is common for the safety analysis set to be broader than the efficacy analysis set and for the integrated safety analyses include relevant data from all patients regardless of region.

b) Please provide the CSRs for SURPASS-J-Mono and SURPASS-J-Combo.

Please find the CSRs for SURPASS-J-Mono and SURPASS-J-Combo in the reference pack.[42] [43]

A 39. Two analysis sets are used regarding the adverse events analysis: the phase 3 Placebo-Controlled Analysis Set and the phase 3 Dose Effect Analysis Set. The first one includes only the placebo-controlled trials (SURPASS-1 and -5) while the second includes all the SURPASS trials.

a) Please provide a rationale on why the phase 3 trials were divided in these specific analysis sets.

The seven Phase 3 studies had the same tirzepatide treatment groups with the same dose escalation schedules, these escalation schedules were different from the Phase 2 studies, it was therefore important to create integrated analysis sets that examined the Phase 3 studies separately from the Phase 2 studies. Consequently, the 2 primary analysis sets to detect drug and dose effects, respectively, are the Phase 3 Placebo-Controlled Analysis Set and the Phase 3 Dose Effect Analysis Set

b) The AEs in the placebo arms of the trials are reported. Please also report the respective AEs in all the intervention arms so that a comparison can be made across all the different types of AE (TEAEs, CV risk, retinopathy, renal safety, hypoglycaemia and SAEs).

Comparisons of tirzepatide groups and active comparators are presented by individual study as needed to discuss specific adverse events of interest in the clinical safety summary, as provided in the reference pack.[21]

A 40. The company states that “A total of 19 completed phase 1, phase 2, and phase 3 studies have contributed safety data with up to 106 weeks of exposure to treatment. A total of 7,769 patients received an intervention in the phase 2 and 3 studies. Of these patients, 5415 received tirzepatide, 312 received placebo, and 2042 received an active comparator. Over the course of these investigations, the safety profile of tirzepatide has been well-characterised and robust management strategies have been developed and

refined for AEs." Nevertheless, none of the phase 1 and phase 2 trials data are presented in the CS. Please provide the appropriate evidence.

For brevity within the submission, and due to the breadth of available phase 3 data for tirzepatide, phase 1 and 2 trial data was not included. However, the phase 1 clinical trial data are available in Section 2.7.4.5.9 of the clinical safety summary document, as provided in the reference pack.[21]

Additionally, there are further data sets available that combine phase 2 and 3 studies presented in the supplied clinical safety summary and summarised below:

- The phase 2/3 analysis set, an uncontrolled integrated analysis set, was created to facilitate identification of the rarer events that require further scrutiny through case reviews. Therefore, this analysis set includes all phase 2 and 3 studies and all tirzepatide doses. In this analysis set, all tirzepatide doses are pooled.
- The phase 2/3 placebo-controlled analysis set includes all placebo-controlled studies, and only includes safety data collected while on treatment. This analysis set provides a means to identify any additional signals that would warrant further scrutiny by including the largest placebo-controlled database possible, and by employing an alternative method for handling intercurrent events.

A 41. Regarding the safety NMA "The degludec treatment arm in SURPASS-3 and glimepiride treatment arm in LEAD-2 studies were not considered to be treatments of interest and they do not inform the network." Please provide an explanation for the exclusion of these studies.

To confirm, the SURPASS-3 and LEAD-2 studies were included in all analyses. However, given they are not comparators of interest, the degludec arm (SURPASS-3) and glimepiride arm (LEAD-2) were not part of the analysis as they do not join the network between any relevant comparators. The other arms for each study did however contribute to the analysis.

A 42. Regarding retinopathy, according to the CS "Worsening of fundoscopic examination results, as recorded on the retinopathy eCRF, was recorded for 18 (0.35%) tirzepatide-treated patients across the SURPASS trials.".

a) Please discuss how this outcome observed during the trial follow-up period might inform progression of diabetic retinopathy?

There is insufficient data from these trials to draw conclusions regarding retinopathy progression. To further evaluate this there is an ongoing cardiovascular outcome trial (SURPASS-CVOT) which includes a sub-study exploring retinopathy, with expected completion date of 2025.

In the absence of detailed outcome data, the SmPC includes the following specific caution:[22]

Diabetic retinopathy

Tirzepatide has not been studied in patients with non-proliferative diabetic retinopathy requiring acute therapy, proliferative diabetic retinopathy or diabetic macular oedema, and should be used with caution in these patients with appropriate monitoring.

b) Please discuss how any other AEs observed during the trial follow-up period might inform the progression of other macro- or micro-vascular complications?

Eye complication data do not necessarily correlate with other macro or microvascular complications and it may be noted that the SUSTAIN-6 study with semaglutide showed an increase in diabetic retinopathy but also demonstrated cardiovascular benefit with a reduction in MACE-3 outcomes (composite of cardiovascular death, nonfatal myocardial infarction, or nonfatal stroke) in patients with type 2 diabetes and high CV risk.[44]

As discussed in A15 and A46, a pre-specified meta-analysis to compare time to first occurrence of confirmed four-component major adverse cardiovascular events (MACE-4; cardiovascular death, myocardial infarction, stroke and hospitalisation for unstable angina) has been conducted between pooled tirzepatide groups and control groups showing no increased risk. SURPASS-CVOT is currently ongoing (with expected completion in 2025) and will provide data on a wider range of outcomes.

A 43. Priority question. In the safety NMA, 23 studies and 16 treatments (nodes) were included.

a) As requested in question A24, please provide the detailed characteristics of these studies in terms of interventions and comparators including any background or concomitant anti-diabetic treatments. Please also consider an updated NMA where each node is identified by the full treatment combination in these studies.

The NMA input data files supplied provide details of the interventions, comparators and background/concomitant anti-diabetic treatments included in all studies.[45] Specifically, the dataset provided for question A24 shows the breakdown of the study by treatment as well as the baseline characteristics of each study, the dataset provided for question A23 shows the treatment and background therapy breakdown very clearly for each study and the dataset provided for question A34 (continuous and binary data) provides clarity on which studies were included in the SBP and nausea NMAs.

b) As requested in question A25, please provide a feasibility assessment for an NMA based on the baseline characteristics of these studies.

The feasibility assessment includes all studies in the NMA, including those in the safety NMA. Please refer to question A25 for further clarification on the feasibility assessment.

A 44. Priority question. Systolic blood pressure (SBP) (mmHg) change from baseline and proportion of patients experiencing nausea were the only two safety outcomes listed in the NMA section (Section B.2.9).

a) **Please explain how these outcomes are meaningful in terms of T2D management and drug safety? Please also explain why no other safety outcomes were considered for the NMA.**

It should be noted that most trials are not designed or powered to detect differences in specific safety endpoints and that it is very common in appraisals not to conduct safety NMAs for this reason. In the context of this appraisal specifically, SBP and nausea safety outcomes were chosen as the focus of the NMA because these were most meaningful for the cost effectiveness analysis (being model inputs) and thus the appraisal decision. This is because SBP is a biomarker for cardiovascular disease risk and nausea is a common adverse event for GLP-1 RAs and potentially has decremental QALY effects. In addition to this, hypoglycaemia endpoints were also investigated, but the analysis of the two attempted hypoglycaemia endpoints could not be conducted due to limited data availability.

b) **Given that the NMA conclusions are expressed in terms of SBP reduction, please consider the value and feasibility of an NMA of the proportion of patients who at follow-up experience a clinically meaningful SBP change e.g., from hypertensive to not hypertensive or vice versa.**

This analysis is not feasible as these endpoints were not generally available in data published for the comparator trials.

c) **Hypoglycaemia is listed as an outcome (distinct from other AEs) in the decision problem (as defined both by NICE and by the company). Despite this the CS states that hypoglycaemia outcomes could not be included in the NMA due to *"limited data availability"* and further states that *"rates of hypoglycaemia were not reported in the NMA as several studies reported zero events; therefore the rate of hypoglycaemia was set to zero for tirzepatide and all comparators in the base case analysis".* Given that reporting of zero events is not the same as an absence of data and that hypoglycaemia rates are available for the SURPSS trials:**

i. **Please provide all hypoglycaemia data for all trials (tirzepatide and comparators)**

Please find data on hypoglycaemia for all trials (tirzepatide and comparators) included in the reference pack.[46]

ii. **Please perform an NMA of hypoglycaemia event rates**

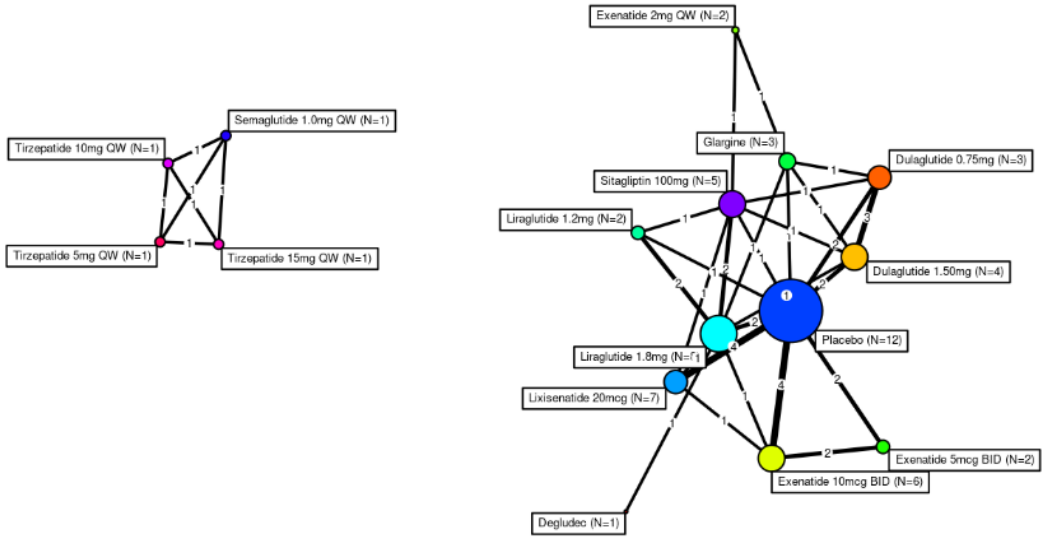
Hypoglycaemia definition of "proportion of patients with at least one hypo with BG <54 mg/dL" had a disconnected network as shown in Figure 17. Hypoglycaemia definition of "proportion of patients experiencing at least one severe hypo event" had a high number of zero count data from studies. This is common in GLP-1 RA treatments.

Therefore, an NMA was performed for the broader hypoglycaemia definition endpoint: "proportion of patients with at least one episode of hypoglycaemia with blood glucose <54 mg/dL or severe

hypoglycaemia." However, it was not feasible to analyse this endpoint despite the existence of a connected network. Studies observed variability and both the random effects and fixed effects models showed major autocorrelation and convergence issues. In order to rectify the issues, continuity correction was applied to the studies with 0 counts for this endpoint. Both the random effects and fixed effects models were run after significantly increasing the thinning, burn-in and sampling runs. The model struggled to estimate the between study standard deviation and there were issues with the sigma distribution in both the models which resulted in very large 95% credible intervals. As such, these results were not meaningful to interpret.

In addition, there was substantial variation in the definition of hypoglycaemia among the trials, making it difficult to find studies with the same or similar definitions of hypoglycaemia, further limiting the number of studies available.

Figure 17: Network diagram for the hypoglycaemia event rates NMA



Abbreviations: BID: twice daily; QW: once weekly.

A 45. The treatment periods vary between the SURPASS trials from 40 weeks to 104 weeks. Furthermore, in SURPASS-4, where the longest treatment period of 104 weeks was observed, it was a variable treatment period, meaning that not all patients were treated for the same amount of time.

a) Please clarify whether a specific time-point was chosen for assessing the AEs in the SURPASS trials or if all of them were pooled together irrespective of treatment times.

For AEs, the assessment timepoint used in studies often included a safety follow-up period (4–5 weeks). Moreover, in the SURPASS trials, AEs were assessed in the time interval between baseline and Week 44 (4-week safety follow-up). Therefore, analysis of AEs allowed for the inclusion of comparator studies with safety windows ending outside the analysis window (26 ± 4 weeks). As described in the CS, SURPASS-4 did not contribute to the safety NMA.

b) If all AEs from all studies for the full treatment periods were pooled together, please provide a justification for that.

Not applicable, please see response to part a).

c) Please consider the feasibility of and NMA with time periods that are more consistent and one where proportions are converted to rates (per unit time) to overcome the problem of follow-up time variation.

Due to there being very limited safety data reported in rates (per unit time), this NMA was not feasible. As described in the CS, SURPASS-4 did not contribute to the safety NMA.

A 46. Priority question. Despite only two safety outcomes being listed in the NMA section, the CS states that a meta-analysis was executed regarding positively adjudicated major adverse cardiovascular events (MACE) - MACE-4.

IMPORTANT: please note that the MACE meta-analyses (not NMA) of tirzepatide trial data is unrelated to the safety NMA and was undertaken specifically for the regulatory process. Further details are provided under the specific sub-questions below.

a) Please provide the definition of MACE and MACE-4 used for this meta-analysis.

The definition of MACE is major adverse cardiovascular events. MACE-4 is a composite analysis including CV death, non-fatal myocardial infarction, non-fatal stroke, or hospitalisation for unstable angina.

b) Please provide all the details of the analysis (data: trial, analysis sets and follow-up periods, and methodology: fixed-, random-effects etc.) as well as the results presented in a forest plot along with the statistical heterogeneity assessment.

In this meta-analysis (not NMA), individual patient data from seven phase 3 studies were pooled (SURPASS 1 to 5, SURPASS J Mono and SURPASS J combo). Analyses were performed on the mITT population (defined as all randomized patients receiving at least 1 treatment dose according to the treatment to which they were randomly assigned). In all summaries and analyses, mITT population patients randomly assigned to any dose of tirzepatide were included in the pooled tirzepatide treatment group, and mITT population patients randomly assigned to comparators (either placebo or an active comparator) were included in the pooled comparator group.

The primary measure for the CV meta-analysis was the time from first dose to the first occurrence of the 4-component MACE endpoints. The primary analysis model was a Cox proportional hazards regression model stratified by study-level CV risk. It was used to derive the associated CI for HR (pooled tirzepatide group versus pooled comparator group). This included treatment as a fixed effect with only 2 levels for the factor (tirzepatide or control).

The primary analysis was a comparison of the distribution of time-to-event between the pooled tirzepatide group and the pooled comparator group. Results from the primary analysis were provided as follows

- Counts and proportions of patients who experienced a primary endpoint event

- Person-years of follow up for the primary endpoint and the incidence rates
- Adjusted Kaplan-Meier estimates of the survival curve for pooled tirzepatide and pooled comparator groups, and
- Relative CV risk based on the HR and CI from a stratified Cox proportional hazards regression model

The objective of this integrated analysis was to demonstrate that tirzepatide is not associated with an unacceptably high risk for MACE-4 in patients with T2DM, following the methodology required by the regulator in the USA, the FDA. This objective as evaluated by comparing the distribution of time from first dose to the first occurrence of MACE-4, for patients receiving any dose of tirzepatide (pooled tirzepatide group) to that in patients administered comparators, placebo, or active control (pooled control group). The primary objective was considered met if the upper bound of the 2-sided 95% confidence interval (CI) for the hazard ration (HR) (pooled TZP vs pooled control) from the meta- analysis is <1.8 (limit as specified by the FDA).

Results from the final analysis

A total of 142 patients were reported with an adjudicated primary endpoint. The pooled tirzepatide groups had 72 patients and the pooled comparator groups had 70 patients who experienced at least 1 component of the MACE-4 composite endpoint (Table 11). The result of the complete analysis (additional supportive analysis) was consistent with the interim analysis (primary analysis), with a HR of [redacted] (95% CI, [redacted] to [redacted]; Table 11, Figure 18).

Table 11: Time-to-Event Analysis of Composite MACE-4 and Individual Components in the meta-analysis

*a - The person-years of follow-up is calculated for each subject as time-to-event divided by 365.25.
Time-to-event is the number of days between the date of first dose and the onset date of the event/censoring date plus 1 day.
*b – [ ] indicates the adjusted estimate to take into account different randomization ratios and differences in patient populations
among strata. Strata are defined as study-level CV risk (GPGM forms one stratum, and all other studies form one stratum).
*c - Derived from a Cox proportional-hazards model with treatment (Pooled Tirzepatide versus Pooled Comparator) as a fixed effect,
stratified by study-level CV risk (GPGM forms one stratum, and all other studies form one stratum). P-value is from Wald test.
When the total number of outcomes is < 10, survival analysis is not performed.
*d - Death due to CV cause includes death due to CV or undetermined cause.
Abbreviations: CEC: clinical endpoint committee; CI: confidence interval; CV: cardiovascular; HR: hazard ratio; MACE: major adverse cardiovascular events; N: number of subjects in population; n: number of subjects in the specified category; TZP: tirzepatide; yrs: years.

Figure 18: Time to first occurrence of MACE-4 in the meta-analysis (CEC confirmed)

Note: MACE-4 includes death due to CV or undetermined cause, myocardial infarction, stroke and hospitalisation for unstable angina. Adjusted Kaplan-Meier is estimated by weighing with inverse probability of randomisation for treatment within stratum.
HR, CI and p-value are derived from a Cox proportional-hazards model with treatment (Pooled Tirzepatide versus Pooled Comparator) as a fixed effect, stratified by study-level CV risk. P-value is from Wald test.
Abbreviations: CEC: clinical endpoint committee; CI: confidence interval; CV: cardiovascular; HR: hazard ratio; MACE: major adverse cardiovascular event; TZP: tirzepatide.

SURPASS 4 results (as this study contributed the most to the MACE-4 events)

Given that SURPASS-4 contributed most of the MACE events in the analysis, results for this study alone are sown in Table 12 and Figure 19.

Table 12: Time-to-event analysis of composite MACE-4 and individual components in SURPASS-4 (CEC confirmed)

*a - The person-years of follow-up is calculated for each subject as time-to-event divided by 365.25.
Time-to-event is the number of days between the date of first dose and the onset date of the event/censoring date plus 1 day.
*b - Derived from a Cox proportional-hazards model with treatment (Pooled Tirzepatide versus Insulin Glargine) as a fixed effect.
P-value is from Wald test. When the total number of outcomes is < 10, survival analysis is not performed.
*c - Death due to CV cause includes death due to CV or undetermined cause.
Abbreviations: CEC: clinical endpoint committee; CI: confidence interval; CV: cardiovascular; HR: hazard ratio; MACE: major adverse cardiovascular events; N: number of subjects in population; n: number of subjects in the specified category; TZP: tirzepatide; yrs: years.

Figure 19: Time to first occurrence of MACE-4 in SURPASS-4 (CEC confirmed)

Note: MACE-4 includes death due to CV or undetermined cause, myocardial infarction, stroke and hospitalisation for unstable angina.
HR, CI and p-value are derived from a Cox proportional-hazards model with treatment (Pooled Tirzepatide versus Insulin Glargine) as a fixed effect. P-value is from Wald test.
Abbreviations: CEC: clinical endpoint committee; CI: confidence interval; CV: cardiovascular; HR: hazard ratio; MACE: major adverse cardiovascular event; TZP: tirzepatide.

c) It appears that all seven trials were included in the meta-analysis irrespective of comparator. Please justify pooling studies with such clinical heterogeneity.

It is an FDA requirement to proceed this way.

d) Please conduct a feasibility assessment of an NMA of MACE.

MACE events are normally studied in dedicated long term studies, cardiovascular outcome trials (CVOT). CVOT trials are particular in terms of design, duration, concomitant antihyperglycaemic treatments allowed during the course of the study and populations. The CVOT for tirzepatide is currently ongoing and is anticipated to conclude in 2025.

In the meta-analysis available, seven phase 3 trials were included, with the SURPASS-4 trial contributing the most in terms of MACE-4 events. The SURPASS-4 study is unique in its design; it is a comparator study in patients at high CV risk with the primary objective of change in HbA1c from baseline at 52 weeks. In contrast, CVOTs published to date permit the addition of treatments to control glycaemia with all patients treated to a target, where glycaemic equipoise is ideal. As such, SURPASS-4 is unique and there is no other trial that we can use to conduct an NMA on MACEs. On the other hand, SURPASS-4 is NOT a CVOT and so cannot be compared to any CVOT available from comparators.

Therefore, at this stage this is not possible to conduct such an NMA.

Section B Clarification on cost-effectiveness data

Model structure

B 1. The CS states that a de novo model was developed because "Models developed prior to 2016, including UKPDS OM1 and OM2 and the IQVIA CORE Diabetes Model, have been shown to under predict CV benefits from the GLP-1 RA class in certain situations. One reason for this could be that models developed earlier than 2016 do not fully capture the benefits of reduced body weight as they tend to be based on cohorts using traditional therapies without any weight loss benefit." This statement is supported by CS reference 140 (Shao et al., Diabetes Care 2020).

a) The EAG did not find compelling evidence from the reference provided (CS reference 140) to support this statement. Please clarify how the findings from this study support these statements (i.e., regarding underpredicting cardiovascular benefits and not fully capturing the benefits of reduced body weight), which would justify the development of a de novo model.

The Shao *et al.* (2020)[47] reference provides evidence supporting the latter part of the statement with an extensive discussion around the role of existing risk factors (versus an inherent cardioprotective effect) and the importance of the derivation dataset in terms of predicting cardiovascular outcomes for CVOTs (and ergo modern type 2 diabetes populations). Moreover, the Shao *et al.* (2020) study was the first time that a type 2 diabetes model was shown to project accurate outcomes for a CVOT without calibration. The second time was the publication of the PRIME T2D Model description and validation in 2022 (Pollock *et al*. 2022)[48]. The most concise source summarizing the shortcomings of other available type 2 diabetes models in terms of predicting the outcomes of CVOTs comes from the Ninth Mount Hood Diabetes Challenge (Si *et al*. 2020), where the authors concluded that: "*commonly used risk equations were generally unable to capture recent CVOT treatment effects but that calibration of the risk equations can improve predictive accuracy.*[49] *Although calibration serves as a practical approach to improve predictive accuracy for CVOT outcomes, it does not extrapolate generally to other settings, time horizons, and comparators. New methods and/or new risk equations for capturing these CV benefits are needed.*" It should be noted that calibration, in this case, generally involves applying risk ratios to the prediction of outcomes in an iterative manner until the modelled outcomes match those from the trial and is, generally speaking, trial specific and not generalisable. We believe this is strong evidence on the shortcomings of earlier models of type 2 diabetes. In contrast, validation analysis with the PRIME T2D Model indicates (see Appendix N3 and Pollock *et al.* 2022) that the model averaging approach used provides an approach that is well suited to modelling outcomes for a modern type 2 diabetes population without the need for calibration.[48] As such, we believe it offers a better approach to evaluating long-term outcomes.

b) Please provide evidence that the developed de novo model, specifically the current implementation as in the CS, has a better performance to predict complications (including cardiovascular events) compared with existing diabetes models.

Without knowing the outcomes for the target population in general practice in England over the next 50 years, it is impossible to provide definitive evidence that the PRIME T2D Model will predict outcomes more accurately than other available T2D model. Similarly, a multi-study validation analysis with all published T2D models is impracticable within the context of the present submission. We are therefore left to consider the available published validation data on the available T2D models. We would suggest that comparison of the Ninth Mount Hood Diabetes Challenge results with the published validation results for the PRIME T2D Model indicate that the PRIME T2D Model may be better placed to predict cardiovascular outcomes in line with those observed from recent CVOTs (Si *et al.* 2020 and Pollock *et al.* 2022).[48] The PRIME T2D Model has also been shown to validate well against the UK-based Lipids in Diabetes Study (Pollock et al. 2022).[48] Shortcomings associated with the UKPDS OM2's ability to predict cardiovascular risk in a modern UK population over 10 years of follow up (ASCEND study) recently highlighted by Keng *et al.* (2022),[50] where the authors outlined, in particular, the lack of a revascularization endpoint and poor performance in older patients as key challenges with the UKPDS OM2. The authors also cite earlier diagnosis and improved risk factor control in modern diabetes care as potential reasons for poorly predicted outcomes (UKPDS data were collected between 1977 and 2007). The approach used with the PRIME T2D Model allows for inclusion of a revascularization endpoint and integration of more recent data for risk evaluation (via model averaging), which are designed to address these shortcomings.

B 2. The model type specified by the company is *"discrete time event"*. It is unclear to the EAG whether a discrete event simulation (DES) is meant here or another model type such as an individual patient state transition model.

a) Please specify the model type of the economic model as described in the CS. Note that DES models can also include annual updates of certain model inputs (such as patient characteristics), see for instance Corro Ramos et al., 2020 https://doi.org/10.1177/0272989x20932145.

Model nomenclature can be challenging given the conventional classifications (many of which originate with much simpler modeling approaches than are suitable for a patient-level simulation of T2D). The original UKPDS OM1 was described as a "probabilistic discrete-time illness–death model with annual cycles," which is analogous to the description we have provided of the PRIME T2D Model as a "discrete time event simulation." We deliberately avoided the term "discrete event simulation (DES)" as it is synonymous with a series of 'events' that occur over time (as opposed to events occurring within an annual cycle) and, perhaps more crucially, assumes no change in the system between events.[51] This is not the case with the PRIME T2D Model as the model runs on an annual cycle length and patient characteristics, treatments and methods of risk evaluation can change between events. The PRIME T2D Model does fit the conventional definition of a microsimulation model, however this term doesn't provide much information beyond implying a complex, patient-level simulation. One could combine the terminology to describe the PRIME T2D Model as a "probabilistic discrete-time event microsimulation model" in an attempt to convey as much information as possible with respect to classification.

b) Please justify the use of alternative model types than a patient-level state-transition model (microsimulation) or traditional DES.

The PRIME T2D Model simulates individual patients that are subject to treatment effects, risk factor progressions, and risks of adverse events and complications on an annual basis. Further to question B 2. a) when the PRIME T2D Model determines — by uniform distribution sampling — that a simulated patient has experienced an event, the individual patient could be described as having undergone a state transition; however, this does not lead us to describe the PRIME T2D Model as a state-transition model (STM) because it is not aligned with the conventional definition of an STM in, e.g. Siebert *et al.* 2012 as patients may be members of multiple "states" simultaneously (e.g. CKD stage 3, heart failure, and history of MI) without the model explicitly incorporating the notion of any composite states.[52] This approach was chosen as it best fits the available published data on diabetes risk prediction and the application of annual costs and utility values from published sources. Furthermore, this structure is more amenable (than many of the alternative model structures) to creating a treatment-agnostic model that can easily be adapted to different healthcare settings for cost-effectiveness analysis (as opposed to being derived from, and specific to, a single population).

c) Please clarify that the challenges, with addressing uncertainty in DES models (also applicable to other patient-level simulation models), specified in the paper by Corro Ramos et al., 2020 https://doi.org/10.1177/0272989x20932145 are appropriately addressed in the CS model:

i. Challenge 1: Remove differences in patient heterogeneity between the intervention and control arms. This includes that the population (as well as the parameters sampled in the deterministic analyses) has to be the same for both arms to get results that differ only because of a treatment effect but not due to a different selection of patients.

ii. Challenge 2: Adjusting remaining life expectancy after the occurrence of an event. This includes that the (re)calculated time to event curves need to be corrected for 1) the time that already had passed since the start of the simulation and 2) for worsening or improvement of the condition.

iii. Challenge 3: Remove stochastic uncertainty from treatment effectiveness. To ensure consistency, the set of random seeds should be fixed per patient. These seeds guarantee that the treatment effect is not removed, increased, or reversed due to randomness.

iv. Challenge 4: Remove heterogeneity and stochastic uncertainty from probabilistic sensitivity analysis (PSA). This includes that input parameters (in the PSA) should be the same across treatment arms. That way, the difference between the 2 arms in the PSA only results from the application of a treatment effect.

In terms of the challenges associated with addressing uncertainty in DES models, we would firstly reiterate that the PRIME T2D Model is not a DES model; however, we agree that some of the challenges of addressing uncertainty are common to all patient-level/microsimulation models and have address these challenges as outlined in the following:

i) The same cohort was used in both treatment arms (of sampled summary baseline characteristics) in all modelling simulation performed. This is achieved by generating the entire simulated cohort at baseline using random sampling from two concurrent but identical pseudo-random number generators (PRNGs) initialised with the same seed.

ii) The challenge around adjusted life expectancy after events is addressed using published risk equations from the UKPDS OM2 to evaluate the risk of mortality after complications in the modelling analysis (see the model technical report for more details, Appendix N3). It should be noted that mortality from other causes is modelled separately as a competing risk using cause-subtracted life tables.

iii) and iv) Stochastic uncertainty is adequately addressed in the model using the approaches outlined in the model technical report (Appendix N3) and in the response to question B 31. We would also like to add here that the concerns raised in the Corro Ramos *et al.* manuscript are substantively mitigated by running the PRIME T2D Model with 300x more patients per arm (300,000 versus 1,000) than the Corro Ramos et al. Chronic Obstructive Pulmonary Disease (COPD) model

B 3. In Appendix N it is described that *"The complications, adverse event and mortality controllers can then trigger various events based on current patient state (Section N.5.3 and N.5.8). To eliminate the risk of a systematic bias towards triggering specific complications, adverse events or death in a given model cycle, a clinical events controller first randomizes the order in which patients pass between the controllers falling within its remit. Based on this random order, the patients are passed between complication, adverse event and background mortality controllers, which are able to inspect the current patient state and trigger events based on their probability for that patient in the current model cycle."* Please explain in more detail, with supporting references where needed, why a random order is required to prevent systematic bias.

Randomising the order in which patients are exposed to each of the complications prevents patients being exposed to the risk of complications in a fixed sequence. As the PRIME T2D Model records patient events "live" within each cycle, and given the interdependencies between

complication risk estimates, this approach prevents the systematic elevation of risk of subsequent complications. For instance, if MI risk were always evaluated ahead of stroke risk, stroke events occurring in the current cycle could never influence risk of MI in the same cycle.

This approach is extremely common in diabetes models, as follows:

Clarke et al. 2004 described this approach in the UKPDS OM1 (emphasis added):[53]

"It is important to note that the order in which the event equations […] are evaluated to determine the occurrence of an event is not predetermined. Further, some of these events are competing risks (e.g. if a patient dies within a cycle of the model, they can have no additional events). To take this into account, the equations are run in random order in each cycle."

Palmer et al. 2004 then described the same approach in the CORE Diabetes Model (now the IQVIA Core Diabetes Model):[54]

"For each cycle the order in which the sub-models run changes randomly."

Hayes et al. 2013 described the same approach in the UKPDS OM2 (emphasis added):[55]

Equations for complications are executed in random order and if an event is predicted to occur in a given cycle it will inform the remaining set of equations still to be estimated in the same cycle.

Finally, event prediction also occurred in a random order in the BRAVO model as reported by Shao et al. 2018:[56]

"All of the events were predicted at a random order to account for event inter-dependency."

B 4. In Appendix N it is described that *"a weighted model averaging approach was used in which each equation was assigned a weight based on the similarity of mean cohort characteristics at baseline between the model cohort and the cohort used to derive the equation (derivation cohort). The greater the similarity between model cohort and derivation cohort, the larger the weight applied to the risk equation from the respective derivation cohort. The model averaging approach was then optimized by running validation simulations to evaluate predictive performance, measured using the Chi-squared statistic, and using a genetic algorithm to minimize Chi squared by adjusting distance coefficients for each characteristic."*

a) Please provide a detailed description of the model averaging implementation for this specific case (including how the weights where exactly calculated and used).

The model approach to model averaging is documented in Pollock et al. 2022 and in the supplied model technical report. The entire implementation of the model averaging algorithm is contained

within the `AveragingController` controller class in the provided source code. The per-patient weights are derived as follows:

$$w_m = \left(1 - \frac{\sum_{j=1}^{N} \frac{\beta_j \frac{|x_{dj} - x_{mj}|}{x_{dj}}}{\sum \beta_j}}{N} \right)^{\lambda}$$

$$\hat{w}_m = \frac{w_m}{\sum w_m}$$

As described in Pollock et al. 2022,[48] in this model, over N cohort characteristics, the sum product of the weight β for characteristic j and the absolute deviation of the simulated characteristic value (xmj) from the derivation cohort value (xdj) is first calculated for each model, normalized to a percentage deviation by dividing by the derivation value. This sum product is adjusted for the number of cohort characteristics to ensure equivalent weighting for models with equal relative deviations from differing numbers of reported characteristics. Finally, the deviation weight is converted to a proximity weight by taking the complement and raising to the power of a distance penalty factor λ, which adjusts how harshly deviations are penalized in the final proximity calculation. The default model weights are included in the `DatabaseController` and the fixed weight penalty is included in the `AveragingController`.

b) Please justify why model averaging is preferred instead of selecting a single predictive model that best matches the decision problem (with alternative models in scenario analyses).

The PRIME T2D Model includes the BRAVO, UKPDS OM2, and Yang et al. macrovascular risk models, which can be parameterised with cohort, risk factor, and treatment effect data from the cohort and trial results of interest (e.g. SURPASS-2). The product and trial-agnostic nature of the PRIME T2D Model necessitates this approach, and model averaging allows the model to derive weights on a per-patient basis to tailor the overall modeling approach to the cohort; in the absence of risk equations derived directly from the trial or trials in question, we consider this approach to be preferable to the selection of a single risk model parameterised from a different population than that under investigation (an approach commonly employed elsewhere in diabetes modeling efforts). In addition to addressing concerns around the structural uncertainty inherent in using a single specific risk model, the approach allows the model to draw on data derived from populations with diverse risk profiles that are, in aggregate, likely to be more representative of any given diabetes population, even when factoring in trial inclusion criteria.

c) Please provide scenario analyses selecting a single predictive model based on the best match of the derivation cohort to the decision problem.

Establishing which model is the “best match” to the decision problem is challenging, hence the default approach of allowing PRIME to weight the use of risk models automatically. The most prominent diabetes risk models (e.g. UKPDS OM1, UKPDS OM2, the IQVIA Core Diabetes Model, and the Cardiff Model) are all based — at least in part — on the UKPDS population, which was a population with newly-diagnosed type 2 diabetes, with the first patients enrolled in 1977, prior to the existence of statins, insulin analogs, SGLT-2 inhibitors, or GLP-1 receptor agonists. The incorporation, through a model averaging framework, of risk models derived from more modern populations of patients such as ACCORD (in the BRAVO model) and the Hong Kong Diabetes Registry (in the Yang *et al.* risk equations) allow the model to tailor the weighting of each model to each simulated patient. We believe this approach to be better suited to the decision problem than selecting a single model as the basis of the analysis.

d) For the model averaging, estimation of “relatedness” only considers the expected value of each covariate in the source population but does not take any (co)variance into account. Hence, it may excessively penalize deviations for variables that tend to (co)vary a lot. Please justify the current implementation of the model averaging focusing on this aspect.

The default weighting implementation captures HbA1c, SBP, BMI and smoking status. While there are known correlations between SBP and BMI, the model was found to perform well in predicting cardiovascular risk in a diverse array of different cohorts — as presented in the validation exercises in Table 2 and Figure 4 of Pollock *et al.* 2022[48] — despite the lack of any explicit correction for correlations or covariance.

e) By implementing model averaging, the analysis considers that sampled patients originate from multiple distinct populations (each with a different baseline risk of developing complications). This setup may be unrealistic if the models being averaged were developed in heterogeneous settings (e.g., different countries; different healthcare systems, etc.). for variables that tend to (co)vary a lot. Please justify the current implementation of the model averaging focusing on this aspect.

The concern around model heterogeneity in a model averaging framework is legitimate; however, we would respond by noting that the model averaging approach is conceptually preferable to the commonly-used alternative approach of using a single risk model (or set of risk equations, e.g. the UKPDS OM2) and populating these with data from a distinct population (e.g. SURPASS-2). In this instance, the heterogeneity between the model derivation cohort and the simulated cohort would be fixed and constant for a given simulation. Conversely, in the PRIME T2D Model approach, every simulated patient is be exposed to complication risk estimates derived from multiple models, tailored by the patient’s similarity to the model derivation cohorts. The other alternative — of deriving risk models and/or equations directly from the trial — is not viable given the short-term nature of the trials (e.g. 40 weeks for SURPASS-2) and the known long-term consequences of poor glycaemic control.

f) To better understand the impact of model averaging, could the company provide the distribution of (normalized) model weights (across all simulated individuals) calculated at baseline.

As the model weights are derived on a per-patient, per-complication basis in each cycle of the model, there would be 1,800,000 baseline weights for heart failure alone. Rather than supplying these weights outside of the model for a single population and for a single timepoint, we would instead recommend outputting the weights directly from the `ModelAveraging` controller. This could be achieved in the ModelAveraging.java on line 73, where the `modelWeights` HashMap contains the weights for each model for the patient currently under evaluation. We would be happy to provide additional support in achieving this in a format that would be most useful.

B 5. Appendix N provides descriptions for the generic PRIME T2D Model. However, the appropriateness of the selected predictive models to estimate the risk of complications in patients with type 2 diabetes is not justified (in detail). Nor is the applicability to the specific decision problem (as specified in the CS) justified.

a) Please provide a justification that the risk models used, both individually and after model averaging are appropriate to estimate the risk of complications in patients with type 2 diabetes. Please provide this separately per risk model.

The justification for using the PRIME T2D Model based on several aspects:

- As outlined in Section B.3.2.1 of the CS, literature identified shortcoming of existing models of T2D with respect to evaluation of cardiovascular endpoints that were shared between many models as they relied on the same risk equations
- In the development of the PRIME Type 1 Diabetes Model, we were able to show that a model averaging approach, when used to evaluate the risk of cardiovascular endpoints, was superior to any individual risk equations alone. The evidence indicated that risk equations performed well in validations against the derivation populations (or similar populations) but poorly in populations with different characteristics or risk profiles. This is the essential tenet of the model averaging approach: risk equations are weighted to match the risk profile of individual patients to avoid the situations where risk equations from low risk populations (e.g. UKPDS) are applied to high risk patients (e.g. patients in a simulation with long duration of diabetes, advanced disease, history of complications and elevated risk factors)
- Previous cost-effectiveness evaluations performed by NICE have relied on risk equations from the UKPDS OM2; however, recent evidence indicates that without calibration (which is not possible unless the outcomes for a population are already known) the evaluation of cardiovascular endpoints using this methodology may have limitations for T2D patients receiving a modern standard of care and/or receiving interventions that have the potential to lower cardiovascular risk (Si *et al.* 2020, Keng *et al.* 2022)[49, 50]
- The model averaging approach used in the PRIME T2D Model is designed to overcome these shortcoming and be adaptable to cardiovascular risk estimation in different

populations. The validation of the PRIME T2D Model performed to date indicates that the model is capable of reliably predicting outcomes for UK populations (c.f. Lipids in Diabetes Study validation) and for populations receiving a modern standard of care and/or receiving interventions that have the potential to lower cardiovascular risk (c.f. REWIND and EMPA-REG validations) as described in the model technical report (Appendix N3).

b) Please provide a justification that the models used, both individually and after model averaging are applicable for the specific decision problem (as specified in the CS). Please provide this separately per model.

As outlined in the response to part a) above, the suitability of the model averaging approach is that it is adaptable as patient characteristics change over time (i.e. as the disease progresses and the risk of complications increases). The risk equations from the UKPDS OM2 have been widely used in the past, have been derived from a UK-specific T2D populations and are likely well-suited for patients with a low risk profile and short duration of disease. Risk equations from the BRAVO Model are better suited to patients with more advanced disease and higher risk profile (derived from the ACCORD trial population which was at high risk of cardiovascular complications). Literature review did not identify any UK-specific risk equations that could be used in a model averaging approach for patients at high risk of complications and therefore BRAVO Model equations were used. Risk equations from the Hong Kong Diabetes Registry present in the PRIME T2D Model are applicable for South East Asian populations and were not influential in the present analysis.

The model averaging approach is relevant to the decision problem because, based on the validation evidence available to date, it can provide reliable outcomes for a T2D population similar to that described in the decision problem as they progress over a long-term time horizon (and the disease advances and risk profiles change over time). This is important for a long-term cost-effectiveness evaluation designed to characterise the impact of changes in risk factors (such as HbA1, blood pressure, serum lipid levels and body mass index) on the risk of diabetes-related complications, as these often take years to develop and have an effect on survival, costs and quality of life.

B 6. Simulated patients in the modelling analysis were assumed to intensify therapy when HbA1c levels rose above 7.5%, in line with NICE NG28. Simulated patients were assumed to switch to basal insulin therapy on intensification and remain on basal insulin therapy for the rest of the simulation. Assumptions regarding the effectiveness of basal insulin therapy were justified based on the absence of evidence.

a) Please provide additional justification for the assumption that BMI returns to baseline levels in the first year of basal insulin therapy.

In terms of the assumption that BMI returns to baseline levels in the first year of basal insulin therapy, this was a conservative assumption for the analysis as no data were available to inform it. There is evidence to suggest that body weight rebounds after stopping GLP-1 receptor agonist therapy (Wilding *et al.* 2022),[57] which formed the based on this assumption. Moreover, the

clinical input data for the modelling analysis (and the SURPASS trial results) showed that tirzepatide was associated with greater BMI reductions that comparators. As a result, any other assumption (e.g. patients regained a percentage of the weight they lost, or patients regained a fixed amount of weight on initiating basal insulin therapy) would have led to a continued benefit for tirzepatide through the remainder of the modelling simulation. To better evaluated cost-effectiveness, it was decided to only model a BMI benefit when on active/comparator treatment, with the assumption that there were basically no differences between treatment arms following intensification to basal insulin therapy.

b) Please provide additional justification for the assumption that all other risk factors (than BMI and HbA1c) were assumed to return to baseline levels upon initiation of insulin therapy.

Similarly to point a) above, the assumption that other risk factors also returned to baseline levels was a conservative assumption made in the absence of evidence to support a continued benefit. Tirzepatide was generally associated with greater improvements in systolic blood pressure and serum lipid levels than comparators in the NMA (and in the SURPASS trial). Assuming a continued benefit in the tirzepatide treatment arms beyond the treatment period would have favoured tirzepatide. It could be considered likely that the benefits in certain risk factors may persist for a period of weeks or months following treatment (as opposed to rebounding back to baseline), however without any direct supporting evidence it was felt that such an assumptions would be inappropriate in the present analysis.

c) Please elaborate on the implications of the assumptions considered in the sub-questions above.

If the benefits associated with tirzepatide treatment were assumed to persist beyond the treatment period in the modelling analysis (as opposed to rebounding to baseline levels), there is the potential to have improved clinical outcomes (due to a reduced risk of diabetes-related complications and improved quality of life associated with lower BMI) at no additional costs. This would have improved the cost-effectiveness of tirzepatide (and lowered ICERs).

d) Please explore the impact of alternative assumptions regarding the effectiveness of basal insulin therapy, in terms of impact on the risk factors (including BMI and HbA1c).

There are many potential assumptions that could have been explored around HbA1c and BMI on intensification to basal insulin therapy in the modelling analysis (leading to a great many simulations). These range from assuming persistent benefits after intensification (favouring tirzepatide therapy over comparators) to assuming no benefits after intensification to insulin (most conservative assumption).

As demonstrated in the sensitivity analysis, the HbA1c and BMI benefits associated with tirzepatide were importance drivers of improved outcomes and, ergo, cost-effectiveness. In Table 106 of the CS, sensitivity analysis reporting an assumption of no HbA1c benefit at all for tirzepatide (ICER for tirzepatide 10 mg versus semaglutide 1.0 mg was £12,510 per QALY gained) and an assumption of no BMI benefit at all for tirzepatide (ICER for tirzepatide 10 mg versus semaglutide 1.0 mg was £15,854 per QALY gained) both showed that tirzepatide remained cost-effective in both of these extreme cases. Changing assumptions of the durability

of effect for HbA1c and BMI, however conservative, would have produced ICERs well within limits outlined in these sensitivity analyses. It should be noted that these results are also reflective of the comparisons of tirzepatide 5 mg with semaglutide 0.5 mg and tirzepatide 15 mg versus semaglutide 1.0 mg. In short, using alternative assumptions around HbA1c and BMI on initiation of basal insulin therapy would not have notably changed the cost-effectiveness profile of tirzepatide.

Population

B 7. Priority question. The baseline characteristics for the model population (provided in CS Table 75) included age, sex, and weight, and were mostly based on the baseline characteristics of patients in the THIN second intensification cohort, which standard care was based on: metformin, sulfonylurea and NPH insulin.

a) Please provide justification for using the THIN second intensification as the main source of the cohort characteristics instead of one of the SURPASS trials, including in relation to applicability to the decision problem.

The rationale for choosing this population was two-fold: 1) it was considered to be representative of the population with T2D in general practice in England (as the data were derived from The Health Improvement Network (THIN) database), and 2) it was considered to be representative of the population initiating second line therapy in clinical practice, after failing diet and exercise plus metformin. This stage in therapy aligns the population broadly with the decision problem (tirzepatide as an adjunct to diet and exercise, in addition to other medicinal products for the treatment of diabetes), the SURPASS-2 population (tirzepatide as add-on therapy to metformin) and the NMA population (CS Section B.2.9.5.1: the analysis population was defined to align with SURPASS-2 and 3 trials, and included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice.). Whilst the decision problem also spans the addition of tirzepatide in subsequent lines of therapy (e.g. third-line), population characteristics from the THIN database are not available for subsequent lines of therapy. Therefore, in line with previous economic evaluations for NICE Guidance NG28, it was considered that population representative of general practice in England was a more suitable choice for the present analysis than a specific trial population (e.g. SURPASS-2 or SURPASS-3), given the multi-country nature of these trials.

b) According to the CS, baseline characteristics were derived from the THIN second intensification cohort, and when values were missing, data from the SURPASS-2 clinical trial cohort were used, given that they had a "comparable duration of diabetes" (mean duration of THIN second intensification cohort: 8.6 years, mean duration of the SURPASS-2 cohort: 8.5 years). Please elaborate on the suitability of

using the SURPASS-2 clinical trial data as a proxy for the THIN second intensification cohort. Please elaborate on the differences in demographics, especially regarding population age (63.95 years vs 56.6 years), ethnic group percentages, and baseline risk factors

Almost all key demographic and baseline risk factor estimates were taken from the THIN second intensification population in the modelling analysis. The exception was total cholesterol levels, which is not a key driver in the risk equations used in the model. In terms of the history of complications in the cohort, estimates for only four out of 13 pre-existing complications at baseline were taken from SURPASS-2. As correctly pointed out by the EAG, the SURPASS-2 cohort was younger than the THIN second intensification cohort, leading to a lower overall risk profile (and therefore lower potential for clinical benefits with risk factor improvements, as evidenced in the QALY benefit of the SURPASS-2 cohort sensitivity analysis relative to the base case). However, the duration of diabetes was comparable (THIN 8.5 years versus SURPASS-2 8.6 years), indicating the that history of complications may be comparable between the two populations. Use of the SURPASS-2 data as a proxy for a small number of baseline cohort characteristics (5 our of 35) can therefore be considered to be a potentially conservative assumption, with a minimal impact on simulation outcomes.

c) **Please provide a table comparing the mean values of CS Table 75 with the values from the remaining SURPASS trials (i.e., SURPASS 3-5), and elaborate on the possible differences (e.g., mean age being higher in the THIN second intensification group).**

A table summarizing available cohort characteristics from SURPASS-3, 4 and 5 and provided alongside the THIN second intensification cohort in the table below. Cohort characteristics are broadly similar across the THIN population and SURPASS-3, as might be expected given that both populations are at a similar stage of treatment intensification. The SURPASS-4 cohort is older than the THIN cohort and has many more cardiovascular complications at baseline, consistent with the “established cardiovascular disease or a high risk of cardiovascular events” definition. Mean baseline HbA1c was also higher in this cohort (8.52%) than in the THIN cohort (7.50%). SURPASS-5 was broadly similar to the THIN cohort, but had a longer duration of diabetes (THIN 8.5 years versus SURPASS-5 13.3 years)

Table 13: Summary of cohort characteristics – SURPASS trials

	THIN second intensification cohort	SURPASS-3	SURPASS-4	SURPASS-5
Demographics				
Percentage male (%)	57.0	55.8	62.5	55.6
Percentage with college education or higher (%)	25.97	n/r	n/r	n/r
Percentage smokers (%)	17.0%	n/r	n/r	n/r
Age (years)	63.95	57.4	63.6	60.6
Duration of diabetes (years)	8.5	8.38	11.78	13.3
Ethnic group				
Percentage White (%)	82.4	91.0	81.8	80.0
Percentage Black (%)	4.5	3.1	3.7	1.3
Percentage Hispanic (%)	0.0	29.3	47.6	4.6
Percentage Southeast Asian (%)	0.0	N/A	N/A	N/A
Percentage Indian (%)	13.1	N/A	N/A	N/A
Percentage Afro/Caribbean (%)	0.0	N/A	N/A	N/A
Percentage Asian (%)	N/A	5.3	3.5	17.9
Percentage American Indian or Alaska Native	N/A	0.3	8.7	0.4
Percentage Native Hawaiian or Other Pacific Islander	N/A	0.3	0.2	N/A
Percentage Multiple	N/A	0.1	2.2	N/A
Percentage Other (%)	0.0	N/A	N/A	0.4

	THIN second intensification cohort	SURPASS-3	SURPASS-4	SURPASS-5
Baseline risk factors				
HbA1c (%)	7.50	8.17	8.52	8.31
Systolic blood pressure (mmHg)	134.44	131.53	134.4	137.9
Total cholesterol (mmol/L)	4.53	n/r	n/r	n/r
Low density lipoprotein cholesterol (mmol/L)	2.29	n/r	n/r	n/r
High density lipoprotein cholesterol (mmol/L)	1.23	n/r	n/r	n/r
Body mass index (kg/m^2)	30.7	33.52	32.550	33.4
Estimated glomerular filtration rate ($mL/min/1.73\ m^2$)	71.37	94.13	81.26	85.5
White blood cell count (10^6 cells/mL)	7.5	n/r	n/r	n/r
Heart rate (beats per minute)	72.0	75.22	72.76	75.2
Haemoglobin (g/dL)	14.5	n/r	n/r	n/r
Complication history				
Percentage with atrial fibrillation at baseline (%)	1.2%	1.5	5.9	2.7
Percentage with urinary albumin ≥50mg/L at baseline (%)	22.6%	0.9 (proteinuria)	n/r	0.6 (proteinuria)
Percentage with peripheral vascular disease at baseline (%)	1.9%	0.3	30.3 (peripheral artery disease)	2.9 (peripheral artery disease)
Percentage with history of myocardial infarction at baseline (%)	2.0%	0.6	32.3	4.2
Percentage with history of stroke at baseline (%)	1.3%	1.7	12.0	3.8
Percentage with ischemic heart	6.0%	n/r	n/r	n/r

	THIN second intensification cohort	SURPASS-3	SURPASS-4	SURPASS-5
disease at baseline (%)				
Percentage with coronary revascularization at baseline (%)	3.0%	2.7	32.3	7.8
Percentage with heart failure at baseline (%)	1.9%	0.1	7.0	0.8
Percentage with foot ulcer at baseline (%)	0.8%	n/r	n/r	n/r
Percentage with amputation at baseline (%)	0.2%	0.1	0.3	0.2
Percentage with blindness at baseline (%)	1.3%	0.1	0.3	0.2
Percentage with renal failure at baseline (%)	0.4%	0.1	0.4	1.7
Percentage with SPSL/neuropathy at baseline (%)	9.0%	14.4	18.6	17.5

Abbreviations: SPSL: severe pressure sensation loss.

d) Please elaborate on the consistency of CS Table 75 with the baseline characteristics of the trials in the NMA described in section B.2.9 of the CS. Please provide an updated CS Table 75, comparing the current model values with the ones obtained for the NMA. Please update this in line with any new NMA as requested in Section A.

The NMA did not report combined cohort characteristics for an analysis population and tabulation of the NMA input studies would involve adding cohort characteristics from 45 studies to the table above (which is impracticable and unlikely to be informative). In line with the points made in response to a) and c) above, we would point out that the population used in the modelling analysis (THIN second intensification population) is well aligned with the analysis population from the NMA, which focused specifically on "studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice."

B 8. Priority question. Question A5 summarises the difference between the decision problem and the NICE scope.

a) The company justified this choice by stating that it would be the anticipated positioning of tirzepatide in the UK clinical practice.

Considering the responses to question A5, please provide further justification for not including the population from the NICE final scope in the economic model and discuss the possible impact on the cost-effectiveness.

As outlined in the response to question A5, Lilly consider that tirzepatide would be an option whenever GLP-1 RAs would otherwise be considered. This anticipated position aligns with current NHS clinical practice in England and reflects the highest unmet need for a more effective treatment option for patients for whom the alternative is a GLP-1 RA, which may not sufficiently control their HbA1c level and/or provide sufficient weight loss. The population used in the modelling analysis is well aligned with this target population. Adjusted the characteristics of the population used in the modelling analysis (without corresponding changes in treatment effects) would have only a minimal impact on the relative cost-effectiveness of tirzepatide versus comparators. This was evidenced in the CS with a sensitivity analysis (CS Table 106) with the SURPASS-2 cohort.

b) **Please provide an updated economic model and scenario analyses including the population in the NICE scope in the economic model.**

In line with the response to part a), Lilly will not be providing these analyses as they maintain that the population included in the decision problem and economic model is the most relevant for UK clinical practice.

As outlined in the response to B7, the population used in the modelling analysis was chosen to be representative of routine clinical practice in England, align with the target population as stated in the decision problem and align with the target population from the NMA. The sensitivity analysis simulations outlined in the CS provide evidence that cohort characteristics are not key drivers of clinical outcomes in the modelling analysis (instead the HbA1c and body weight benefits associated with tirzepatide are important drivers of cost-effectiveness). This simulation has therefore not been run.

Intervention and comparator

B 9. Priority question. The final scope mentions the following treatments as comparators: sulfonylureas, DDP-4 inhibitors, pioglitazone, GLP-1 mimetics, SGLT-2 inhibitors, and insulin, as monotherapy or in a combination regimen (in accordance with NICE guidance). In the CS, the comparator defined were only a series of GLP-1 RAs in combination regimens.

a) **Please provide a justification for not including all the comparators described in the final scope (neither in monotherapy nor in combination), as a comparator in the economic model.**

Please refer to the response given in Question A5 for justification as to why these comparators were excluded from the analysis.

b) Please provide an updated economic model and scenario analyses, including all comparators described in the final scope, as a comparator in the economic model.

As discussed in previous responses, not all of the comparators included in the final scope are relevant to this submission, due to the anticipated positioning of tirzepatide in UK clinical practice. An updated economic model and associated scenario analyses have therefore not been presented.

c) Please provide the results of a fully incremental analysis (and updated economic model used for this analysis) with all comparators listed in the scope as comparators modelled separately, considering the responses to question A5.

As not all of the comparators included in the final scope are relevant to this submission, a fully incremental analysis has not been presented. A fully incremental analysis versus relevant comparators is presented in Question B29.

B 10. Priority question. As described in question A6, the recommended maintenance doses of the intervention (i.e., tirzepatide) are 5 mg, 10 mg and 15 mg. As per Table 2 in the CS, tirzepatide should be initiated at 2.5 mg via injection every week (QW). After 4 weeks, the dose is increased to 5 mg QW. If needed, the dose could be increased in 2.5 mg increments every 4 weeks up to 15 mg. In CS, B.3.2.4., a series of GLP-1 RAs are listed as comparators in the base case; including diverse dosages and formats (e.g., oral or injectable) of dulaglutide, semaglutide, liraglutide. Nonetheless, exenatide and lixisenatide are not included in the main analysis due to limited market share in the UK.

a) Please justify why comparisons were made within each recommended maintenance dose step (5, 10 and 15 mg), rather than between the recommended dose steps (CS, Table 76). Please also provide further evidence, if necessary, to justify your response.

As outlined in the response to question A6, in seeking to meaningfully interpret the comparative efficacy results for tirzepatide 5, 10 and 15 mg, it is more meaningful to compare within dose steps rather than between, as in any given individual patient the dose will have been titrated according to the balance of patient tolerability and observed treatment effect. In clinical practice, patients are expected to be titrated up the recommended maintenance doses as required, and

the most appropriate dose will be determined by the clinician based on clinical characteristics and patient tolerability, aligned to the SmPC.

However all comparators are included in the NMA and may be analysed in the economic model if desired.

b) As stated in question A6, please compare all comparators currently included (listed in CS Table 89; analyses 1-12), not only within a specific 'maintenance dose step'.

As outlined in point a), only the maintenance dose steps are relevant for a long-term cost-effectiveness analysis of tirzepatide (as intermediate doses will only be used transiently for titration).

c) A different escalation time was observed in the SURPASS trials for the intervention (i.e., tirzepatide), which ranged from 0-20 weeks, and the comparators, which ranged from 0-12 weeks. Please clearly explain and show how this was modelled in the PRIME T2D model, showing the exact code and how it could be modified in the online interface.

Dose escalation steps were not explicitly modelled in the long-term cost-effectiveness analysis. The effects of titration/dose escalation would be captured in the treatment effects and adverse event rates reported in the trial data contributing to the NMA results (and are therefore implicitly captured in the modelling analysis).

d) Please elaborate on the decision to exclude exenatide and lixisenatide due to a 'limited market share in the UK', explain the possible reasoning behind the limited market share, and provide further justification on why they should be excluded apart from the market share. Please provide supporting evidence showing that these assumptions are consistent with relevant external data and/or expert opinion. In case of expert opinion, please provide a full description of the methods and results of the expert consultation conducted.

The estimated market share of exenatide is 1.2% for the once weekly dose and 0.8% for the twice daily dose. For lixisenatide, the estimated market share is 0.5%. These estimates were based on Prescribing Based Services (a sample of 82% of patient medical records (PMR) systems, covering 96% of active UK GPs) and Hospital Pharmacy Audit (covering 95% of drug usage/prescription in UK Hospitals).[58] As such, exenatide and lixisenatide do not represent relevant comparators for this submission.

e) Please update the model base case to include exenatide and lixisenatide as well.

In line with the response given in part d), the model base case has not been updated to include exenatide and lixisenatide as these are not relevant comparators in UK clinical practice.

B 11. Priority question. The NMA performed by the company included different comparators than the ones included in the base case. In particular dulaglutide (0.75mg), exenatide (5 and 10 µg) and semaglutide (2 mg).

a) Please elaborate on the impact of including all these comparators in the NMA but not in the base-case analysis and provide further justification for their exclusion in the base case.

The comparators listed above were not included in the base case analysis of this submission because injectable semaglutide 2.0 mg is not licenced in UK clinical practice, dulaglutide 0.75 mg is currently only licensed as monotherapy (which is not recommended by NICE NG28) and as a starting dose for patients who may be considered more vulnerable, therefore only relevant to a sub-population in UK clinical practice and because exenatide (5 and 10 µg) only have a very small market share. The NMA however was designed at a global level to meet the needs of multiple countries, and therefore extends beyond UK clinical practice and the decision problem addressed within this submission.

b) Please update the base-case analysis with the corresponding comparators.

In line with the response to part a), the base-case analysis will not be updated.

B 12. As described in question A9, NG28 recommend GLP-1 RAs in adults with a higher BMI of 35kg/m2 and adults with a lower BMI than 35kg/m2, but for whom insulin therapy would have significant occupational implications or weight loss would benefit other significant comorbidities. Please clearly explain how this was modelled in the PRIME T2D model, showing the exact code used to apply this and how it can be modified in the online interface.

No sub-group analysis based on BMI (e.g. above or below 35 kg/m^2) was performed as part of the modelling analysis. The rationale was simply that the treatment effect data did not exist for these sub-groups. Treatment effect data was only available for the overall target population from the NMA (as described in Section B.2.9). Specific cohort characteristics and treatment effects for all comparators would be needed for each sub-group to be able to perform the cost-effectiveness analysis alluded to in the question.

B 13. As per NICE NG28, the intensification threshold is a percentage of HbA1c of 7.5%. Given that the mean HbA1c level of the cohort in the first year is 7.5%, one could assume that some patients would lower the HbA1c after the first years.

a) Please justify whether patients that have lowered their HbA1c could use or not the treatments recommend in NICE NG28 for patients under the 7.5% threshold. Please provide supporting evidence showing that these assumptions are consistent with relevant external data and/or expert opinion.

NICE recommends that targets are individualised to the patient and provides a tool to support decision making. Additionally, the guidelines suggest:

- For adults whose T2D is managed either by lifestyle and diet, or lifestyle and diet combined with a single drug not associated with hypoglycaemia, support them to aim for an HbA1c level of 48 mmol/mol (6.5%). For adults on a drug associated with hypoglycaemia, support them to aim for an HbA1c level of 53 mmol/mol (7.0%). [2015]
- In adults with T2D, if HbA1c levels are not adequately controlled by a single drug and rise to 58 mmol/mol (7.5%) or higher:
  - reinforce advice about diet, lifestyle and adherence to drug treatment and
  - support the person to aim for an HbA1c level of 53 mmol/mol (7.0%) and
  - intensify drug treatment
- If adults with T2D reach an HbA1c level that is lower than their target and they are not experiencing hypoglycaemia, encourage them to maintain it. Be aware that there are other possible reasons for a low HbA1c level, for example deteriorating renal function or sudden weight loss.

Given the position in the treatment algorithm it is anticipated that patients will already be on multiple oral therapies and therefore the only escalation option likely to provide efficacy remaining is insulin.

b) Please elaborate on why basal insulin was considered the only treatment option after intensification, elaborate on the implications of this assumption and provide supporting evidence on other possible treatments that could be used after intensification.

When planning the cost-effectiveness analysis, different treatment options were explored following intensification from the intervention/comparator treatments. It was noted that, provided there was an assumption that the treatment effects were equivalent in both arms, that post-intensification treatments had very little effect on cost-effectiveness. This is evidenced in the scenario analysis presented in Table 107 in the CS, where a second treatment intensification to basal-bolus therapy was shown to have almost no effect on the ICERs for tirzepatide versus semaglutide. Alternative treatment intensification assumptions could have been triple therapy (same results as above), or potentially continuing the intervention/comparator with basal insulin therapy. This latter scenario was explored in a scenario analysis of tirzepatide 10 mg versus semaglutide 1.0 mg (CS Table 107) where improvements in risk factors were maintained until the second intensification step (switch to basal-bolus insulin therapy with risk factors returning to baseline). This scenario also provided cost-effectiveness outcomes similar to the base case analysis.

Effectiveness

B 14. Priority question: In the CS it is stated that "For all other risk factors [besides HbA1c], it was assumed that no long-term changes would be applied in the modelling analysis. In effect, risk factor changes associated with treatment were applied in year 1 of the simulation and remained constant until treatment intensification when they returned to baseline levels. This simplifying, conservative assumption was used because the clinical benefits associated with therapy were applied for the duration of treatment (therefore balancing costs and effects for all comparators in the analysis) and there was little long-term data on the durability of treatment effects and effects of switching to basal insulin from GLP-1 RA on individual risk factors."

a) Please provide justifications (in addition to sparsity of data) for this assumption of constant risk factors after year 1 up to treatment intensification.

The assumption of no risk factor progression over time is aligned with treatment goals of not allowing hypertension to develop/worsen or allowing dyslipidaemia to develop/worsen. For most risk factor progressions other than HbA1c (e.g. BMI, SBP and serum lipid levels), only very modest changes are observed over time (Leal *et al.* 2021).[59] As a result, the differences between the assumption of no change over time and the application of risk factor progression equations (e.g. from UKPDS OM2) is negligible whilst patients on tirzepatide or comparator therapy. After intensification to basal insulin, provided the assumptions around risk factor progression are the same in both treatment arms, there would be no difference between the arms and the assumption would have little or no impact on cost-effectiveness. Sensitivity analysis using UKPDS OM2 risk factor progressions for all risk factors produced outcomes very similar to the base case analysis (as outlined in the CS Table 106 for tirzepatide 10 mg versus semaglutide 1.0 mg).

b) Please explore alternative scenarios, including scenarios assuming:

i. after year 1 the risk factor values returned to baseline levels

ii. after year 2 the risk factor values returned to baseline levels

iii. after year 3 the risk factor values returned to baseline levels

iv. after year 4 the risk factor values returned to baseline levels

v. after year 5 the risk factor values returned to baseline levels

Outcomes for the requested simulations with all risk factors returning to baseline after 1-5 years in the simulation are provided in the table below. Broadly speaking, the simulations show that in the scenarios with treatment effects lasting 2 years or more, tirzepatide was cost-effective versus semaglutide. It should be noted that the 1-year and 2-year return to baseline scenarios are extremely conservative, as they assume treatment costs were accrued over several years but the corresponding clinical benefits are only realised for a much shorter time period. There are a few points to note about the results presented below. The requested scenarios are not compatible with the base case assumption of treatment intensification with HbA1c over 7.5% (modelled using HbA1c progression from UKPDS OM2). For the requested scenarios, therefore, no HbA1c creep over time was assumed (HbA1c was lower on treatment based on the NMA change from baseline estimates and then returned to baseline level at year 1-5 and remained there until the end of the simulations). The same assumption was applied to other risk factors. Total direct costs were estimated by taking treatment costs from the base case analysis (to approximate treatment costs based on switching with HbA1c at 7.5%) and complication costs from simulations with clinical benefits lasting only 1-5 years. Differences in survival between the base case analysis and truncated clinical benefits simulations may lead to a small underestimation of lifetime treatment costs (that is approximately equivalent in both treatment arms and therefore will have little impact on cost-effectiveness).

Table 14: Summary of base case results for tirzepatide 5 mg versus comparators

	Estimated direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Estimated incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg versus semaglutide 0.5 mg							
Clinical benefit for 1 year only - TZP	[redacted]	13.898	9.539				
Clinical benefit for 1 year only - SEMA	[redacted]	13.88	9.582	1,178	0.018	0.042	28,048
Clinical benefit for 2 years only - TZP	[redacted]	13.966	9.687				
Clinical benefit for 2 years only - SEMA	[redacted]	13.933	9.624	1,120	0.033	0.064	17,500
Clinical benefit for 3 years only - TZP	[redacted]	13.992	9.762				
Clinical benefit for 3 years only - SEMA	[redacted]	13.956	9.687	1,056	0.035	0.075	14,080
Clinical benefit for 4 years only - TZP	[redacted]	14.062	9.863				
Clinical benefit for 4 years only - SEMA	[redacted]	14.017	9.773	973	0.045	0.09	10,811
Clinical benefit for 5 years only - TZP	[redacted]	14.112	9.948				
Clinical benefit for 5 years only - SEMA	[redacted]	14.051	9.838	923	0.061	0.111	8,315

	Estimated direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Estimated incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 10 mg versus semaglutide 1.0 mg							
Clinical benefit for 1 year only - TZP	[redacted]	13.919	9.566				
Clinical benefit for 1 year only - SEMA	[redacted]	13.906	9.606	1,178	0.014	0.04	29,450
Clinical benefit for 2 years only - TZP	[redacted]	13.98	9.714				
Clinical benefit for 2 years only - SEMA	[redacted]	13.936	9.643	1,120	0.044	0.071	15,775
Clinical benefit for 3 years only - TZP	[redacted]	14.018	9.806				
Clinical benefit for 3 years only - SEMA	[redacted]	13.993	9.734	1,056	0.025	0.072	14,667
Clinical benefit for 4 years only - TZP	[redacted]	14.071	9.901				
Clinical benefit for 4 years only - SEMA	[redacted]	14.045	9.816	973	0.025	0.085	11,447
Clinical benefit for 5 years only - TZP	[redacted]	14.136	10.002				
Clinical benefit for 5 years only - SEMA	[redacted]	14.064	9.878	923	0.072	0.124	7,444

	Estimated direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Estimated incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 15 mg versus semaglutide 1.0 mg							
Clinical benefit for 1 year only - TZP	[redacted]	13.922	9.614				
Clinical benefit for 1 year only - SEMA	[redacted]	13.906	9.566	1,233	0.016	0.047	26,234
Clinical benefit for 2 years only - TZP	[redacted]	13.972	9.721				
Clinical benefit for 2 years only - SEMA	[redacted]	13.936	9.643	1,104	0.036	0.079	13,975
Clinical benefit for 3 years only - TZP	[redacted]	14.052	9.845				
Clinical benefit for 3 years only - SEMA	[redacted]	13.993	9.734	1,031	0.059	0.111	9,288
Clinical benefit for 4 years only - TZP	[redacted]	14.104	9.945				
Clinical benefit for 4 years only - SEMA	[redacted]	14.045	9.816	978	0.058	0.128	7,641
Clinical benefit for 5 years only - TZP	[redacted]	14.164	10.049				
Clinical benefit for 5 years only - SEMA	[redacted]	14.064	9.878	915	0.1	0.171	5,351

Abbreviations: QALY: quality-adjusted life year; TZP; tirzepatide; SEMA; semaglutide. * incremental cost-effectiveness ratio for tirzepatide versus comparator. Note: all ICERs are calculated. Estimated direct costs were calculated by taking treatment costs from the base case analysis (to approximate treatment costs based on switching with HbA1c at 7.5%) and complication costs from simulations with truncated clinical benefits. Differences in survival between the base case analysis and truncated clinical benefits simulations may lead to a small underestimation of lifetime treatment costs (that is approximately equivalent in both treatment arms and therefore will have little impact on cost-effectiveness)

B 15. In the CS it is stated that: “For the model inputs, changes from baseline in BMI were calculated based on changes in body weight reported in the NMA. This is because whilst BMI was included in the NMA outputs, values were not available for all comparators whereas changes in body weight were available for all comparators. To avoid the use of proxy inputs from "nearest neighbour" comparators and as done in the NG28 economic analysis, BMI changes were calculated from the NMA-reported body weight changes and an assumed cohort height of 1.68 m, in line with the mean value reported for the THIN population.”

a) Please provide an overview of missing BMI values for the comparators.

BMI estimates from the NMA were not available for four out of the nine comparators in the analysis. These were:

- Dulaglutide 3.0 mg
- Dulaglutide 4.5 mg
- Oral semaglutide 7 mg
- Liraglutide 1.2 mg

b) Please provide an overview and comparison of the BMI 1) as calculated from the NMA and; 2) as calculated based on body weight (i.e. as in the CS base-case).

An overview of change in BMI values from the NMA and those calculated based on body weight is provided in Table 15. Multiple sensitivity analyses on change from baseline in BMI, including a scenario using the value directly from the NMA, are provided in Table 106 of the CS.

Table 15: Change in BMI values from the NMA and calculated mean change

Intervention	NMA mean (SD) change in BMI	Calculated mean (SD) change in BMI
Tirzepatide 5 mg	−2.44 (0.38)	−2.48 (0.23)
Tirzepatide 10 mg	−3.44 (0.38)	−3.46 (0.23)
Tirzepatide 15 mg	−4.19 (0.38)	−4.18 (0.23)
Dulaglutide 1.5 mg	−0.93 (0.26)	−0.82 (0.14)
Dulaglutide 3.0 mg		−1.11 (0.23)
Dulaglutide 4.5 mg		−1.25 (0.23)
Semaglutide 0.5 mg	−1.32 (0.35)	−1.19 (0.21)
Semaglutide 1.0 mg	−1.89 (0.30)	−1.81 (0.17)
Oral semaglutide 7 mg		−0.92 (0.22)
Oral semaglutide 14 mg	−1.61 (0.29)	−1.41 (0.18)
Liraglutide 1.2 mg		−0.85 (0.16)
Liraglutide 1.8 mg	−1.07 (0.27)	−1.03 (0.13)

c) Please justify the appropriateness of calculating the BMI changes based on an assumed average cohort height of 1.68m.

BMI changes were estimated based on body weight changes to resolve the issue of missing data (as the results for four comparators were not available from the NMA). It was assumed that this approach would provide an accurate reflection of the impact of weight loss as estimated in the NMA on the modelled populations (THIN second intensification cohort), which was taken to be representative of a T2D population in general practice in England. The value of 168 cm (derived from the THIN second intensification cohort) is also well aligned with average height of the general population based on data from the National Health Survey for England 2019 (https://digital.nhs.uk/supplementary-information/2022/hse19-mean-and-median-heights).[60]

d) The scenario “Change in BMI values taken directly from NMA” (CS Table 106) indicated a relatively substantial increase in ICER. Please explain what drives this increase in estimated ICER.

The relative difference in terms of change from base line in BMI between these two scenarios was approximately 0.1 kg/m^2 with a slightly smaller decrease in BMI with tirzepatide 10 mg and slightly great decrease with semaglutide 1.0 mg. This led to a difference in incremental QALYs of 0.006 relative to the base case analysis, which along side marginally higher costs on tirzepatide (+£51 over a lifetime) and slightly lower costs on semaglutide (-£43) due to modest differences in complication rates, led to an ICER that was approximately £1,302 higher than the base case (at £10,009 per QALY gained).

B 16. Please clarify that the mean values reported in CS Table 80 are the treatment effects (for body weight) applied in the first year of the simulation (as specified in the headers for Tables 77-79).

We can confirm that the values in Table 80 were the changes applied in year 1 of the modelling simulations. It should be noted that the model on has inputs for BMI. No changes in body weight were entered directly into the model. These values were used only to calculate the BMI changes summarized in Table 80 of the CS (and Tables 77-79).

B 17. In the CS it is stated that: “In the absence of missing inputs from the NMA, a conservative "nearest neighbour" approach was used to fill data gaps. Where inputs were missing, the corresponding input from the same compound was used as a proxy, wherever possible using higher (more efficacious) doses of comparator.”

a) Please provide an overview of parameters where this “nearest neighbour” approach was used.

All values for treatment effects that were not directly available from the NMA and were populated with surrogate values using the nearest neighbour approach were highlighted in Tables 77-79 in the CS and described in the table footnotes. The tables are reproduced below with surrogate

values in red font for easy identification. Out of the 75 treatment effects values in the modelling analysis, 14 missing parameters were populated in this way.

Table 16: Treatment effects applied in the first year of the simulation for tirzepatide 5 mg and comparators

	Tirzepatide 5 mg mean (SD)	DULA 1.5 mg mean (SD)	SEMA 0.5 mg mean (SD)	ORAL SEMA 7 mg mean (SD)	LIRA 1.2 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with semaglutide 0.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA: liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: semaglutide.

Table 17: Treatment effects applied in the first year of the simulation for tirzepatide 10 mg and comparators

	Tirzepatide 10 mg mean (SD)	DULA 3.0 mg mean (SD)	SEMA 1.0 mg mean (SD)	ORAL SEMA 14 mg mean (SD)	LIRA 1.8 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with dulaglutide 1.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA, liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: semaglutide.

Table 18: Treatment effects applied in the first year of the simulation for tirzepatide 15 mg and comparators

	Tirzepatide 15 mg mean (SD)	DULA 4.5 mg mean (SD)	SEMA 1.0 mg mean (SD)	ORAL SEMA 14 mg mean (SD)	LIRA 1.8 mg mean (SD)	Source
HbA1c change from baseline (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
SBP change from baseline (mmHg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
BMI change from baseline (kg/m^2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
HDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA
LDL change from baseline (mmol/L)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NMA

* value not available from the NMA, change from baseline associated with semaglutide 1.0 mg used as a proxy. ** value not available from the NMA, change from baseline associated with dulaglutide 1.5 mg used as a proxy.
Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA: liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: Semaglutide.

b) Please provide a justification of the appropriateness of this “nearest neighbour” approach, individually for all occurrences.

The nearest neighbour approach involved substituting in change from baseline estimates from another comparator in the even of missing data. This was done hierarchically by 1) selecting a surrogate value from the same treatment (e.g. using a value from a different dose of the same comparator or, in the case of oral semaglutide using a value from injectable semaglutide, and 2) selecting a surrogate value from the nearest, more efficacious dose where ever possible (i.e. conservative assumption). The only cases where conservative nearest neighbour substitutions were not possible were for the 3.0 and 4.5 mg doses of dulaglutide in terms of the HDL and LDL parameters (where values from dulaglutide 1.5 mg were substituted in). Sensitivity analysis (CS Table 106) showed that the changes from baseline in serum lipid levels had a very modest impact on cost-effectiveness and it can be assumed that the dulaglutide nearest neighbour substitutions would have a negligible impact on the outcomes of the modelling analysis. All other nearest neighbour substitutions were conservative (and would have favoured the comparator over tirzepatide).

c) Please provide a detailed, step by step, description of the “nearest neighbour” approach used.

Please see point b) above for a description of the nearest neighbour approach.

B 18. In the CS it is stated that: “It was assumed for the modelling analysis that tirzepatide and all comparators had an equivalent effect on renal function, with changes from baseline in estimated glomerular filtration rate (eGFR) set to zero for all treatments.” and “change from baseline in white blood cell count and haemoglobin levels were set to zero for tirzepatide and all comparators, as these endpoints were not included in the NMA”.

a) Please elaborate on the implications of assuming tirzepatide and all comparators had an equivalent effect on renal function (assuming baseline values for all treatments).

In the absence of evidence of a differential effect on renal function for tirzepatide versus comparators from the NMA or from the SURPASS trial program, it was assumed that change from baseline in eGFR was the same for all treatments so this would not have a direct effect on renal outcomes (and therefore cost-effectiveness in the modelling analysis). In the SURPASS-2 scenario analysis, changes from baseline in eGFR were modelled as reported in CSR and had very little impact on cost-effectiveness (with the scenarios producing very similar ICERs to the base case analysis CS Table 107). Similarly, sensitivity analysis (CS Table 106) using the UKPDS OM2 risk formula to estimate the risk of renal disease (and therefore capturing the impact of systolic blood pressure and BMI on the risk of renal failure as well as eGFR) similarly had little impact on cost-effectiveness.

b) Please elaborate on the implications of assuming tirzepatide and all comparators had an equivalent effect on white blood cell count and haemoglobin levels (assuming baseline values for all treatments).

Similarly, in the absence of evidence of a differential effect on white blood cell count (WBC) or haemoglobin levels for tirzepatide versus comparators from the NMA or from the SURPASS trial program, it was assumed that these parameters were the same for all treatments. These were set to the index value for the UKPDS OM2 risk equations so they would have no influence on outcomes. Differential effects on WBC can potentially have a small impact on the risk of multiple complications (including myocardial infarction, stroke, blindness, amputation and renal failure) whereas differential effects on haemoglobin can influence the risk of renal failure when using the UKPDS OM2 risk equations (not applied in the base case analysis). In light of the paucity of evidence on any differential effects on either of these parameters combined with the very modest impact of WBC on complication risk and the fact that haemoglobin did not influence risk in the base case analysis, these parameters were essentially set to "no effect" in the modelling analysis.

B 19. In the CS it is stated that "Simulated patients in the modelling analysis were assumed to intensify therapy when HbA1c levels rose above 7.5%, in line with NICE guidance for the management of T2D". In the CS base-case treatment intensification implies discontinuing the initial treatment and switching to basal insulin therapy.

a) Please clarify and justify whether other causes for treatment discontinuation were incorporated.

No other causes of treatment discontinuation were included in the modelling analysis.

b) Please provide scenario analyses incorporating other causes for treatment discontinuation.

Appropriately modelling discontinuation rates can be challenging in a diabetes model primarily because no treatment is no treatment is not a viable option. Therefore, discontinuation events need to have a rescue therapy, which comes with corresponding costs and effects, and can easily skew the results of a cost-effectiveness analysis. For example, a treatment can have a good ICER if there is a high discontinuation rate and a low cost rescue medication, particularly if that medication lowers HbA1c. In addition, it can lead to complex treatment arms (where some patients intensify therapy and others switch to rescue medication) for the intervention and comparators, making it challenging to ascertain cost-effectiveness due to the interaction of costs and effects from the different therapies involved.

To avoid the potential for rescue medication influencing the outcomes of the present analysis, treatment intensification was only modelled using an HbA1c threshold. This is aligned with an assumption that patients who do not tolerate the interventions well are likely to miss doses, leading to poorer glycaemic control and meeting the criterion for intensification (see response to question B29 for additional details on intensification). It should be noted that changing intensification criteria had a generally modest effect on cost-effectiveness (CS Table 106) and it

can be assumed that modelling discontinuation would similarly have a modest impact on cost-effectiveness provided that the a balanced approach to costs and effects was applied to the rescue medication.

c) According to amongst others Appendix N, Figure 38, treatment discontinuation occurs within the first year. Please clarify how this is possible since the long-term progression of HbA1c only starts after the first year (as also illustrated by CS Figure 70).

Patients intensify therapy in year 1 of the simulation because they have HbA1c levels above 7.5%. This is because baseline HbA1c for each simulated patient is sampled from a normal distribution (based in mean and SD defined by the user and truncated withing physiological limits) and the treatment effects applied in year 1 are also sampled from a normal distribution (based on the mean and SD values defined by the user and truncated within plausible limits). As a result, a proportion of patients will have high HbA1c at baseline and/or have a small HbA1c reduction in year 1, leading to them meeting the criterion for intensification to basal insulin therapy of HbA1c above 7.5%.

d) Please provide more details regarding the long term HbA1c progression in the modelling analysis (based on the UKPDS Outcomes Model 2 equation published by Leal et al. in 2021) and justify the appropriateness of this approach.

As outlined in Table 4-29 of the model technical report (Appendix N3) and in the Leal *et al.* 2021 publication,[59] the HbA1c progression equation takes the form of a panel regression with the coefficients described in Table 19. This HbA1c progression approach was used in the NICE cost-effectiveness evaluation that supported NG28 published in 2022, and represents and update on the UKPDS OM1 approach used in the NICE evaluation in 2015, the TA288 submission on dapagliflozin and the TA336 submission on empagliflozin.[61, 62] The approach was derived from a UK population (UKPDS) and mimics an increasing HbA1c over time that is aligned with disease progression, diminishing beta-cell function and the need for treatment intensification. For a cost-effectiveness analysis, it provides a mechanism for treatment intensification and the diminishing effects of treatment over time (i.e. the differences in mean HbA1c between tirzepatide and comparators decrease over time using this approach). One potential weakness of the approach is that simulated patients have HbA1c levels well above target after many years in a simulation, but as this effect is the same in both treatment arms it has little impact on cost-effectiveness.

Table 19: HbA1c panel regression with coefficients

Coefficient	Mean	Standard error
Constant	1.419	0.041
Female	0.054	0.012
African Caribbean	0.066	0.026
Asian-Indian	0.046	0.020
Value of HbA1c in previous year	0.724	0.005
Ln (year since diagnosis)	0.141	0.007
First recorded value	0.081	0.007

Abbreviations:

e) Please provide a step by step explanation of the HbA1c progression option defined as "PRIME" approach in the JAVA source code, justify its appropriateness and clarify whether this approach is adopted in the CS base-case.

The PRIME approach to the progression of HbA1c over time was not used in the modelling analysis. The approach is designed to mimic a treat-to-target methodology in the modelled population. In summary, HbA1c tracks towards a user specified goal (default 7.5%) over a user-defined number of years (default 7 years). The annual change in HbA1c is then calculated taking into account user-defined HbA1c target adjustments for age and severe hypoglycaemia, and a proportion of patients not trending to target, as follows:

The number of years to target [YTG] is derived:

$$YTG = \max(15 - Year_{current} | 2.0)$$

The target HbA1c [tHB] is modified by the patient age and the number of severe hypoglycemia events in the past year:

$$tHB = 7.0 + \left(0.1 \times PC^{12\,m}_{Severe\ Hypo}\right) + \left(\underset{Age>40}{\text{if}} \left(\frac{PC_{age} - 40}{10} \middle| 0\right)\right)$$

The required change in HbA1c is then given by:

$$\Delta HbA1c = \frac{tHB - PC_{HbA1c}}{YTG}$$

This value is covaried with current total cholesterol, HbA1c, age and the previous change in HbA1c. The value returned is used as the mean for the change in HbA1c sampling distribution (a normal distribution with a standard deviation of 0.05%).

B 20. Please provide scenario analyses using alternative NMA results.

a) Please provide scenario analyses using the sensitivity analyses reported in Appendix D, including the model-based NMA.

b) Please provide scenario analyses using the NMA results provided in response to clarification questions in Section A of this document.

The NMA results used for the base case analysis were chosen because they reflected the most appropriate data for the target population (see CS Section B.2.9). Alternative NMA analyses presented in Appendix D of CS were as follows:

- Consideration of Asian population studies (8 additional studies) – produced similar results to main NMA (and therefore would produce similar cost-effectiveness results in the modeling analysis) and not particularly relevant for the target population in the present study
- Inclusion of Phase 2 study data – produced similar results to the main NMA for HbA1c and BMI (and therefore would produce similar cost-effectiveness results in the modeling analysis)
- Modification of the network definition – produced similar results to the main NMA for HbA1c and BMI (and therefore would produce similar cost-effectiveness results in the modeling analysis)
- Exclusion of studies with insulin glargine trials (9 studies) – produced similar results to the main NMA for HbA1c and BMI (and therefore would produce similar cost-effectiveness results in the modeling analysis)
- Change in analysis time window – produced similar results to the main NMA for HbA1c and BMI (and therefore would produce similar cost-effectiveness results in the modeling analysis)
- Meta-regression approach - – produced similar results to the main NMA for HbA1c and BMI (and therefore would produce similar cost-effectiveness results in the modeling analysis)

To reproduce the base case analysis for all alternative NMA scenarios would mean running an additional 96 modeling simulations, which would represent a significant time investment that was not possible within the timelines for responses. Moreover, the NMA sensitivity analyses were conducted on a limited number of endpoints which does not provide a full set of model inputs. Given that these analyses would provide cost-effectiveness outcomes very closely aligned with the existing base case analysis (based on the data presented in Appendix D, these simulations were not performed. No additional simulations were identified from the responses in Section A with respect to the NMA.

Adverse events

B 21. Priority question. Only nausea is incorporated (hypoglycaemia only for basal insulin therapy) as adverse event.

a) Please justify the current inclusion of adverse events.

As outlined in Section B.2.9 of the CS, GLP-1 RAs are known to be associated with GI AEs, including nausea and vomiting, in the early months of treatment. Nausea rates for tirzepatide and all comparators were derived from the NMA and were assumed to negatively impact quality of life in year 1 of the simulation (CS Section B.3.4.5 and Table 81) in the analysis as this aspect of tolerability may have been a differentiator between different GLP-1 RA agonists in the cost-effectiveness evaluation. The NMA provided separate rates of nausea and vomiting with no information on the combined “nausea and vomiting” endpoint. For the base case analysis, it was

assumed that: 1) the rate of nausea reported from the NMA would be represent the rate of the combined nausea and vomiting endpoint, and 2) a disutility representing the more severe health state of nausea and vomiting would be applied to nausea rates in the analysis (conservative assumption). Sensitivity analysis showed that the impact of nausea utilities on cost-effectiveness outcomes was minimal (CS Table 106). Including both the nausea and vomiting rates from the NMA in the same simulation would have created a risk of double-counting events.

Although the proportion of patients experiencing diarrhoea was included amongst the NMA outputs, this was not included in the base case analysis (as it was expected to be comparable across interventions and have little impact on cost-effectiveness). This adverse event was not included in previous evaluations by NICE (see response to b) below for more details).

As outlined in CS Section B.3.4.4, rates of hypoglycaemia were not reported in the NMA due to many studies reporting zero events; therefore rates of hypoglycaemia were set to zero for tirzepatide and all comparators in the base case analysis. This assumption is likely to be a reasonable approximation for the interventions included in the present analysis based on the very low hypoglycaemia rates observed in the SURPASS trial programme and clinical studies of other T2D medications such as GLP-1 RAs. For basal insulin therapy, hypoglycaemic event rates were aligned with those used in the NICE 2022 health economic report used to inform NG28.

b) Please provide scenario analyses including other relevant adverse events for the intervention and comparators, including hypoglycaemia (see question A44c), gastrointestinal adverse events such as diarrhoea and vomiting.

No additional analyses were performed incorporating hypoglycaemia rates as reported rates from the SURPASS trials were sufficiently low as to not influence cost-effectiveness outcomes. It should be noted that hypoglycaemia rates associated with tirzepatide and semaglutide were included in the SURPASS-2 scenario analysis and had a negligible impact on projected outcomes.

Simulations were run incorporating rates of diarrhoea from the NMA (see results table below) and showed only modest QALY differences from the base case analysis. Literature review failed to identify appropriate utilities for diarrhoea in the target population and therefore the nausea and vomiting utility published by Matza *et al*. and used in the base case analysis was used as a proxy (-0.04 for each patient experiencing diarrhoea). This was applied to the proportion of patients who experienced diarrhoea and to the proportion of patients who experiencing nausea based on the NMA in year 1 of the simulations. The total proportions for each treatment were as follows:

Intervention	Proportion of patients experiencing nausea	Proportion of patients experiencing diarrhoea	Combined proportion to receive -0.04 disutility
Tirzepatide 5 mg	25.8	17.1	42.8
Tirzepatide 10 mg	34.3	19.5	53.8
Tirzepatide 15 mg	37.2	17.7	55.0
Dulaglutide 1.5 mg	28.1	15.1	43.2
Dulaglutide 3.0 mg	28.1*	15.1*	43.2
Dulaglutide 4.5 mg	28.1*	15.1*	43.2
Semaglutide 0.5 mg	24.9	12.3	37.3
Semaglutide 1.0 mg	28.1	14.3	42.4
Oral semaglutide 7 mg	24.9*	12.3*	37.3
Oral semaglutide 14 mg	28.1*	14.3*	42.2
Liraglutide 1.2 mg	20.3	7.7	28.1
Liraglutide 1.8 mg	25.3	12.5	37.8

Any apparent discrepancies in the combined proportion column are due to rounding.

Table 20: Summary of simulation results with diarrhoea disutility included for tirzepatide 5 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg	[redacted]	9.481			
Dulaglutide 1.5 mg	[redacted]	9.342	+923	+0.139	6,626
Semaglutide 0.5 mg	[redacted]	9.369	+1,007	+0.112	9,009
Oral semaglutide 7 mg	[redacted]	9.315	+474	+0.166	2,852
Liraglutide 1.2 mg	[redacted]	9.307	+1,069	+0.174	6,151

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Table 21: Summary of simulation results with diarrhoea disutility included for tirzepatide 10 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 10 mg	[redacted]	9.527			
Dulaglutide 3.0 mg	[redacted]	9.371	+873	+0.155	5,776
Semaglutide 1.0 mg	[redacted]	9.416	+981	+0.111	8,880
Oral semaglutide 14 mg	[redacted]	9.385	+970	+0.142	6,870
Liraglutide 1.8 mg	[redacted]	9.316	−110	+0.211	Dominant

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Table 22: Summary of simulation results with diarrhoea disutility included for tirzepatide 15 mg versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 15 mg	[redacted]	9.574			
Dulaglutide 4.5 mg	[redacted]	9.401	+873	+0.173	5,046
Semaglutide 1.0 mg	[redacted]	9.416	+981	+0.158	6,212
Oral semaglutide 14 mg	[redacted]	9.383	+970	+0.191	5,087
Liraglutide 1.8 mg	[redacted]	9.316	−110	+0.258	Dominant

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Quality of life

B 22. Priority question: Please provide additional information on the health-related quality of life data.

a) Health-related quality-of-life data was sourced from a literature review and displayed in CS Table 82. Multiple studies reporting on utility values are available for most outcomes (as reported in appendix N). Please justify for each outcome why (or why not) a particular study was selected.

Health-related quality-of-life utility data for the modelling analysis were principally selected to be consistent with the 2022 health economic analysis to support NG28, as it was assumed that NICE would consider these estimates to be appropriate for T2D patients in clinical practice in England (and given that the literature review did not provide more suitable estimates for use in the present analysis). This accounted for the utilities associated with all end-stage complications in the base case modelling analysis. Where utilities were not available from the NICE health economic analysis for intermediate endpoints, they were taken from other sources. Preference was given to utilities 1) derived using the EQ-5D instrument and 2) derived from UK populations wherever a choice a was available from the literature review. These utilities are summarized in the following table:

Complication / health state / adverse event	Disutility	Source	Justification
Lower extremity amputation (subsequent years)	−0.122	Nauck *et al.* Diabetes Obes Metab. 2019[63]	EQ-5D derived utility from a population that included UK-based patients
Macular oedema (first year)	−0.047	Mitchell *et al.* Br J Ophthalmol 2012;96: 688-93[64]	EQ-5D health scores were converted into utility scores using preferences from a UK population survey, corresponding to best corrected visual acuity change from 76-85 to 66-75
Neuropathy / SPSL (each year)	−0.066	Shao *et al.* Pharmacoeconomics 2019;37:921-929[65]	Only utility identified matching the neuropathy / SPSL endpoint, derived using HUI-3 instrument
KDIGO CKD eGFR stage 1	0	Assumed	Stage 1 eGFR is essentially asymptomatic
KDIGO CKD eGFR stage 2	0	Assumed	Stage 1 eGFR is essentially asymptomatic
KDIGO CKD eGFR stage 3	−0.004	Nauck *et al.* Diabetes Obes Metab. 2019; 21(3): 525-532[63]	EQ-5D derived utility from a population that included UK-based patients

KDIGO CKD eGFR stage 4	−0.004	Nauck *et al.* Diabetes Obes Metab. 2019; 21(3): 525-532[63]	EQ-5D derived utility from a population that included UK-based patients
Non-severe hypoglycaemic event	−0.005	Evans *et al.* Health Qual Life Outcomes 2013;11:90[66]	Time trade off derived utility from a population that included UK-based patients, frequently used in published cost-effectiveness studies
For each patient experiencing nausea, a disutility of 0.04 was applied in the first year of the simulation	−0.04	Matza *et al.* Qual Life Res 2007;16:1251-65. [67]	Only utility estimate identified by literature review for patients on GLP-1 RAs experiencing nausea and vomiting adverse events
For patients receiving tirzepatide and dulaglutide, a device utility of 0.007 was applied in the first year on treatment in comparisons with other injectables	+0.007	Boye *et al.* J Med Econ 2019;22:806-813.[68]	Only utility estimate available aligned with the observation that tirzepatide will be administered using the same pen device as dulaglutide, which has shown a utility benefit over the semaglutide administration device
To capture the improvement in quality of life associated with bodyweight reductions in the first year of GLP-1 RA therapy	Variable, indexed by weight loss	Boye *et al.* (J Med Econ 2022;25:14-25.[69]	UK-specific utilities for weight change (see response to B 25) in a population with type 2 diabetes and obesity
In years 2+ of the simulations, the impact of bodyweight/BMI on quality of life	-0.0061 for each unit of BMI over 25 kg/m^2	Bagust and Beale in 2005[70]	Utility associated with BMI (or body weight) state previously used in the NICE NG28 modelling analysis
For oral semaglutide, the only non-injectable comparator in the modelling analysis, a utility of +0.004 was applied for each year on therapy (to improve quality of life versus injectable comparators)	+0.004	NICE NG28[71]	Utility was estimated based on the single daily injection utility of 0.029, divided by 7 to compare with weekly injectables, derived from the NICE 2022 health economic report for NG28

b) In case multiple studies were eligible, please provide reasoning for not pooling quality-of-life data.

The heterogeneous nature of the quality-of-life data identified from literature review was the principle reason for not pooling the data (see table below). In addition it was assumed that utility values recently used by NICE, unless more recent, appropriate and robust estimates were identified in literature review, would represent the most appropriate utility values for the present submission.

c) Please provide a table in which all utility values used are summarized including a measure of uncertainty, duration of disutilities, the distributions applied in the model, and the sources. Also include the extra information that is provided below CS Table 83, which refers to

the utilities used for nausea, BMI and use of the device. Currently it is not clear how these relate to the utility values presented in CS Table 82.

The requested table is provided below and the approaches to sampling in the model are summarized in the response to question B31. With respect to CS Tables 82 and 83, patients were assigned utilities (and can be assigned costs) based on renal function status, which is defined by eGFR in the model. Table 82 shows the utilities applied based on renal function (KDIGO stages) and Table 83 shows how simulated patients' eGFR is mapped to the KDIGO stages in the model. With respect to nausea, weight loss, BMI and device utilities, the model features a treatment-related utility function that is editable by the user and can be used to define separate utilities to be applied in year 1 and years 2+ of any given simulation. The treatment related utilities are added to the annual utility score for each patient as calculated based on the inputs in the Utilities element. In the current set of simulations, the treatment-related utility function was used to capture the following utilities:

- Year 1: body weight change utility (no separate BMI utility), device utility and the nausea and vomiting utility
- Years 2+: BMI utility only (no body weight change utility)

Table 23: Utilities and disutilities used in the modelling analysis for diabetes-related complications and hypoglycaemic events

Baseline	Utility	Reported measure of uncertainty	Duration of utility	Distributions	Source
T2D with no complications	+0.815	Not reported	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)[72] Error! Bookmark not defined.
Complication / adverse event	**Disutility**	**Measure of uncertainty**	**Duration of disutility**	**Distributions**	**Source**
Macrovascular complications					
Myocardial infarction event	−0.055	95% CI: −0.067, −0.042	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
History of myocardial infraction	−0.055	95% CI: −0.067, −0.042	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027) Error! Bookmark not defined.
Stroke event	−0.164	95% CI: −0.222, −0.105	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
History of stroke	−0.164	95% CI: −0.222, −0.105	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Ischemic heart disease (each year)	−0.090	95% CI: −0.126, −0.054	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Congestive heart failure (each year)	−0.108	95% CI: −0.169, −0.048	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Microvascular complications					
Foot ulcer (year of event)	−0.170	SE: 0.19	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Lower extremity amputation (year of event)	−0.280	95% CI: −0.389, −0.170	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Lower extremity amputation (subsequent years)	−0.122	SE: 0.011	1 year	Normal (sampled during PSA)	Bagust and Beale (2005)

Baseline	Utility	Reported measure of uncertainty	Duration of utility	Distributions	Source
Blindness (each year)	−0.074	95% CI: −0.124, −0.025	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Macular oedema (first year)	−0.047	SE: visual acuity 66-75: 0.012; 76-85: 0.014	1 year	Normal (sampled during PSA)	Mitchell *et al.* (2012) assumed, corresponding to best corrected visual acuity change from 76-85 to 66-75
Macular oedema (subsequent years)	0	N/A	N/A	Not used	Assumed
Neuropathy / SPSL (each years)	−0.066	SE: 0.007	1 year	Normal (sampled during PSA)	Shao *et al.* (2019)
Renal complications					
KDIGO CKD eGFR stage 1	0	N/A	N/A	Not used	Assumed
KDIGO CKD eGFR stage 2	0	N/A	N/A	Not used	Assumed
KDIGO CKD eGFR stage 3	−0.004	95% CI: −0.024, 0.016	1 year	Normal (sampled during PSA)	Assumed based on Nauck *et al.* (2019)
KDIGO CKD eGFR stage 4	−0.004	95% CI: −0.024, 0.016	1 year	Normal (sampled during PSA)	Assumed based on Nauck *et al.* (2019)
KDIGO CKD eGFR stage 5	−0.164	95% CI: −0.274, −0.054	1 year	Normal (sampled during PSA)	NICE HE Report 2022 (Table HE027)
Adverse events					
Severe hypoglycaemic event	−0.062	95% CI: 0.054, 0.071	Per event utility	Normal (sampled during PSA)	NICE HE Report 2022 (Section 2.3.5)
Non-severe hypoglycaemic event	−0.005	95% CI: 0.004, 0.007	Per event utility	Normal (sampled during PSA)	Evans *et al.* (2013)
Nausea	−0.04	SD: 0.07	1 year	Not sampled	Matza et al. (2007)

Baseline	Utility	Reported measure of uncertainty	Duration of utility	Distributions	Source
Other					
Device utility associated with pen	0.007	SD: 0.009	1 year	Treatment-related utility (not sampled)	Boye et al. (2019)
Oral administration (no injection)	0.004	Not reported	1 year	Treatment-related utility (not sampled)	NICE HE Report 2022 (Section 2.3.5.2)
Bodyweight reduction	Linear interpolation of utilities in Table 84	See Table 84	1 year	Treatment-related utility (not sampled)	Boye et al. (2022)
BMI (per unit over 25 kg/m^2)	−0.0061	SE: 0.001	1 year	Treatment-related utility (not sampled)	Bagust and Beale (2005)

Abbreviations: CI; confidence interval: PSA, probabilistic sensitivity analysis: SE; standard error: SD; standard deviation

B 23. The baseline utility for patients with T2D without complications (0.815) seems to be derived from the study of Alva et al. (2013) which describes quality-of-life in a long-term follow up cohort of the UKPDS and is only slightly lower than the UK general population norm for this age group (0.819 for 55-64 years, Szende et al. 2014).

a) Please compare the utility value to the age and sex adjusted utility value of the general population in the UK and justify why 0.815 is an appropriate utility value for patients with T2D without complications, especially given that as per CS Table 75, the mean age of the cohort is 63.95 years and had a mean duration of diabetes of 8.5 years.

As outlined in the response to question B22, health-related quality-of-life utility data for the modelling analysis were principally selected to be consistent with the 2022 health economic analysis to support NG 28, which also used the Alva *et al.* (2014) utility for diabetes with no complications, which was derived from a UK population using the EQ-5D instrument. The value represents diabetes with no complications in a UK T2D populations and is therefore well suited to modelling studies, where disutilities associated with diabetes-related complications are used to adjust simulated patients quality of life. Literature review did not identify a more robust utility value for diabetes with no complications. It should be noted that the utility is adjusted down based on the history of diabetes-related complications (as well as their incidence during the simulation). Therefore, in a population with long duration of diabetes, one would expect a higher incidence of complications at baseline, which would lower the utility score for simulated patients.

It is also noteworthy that this utility is used equally in both treatment arms in the cost-effectiveness evaluation. Decreasing to adjust for age would have a negligible impact on incremental outcomes and therefore cost-effectiveness (the effect would be limited to the very end of the simulation, at which time discounting minimises the impact of any incremental differences, where small survival benefits for more efficacious interventions would be captured). The minimal impact of age-adjustment on incremental outcomes was demonstrated in the sensitivity analysis (CS Table 106) where age-adjusted utilities were used based on the methodology of Ara and Brazier (2010). See also the response to question B26 with respect to age-adjustment of utilities.

b) Was the option of a utility decrement to adjust for having T2D explored? Please elaborate on why or why not and justify not including a T2D related utility decrement in the CS base-case analyses.

As outlined above, the choice of T2D utility was aligned with the previous NICE approach in this area. Sensitivity analysis showed that adjusting this value had little impact on incremental outcomes. Other approaches to adjust the T2D with no complications utility would produce similar outcomes (the effect would essentially be the same in both treatment arms, leading to no notable difference in incremental outcomes and therefore cost-effectiveness).

B 24. Regarding the utility increments associated with the modes of administration, it seems that drug administration using the tirzepatide and dulaglutide device results in a higher utility than oral administration (tirzepatide and dulaglutide device utility of 0.007 as compared to the utility used for oral semaglutide (0.004)). Please justify why administration using the tirzepatide and dulaglutide device has a greater utility benefit as compared to oral administration. In addition, please explain why this effect (device utility of 0.007) is only present in the first year of treatment.

The administration utilities used were based on published evidence and for the comparisons between injectables can be assumed to be a conservative approach. The device utility of 0.007 associated with tirzepatide and dulaglutide was only applied in year 1 of the simulations as there was no evidence that the perceived benefit of the administration device leads to the same quality of life improvement in every year of use (so it was conservatively assumed only to last for one year). With respect to the comparison with oral semaglutide, the EAG is right to point out the potential shortcomings of this approach. Unfortunately, there is not quality of life data directly comparing administration of oral semaglutide with tirzepatide or dulaglutide to inform the analysis. We have therefore run simulations assuming that there is no device utility associated with tirzepatide in the comparison with oral semaglutide (which has an administration utility of +0.004) and the results are summarized in the table below for all three doses of tirzepatide. The findings show that removing the device related utility for tirzepatide had little impact on overall cost-effectiveness with ICERs between £2,926 and £6,993 per QALY gained for tirzepatide 5 mg versus oral semaglutide 7 mg and tirzepatide 10 mg and oral semaglutide 14 mg, respectively.

It is also worth noting that a sensitivity analysis was performed (CS Table 106) with no device utility associated with tirzepatide and the analysis showed that device utility had only a very

modest impact on incremental QALYs and therefore cost-effectiveness (ICER for tirzepatide 10 mg versus semaglutide 1.0 mg in this sensitivity analysis was GBP 9,270 per QALY gained).

Table 24: Summary of simulation results with no device utility included for tirzepatide 5 mg versus oral semaglutide

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg	[redacted]	9.481			
Oral semaglutide 7 mg	[redacted]	9.319	+474	+0.162	2,926

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Table 25: Summary of simulation results with no device utility included for tirzepatide 10 mg versus oral semaglutide

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 10 mg	[redacted]	9.528			
Oral semaglutide 14 mg	[redacted]	9.388	+977	+0.140	6,993

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Table 26: Summary of simulation results with no device utility included for tirzepatide 15 mg versus oral semaglutide

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 15 mg	[redacted]	9,574			
Oral semaglutide 14 mg	[redacted]	9.388	+970	+0.186	5,224

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

B 25. Priority question: Two measures for impact of weight on quality of life were used: a disutility of -0.0061 for each unit of BMI over 25 and a utility gain when BMI is changed, as displayed in CS Table 84.

a) Please justify why both measures are appropriate and do not result in double counting of utility effects, i.e. when a patient loses weight there is a utility gain for 1) the reduction in BMI units times 0.0061 based on Bagust and Beale 2005, and 2) a change in BMI based on Boye et al. 2022.

The use of two different utilities is based on observations from the literature that there the effects of weight change versus being at a specific body weight or BMI level are different in terms of quality of life (Dennett *et al*. 2008).[73] Therefore in the present analysis a *weight change* utility was applied in year 1 of the simulations (i.e. when the changes in body weight associated with GLP-1 RA therapy were applied in the modelling analysis) and a *BMI level* utility was applied in each subsequent year (i.e. years 2+) (i.e. when body weight was assumed to be stable in the modelling analysis). As the utilities were applied at different times in the modelling analysis (never both in the same year) and applied to different aspects of body weight/BMI (the first for change in body weight and the second for living with a BMI over 25 kg/m^2), there was no double-counting. These utilities were applied in the modelling analysis as described in the response to question B22.

b) Please add a scenario analysis where the utility gain in change in BMI based on Boye et al. 2022 is not included and consider this to be the base case scenario

The results of the requested scenario (with not weight loss utility applied in year 1 of the simulations) are summarized in the following table. As per the sensitivity analysis in CS Table 106, omitting the body weight change utility from the analysis led to only a small decrease in the QALY benefits associated with tirzepatide, leading to only slightly higher ICERs than in the base case.

Table 27: Summary of results with no body weight change utility for tirzepatide versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg versus comparators					
Tirzepatide 5 mg	[redacted]	9.435			
Dulaglutide 1.5 mg	[redacted]	9.311	+923	+0.125	7,401
Semaglutide 0.5 mg	[redacted]	9.333	+1,007	+0.102	9,860
Oral semaglutide 7 mg	[redacted]	9.285	+474	+0.151	3,146
Liraglutide 1.2 mg	[redacted]	9.275	+1,069	+0.160	6,667
Tirzepatide 10 mg versus comparators					
Tirzepatide 10 mg	[redacted]	9.478			
Dulaglutide 3.0 mg	[redacted]	9.336	+898	+0.073	6,345
Semaglutide 1.0 mg	[redacted]	9.374	+988	+0.104	9,350
Oral semaglutide 14 mg	[redacted]	9.346	+977	+0.132	7,410
Liraglutide 1.8 mg	[redacted]	9.238	−103	+0.195	Dominant
Tirzepatide 15 mg versus comparators					
Tirzepatide 15 mg	[redacted]	9.522			
Dulaglutide 4.5 mg	[redacted]	9.363	+873	+0.158	5,512
Semaglutide 1.0 mg	[redacted]	9.374	+981	+0.148	6,634
Oral semaglutide 14 mg	[redacted]	9.346	+970	+0.176	5,510
Liraglutide 1.8 mg	[redacted]	9.283	−110	+0.239	Dominant

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

c) Currently, the utility seems to deteriorate with every point increase in BMI at a constant rate. Please elaborate on why this is and whether or not a plateau effect can be expected at a certain BMI.

A recent review of health-related quality of life and BMI/body weight by Dennett *et al.*[73] did not provide any evidence of a plateau effect. The review described 18 articles investigating either: 1) utility values by body-mass index (BMI) or body weight, or 2) the change in utility scores or quality-adjusted life-years based on unit changes in BMI or body weight. Regardless of the study population or methodology used to elicit utility scores, all studies reviewed found that as body weight increased, patient utility decreased. It is entirely possible that at very high BMI levels there is a plateau effect with respect to quality of life, but it is likely to be levels higher than those relevant to the present health economic analysis. Dennett *et al.* report studies up to the superobese range (50–90 kg/m2) without evidence of a plateauing and baseline BMI in the simulation population is 30.7 kg/m^2 (SD 6.90), suggesting that there would be very few simulated patients in the superobese range in the modelling analysis.

d) "*To capture the improvement in quality of life associated with bodyweight reductions in the first year of GLP-1 RA therapy, utilities from the Boye et al. (2022) study were used.161 Linear interpolation of the utilities summarised in Table 84 were used to evaluate the impact of weight loss in year 1 of the simulation for tirzepatide and comparator treatments*" Please consider the appropriateness of other relationships, like the logarithmic relationship as was presented by Boye et al., and whether these should be used in the base case analysis.

The differences in utilities associated with weight loss using a linear interpolation method and alternate methods of curve fitting (e.g. logarithmic or polynomial) would be very small and would not have a notable impact on the cost-effectiveness evaluation. The utilities were originally estimated based on the congress publication of the Boye *et al.* 2022 data,[69] which included a tabular summary but no regression function (it only became available later with the full manuscript publication). Linear interpolation and polynomial curve fitting were explored to best fit the data, with the former approach selected when both methods produced similar values. Approximate utility values based on the Boye *et al.* (2022) data using three different approaches are summarized in the table below (liner interpolation, polynomial curve fitting and the log-linear regression approach from the full publication). It is perhaps notable that the published log-linear regression function fits the known data points more poorly than the other two approaches when matched to the known data points. Given the modest differences between approaches and in light of the request to run simulations without any weight change utility in year 1 (point b, above), no further simulations have been run to investigate the impact of using utilities derived in other ways.

Table 28: Utility estimates associated with body weight changes derived from Boye et al. (2022) using different approaches

Intervention	Weight loss	Linear interpolation utility	Polynomial curve utility	Log-linear regression utility*
Tirzepatide 5 mg	8.0%	0.034	0.035	0.036
Tirzepatide 10 mg	11.1%	0.044	0.044	0.044
Tirzepatide 15 mg	13.4%	0.050	0.051	0.048
Dulaglutide 1.5 mg	2.6%	0.011	0.011	0.012
Dulaglutide 3.0 mg	3.5%	0.016	0.016	0.019
Dulaglutide 4.5 mg	4.0%	0.018	0.018	0.021
Semaglutide 0.5 mg	3.8%	0.017	0.017	0.020
Semaglutide 1.0 mg	5.8%	0.026	0.026	0.030
Oral semaglutide 7 mg	3.0%	0.013	0.014	0.015
Oral semaglutide 14 mg	4.5%	0.021	0.021	0.024
Liraglutide 1.2 mg	2.7%	0.012	0.012	0.013
Liraglutide 1.8 mg	3.3%	0.015	0.016	0.017

* As published by Boye *et al.* (2022) for patients with diabetes
y = -0.00860970799921508 + 0.0216883437387603*log(x)

B 26. "No age-adjustment was used in the base case analysis; the inclusion of age-adjustment was explored in sensitivity analyses using the methodology of Ara and Brazier (2010) and was found to have little impact." Although age-adjustment has little impact on cost-effectiveness, NICE recommends adjusting baseline utility values for age when they are extrapolated over long time horizons. With the current assumption it is likely that utility values for older patients with T2D will exceed the utility values for the general population as reported by Szende et al. 2014. Please reconsider including age-adjustment in the base case analysis or justify why age-adjustment was not used in the base case.

The approach to the estimation of quality-adjusted life expectancy in the present analysis was analogous to the approach used by NICE in the health economic evaluation to support NG 28 in 2022. When using utility scores derived from T2D populations and subsequently adjusting for age, there is a risk of double-counting the effect of age on quality of life, as the unadjusted utilities already reflect the impact of complications on an aging population (as T2D populations are, by definition, relatively old). To quote from the NICE Health Economic Model report directly (Section 2.3.5.1):

"…given that the baseline population utility was sourced from a type 2 diabetes population, the changes in utility with age have been partially accounted for. Furthermore, accounting for changes in utility with increasing age is unlikely to have a significant impact on the treatment decision given that this would apply across all treatment arms and would only have an impact if

there were substantial differences between treatments in the time spent living in the model, which is not the case for all analyses."

For these reasons, age-adjustment was not included in the base case for the present analysis. In line with the EAG request, however, the results of an age-adjusted base case analysis are summarized in the following table. As per the sensitivity analysis in CS Table 106, using age-adjusted utilities in the analysis led to only a small decrease in the QALY benefits associated with tirzepatide, leading to only slightly higher ICERs than in the base case.

Table 29: Summary of results for tirzepatide versus comparators with age-adjustment applied to utilities

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg versus comparators					
Tirzepatide 5 mg	[redacted]	8.786			
Dulaglutide 1.5 mg	[redacted]	8.653	+923	+0.133	6,925
Semaglutide 0.5 mg	[redacted]	8.678	+1,007	+0.108	9,319
Oral semaglutide 7 mg	[redacted]	8.627	+474	+0.159	2,974
Liraglutide 1.2 mg	[redacted]	8.619	+1,069	+0.167	6,386
Tirzepatide 10 mg versus comparators					
Tirzepatide 10 mg	[redacted]	8.832			
Dulaglutide 3.0 mg	[redacted]	8.682	+898	+0.151	5,965
Semaglutide 1.0 mg	[redacted]	8.724	+988	+0.109	9,101
Oral semaglutide 14 mg	[redacted]	8.692	+977	+0.141	6,933
Liraglutide 1.8 mg	[redacted]	8.629	−103	+0.204	Dominant
Tirzepatide 15 mg versus comparators					
Tirzepatide 15 mg	[redacted]	8.875			
Dulaglutide 4.5 mg	[redacted]	8.71	+873	+0.166	5,273
Semaglutide 1.0 mg	[redacted]	8.724	+981	+0.151	6,486
Oral semaglutide 14 mg	[redacted]	8.692	+970	+0.184	5,283
Liraglutide 1.8 mg	[redacted]	8.629	−110	+0.246	Dominant

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

B 27. In appendix N.5.9 it is described that *"Quality-adjusted life expectancy is evaluated in the model using an additive approach."* There is currently no consensus about which method (additive, multiplicative or minimum) is best, but Ara and Brazier (2011) state that the multiplicative method appears to be most accurate overall. The method used has influence on the cost-effectiveness of Tizerpatide. Please reconsider the multiplicative method for the base case or justify the use of an additive method of combining disutility values in the base case analysis.

An additive approach was adopted for the present analysis as this is best aligned with utility and disutility input data used in the modelling analysis. Specifically, disutility data used were reported as a decrease in absolute utility associated with the presence of a diabetes-related complication, not a relative decrease as would better suit a multiplicative approach. As previously mentioned, the approach to the estimation of quality-adjusted life expectancy in the present analysis was analogous to the approach used by NICE in the health economic evaluation to support NG 28 in 2022. As pointed out by the EAG, there is no consensus on the best approach to combining utilities estimates when direct estimates are not available for the presence of multiple disease conditions. The results of the requested analysis (using a multiplicative approach to combining utilities) is provided in the following table:

Table 30: Summary of results for tirzepatide versus comparators with a multiplicative approach to combining quality of life utilities

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg versus comparators					
Tirzepatide 5 mg	[redacted]	9.397			
Dulaglutide 1.5 mg	[redacted]	9.292	+923	+0.106	8,736
Semaglutide 0.5 mg	[redacted]	9.313	+1,007	+0.084	11,925
Oral semaglutide 7 mg	[redacted]	9.272	+474	+0.125	3,793
Liraglutide 1.2 mg	[redacted]	9.263	+1,069	+0.134	7,952
Tirzepatide 10 mg versus comparators					
Tirzepatide 10 mg	[redacted]	9.428			
Dulaglutide 3.0 mg	[redacted]	9.311	+898	+0.117	7,699
Semaglutide 1.0 mg	[redacted]	9.345	+988	+0.083	11,975
Oral semaglutide 14 mg	[redacted]	9.325	+977	+0.103	9,485
Liraglutide 1.8 mg	[redacted]	9.269	−103	+0.159	Dominant
Tirzepatide 15 mg versus comparators					
Tirzepatide 15 mg	[redacted]	9.464			
Dulaglutide 4.5 mg	[redacted]	9.335	+873	+0.129	6,765
Semaglutide 1.0 mg	[redacted]	9.345	+981	+0.118	8,285
Oral semaglutide 14 mg	[redacted]	9.325	+970	+0.139	6,983
Liraglutide 1.8 mg	[redacted]	9.269	−110	+0.195	Dominant

Abbreviations: QALY: quality-adjusted life year; * for tirzepatide versus comparator. Only quality-adjusted life expectancy and costs are shown as only utility values were modified from the base case simulations.

Costs and resource use

B 28. Please clarify the following regarding costs and resource use:

a) In CS Table 87 the costs for complications and adverse events are only present in the first two years after the event. However, it seems reasonable to the EAG that yearly costs over lifetime may apply for some complications, such as blindness. Please justify this choice and include scenario analyses including lifetime costs for individual complications and complications combined.

The labelling in CS Table 87 is a little misleading inasmuch as the “year 2” costs would be better labelled “years 2+” in line with the PRIME T2D Model interface. “Year 1” costs are applied in the year the event occurs in the simulation and “years 2+” costs (represented as “year 2” costs in CS Table 87) are applied in every subsequent year of the simulation (when the patient is alive with a history of the event). This is the standard approach used in the model (see model technical report, Appendix N3) to capture lifetime costs associated with complications (in line with the EAG’s expectations on this point).

b) Justify why there are no specific T2D health state costs included? For example, to account for standard GP/hospital checkups as mentioned in appendix I.4.1 Table 90? Please also provide a scenario analysis including these costs.

The approach to cost estimation is aligned with previous publications in this area and approach recently used by NICE in the preparation of NG28. No specific health state costs for T2D, or other costs that would be the same across all treatment arms were included. The rationale for this was simply that inclusion of any such costs would not have a significant impact on the treatment decision, given that this would apply across all treatment arms and could only have an impact if there were substantial differences between treatments in the time spent living in the model, which is not the case in the present analysis. In line with this logic, no additional cost scenario simulations have been run as changing the annual costs associated with T2D management would not impact cost-effectiveness.

c) Please justify the exclusion of diet and exercise costs.

Please see response to b) above.

d) Please provide scenario analyses including diet and exercise costs

Please see response to b) above.

Cost effectiveness results and sensitivity analysis

B 29. Priority question: Please provide additional information on the cost effectiveness results.

a) Only pairwise cost effectiveness results are provided. Please also provide fully incremental analysis as well as net health benefits for each analysis (dosage).

A fully incremental set of cost-effectiveness results with net health benefit results (assuming a willingness to pay of £20,000 per QALY gained) is presented in the tables below for each dose of tirzepatide and comparators. The summary is based on the data provided in CS Tables 95-97 (and therefore ICERs in non-tirzepatide comparison may be subject to rounding errors). In the submission, each dose of tirzepatide was compared with relevant comparators and, based on these results, a cost-effectiveness frontier was presented. The goal of the cost-effectiveness frontier figures was to highlight the most cost-effective comparator(s) for tirzepatide. In Figures 83 and 84 in the CS, it can be seen that semaglutide 1.0 mg is associated with greater effectiveness and lower costs than any of the other comparators and represents the most appropriate comparator for tirzepatide 10 and 15 mg. In Figure 82, liraglutide 1.2 mg and semaglutide 0.5 mg are associated with similar costs, but semaglutide 0.5 mg has much better effectiveness making it the most appropriate comparator for tirzepatide 0.5 mg. These comparisons were the focus of the cost-effectiveness evaluation (as a fully incremental analysis would provide little additional insight into the cost-effectiveness of tirzepatide).

b) Please provide information on time to treatment discontinuation in table overviews, including average time on treatment per comparator and proportion of patients on treatment over time (based on the Figures provided in Appendix J).

Please find below tables summarizing the proportion of patients on treatment over the first 10 years of each of the base case simulations as requested. The model does not mean times on treatment but plotting the points from these tables allows a simple estimation of the time until 50% of simulated patients have intensified. The estimates are as follows: TZP 5 mg (3.6 years), DULA 1.5 mg (2.9 years), SEMA 0.5 mg (3.1 years), ORAL SEMA 7 mg (2.5 years), LIRA 1.2 mg (2.6 years), TZP 10 mg (3.8 years), DULA 3 mg (3.2 years), SEMA 1.0 mg (3.4 years), ORAL SEMA 14 mg (2.9 years), LIRA 1.8 mg (2.7 years), TZP 15 mg (4.0 years) and DULA 4.5 mg (3.4 years).

c) Please provide information on disaggregated QALYs and LYs in table overviews, to be able to assess the drivers of benefit.

All of the simulations provided to the EAG in the model have a breakdown of average QoL decrements per patient over time. A summary table is provided below to outline the QoL decrements by category at a 50-year time horizon for each comparator in the base case simulations. A tabular breakdown over time is impracticable (as each simulation produced 600 data points), hence the inclusion of interactive figures in the model interface. The request for incremental life expectancy is an unusual one in diabetes modelling as clear delineation of the cause of death is practically impossible given the risk equations available to date. This breakdown is not provided as part of the present analysis (nor would it be available from any of the other published T2D model that we're aware of from the Mount Hood Diabetes Challenge and previous experience in this area) (Si *et al.* 2020)[49]

Table 31: Summary of fully incremental base case results for tirzepatide 5 mg versus comparators

Intervention	Semaglutide 0.5 mg				Dulaglutide 1.5 mg				Oral semaglutide 7 mg				Liraglutide 1.2 mg			
	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB
Tirzepatide 5 mg	[redacted]	0.114	8,839	0.064	[redacted]	0.14	6,571	0.094	[redacted]	0.169	2,808	0.145	[redacted]	0.178	6,012	0.125
Semaglutide 0.5 mg					[redacted]	0.027	DOM	0.031	[redacted]	0.055	DOM	0.082	[redacted]	0.064	984	0.061
Dulaglutide 1.5 mg									[redacted]	0.028	DOM	0.050	[redacted]	0.037	3,946	0.030
Oral semaglutide 7 mg													[redacted]	0.009	66,111	-0.021
Liraglutide 1.2 mg																

Abbreviations: Δ: incremental; QALY: quality-adjusted life year; ICER, incremental cost-effectiveness ratio in £ per QALY gained; DOM: dominant (no ICER calculated); NHB: net health benefit in QALYs (assuming a willingness to pay of £20,000 per QALY gained). Note: ICERs for non-tirzepatide comparison are estimated based on the data from CS tables 95-91 and may be subject to roundsnipping errors.

Table 32: Summary of fully incremental base case results for tirzepatide 10 mg versus comparators

Intervention	Semaglutide 1.0 mg				Oral semaglutide 14 mg				Dulaglutide 3.0 mg				Liraglutide 1.8 mg			
	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB
Tirzepatide 10 mg	[redacted]	0.114	8,707	0.065	[redacted]	0.147	6,665	0.098	[redacted]	0.157	5,702	0.112	[redacted]	0.214	DOM	0.219
Semaglutide 1.0 mg					[redacted]	0.033	DOM	0.034	[redacted]	0.044	DOM	0.049	[redacted]	0.101	DOM	0.156
Oral semaglutide 14 mg									[redacted]	0.011	DOM	0.016	[redacted]	0.068	DOM	0.122
Dulaglutide 3 mg													[redacted]	0.057	DOM	0.106
Liraglutide 1.8 mg																

Abbreviations: Δ: incremental; QALY: quality-adjusted life year; ICER, incremental cost-effectiveness ratio in £ per QALY gained; DOM: dominant (no ICER calculated); NHB: net health benefit in QALYs (assuming a willingness to pay of £20,000 per QALY gained). Note: ICERs for non-tirzepatide comparison are estimated based on the data from CS tables 95-91 and may be subject to rounding errors.

Table 33: Summary of fully incremental base case results for tirzepatide 15 mg versus comparators

Intervention	Semaglutide 1.0 mg				Oral semaglutide 14 mg				Dulaglutide 4.5 mg				Liraglutide 1.8 mg			
	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB	Δ Cost	Δ QALYs	ICER	NHB
Tirzepatide 15 mg	[redacted]	0.16	8,707	0.111	[redacted]	0.193	6,665	0.145	[redacted]	0.174	5,702	0.130	[redacted]	0.26	DOM	0.265
Semaglutide 1.0 mg					[redacted]	0.033	DOM	0.034	[redacted]	0.015	DOM	0.020	[redacted]	0.101	DOM	0.156
Oral semaglutide 14 mg									[redacted]	-0.018	DOM	-0.013	[redacted]	0.068	DOM	0.122
Dulaglutide 4.5 mg													[redacted]	0.086	DOM	0.135
Liraglutide 1.8 mg																

Abbreviations: Δ: incremental; QALY: quality-adjusted life year; ICER, incremental cost-effectiveness ratio in £ per QALY gained; DOM: dominant (no ICER calculated); NHB: net health benefit in QALYs (assuming a willingness to pay of £20,000 per QALY gained). Note: ICERs for non-tirzepatide comparison are estimated based on the data from CS tables 95-91 and may be subject to rounding errors.

Table 34: Percentage of patients by treatment step for tirzepatide 5 mg and comparators

Year of simulation	Tirzepatide 5 mg		Dulaglutide 1.5 mg		Semaglutide 0.5 mg		Oral semaglutide 7 mg		Liraglutide 1.2 mg	
	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin
1	96.9%	3.1%	88.9%	11.1%	90.4%	9.6%	80.3%	19.7%	83.1%	16.9%
2	87.8%	12.2%	72.6%	27.4%	75.0%	25.0%	61.4%	38.6%	64.5%	35.5%
3	66.7%	33.3%	48.2%	51.8%	50.7%	49.3%	38.1%	61.9%	40.6%	59.4%
4	38.7%	61.3%	25.0%	75.0%	26.6%	73.4%	18.9%	81.1%	20.4%	79.6%
5	17.2%	82.8%	10.5%	89.5%	11.2%	88.8%	7.8%	92.2%	8.5%	91.5%
6	6.3%	93.7%	3.8%	96.2%	4.1%	95.9%	2.9%	97.1%	3.1%	96.9%
7	2.0%	98.0%	1.3%	98.7%	1.3%	98.7%	1.0%	99.0%	1.0%	99.0%
8	0.6%	99.4%	0.4%	99.6%	0.4%	99.6%	0.3%	99.7%	0.3%	99.7%
9	0.2%	99.8%	0.1%	99.9%	0.1%	99.9%	0.1%	99.9%	0.1%	99.9%
10	0.06%	99.94%	0.05%	99.95%	0.05%	99.95%	0.04%	99.96%	0.04%	99.96%

Percentages reflect the proportion of patients pre- and post-intensification in each of the first 10 years of the base case simulations.

Table 35: Percentage of patients by treatment step for tirzepatide 10 mg and comparators

Year of simulation	Tirzepatide 10 mg		Dulaglutide 3.0 mg		Semaglutide 1.0 mg		Oral semaglutide 14 mg		Liraglutide 1.8 mg	
	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin
1	98.5%	1.5%	92.7%	7.3%	94.4%	5.6%	89.2%	10.8%	85.7%	14.3%
2	92.4%	7.6%	79.1%	20.9%	82.6%	17.4%	73.2%	26.8%	68.0%	32.0%
3	74.4%	25.6%	55.5%	44.5%	59.5%	40.5%	48.9%	51.1%	43.7%	56.3%
4	45.8%	54.2%	30.0%	70.0%	33.0%	67.0%	25.5%	74.5%	22.2%	77.8%
5	21.1%	78.9%	12.8%	87.2%	14.3%	85.7%	10.8%	89.2%	9.4%	90.6%
6	7.8%	92.2%	4.7%	95.3%	5.2%	94.8%	3.9%	96.1%	3.4%	96.6%
7	2.5%	97.5%	1.5%	98.5%	1.7%	98.3%	1.3%	98.7%	1.1%	98.9%
8	0.7%	99.3%	0.5%	99.5%	0.5%	99.5%	0.4%	99.6%	0.4%	99.6%
9	0.2%	99.8%	0.1%	99.9%	0.2%	99.8%	0.1%	99.9%	0.1%	99.9%
10	0.07%	99.93%	0.05%	99.95%	0.06%	99.94%	0.04%	99.96%	0.04%	99.96%

Percentages reflect the proportion of patients pre- and post-intensification in each of the first 10 years of the base case simulations.

Table 36: Percentage of patients by treatment step for tirzepatide 15 mg and comparators

Year of simulation	Tirzepatide 15 mg		Dulaglutide 4.5 mg		Semaglutide 1.0 mg		Oral semaglutide 14 mg		Liraglutide 1.8 mg	
	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin	Active treatment	Basal insulin
1	99.1%	0.9%	94.6%	5.4%	94.4%	5.6%	89.2%	10.8%	85.7%	14.3%
2	94.5%	5.5%	82.8%	17.2%	82.6%	17.4%	73.2%	26.8%	68.0%	32.0%
3	78.7%	21.3%	59.8%	40.2%	59.5%	40.5%	48.9%	51.1%	43.7%	56.3%
4	50.4%	49.6%	33.2%	66.8%	33.0%	67.0%	25.5%	74.5%	22.2%	77.8%
5	23.8%	76.2%	14.4%	85.6%	14.3%	85.7%	10.8%	89.2%	9.4%	90.6%
6	8.8%	91.2%	5.2%	94.8%	5.2%	94.8%	3.9%	96.1%	3.4%	96.6%
7	2.8%	97.2%	1.7%	98.3%	1.7%	98.3%	1.3%	98.7%	1.1%	98.9%
8	0.8%	99.2%	0.5%	99.5%	0.5%	99.5%	0.4%	99.6%	0.4%	99.6%
9	0.2%	99.8%	0.2%	99.8%	0.2%	99.8%	0.1%	99.9%	0.1%	99.9%
10	0.08%	99.92%	0.05%	99.95%	0.06%	99.94%	0.04%	99.96%	0.04%	99.96%

Percentages reflect the proportion of patients pre- and post-intensification in each of the first 10 years of the base case simulations.

Table 37: Average QoL utility decrement breakdown per patient by comparator at the end of each base case simulation (50-year time horizon)

Intervention	Treatment-related*	Cardiovascular complications	Renal disease	Neuropathy and diabetic foot complications	Ocular complications	Hypoglycaemia
Tirzepatide 5 mg	-0.384	-0.362	-0.018	-0.369	-0.046	-0.358
Tirzepatide 10 mg	-0.360	-0.357	-0.018	-0.366	-0.045	-0.350
Tirzepatide 15 mg	-0.343	-0.354	-0.017	-0.364	-0.045	-0.347
Dulaglutide 1.5 mg	-0.430	-0.368	-0.018	-0.375	-0.047	-0.376
Dulaglutide 3.0 mg	-0.421	-0.366	-0.018	-0.373	-0.046	-0.368
Dulaglutide 4.5 mg	-0.416	-0.366	-0.018	-0.372	-0.046	-0.364
Semaglutide 0.5 mg	-0.426	-0.367	-0.018	-0.374	-0.047	-0.373
Semaglutide 1.0 mg	-0.409	-0.364	-0.017	-0.370	-0.046	-0.364
Oral semaglutide 7 mg	-0.429	-0.368	-0.018	-0.377	-0.047	-0.388
Oral semaglutide 14 mg	-0.413	-0.366	-0.018	-0.374	-0.047	-0.376
Liraglutide 1.2 mg	-0.435	-0.368	-0.018	-0.376	-0.047	-0.384
Liraglutide 1.8 mg	-0.431	-0.368	-0.018	-0.376	-0.047	-0.380

* Treatment-related utility decrements include utilities for weight year in year 1, BMI state in years 2+, utilities associated with administration, and disutilities associated with nausea and vomiting.

d) Please also provide all results of all presented analyses with all discounted comparator costs split into treatment cost and other costs (to enable the calculation of ICERs with potential comparator price discounts).

A table is provided below summarizing the discounted treatment costs and other costs as requested for the base case analysis. It should be noted that treatment costs from the model include the intervention/comparator costs, the costs of background therapy and the cost of basal insulin (after intensification) as outlined in Table 85 of the CS. Time on therapy is an important driver for total treatment costs in the simulations. To assess the impact of cost reductions on comparators, annual treatment costs should be recalculated (as per the method in Table 85 of the CS) to run new simulations in the model.

Table 38: Average treatment cost per patient by comparator from the base case simulations (50-year time horizon)

Intervention	Treatment costs (£)	Complication and adverse event costs (£)	Total costs (£)
Tirzepatide 5 mg	[redacted]	24,635	32,876
Tirzepatide 10 mg	[redacted]	24,343	32,811
Tirzepatide 15 mg	[redacted]	24,200	32,804
Dulaglutide 1.5 mg	[redacted]	25,088	31,953
Dulaglutide 3.0 mg	[redacted]	24,902	31,913
Dulaglutide 4.5 mg	[redacted]	24,826	31,931
Semaglutide 0.5 mg	[redacted]	24,946	31,870
Semaglutide 1.0 mg	[redacted]	24,720	31,822
Oral semaglutide 7 mg	[redacted]	25,095	32,402
Oral semaglutide 14 mg	[redacted]	24,945	31,833
Liraglutide 1.2 mg	[redacted]	25,079	31,807
Liraglutide 1.8 mg	[redacted]	25,057	32,913

* Treatment-related utility decrements include utilities for weight year in year 1, BMI state in years 2+, utilities associated with administration, and disutilities associated with nausea and vomiting.

e) We note that the extrapolation period is influential. Please provide a comparison of the proportion of the modelled observed benefit within the observed period (1 year) versus beyond the observed period, by filling in this template for relevant benefits (as many with meaningful incremental differences as needed) and all comparators.

	Observed period	Modelled period	
	Outcome after first year	Outcome with modelled lifetime horizon	Proportion beyond observed data
Life years (undiscounted)			
Comparator (multiple rows needed)			
Tirzepatide			
Increment			
QALYs (undiscounted)			
Comparator (multiple rows needed)			
Tirzepatide			
Increment			

The requested tables of year 1 and long-term outcomes are provided below for tirzepatide 5 mg, 10 mg and 15 mg and the corresponding comparators. It is noteworthy that the improvements in risk factors such as HbA1c are known to reduce the risk of long-term, end-stage complication that rarely occur at short time horizons. For this reason, a long-term time horizon is required to adequately capture the benefits associated with diabetes interventions that improve glycaemic control and other risk factors, as reflected in published guideline for diabetes modelling and in NICE modelling evaluations to support the development of guideline NG28.[71]

Table 39: Undiscounted life expectancy and quality-adjusted life expectancy after year 1 and after patients' lifetimes from the base case analysis of tirzepatide 5 mg versus comparators

	Year 1	Lifetime	Modelled (lifetime–year 1)	Incremental in year 1	Incremental over lifetime	Modelled incremental (lifetime–year1)
Undiscounted life expectancy (years)						
Tirzepatide 5 mg	0.98	19.32	18.34			
Semaglutide 0.5 mg	0.98	19.22	18.24	0	0.095	0.095
Dulaglutide 1.5 mg	0.98	19.19	18.21	0	0.126	0.126
Oral semaglutide 7 mg	0.98	19.15	18.17	0	0.170	0.170
Liraglutide 1.2 mg	0.98	19.13	18.15	0	0.185	0.185
Undiscounted quality-adjusted life expectancy (QALYs)						
Tirzepatide 5 mg	0.812	13.800	12.988			
Semaglutide 0.5 mg	0.786	13.651	12.865	0.026	0.149	0.123
Dulaglutide 1.5 mg	0.785	13.614	12.829	0.027	0.186	0.159
Oral semaglutide 7 mg	0.781	13.572	12.791	0.031	0.228	0.197
Liraglutide 1.2 mg	0.780	13.558	12.778	0.032	0.242	0.210

Incremental values show tirzepatide value minus the comparator value for a given outcome. QALYs; quality-adjusted life years

Table 40: Undiscounted life expectancy and quality-adjusted life expectancy after year 1 and after patients' lifetimes from the base case analysis of tirzepatide 10 mg versus comparators

	Year 1	Lifetime	Modelled (lifetime–year 1)	Incremental in year 1	Incremental over lifetime	Modelled incremental (lifetime–year1)
Undiscounted life expectancy (years)						
Tirzepatide 10 mg	0.985	19.316	18.331			
Semaglutide 1.0 mg	0.984	19.221	18.237	0.001	0.095	0.094
Oral semaglutide 14 mg	0.984	19.190	18.206	0.001	0.126	0.125
Dulaglutide 3 mg	0.984	19.146	18.162	0.001	0.170	0.169
Liraglutide 1.8 mg	0.984	19.131	18.147	0.001	0.185	0.184
Undiscounted quality-adjusted life expectancy (QALYs)						
Tirzepatide 10 mg	0.812	13.800	12.988			
Semaglutide 1.0 mg	0.786	13.651	12.865	0.026	0.149	0.123
Oral semaglutide 14 mg	0.785	13.614	12.829	0.027	0.186	0.159
Dulaglutide 3 mg	0.781	13.572	12.791	0.031	0.228	0.197
Liraglutide 1.8 mg	0.780	13.558	12.778	0.032	0.242	0.210

Incremental values show tirzepatide value minus the comparator value for a given outcome. QALYs; quality-adjusted life years

Table 41: Undiscounted life expectancy and quality-adjusted life expectancy after year 1 and after patients' lifetimes from the base case analysis of tirzepatide 15 mg versus comparators

	Year 1	Lifetime	Modelled (lifetime–year 1)	Incremental in year 1	Incremental over lifetime	Modelled incremental (lifetime–year1)
Undiscounted life expectancy (years)						
Tirzepatide 15 mg	0.985	19.386	18.401			
Semaglutide 1.0 mg	0.985	19.241	18.256	0.000	0.145	0.145
Oral semaglutide 14 mg	0.984	19.226	18.242	0.000	0.159	0.159
Dulaglutide4.5 mg	0.984	19.248	18.264	0.000	0.137	0.137
Liraglutide 1.8 mg	0.984	19.139	18.155	0.001	0.247	0.246
Undiscounted quality-adjusted life expectancy (QALYs)						
Tirzepatide 15 mg	0.825	13.928	13.103			
Semaglutide 1.0 mg	0.795	13.706	12.911	0.030	0.221	0.191
Oral semaglutide 14 mg	0.791	13.668	12.877	0.034	0.259	0.225
Dulaglutide 4.5 mg	0.794	13.697	12.903	0.031	0.231	0.200
Liraglutide 1.8 mg	0.782	13.574	12.792	0.044	0.354	0.310

Incremental values show tirzepatide value minus the comparator value for a given outcome. QALYs; quality-adjusted life years

B 30. Further sensitivity analyses / clarification on existing sensitivity analyses would be desirable.

a) Sensitivity and scenario analyses were only provided for the semaglutide comparison but should be provided for all comparisons. Please provide scenario analyses based on the fully incremental analysis (for all other comparators).

The request for full sensitivity analysis is impracticable within the time frame permitted for response. The original submission contained the results of 73 simulations, of which 61 were sensitivity analyses. The request for sensitivity analysis for all comparators would involve another 300 simulations, in addition to 75 simulations already performed in response to other questions. Moreover, for the reasons outlined in the response to question B29, sensitivity analysis for all comparators would provide little or no additional data that will help answer the decision question (c.f. comments on the cost-effectiveness frontier and semaglutide being less costly and more effective than other comparators). For these reasons, the full sensitivity analysis for all comparators requested above was not performed.

b) Please provide sensitivity analysis for all input parameters individually and present results in tornado diagrams.

Similarly, the request to provide sensitivity analysis for all input parameters is impracticable. A standard simulation has over 185 input parameters (not including life tables). To do this for all comparators would be approximately 2,200 simulations. All key model inputs that have an influence on cost-effectiveness were explored in sensitivity analysis in the CS (Table 106). An exhaustive analysis in line with the request will not provide additional useful information with respect to the decision question. In response the EAG request, tornado diagrams have been provided for the sensitivity analyses included in the CS (see figures below).

c) Some scenario analyses are only provided for the 10mg tirzepatide dose (SURPASS-2 model inputs and intensification of therapy by adding basal insulin). Please comment on whether the results of these analyses are generalisable to the other dosages.

The sensitivity analyses provided on the 10mg tirzepatide dose (SURPASS-2 model inputs and intensification of therapy by adding basal insulin) are considered generalizable to the other dosages. Similar patterns of results would be observed with respect to cost-effectiveness for analogous simulations with other tirzepatide doses.

d) Please comment on the plausibility of assuming only HbA1c and/or BMI differences between treatments (scenario analysis in CS Table 106).

The sensitivity analysis on clinical drivers showed that HbA1c and BMI were the most important risk factor changes in terms of the cost-effectiveness of tirzepatide relative to semaglutide. The scenario assuming only HbA1c and BMI changes together was designed to show that most of the benefits associated with tirzepatide in the base case analysis was driven by these two risk factors (i.e. blood pressure and serum lipid levels were less important). Data from the SURPASS

trial program shows that tirzepatide was associated with significant improvements in HbA1c and body weight versus all comparators tested to date, strengthening the contention that the clinical benefits associated with tirzepatide could lead to cost-effective outcomes.

e) Please comment on the appropriateness of an intensification threshold of 7.5%, rather than for instance 8.5% or 9.5% (scenario analyses in CS Table 106).

The intensification threshold of 7.5% was chosen for the base case analysis in line with NICE recommendations as well as the approach used by NICE in the recent economic evaluation to support NG28. Data from the National Diabetes Audit has shown that approximately one-third of people with T2D in England and Wales had an HbA1c above 7.5% (58 mmol/mol) in 2020-21.[74] This suggests that the majority of patients (two-thirds) are being managed in a way that is broadly consistent with the published guidance and, therefore, higher intensification thresholds of 8.5% or 9.5% may not be aligned with the majority of the T2D population. Using UKPDS OM2 HbA1c progression mean that, when higher intensification thresholds were assumed, patients were on the intervention/comparators for much longer. For example, in the 8.5% intensification scenario, it took approximately 8 years for 50% of patients to have intensified therapy from tirzepatide 10 mg. The corresponding value in the semaglutide 1.0 mg treatment arm was approximately 7.4 years. It is difficult to know if these longer duration of therapy estimates are realistic for tirzepatide or semaglutide based on currently available evidence. Discontinuation rates from SURPASS-2 were low (all below 8%) and published evidence on the durability of other GLP-1 receptor agonist treatment effects is positive (Courtney *et al.* 2017).[75] However, evidence on older GLP-1 receptor agonists suggests that persistence maybe lower in general practice (Wilke *et al.* 2016).[76] Based on this uncertainty, we would endorse the approach (7.5% intensification) used in the base case analysis and use higher thresholds only for exploratory analysis.

f) Please comment on the plausibility of treatment intensification for all patients after 3 or 5 years (scenario analyses in CS Table 106).

The scenarios with treatment intensification at 3 years or 5 years were run primarily to provide a more transparent cost-effectiveness analysis (i.e. treatment costs and effects for intervention and comparator are applied for an equivalent time period, after which there are no differences between treatment arms, allowing a balanced and clear evaluation of additional costs and additional benefits). The shortcoming of these scenarios is that they are likely to represent a less realistic interpretation of the management of T2D patients in routine clinical practice (relative to the base case). The 3-year intensification scenario provide an approximation of the average duration of therapy across comparators (see response to question B29) but the 5-year scenario assumes a longer duration of therapy.

g) Please explain why, in the intensification of therapy by adding basal insulin scenario the incremental costs are reduced (and not increased) compared to the base-case (CS Table 107).

As intensification of therapy is triggered by HbA1c (over 7.5%), greater HbA1c improvements with tirzepatide delay intensification relative to semaglutide therapy (also true of other comparators). When a second intensification step is included, the second intensification also occurs later in the tirzepatide arm than in the comparator arm. As basal-bolus therapy is more

costly than basal insulin alone, later intensification in the tirzepatide treatment arm leads to lower incremental costs in this scenario relative to the base case.

Figure 20: Tornado diagram of general sensitivity analysis for tirzepatide 5 mg versus semaglutide 0.5 mg

Figure 21: Tornado diagram of general sensitivity analysis for tirzepatide 10 mg versus semaglutide 1.0 mg

Figure 22: Tornado diagram of general sensitivity analysis for tirzepatide 15 mg versus semaglutide 1.0 mg

Figure 23: Tornado diagram of extensive sensitivity analysis for tirzepatide 10 mg versus semaglutide 1.0 mg

B 31. Priority question: The probabilistic sensitivity analysis (PSA) appears to be based upon bootstrapping with replacement. Bootstrapping is more commonly used in trial-based economic evaluation and it represents first-order uncertainty, i.e. the variability of a statistic. Furthermore, when running the PSA we encountered unexpected results: 1) PSA results were the same regardless of the number of simulations. 2) The cost-effectiveness plane shows a large number of dots even when only 10 simulations are run.

a) Could the company provide more clarity on the methods used for the PSA (including an explanation of the unexpected results described above, a general description of the PSA approach adopted, a step by step explanation of the PSA implementation and all parameters considered in the PSA)?

The ability for the user to change the number PSA iterations was added in response to PRIMA review. The approach to specifying the number of bootstrap iterations was verified and tested in the model directly on the command line using JSON files; however, following the comments from the EAG, it was established that, when specified through the web interface, the “iterations” input was not being honored by the model. This has now been corrected by adding the following code to the DatabaseController.java:

This change to the database code has been deployed to https://hta.prime-diabetes-model.com/index.html#!/t2d/home and the revised code has been made available for download from https://lilly.covalence-research.com/PRIME/.

b) Please clarify whether PSA involves sampling of patient characteristics, treatment effects, costs and utilities, and explain how this is achieved.

When PSA is active, patient characteristics, treatment effects, costs, and utilities are all sampled, in addition to model coefficients.
Regardless of whether PSA is active, on generating the simulated cohort, patient characteristics are sampled in the CohortController.java file, drawing from uniform distributions to establish the

patient-level presence (or absence) of binary characteristics such as smoking status (e.g. line 100 of CohortController.java) or history of events (e.g. line 212 of CohortController.java), and from user-specified distributions to establish the patient-specific baseline value of continuous distributions (e.g. line 178 of CohortController.java for HbA1c).
Treatment effects are sampled in TreatmentController.java. All treatment effect sampling is contained within the private function `applyTreatmentEffectsToPatient()`. For example, the sampled HbA1c treatment effect is applied on line 216 of TreatmentController.java.
Costs are sampled in EconomicsController.java. Cost values are drawn from the specified distributions using the `sample()` method of the `MeanCostWithSD` class (line 29 of MeanCostWithSD.java). For instance, the cost of a heart failure event is sampled on line 138 of EconomicsController.java.
Similarly, utilities are sampled in QualityOfLifeController.java. Utility values are drawn from the specific distributions using the `sample()` method of the `MeanQoLUtilityWithSD` class (line 27 of MeanQoLUtilityWithSD.java). For instance, the utility associated with the onset of heart failure is sampled on line 181 of QualityOfLifeController.java.
Model coefficients are sampled on line 16 of the base `PatientController` class, from which all patient complication "controllers" inherit. Sampling is implemented using the `sample()` method on the `SampledDouble` class (line 43 of SampledDouble.java).

c) Please justify why bootstrapping is applied to a sub-sample of patients rather than the complete cohort. Please clarify what the consequences of this implementation are on estimated standard errors derived using PSA (as compared to when a full bootstrap approach would be implemented).

When PSA is active, bootstrap samples are indeed drawn from the whole cohort. Lines 201-210 of the ResultsController are responsible for randomising the order of the patients in the entire simulated cohort and lines 213, 237, and 248 are then responsible for drawing individual bootstrap samples to calculate costs, life expectancy, and quality-adjusted life expectancy, respectively.

d) Please reflect on whether the chosen approach truly reflects second-order uncertainty.

We are confident that second order uncertainty is thoroughly captured in the model when PSA is active. When the PSA iteration count is non-zero, every single "controller" parameter (i.e. all of those parameters specified using the `SampledDouble` class and stored in the "params" Enumerated Map of each controller) is sampled. The sample is implemented in the base `PatientController` class in the "v" method (abbreviation of "value").

e) Please implement a different method for the PSA by sampling from the (joint) probability distributions for each parameter (see for example Corro-Ramos et al https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7401182/pdf/10.1177_0272989X20932145.pdf), and provide the results.

As described in the response to question B 83. b) above, all cohort, treatment effect, cost, utility, and model parameter distributions are already sampled during PSA. We are unsure as to how joint distributions of the parameters could be sampled given the intervention- and trial-agnostic nature of the model and the lack of data on how distributions covary; covariance between baseline characteristic and treatment effect distributions could be captured, but this was found to result in negligible differences in modelled outcomes during the development of the PRIME Diabetes Model for Type 1 Diabetes based on covariance matrices derived from patient-level data from the DCCT; we would be surprised if such covariance made a meaningful difference in the context of a T2D model.

f) Please explain and correct the issues with the PSA mentioned above.

We consider the PSA implementation in PRIME to be robust and hope that the above clarifications and the fix to the database code responsible for reading the desired number of iterations addresses the EAG's concerns.

Validation

B 32. Priority question: Further information on validation efforts would be desirable, focusing on this specific implementation of the PRIME T2D model.

a) Please complete the TECH-VER checklist (Büyükkaramikli et al. 2019, https://pubmed.ncbi.nlm.nih.gov/31705406/) and provide the results.

The TECH-VER checklist represents an extensive checklist of more than 80 points which appears designed to accompany the development of new health economic, principally in Microsoft Excel based on the checklist content, that would add little additional value in the current circumstances. Most of the key areas around model verification and validation have already been addressed elsewhere in the submission (see the external code audit and validation as described in Appendix N3) and many of the other points on the checklist are not well suited to the development of a complex patient level simulation of T2D. Therefore, given the limited timeframe for response to questions and the fact that this checklist is not part of the guidance to manufacturers prior to submission, the checklist has not been retrospectively completed for the PRIME T2D Model.

b) Please provide a tabulated overview of all parameters used in the model, including se / sd / CIs, the probability distribution used, the source, justification for the source, and a specific description of how the parameter was implemented in the model.

The request to provide tabulated input parameters for a model of this complexity is impracticable. For each simulation, this would involve over 300 input parameters and 73 simulations in the submission, producing a table with approximately 22,000 rows. Even for only the base case simulations, there would be around 3,600 rows in the proposed table. It should be noted, however, that JSON files of all model inputs have been provided to the EAG providing most of data requested. The response to question B 40 provides details on sampling and distributions

used in the model. And the justification for all key inputs (that would have a bearing on cost-effectiveness) has been provided in the CS.

c) Please report on the face validity of the model structure, model assumptions, model inputs, intermediate outcomes as well as final outcomes of this specific application of the PRIME T2D model in more detail (including what aspects were assessed, clinical expert opinion, their considerations as well as conclusions).

The face validity of the model was developed and tested in three main ways:

- Literature review of existing model of T2D was used to inform the overall model concept by identifying strengths and weakness of different modelling approaches, complications and risk factors for consideration and key outcomes to be report
- Advisory boards in 2014, 2015 and 2019 were used to get expert clinical and health economic input for the development of the model on the statistical approach for handling uncertainty, modelling complications and risk factors, and approaches to risk estimation
- PRIMA review in 2022 was sought to test the face validity of the PRIME T2D Model and evaluate its suitability to support a submission to NICE. Details of the PRIMA review have been provided to the EAG.

d) Please provide the technical development report mentioned in step 5 of Appendix N3.

The technical development report mentioned in step 5 was a working document that became the PRIME T2D Model Technical Report (that is provided in Appendix N3).

B 33. Priority question: Please provide a tabulated overview of a subsample of simulated patients (n=50) with their baseline characteristics, assignment to treatment, treatment duration and the events and timepoints of events that patients experience over the time horizon.

We have attached a comma-separated variables (CSV) file including the baseline characteristics, risk factor trajectories, treatment assignment, treatment progression, and event histories of 50 patients from the simulated cohort in the analysis titled “nice_2_d TZP5_SEMA0.5 NMA BC”.

B 34. Priority question: Please provide a cross validation, i.e. a comparison with the NICE Guideline 28, as well as other relevant technology appraisals (e.g. those mentioned in the scope) of the following in a tabular overview:

a) Model structure and assumptions, input parameters related to clinical effectiveness, health state utility values, resource use and costs

Please find below a table summarizing key aspects of other recent technology appraisals in line with the EAG request.

Table 42: Overview of other health economic analyses in type 2 diabetes preceding the present analysis

	NICE NG 28 (2022)	NICE NG 28 (2015)	TA315 Canagliflozin	TA288 Dapagliflozin	TA336 Empagliflozin
Model used	UKPDS Outcomes Model 2 to model the standard of care arm, and a multi-state model for medications evaluated in CVOTs (comprising all possible events, event histories and combination of events/histories modelled in the UKPDS)	UKPDS Outcomes Model 1, with added functionality (hypoglycemia, utilities relating to weight changes and nausea)	ECHO-T2DM	De novo model based in C++ with Excel front-end	IQVIA (formerly IMS) CORE Diabetes Model
Validation	Comparison of UKPDS Outcomes Model 2 and LEADER standard of care arms	No primary validation performed (internal and external validation previously performed)	No primary validation reported (internal and external validation previously performed)	Results compared with those produced by the IQVIA CORE Diabetes Model	No primary validation performed (internal and external validation previously performed)
Clinical inputs	Cohort: THIN Treatment effects: Previous NG28 analysis for standard of care analyzed using the UKPDS Outcomes Model 2; CVOTs for hazard ratios for CV outcomes; Dunkley et al. Diabetes Obes Metab. 2019;21(7):1585-95 and CVOTs for hypoglycemia	Cohort: THIN Treatment effects: Primary NMA	Cohort: Canagliflozin clinical trials Treatment effects: Canagliflozin clinical trials	Cohort: Dapagliflozin clinical trial and NMA Treatment effects: Dapagliflozin clinical trial and NMA	Cohort: Not reported Treatment effects: Primary NMA

	NICE NG 28 (2022)	NICE NG 28 (2015)	TA315 Canagliflozin	TA288 Dapagliflozin	TA336 Empagliflozin
Cost inputs	Unit costs: NHS Drug Tariff Complications: UKPDS 84 (Alva et al. Diabet Med. 2015;32(4):459-66) Hypoglycemia: Hammer et al. J Med Econ. 2009;12(4):281-90	Unit costs: NHS Drug Tariff Complications: UKPDS 65 (Clarke et al. Diabet Med. 2003;20(6):442-50) Hypoglycemia: Hammer et al. J Med Econ. 2009;12(4):281-90	Unit costs: Not reported Complications: Not reported	Unit costs: England and Wales Drug Tariff Complications: Primarily UKPDS 65 (Clarke et al. Diabet Med. 2003;20(6):442-50), with other UK-specific studies informing ESRD and hypoglycemia	Unit costs: NHS list prices Complications: UKPDS and previous NICE appraisals (no further detail reported) ERG performed sensitivity analyses with UKPDS 65 (Clarke et al. Diabet Med. 2003;20(6):442-50) costs
HRQoL utility inputs and estimation method	Sources: Baseline from Alva et al. Health Econ. 2014;23(4):487-500 Complications from Beaudet et al. Value Health. 2014;17(4):462-70 Changes in BMI from Bagust and Beale Health Econ. 2005;14(3):217-30 Injections from Olofsson et al. J Med Econ. 2016;19(10):945-58 Hypoglycemia from Evans et al. Health Qual Life Outcomes. 2013;11(1):90 Estimation method: Additive	Sources: Baseline and complications from UKPDS 62 (Clarke et al. Med Decis Making. 2002;22(4):340-9) Changes in BMI from Bagust and Beale Health Econ. 2005;14(3):217-30 Hypoglycemia from Currie et al. Curr Med Res Opin. 2006;22(8):1523-34 Estimation method: Additive	Source: CODE-2 (non-interventional, observational study) Estimation method: Multivariate regression	Sources: Baseline from Department of Health Survey for England Complications from UKPDS 62 (Clarke et al. Med Decis Making. 2002;22(4):340-9), the Health Outcomes Data Repository, and Currie et al. Curr Med Res Opin. 2006;22(8):1523-34 Changes in BMI from a Canada-specific, manufacturer-endorsed study Estimation method: Not reported	Source: UKPDS 62 (Clarke et al. Med Decis Making. 2002;22(4):340-9) and Sullivan et al. Med Decis Making. 2011;31(6):800-4 Estimation method: Not reported

	NICE NG 28 (2022)	NICE NG 28 (2015)	TA315 Canagliflozin	TA288 Dapagliflozin	TA336 Empagliflozin
Complication risk estimates	UKPDS Outcomes Model 2 for standard of care CVOT hazard ratios for medications evaluated in CVOTs	UKPDS Outcomes Model 1	Microvascular complications: Wisconsin Epidemiologic Study of Diabetic Retinopathy; Rochester Epidemiology Project; CDC model of chronic kidney disease Macrovascular complications: UKPDS Outcomes Model 1	UKPDS Outcomes Model 1	Not reported, likely primarily UKPDS Outcomes Model 1 (default in the IQVIA CORE Diabetes Model)
Progression of risk factors over time	UKPDS Outcomes Model 2	UKPDS Outcomes Model 1	Annual, class-specific drift (source not reported)	UKPDS Outcomes Model 1	Not reported, likely primarily UKPDS Outcomes Model 1 (default in the IQVIA CORE Diabetes Model)
Adverse events modeled	Hypoglycemia Severe adverse events from CVOTs modeled as one parameter in a sensitivity analysis	Hypoglycemia Nausea	Hypoglycemia Other adverse events modeled, but not specifically reported	Hypoglycemia Urinary tract infections Genital infections	Hypoglycemia Urinary tract infections Genital infections
Notable assumptions	Treatment intensification occurred at a 7.5% HbA1c threshold Use of unadjusted hazard ratios from CVOTs	Treatment intensification occurred at a 7.5% HbA1c threshold	Treatment intensification threshold/regimen not reported	Treatment intensification occurred at unique, study/NMA-specific HbA1c thresholds (different thresholds for different comparisons)	Treatment intensification occurred at a 7.5% HbA1c threshold
CVOT calibration approach	Unadjusted hazard ratios directly applied from CVOTs	No CVOT data incorporated	No CVOT data incorporated	No CVOT data incorporated	No CVOT data incorporated

	NICE NG 28 (2022)	NICE NG 28 (2015)	TA315 Canagliflozin	TA288 Dapagliflozin	TA336 Empagliflozin
Source	https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845	https://www.nice.org.uk/guidance/ng28/evidence/appendix-f-full-health-economics-report-pdf-2185320355	https://www.nice.org.uk/guidance/ta315/chapter/3-The-manufacturers-submission#cost-effectiveness	https://www.nice.org.uk/guidance/ta288/chapter/3-The-manufacturers-submission#cost-effectiveness	https://www.nice.org.uk/guidance/ta336/chapter/3-The-companys-submission#cost-effectiveness

BMI, body mass index; CVOT, cardiovascular outcomes trial; ERG, Evidence Review Group; ESRD, end-stage renal disease; HbA1c, glycated haemoglobin; HRQoL; health-related quality of life; THIN, The Health Improvement Network.

b) And how these differences affect estimated outcomes per comparator / interventions (life years, QALYs, costs)

Without running comparable analysis across all of the models (and using different approaches) summarized in the table below, it is difficult to comment extensively on the influence of each assumption on the outcomes for each parameter. The following general comments may provide some insight:

- **Model used and validation:** The models used vary between all six analyses considered. Ostensibly, the risk equations used should be the main determinant of different outcomes between model (see comments below). With respect to validation, The PRIME T2D Model has more recent and, in most cases, more extensive validation than the other models summarized in the table below (see Appendix N3), including validation against CVOTs.
- **Clinical inputs:** The modelling approaches were broadly similar with respect to clinical inputs used, with all analyses relying on short-term trial data or NMA estimates for treatment effects inputs.
- **Cost and HRQoL utility inputs:** the present analysis is well aligned with the cost input and utilities used in the previous NICE analyses (2022 and 2015) and the assessments on dapagliflozin (TA28) and empagliflozin (TA336). Only the assessment on canagliflozin used different approaches (TA315); costs were not well reported and utilities relied on estimates from the CODE-2 study which are not specific to the UK setting.
- **Complication risk estimation:** The present analysis uses UKPDS OM2 risk formulae with BRAVO Model risk formulae (model averaging). The 2022 NICE evaluation used UKPDS OM2 in combination with hazard ratios from CVOTs (see comments below). The other analyses relied on risk equations from UKPDS OM1, which has been shown to overestimate complication rates in higher risk populations and was derived from older data collected between 1977 and 1997 (Hayes *et al.* 2013).[55] Validation of the PRIME T2D Model would suggest that calibration with hazard ratios, which represents a problematic approach (see comments below), is not necessary. Without a head-to-head comparison on the same dataset, it's difficult to comment on how the present approach would compare directly with a CVOT-calibrated UKPDS OM2 modelling approach.
- **Progression of risk factors over time:** The approaches were broadly aligned across all evaluations with respect to progression of risk factors (with the exception of the canagliflozin analysis, which may lead to very high risk factor estimates in later years of a long-term simulation). We would not expect the differences in risk factor approaches to be a notable differentiator between the analyses.
- **Adverse events:** modelling analyses were broadly aligned in terms of adverse events. Most other analyses incorporated hypoglycaemia on treatment, but given the very low rates of hypoglycaemia with tirzepatide and comparators in the present analysis including hypoglycaemia rates for tirzepatide and comparators would have had a very modest effect on outcomes. Hypoglycaemia was included in the modelling of insulin therapy in the present analysis.
- **Intensification assumptions:** the present analysis was well aligned with the intensification thresholds used by NICE and in the submission on empagliflozin. The other two assessments used more specialized intensification assumptions, which may

have influence outcomes in these analyses depending on the assumptions around the progression of risk factors applied with each intensification step.

- **Use of CVOT hazard ratios / calibration:** Whilst the PRIME T2D Model has validated against several CVOTs without calibration, the 2022 NICE health economic evaluation relied in a calibration approach using data from different CVOTs in combination. The earlier assessments described in the table relied on older methods to evaluate the risk of complications and did not use CVOT data. The calibration of existing T2D model with hazard ratios from CVOTs is a complex challenge with considerable potential to provide misleading results when comparing multiple interventions as recently summarized by Evans *et al.* (2023).[77] A main concern focuses on the heterogeneity of the trials, with different study durations, inclusion criteria, rescue medication protocols and endpoint definitions, which results in significant uncertainty when comparing two or more interventions evaluated in separate CVOTs, as robust adjustment for these differences is very challenging. This is compounded by differences in endpoint definitions in a given model (which need to match those in the CVOT to be suitable for calibration) and the challenge of double-counting treatment effects (the hazard ratios from CVOTs are typically not adjusted for improvements in conventional risk factors such as HbA1c). The use of unadjusted hazard ratios from multiple CVOTs in a long-term cost-effectiveness analysis has considerable potential to skew the outcomes if these challenges are not appropriately addressed. As outlined by Evans *et al.* it is likely that these challenges can only be overcome by combining patient-level data from CVOTs to prepare novel risk equations that can better model modern therapies for T2D.

c) Please elaborate on the identified differences.

See response b) above

B 35. Priority question: Further external validation of modelled estimates against the SURPASS trials and (potentially available) alternative evidence would be desirable. Please assess the external validity of model inputs, intermediate outcomes and (long-term) disaggregated results (as provided in Appendix J) as well as final outcomes using the SURPASS trials and available alternative evidence sources.

External validation of the outcomes of a long-term diabetes model, such as the PRIME T2D Model, against short-term trial data from a program such as SURPASS would be provided little insight into the model's validity. The SURPASS outcomes reported to date are all based on follow up of 1 year or less (SURPASS-2 – 40 weeks, SURPASS-3 – 52 weeks, SURPASS-4 – 52 to 104 week, SURPASS-5 – 40 weeks). This short-term trial data would be an input to the model (to be applied in year 1 of simulations) and therefore would be closely matched to model outputs at 1 year for risk factors and cohort characteristics. Comparison of end-stage complication rates between the modelling analysis at 1 year and the trial outcomes at 40 or 52 weeks, would similarly provide little insight into validity as event rates would be very low in cases where comparable endpoints were reported, leading to substantial random variation between treatment groups (e.g. after 40 weeks in SURPASS-2, 1.1% of patients in the TZP 5 mg treatment group experienced SAEs of cardiac disorders, compared with 0.2% in the TZP 10 mg group, 0.9% on TZP 15 mg and 0.2% on semaglutide 1.0 mg). Validation against these data would provide little evidence that the model was capable of simulating long-term outcomes with discrimination between treatment arms.

There is a paucity of suitable long-term study data for model validation. The majority of studies that report long-term outcomes (e.g. incidence of diabetes-related complications), often lack sufficient detail around cohort characteristics, medical history, treatments and risk factors to be useful for validation exercises. Whilst we would always agree with the reviewers' assertion that more validation is always advantageous, given the timelines for response to clarification questions, a literature review to identify new studies for additional validation analyses is impracticable. If the reviewers have specific studies in the mind for validation, we would be happy to undertake the analysis and share the results. It is noteworthy, however, that in comparison with most other diabetes models, more extensive validation has been published on PRIME than other models in the last 5-10 years (for example, the UKPDS Outcomes Model 2 or the IQVIA CORE Diabetes Model).

PRIME T2D Model interface and source code

B 36. Priority question: There seem to be some discrepancies between the analyses described in the CS and the possibilities in the PRIME T2D Model interface.

a) Please provide for all EAG accounts the functionality and simulations in the "simulation list" to reproduce all analyses reported in the CS (including all scenario analyses reported in CS Tables 103-107)

We will be happy to provide EAG with access to the sensitivity analyses provided in the CS: The EAG accounts already have full access to the same functionality as all other user accounts in the model.

b) Please provide for all EAG accounts the functionality and simulations in the "simulation list" to reproduce all analyses reported in the responses to the clarification questions.

We will provide EAG access to the simulations described in this response document.

c) The possibilities for including uncertainty in the cohort characteristics in the "Cohorts" tab are limited. Please add the possibility of adding an uncertainty measure for the demographics, race and complication history.

For categorical variables (such as male/female, ethnic group or history of complications), there is no conventional measure of uncertainty associated with these measures. Simulated patients are either in the category or not. This would also be problematic in terms of mutually exclusive categorical variables (e.g. the total of male and female needs to add to 100%) where introducing a measure of uncertainty would make combining sampled results very challenging. For these reasons, no measures of uncertainty have been added to categorical variables in the model. It should be noted though that changing cohort characteristics (within a plausible) without a corresponding treatment effect change has a very limited impact on relative cost-effectiveness between interventions.

d) **In the "Countries" tab it is possible to select either the PRIME default or BRAVO as complication risk models. We assume that the PRIME default comprises the model averaging approach, please make this clear in the interface. Please also list all possible complication risk models here so that sensitivity analyses can be performed.**

The request to allow selection of multiple different risk model in the *Countries* element of the model interface is not practicable within the timeframe for response. It would require significant changes to the model code and to the model interface (as not all endpoints can be evaluated). The PRIME T2D Model is set up to use model averaging to evaluate the risk of most diabetes related endpoints. We believe, based on validation analyses, that this is the optimal approach for modern diabetes management in most country settings, including the UK.

e) **It is unclear how the adverse events related to treatment were implemented in the model, for example the percentage of patients that experienced nausea due to treatment. Please add the adverse events parameters to the model interface.**

As outlined in the response to question B22, the model features a treatment-related utility function that is editable by the user and can be used to define separate utilities to be applied in year 1 and years 2+ of any given simulation. The treatment related utilities are added to the annual utility score for each patient as calculated based on the inputs in the Utilities element. In the current set of simulations, the treatment-related utility function was used to capture the following utilities:

- Year 1: body weight change utility (no separate BMI utility), device utility and the nausea and vomiting utility
- Years 2+: BMI utility only (no body weight change utility)

f) **It is unclear how the utility gain associated with weight loss and the disutility associated with a BMI above 25 were implemented in the model. Please add the possibility to change these parameters to the model interface to allow for replication of scenario analyses 58 and 59.**

See point e) above.

g) **In scenario analysis 63, QALY values were adjusted for age. It is unclear how this was implemented in the PRIME T2D model interface. Please add this possibility to the model interface to allow for replication of scenario analysis 63.**

In the *Utility* element in the model interface, the user can select the calculation method from the drop-down menu labelled “Quality of life calculation approach” allowing for additive or multiplicative approaches, with or without age-adjustment, or use of the Shao et al. OLS regression formula.

B 37. Please add functionality to the PRIME T2D Model interface to extract discounted comparator costs split into treatment cost and other costs (to enable the calculation of ICERs with potential comparator price discounts)

This is already available in the model interface using the “Show tables” checkbox at the top of the simulation results page. A breakdown of costs table is provided below the bar chart of cumulative incidence of complications.

B 38. Considering the options to specify the PSA in the dashboard:

a) The EAG could not find any option in the dashboard to specify SDs for costs, and to enable sampling of costs. Please clarify how to run a PSA with random annual costs.

In the top right corner of the *Costs* element page in the model, there is a checkbox marked “Advanced.” When this is checked by the user, fields for SDs are made visible in the user interface.

b) The EAG could not find any option in the dashboard to specify SDs for utilities, and to enable sampling of utilities. Please clarify how to run a PSA with random utilities.

In the top right corner of the *Utilities* element page in the model, there is a checkbox marked “Advanced.” When this is checked by the user, fields for SDs are made visible in the user interface.

B 39. After inspecting the source model the following questions arose

a) The JAVA source code does not contain any executables for the user interface. Please clarify how the code can be used to run specific simulations and to evaluate the corresponding output. In particular, the executable “Main.java” in the package com.ossianconsulting.controllers requires to specify a list of arguments and JSON files to run the simulation; it is however not clear how these arguments should be defined. Could you please provide a worked example (including any commands and/or arguments as well as configuration files) to run a specific simulation (e.g. the base case of the report) using the JAVA executables?

We have provided a shell script along with a compiled Java archive (JAR) file, a series of example JSON files and the corresponding UK life table file that can be used to run the model. The shell script has been written to run in a Linux/Unix-like environment, but should be easily adaptable to other environments such as a Windows batch file. The shell script includes instructions in the comments from line 3-33 describing the directory structure and structure of the command.

b) Please share the JSON configuration files for running the simulations (including base case simulation) and clarify how they should be provided to the JAVA application using the command line.

As noted in the response to point a), we have provided the compiled JAR file, JSON files, shell script, and UK lifetable file that collectively allow the model to be run locally from the command line.

c) The JAVA source code does not contain any executables to visualize simulation results. Could you please provide a working example to export simulation results to an output that is similar to the output from the online dashboard?

At present, the web interface (available at https://hta.prime-diabetes-model.com/#!/t2d/home) is the only way to visualise the outputs from the PRIME T2D Model. The results presented in the web interface are based on exactly the same JSON file format as produced by the model, and we can provide access to the source code for the web interface and support for configuring it to run locally; in brief, the web interface consists of a static AngularJS application that can be served from nginx or Apache, while an application programming interface (API) to the MongoDB document store runs as a simple Express/Passport app on node.js.

B 40. To allow the EAG to scrutinise and potentially adjust existing analyses. Please provide the following information:

a) All relevant JSON files of all simulations that were run, including new simulations based on the clarification.

We have provided the JSON files to allow a full example simulation to be run, and would recommend that the EAG use these files as the basis for configuring other simulations as needed.

b) Instructions on how to “pass” the JSON files to the main java executable.

Instructions for how to run the model, including how to pass the JSON file paths to the Java executable, are included in the shell script provided in response to question B 39 a) and b).

c) Instructions on how to store results from the simulations in the database or a JSON file (currently all output is printed to the command line)

The results from the model can be stored to a JSON file by appending the `-output results.json` option to the shell script. Reading from and saving to a suitably configured local MongoDB document store (running on localhost on port 27017) can be achieved by adding the `-useDatabase true` option.

d) Instructions to summarize and visualise results from the simulations (which are stored in a database or as a JSON file). The available source code does not have any executables to visualize simulation study results.

As noted in the response to B 38 c), at present, the web interface (available at https://hta.prime-diabetes-model.com/#!/t2d/home) is the only way to visualise the outputs from PRIME T2D Model. The results presented in the web interface are based on exactly the same JSON file format as output by the model, and we can provide access to the source code for the web interface and support for configuring it to run locally.

References

1. Ozempic Summary of Product Characteristics (SmPC). Available at https://www.ema.europa.eu/en/documents/product-information/ozempic-epar-product-information_en.pdf. [Accessed 13 February 2023].
2. Ahmann AJ, Capehorn M, Charpentier G, et al. Efficacy and safety of once-weekly semaglutide versus exenatide ER in subjects with type 2 diabetes (SUSTAIN 3): a 56-week, open-label, randomized clinical trial. Diabetes care 2018;41:258-266.
3. Ahrén B, Masmiquel L, Kumar H, et al. Efficacy and safety of once-weekly semaglutide versus once-daily sitagliptin as an add-on to metformin, thiazolidinediones, or both, in patients with type 2 diabetes (SUSTAIN 2): a 56-week, double-blind, phase 3a, randomised trial. The lancet Diabetes & endocrinology 2017;5:341-354.
4. Aroda VR, Bain SC, Cariou B, et al. Efficacy and safety of once-weekly semaglutide versus once-daily insulin glargine as add-on to metformin (with or without sulfonylureas) in insulin-naive patients with type 2 diabetes (SUSTAIN 4): a randomised, open-label, parallel-group, multicentre, multinational, phase 3a trial. The Lancet Diabetes & Endocrinology 2017;5:355-366.
5. Rodbard HW, Lingvay I, Reed J, et al. Semaglutide added to basal insulin in type 2 diabetes (SUSTAIN 5): a randomized, controlled trial. The Journal of Clinical Endocrinology & Metabolism 2018;103:2291-2301.
6. Sorli C, Harashima S-i, Tsoukas GM, et al. Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. The lancet Diabetes & endocrinology 2017;5:251-260.
7. Dungan KM, Povedano ST, Forst T, et al. Once-weekly dulaglutide versus once-daily liraglutide in metformin-treated patients with type 2 diabetes (AWARD-6): a randomised, open-label, phase 3, non-inferiority trial. Lancet 2014;384:1349-57.
8. Blonde L, Jendle J, Gross J, et al. Once-weekly dulaglutide versus bedtime insulin glargine, both in combination with prandial insulin lispro, in patients with type 2 diabetes (AWARD-4): a randomised, open-label, phase 3, non-inferiority study. Lancet 2015;385:2057-66.
9. Giorgino F, Benroubi M, Sun JH, et al. Efficacy and Safety of Once-Weekly Dulaglutide Versus Insulin Glargine in Patients With Type 2 Diabetes on Metformin and Glimepiride (AWARD-2). Diabetes Care 2015;38:2241-9.
10. Skrivanek Z, Gaydos BL, Chien JY, et al. Dose-finding results in an adaptive, seamless, randomized trial of once-weekly dulaglutide combined with metformin in type 2 diabetes patients (AWARD-5). Diabetes Obes Metab 2014;16:748-56.
11. Umpierrez G, Tofe Povedano S, Perez Manghi F, et al. Efficacy and safety of dulaglutide monotherapy versus metformin in type 2 diabetes in a randomized controlled trial (AWARD-3). Diabetes Care 2014;37:2168-76.
12. Wysham C, Blevins T, Arakaki R, et al. Efficacy and safety of dulaglutide added onto pioglitazone and metformin versus exenatide in type 2 diabetes in a randomized controlled trial (AWARD-1). Diabetes Care 2014;37:2159-67.
13. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2022;45:2753-2786.
14. National Institute of Health and Care Excellence (NICE). Semaglutide. British National Formulary. Available at: https://bnf.nice.org.uk/drugs/semaglutide/. [Accessed 14 February 2023].
15. Chubb B, Gupta P, Gupta J, et al. Once-Daily Oral Semaglutide Versus Injectable GLP-1 RAs in People with Type 2 Diabetes Inadequately Controlled on Basal Insulin: Systematic Review and Network Meta-analysis. Diabetes Ther 2021;12:1325-1339.

16. Weiss T, Iglay K, Carr R, et al. 987-P: Real-World Weight Change, Adherence, and Discontinuation among Type 2 Diabetes (T2D) Patients Initiating GLP-1 Receptor Agonists (GLP-1RAs) in the UK. Diabetes 68 (Supplement 1). doi: 10.2337/db19-987-P. 2019.
17. Mulnier HE, Seaman HE, Raleigh VS, et al. Mortality in people with type 2 diabetes in the UK. Diabet Med 2006;23:516-21.
18. Carey VJ, Walters EE, Colditz GA, et al. Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. Am J Epidemiol 1997;145:614-9.
19. Wu Y, Ding Y, Tanaka Y, et al. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci 2014;11:1185-200.
20. Bellou V, Belbasis L, Tzoulaki I, et al. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. PLoS One 2018;13:e0194127.
21. Eli Lilly Data on File. SURPASS Clinical Safety Summary.
22. Eli Lilly Data on File. Tirzepatide Draft Summary of Product Characteristics (SmPC).
23. Kwan AYM MJ, Wang H, Rasouli N, Wilding JPH. Tirzepatide Induces Weight Loss in Patients with Type 2 Diabetes Regardless of Baseline BMI: A Post Hoc Analysis of SURPASS-1 through -5 Studies. American Diabetes Association 82nd Annual Scientific Sessions New Orleans USA 2022.
24. Urva S, Quinlan T, Landry J, et al. Effects of Renal Impairment on the Pharmacokinetics of the Dual GIP and GLP-1 Receptor Agonist Tirzepatide. Clin Pharmacokinet 2021;60:1049-1059.
25. Heerspink HJ, Sattar N, Pavo I, et al. Effects of tirzepatide versus insulin glargine on kidney outcomes in type 2 diabetes in the SURPASS-4 trial: post-hoc analysis of an open-label, randomised, phase 3 trial. The Lancet Diabetes & Endocrinology 2022;10:774-785.
26. Frías JP, Davies MJ, Rosenstock J, et al. Tirzepatide versus Semaglutide Once Weekly in Patients with Type 2 Diabetes. N Engl J Med 2021;385:503-515.
27. Ludvik B, Giorgino F, Jódar E, et al. Once-weekly tirzepatide versus once-daily insulin degludec as add-on to metformin with or without SGLT2 inhibitors in patients with type 2 diabetes (SURPASS-3): a randomised, open-label, parallel-group, phase 3 trial. Lancet 2021;398:583-598.
28. Del Prato S, Kahn SE, Pavo I, et al. Tirzepatide versus insulin glargine in type 2 diabetes and increased cardiovascular risk (SURPASS-4): a randomised, open-label, parallel-group, multicentre, phase 3 trial. Lancet 2021;398:1811-1824.
29. Eli Lilly Data on File. SURPASS-5 CSR.
30. Dahl D, Onishi Y, Norwood P, et al. Effect of Subcutaneous Tirzepatide vs Placebo Added to Titrated Insulin Glargine on Glycemic Control in Patients With Type 2 Diabetes: The SURPASS-5 Randomized Clinical Trial. JAMA 2022;327:534-545.
31. Eli Lilly Data on File. SURPASS-2 CSR.
32. Eli Lilly Data on File. SURPASS-3 CSR.
33. Eli Lilly Data on File. SURPASS-4 CSR.
34. Trulicity Summary of Product Characteristics (SmPC). Available at https://www.ema.europa.eu/en/documents/product-information/trulicity-epar-product-information_en.pdf. [Accessed 13 February 2023].
35. Eli Lilly Data on File. A23 list of treatments and background therapies.
36. Dias S, Sutton, A. J., Welton, N. J., & Ades, A. E.,. NICE DSU Technical Support Document 3: Heterogeneity: Subgroups, Meta-Regression, Bias and Bias-Adjustment. National Institute for Health and Clinical Excellence. 2011.
37. Dias S, Welton NJ, Sutton AJ, et al. NICE Decision Support Unit Technical Support Documents. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. London: National Institute for Health and Care Excellence (NICE) Copyright © 2014 National Institute for Health and Clinical Excellence, unless otherwise stated. All rights reserved., 2014.
38. Eli Lilly Data on File. NMA: Baseline Characteristics by Treatment Arm.

39. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018;41:2669-2701.
40. Eli Lilly Data on File. A34: R Code and Data.
41. Eli Lilly Data on File. NMA FitModel.R MCMC input files.
42. Eli Lilly Data on File. SURPASS-J-Mono CSR.
43. Eli Lilly Data on File. SURPASS-J-Combo CSR.
44. Marso SP, Bain SC, Consoli A, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. New England Journal of Medicine 2016;375:1834-1844.
45. Eli Lilly Data on File. A43: NMA input data.
46. Eli Lilly Data on File. A44: Hypoglycaemia Data.
47. Shao H, Shi L, Fonseca VA. Using the BRAVO Risk Engine to Predict Cardiovascular Outcomes in Clinical Trials With Sodium-Glucose Transporter 2 Inhibitors. Diabetes Care 2020;43:1530-1536.
48. Pollock RF, Norrbacka K, Boye KS, et al. The PRIME Type 2 Diabetes Model: a novel, patient-level model for estimating long-term clinical and cost outcomes in patients with type 2 diabetes mellitus. J Med Econ 2022;25:393-402.
49. Si L, Willis MS, Asseburg C, et al. Evaluating the Ability of Economic Models of Diabetes to Simulate New Cardiovascular Outcomes Trials: A Report on the Ninth Mount Hood Diabetes Challenge. Value Health 2020;23:1163-1170.
50. Keng MJ, Leal J, Mafham M, et al. Performance of the UK Prospective Diabetes Study Outcomes Model 2 in a Contemporary UK Type 2 Diabetes Trial Cohort. Value Health 2022;25:435-442.
51. Fishman GS. Discrete-Event Simulation - Modeling, Programming, and Analysis: Springer, 2001.
52. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. Value Health 2012;15:812-20.
53. Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). Diabetologia 2004;47:1747-59.
54. AJ P, S R, WJ V, et al. The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. . Curr Med Res Opin 2004;20 Suppl 1:S5–26.
55. Hayes AJ, Leal J, Gray AM, et al. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia 2013;56:1925-33.
56. Shao H, Fonseca V, Stoecker C, et al. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). Pharmacoeconomics 2018;36:1125-1134.
57. Wilding JPH, Batterham RL, Davies M, et al. Weight regain and cardiometabolic effects after withdrawal of semaglutide: The STEP 1 trial extension. Diabetes Obes Metab 2022;24:1553-1564.
58. Eli Lilly Data on File. XBPI and HPA, IQVIA Solutions UK Ltd, December 2022.
59. Leal J, Alva M, Gregory V, et al. Estimating risk factor progression equations for the UKPDS Outcomes Model 2 (UKPDS 90). Diabet Med 2021;38:e14656.
60. NHS Digital. HSE 2019: mean and median heights in adults aged 16 and over. Available at: https://digital.nhs.uk/supplementary-information/2022/hse19-mean-and-median-heights. [Accessed 20 February 2023].

61. National Institute of Health and Care Excellence (NICE). Dapagliflozin in combination therapy for treating type 2 diabetes. Available at: https://www.nice.org.uk/Guidance/TA288. [Accessed 20 February 2023].
62. National Institute of Health and Care Excellence (NICE). Empagliflozin in combination therapy for treating type 2 diabetes. Available at: https://www.nice.org.uk/guidance/ta336. [Accessed 20 February 2023].
63. Nauck MA, Buse JB, Mann JFE, et al. Health-related quality of life in people with type 2 diabetes participating in the LEADER trial. Diabetes Obes Metab 2019;21:525-532.
64. Mitchell P, Annemans L, Gallagher M, et al. Cost-effectiveness of ranibizumab in treatment of diabetic macular oedema (DME) causing visual impairment: evidence from the RESTORE trial. Br J Ophthalmol 2012;96:688-93.
65. Shao H, Yang S, Fonseca V, et al. Estimating Quality of Life Decrements Due to Diabetes Complications in the United States: The Health Utility Index (HUI) Diabetes Complication Equation. Pharmacoeconomics 2019;37:921-929.
66. Evans M, Khunti K, Mamdani M, et al. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. Health Qual Life Outcomes 2013;11:90.
67. Matza LS, Boye KS, Yurgin N, et al. Utilities and disutilities for type 2 diabetes treatment-related attributes. Qual Life Res 2007;16:1251-65.
68. Boye KS, Matza LS, Stewart KD, et al. Patient preferences and health state utilities associated with dulaglutide and semaglutide injection devices among patients with type 2 diabetes in Italy. J Med Econ 2019;22:806-813.
69. Boye KS, Matza LS, Stewart KD, et al. Health state utilities associated with weight loss in type 2 diabetes and obesity. J Med Econ 2022;25:14-25.
70. Bagust A, Beale S. Modelling EuroQol health-related utility values for diabetic complications from CODE-2 data. Health economics 2005;14:217-230.
71. National Institute for Health and Care Excellence (NICE). Resource impact template (type 2 diabetes and cardiovascular risk update). 2022 [NG28]. Available at: https://www.nice.org.uk/guidance/ng28/resources/ [Accessed 02 May 2022]. 2022.
72. National Institute of Health and Care Excellence (NICE). NG28. Type 2 diabetes in adults: management. Health Economic Model Report. Available at: https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845/ [Accessed 10 April 2022].
73. Dennett SL, Boye KS, Yurgin NR. The impact of body weight on patient utilities with or without type 2 diabetes: a review of the medical literature. Value Health 2008;11:478-86.
74. NHS Digital. National Diabetes Audit, 2020-21. Report 1: Care Processes and Treatment Targets. England and Wales. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/core-report-1-2020-21#. [Accessed 20 February 2023]. .
75. Courtney H, Nayar R, Rajeswaran C, et al. Long-term management of type 2 diabetes with glucagon-like peptide-1 receptor agonists. Diabetes Metab Syndr Obes 2017;10:79-87.
76. Wilke T, Mueller S, Groth A, et al. Non-Persistence and Non-Adherence of Patients with Type 2 Diabetes Mellitus in Therapy with GLP-1 Receptor Agonists: A Retrospective Analysis. Diabetes Ther 2016;7:105-24.
77. Evans M, Berry S, Nazeri A, et al. The challenges and pitfalls of incorporating evidence from cardiovascular outcomes trials in health economic modelling of type 2 diabetes. Diabetes Obes Metab 2023;25:639-648.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal
Tirzepatide for treating type 2 diabetes [ID3938]
Patient Organisation Submission

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type. [Please note that declarations of interests relevant to this topic are compulsory].

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you

1.Your name	[redacted]
2. Name of organisation	Diabetes UK
3. Job title or position	[redacted]
4a. Brief description of the organisation (including who funds it). How many members does it have?	Diabetes UK is the country's leading diabetes charity representing the 4.9 million people living with diabetes in the UK. We help people manage their diabetes effectively by providing information, advice and support. We campaign with people with diabetes and healthcare professionals to improve the quality of diabetes care across the UK's health services. We fund pioneering research into care, cure and prevention for all types of diabetes. The majority of Diabetes UK's income is from legacies and donations. We also earn income from activities which support our charitable mission, such as our Diabetes UK Professional Conference. A small percentage of our income is from support for specific programmes of work from or sponsorship of events by the pharmaceutical industry. We are a growing community with more than 300,000 supporters nationwide – including people with diabetes, their friends and families – and more than 100,000 lay and healthcare professional members.

4b. Has the organisation received any funding from the company bringing the treatment to NICE for evaluation or any of the comparator treatment companies in the last 12 months? [Relevant companies are listed in the appraisal stakeholder list.] **If so, please state the name of the company, amount, and purpose of funding.**	Diabetes UK receives some funding from the pharmaceutical industry to support specific programmes of work and for conferences we run including: £108,100 - Lilly
4c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No
5. How did you gather information about the experiences of patients and carers to include in your submission?	• Conversations with people living with type 2 diabetes via our Helpline and other channels • Surveys • Our online forum community • National diabetes audits

Living with the condition

6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	Type 2 diabetes is a relentless condition to live with and diabetes-related complications such as sight loss, cardiovascular disease and kidney failure can have a devastating impact on the lives of people living with type 2 diabetes and their loved ones. It can be a progressive condition and people tell us of their concern that their diabetes is developing; becoming more difficult to manage effectively and increasing their risk of complications. This causes a lot of anxiety which, in turn, impacts on a person's ability to self-manage their diabetes. This anxiety can be further exacerbated when people feel their blood glucose levels are not well controlled. Many people with diabetes tell us of feeling overwhelmed by the pressures of having the condition over a long period of time. The constant need to carefully manage blood glucose levels, medications, diet etc. tied to the emotional impact of, for example, being told they have a higher than expected HbA1c level at an appointment despite their best efforts can commonly lead to diabetes distress. Our insights indicate that one in five people with type 2 diabetes experience diabetes distress, and the most common reason is worrying about getting complications in the future or feeling anxious that management isn't good enough. If left unchecked, sustained feeling of diabetes distress over time can lead to burnout and a person with diabetes giving up on their care by skipping medication or routine appointments which is of serious concern. There is also a close association between having type 2 diabetes and living with overweight or obesity. Carrying excess weight is strongly tied to difficulties managing blood glucose levels and an increased risk of complications in those diagnosed with type 2 diabetes. As a result, weight loss is a primary goal

	in managing diabetes for people living with overweight or obesity. This can be very challenging in the context of rising obesity levels across society and particularly with the feelings of weight-related stigma that are commonly reported and physical and psychological harm which often accompanies them. Importantly, those who experience weight stigma are less likely to receive good care and seek help from a healthcare professional to support weight loss.

Current treatment of the condition in the NHS

7. What do patients or carers think of current treatments and care available on the NHS?	There is a wide range of treatments available for type 2 diabetes on the NHS and people should develop an individualised plan with their healthcare professional to meet their needs and preferences. If blood glucose levels cannot be managed by diet and lifestyle then Metformin is usually the first drug to be prescribed, and if tolerated is available as monotherapy or in combination with other medicines as treatment is reviewed and adjusted over time. It can be difficult for people to stay well-informed about the potential benefits and side effects of various different drug types and classes and follow complicated regimens. Generally, whilst few want to take more medicines and risk having associated side effects, most understand the need for treatments that can control blood glucose levels, reduce the risk of complications and improve outcomes and are interested in learning more about their options in a person-centred approach. The Covid-19 pandemic also heavily impacted routine diabetes care with just 37% of people in England with type 2 diabetes receiving all their recommended checks – including HbA1c, BMI and cholesterol - in 2020-21 compared with 58% in 2019-2020. In a recent survey of 10,000 people with diabetes we conducted almost half told us they had difficulties managing their condition during the pandemic and over 60% of them attributed this in part to not having sufficient access to their healthcare team. There is deep concern about the negative effects disruption of routine checks will have on management of diabetes and the care and treatments people receive.

	For people with type 2 diabetes who particularly want to lose weight, the current provision of specialised weight management services is uneven across England and unable to meet the growing numbers of people who could benefit from specialised services for obesity. This unfairly excludes many people from accessing treatment for reasons outside of their control. There is growing awareness of wider benefits of weight loss for people with type 2 diabetes but less clarity about how to achieve effective and sustainable weight loss and access the treatments and care currently available on the NHS.

8. Is there an unmet need for patients with this condition?	There is an unmet need given the increasing prevalence of type 2 diabetes in the population. Our statistics show that the prevalence of diabetes has more than doubled in the last 15 years with 4.9 million now living with the condition in the UK and 90% of these having type 2 diabetes. It is also estimated that 90% of adults with type 2 diabetes are living with overweight or obesity at diagnosis. The 2020-2021 National Diabetes Audit showed that almost half of adults with type 2 diabetes in England had an HbA1c level above 53mmol/mol. Furthermore, 36.6% had an HbA1c level higher than 58 mmol/mol, which is the treatment target threshold for intensifying drug treatment if target HbA1c is not met by a single drug as advised in the 'Type 2 Diabetes in Adults' guideline [NG28]. This suggests that further treatment options are needed which can help people improve their blood glucose management and therefore lower their risk of developing devastating and potentially life-threatening diabetes complications. This would also offer patients more choice for their care, reduce the risk of people feeling a sense of hopelessness and could encourage a shared-decision making approach where a full range of options are discussed. Losing weight is another key goal for many people living with type 2 diabetes and weight loss of around 5% or more has been shown to reduce HbA1c, cholesterol and blood pressure in people living with the condition. There are other significant improvements in quality of life such as mobility, physical and sexual function reported following weight loss. This further demonstrates the clear need for further

	treatments that are well-evidenced to promote weight loss as well as reducing blood glucose levels with minimal side effects.

Advantages of the technology

9. What do patients or carers think are the advantages of the technology?	• Potential to improve glycaemic outcomes • Potential to improve quality of life • Potential for weight loss • Reduction in risk of developing diabetes-related complications • Additional choices available when it comes to treatment options

Disadvantages of the technology

10. What do patients or carers think are the disadvantages of the technology?	• Administered as an injection instead of orally

Patient population

11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	Some people living with diabetes who have a fear of needles or dexterity problems may find administering this technology a challenge. The reduced risk of hypoglycemia from this treatment compared to other alternatives may also make it more beneficial for other people.

Equality

12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	There is a higher risk of being diagnosed with type 2 diabetes, and at a younger age, for people of South Asian, Black Caribbean and Black African ethnic background. There is also a higher prevalence of the condition amongst those in more deprived areas and they receive poorer care which is borne out in consistently poorer achievement of care processes and treatment targets. Obesity also disproportionately impacts these groups.

Other issues

13. Are there any other issues that you would like the committee to consider?	

Key messages

14. In up to 5 bullet points, please summarise the key messages of your submission.	• Type 2 diabetes is a serious and sometimes progressive condition that deeply impacts health and wellbeing and can cause devasting, life-changing complications • Prevalence of the condition is also growing, closely tied to fast-rising levels of obesity in society • Reduction of blood glucose levels and weight loss are proven ways to improve condition and reduce risk of complications so an additional treatment with these benefits is very important for many people living with type 2 diabetes • This offers another welcome option for people with type 2 diabetes when developing an individualised treatment plan with their healthcare team

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

NICE National Institute for Health and Care Excellence

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

Please select YES if you would like to receive information about other NICE topics - YES or NO

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]

Clinical expert statement

Information on completing this form

In part 1 we are asking for your views on this technology. The text boxes will expand as you type.

In part 2 we are asking you to provide 5 summary sentences on the main points contained in this document.

Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable. Please type information directly into the form.

Do not include medical information about yourself or another person that could identify you or the other person.

We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.

Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.

Clinical expert statement

Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]

NICE National Institute for Health and Care Excellence

Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information redacted. See the NICE health technology evaluation guidance development manual (sections 5.4.1 to 5.4.10) for more information.

The deadline for your response is by **5pm** on **Tuesday 30 May 2023.** Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Thank you for your time.

We reserve the right to summarise and edit comments received, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

Clinical expert statement

Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]

NICE National Institute for Health and Care Excellence

Part 1: Treating type 2 diabetes and current treatment options

Table 1 About you, aim of treatment, place and use of technology, sources of evidence and equality

1. Your name	[redacted]
2. Name of organisation	Association of British Clinical Diabetologists (ABCD)
3. Job title or position	Professor of Medicine (Diabetes), Swansea University Medical School and Swansea Bay University Health Board
4. Are you (please tick all that apply)	☒ A representative of a healthcare professional organisation that represents clinicians ☒ A specialist in the treatment of people with type 2 diabetes ☒ A specialist in the clinical evidence base for type 2 diabetes or technology
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☒ Other; I am not aware that ABCD have submitted a submission.
6. If you wrote the organisation submission and/or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission)	☐
7. Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	None

8. What is the main aim of treatment for type 2 diabetes? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability)	The initial aim is to reverse the symptoms of high glucose levels (hyperglycaemia). Thereafter, there are several aims: Avoidance of deleterious effects of glucose lowering therapies (e.g. hypoglycaemia, weight gain) Reduce the risk of diabetes microvascular complications (glucose-related, such as diabetic retinopathy, neuropathy) Reduce the risk of macrovascular complications (not specific to diabetes but seen more frequently in people with type 2 diabetes) such as heart disease, cerebrovascular disease and peripheral arterial disease Reduction of obesity is now cited as a major aim of therapy by the American (ADA) and European (EASD) guidelines, updated September 2022
9. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount)	Clinically significant would mean achieving HbA1c targets, which would be individually set but typically between 48-64 mmol/mol (6.5-8.0%) without hypoglycaemia. For those individuals with type 2 diabetes who are overweight or obese, then achievement of a normal body mass index (BMI), corrected for ethnicity, would be a good treatment response.
10. In your view, is there an unmet need for patients and healthcare professionals in type 2 diabetes?	There is clearly an unmet need since, despite the availability of eight different classes of glucose lowering therapies (in additional to lifestyle interventions), less than 2/3rds of people with type 2 diabetes in the UK achieve an HbA1c <53 mmol/mol (<7%). In addition, over 90% of people with type 2 diabetes are classified as overweight or obese.
11. How is type 2 diabetes currently treated in the NHS?	Type 2 diabetes in England and Wales is currently managed according to NICE Guideline (NG) 28, which was first published in 2015 and most recently updated in March 2022. A further update, to take into account cardiovascular trial data, is on-going and anticipated to be published at the end of 2024. The pathway of

• Are any clinical guidelines used in the treatment of the condition, and if so, which? • Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.) • What impact would the technology have on the current pathway of care?	care is well-defined but there has been a change in practice recently with sodium-glucose co-transporter 2 inhibitors being recommended as a co-first line glucose lowering medicine. This follows positive data from cardiovascular outcome trials (CVOTs) for the class, followed by evidence for protection from heart failure and chronic kidney disease, conditions which are seen more frequently in people with type 2 diabetes. These latter benefits appear to be independent of glucose lowering. The NICE guidelines are out of kilter with most type 2 diabetes guidance from Europe, North America and elsewhere, in that they do not prioritise the use of glucagon-like peptide 1 (GLP-1) receptor agonists. Most of the GLP-1 receptor agonists have shown superiority over placebo in large CVOTs but these outcomes did not influence the most recent update of NG28. As a result, GLP-1 receptor agonists are typically the last line of glucose lowering therapy in the UK, versus second- (or even first-) line for people with type 2 diabetes at high cardiovascular risk in Europe. Tirzepatide is the first dual incretin agonist glucose lowering medicine, interacting with both the GLP-1 receptor and the receptor for glucose-dependent insulinotrophic polypeptide (GIP). It is a peptide injection, self-administered on a weekly basis. It is likely, therefore, that it will be positioned in the same place as the current GLP-1 receptor agonists.
12. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice? • How does healthcare resource use differ between the technology and current care?	The technology has not been launched in the UK and so has not been used in the NHS. The resource use of Tirzepatide is likely to be similar to that of the currently available (albeit in limited supplies) once weekly GLP-1 receptor agonist

• In what clinical setting should the technology be used? (for example, primary or secondary care, specialist clinic) • What investment is needed to introduce the technology? (for example, for facilities, equipment, or training)	injectables, although the titration schedule (a starting dose and five potential escalations) is more onerous. Given that this is being promoted as the first of a new class of glucose lowering therapies (incretin dual agonists), it is likely that it will be initiated in specialist care only to begin with. Going forwards, those primary care practices which are familiar with, and currently initiate, GLP-1 receptor agonists are likely to take this on. There is no additional training or extra facilities needed for those areas (in both primary and specialist care) already involved in the prescribing of GLP-1 receptor agonist injectables.
13. Do you expect the technology to provide clinically meaningful benefits compared with current care? • Do you expect the technology to increase length of life more than current care? • Do you expect the technology to increase health-related quality of life more than current care?	The SURPASS clinical trial programme in people with type 2 diabetes has reported superior glucose lowering and weight reduction for Tirzepatide versus other glucose lowering classes, including basal insulin and a GLP-1 receptor agonist injection (semaglutide 1mg QW). Although a higher dose of semaglutide is now licenced for glucose lowering (2mg QW), it is likely that the higher doses of Tirzepatide will achieve better HbA1c and weight reductions and one might expect these to provide clinically meaningful benefits. A head-to-head CVOT is being performed for Tirzepatide versus the GLP-1 receptor agonist dulaglutide (1.5mg QW). This aims to demonstrate CV safety and show that Tirzepatide does not shorten length of life in comparison with a drug which has demonstrated CV superiority over placebo in people with type 2 diabetes.

14. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	Tirzepatide, in common with GLP-1 receptor agonists, is unlikely to be beneficial for those people with type 2 diabetes who have a low BMI (and are likely to be insulin deficient). It should be noted that any BMI cut-off would be much lower than the original threshold proposed by NICE for the first GLP-1 receptor agonist (exenatide) and which has remained in NICE guidance ever since; an ethnically adjusted BMI of 35 Kg/m2 was never based on clinical trial evidence and was purely a cost issue (being the BMI at which exenatide was the same price as once daily U100 glargine insulin, at that time 'Lantus').
15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use? (For example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed)	There will be no difference in managing Tirzepatide to any of the other currently used injectable GLP-1 receptor agonists.
16. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	NICE currently has restrictive starting rules for the GLP-1 receptor agonists, which contributes to low levels of prescribing in the UK. The stopping rules for this class, requiring an HbA1c decline of 11mmol/mol (1%) over six months and weight loss of 3% over the same period are also restrictive. Most specialists would continue a GLP-1 receptor agonist at this point and add basal insulin, rather than stopping it and starting 'from scratch' with an insulin regime. Clinical trials support the specialist clinician view. If NICE adopted the same starting and stopping rules to Tirzepatide as it does for GLP-1 receptor agonists, then no additional testing would be necessary.

17. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation? • Do the instruments that measure quality of life fully capture all the benefits of the technology or have some been missed? For example, the treatment regimen may be more easily administered (such as an oral tablet or home treatment) than current standard of care	The QALY calculations will capture both HbA1c and weight reductions but may miss more subtle PROMs data, especially related to the substantial weight losses seen with Tirzepatide.
18. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met? • Is the technology a 'step-change' in the management of the condition? • Does the use of the technology address any particular unmet need of the patient population?	The creation of the first incretin dual-receptor agonist may be a step-change improvement in the management of type 2 diabetes. Unmet needs of inadequate glucose control and limited weight loss will be partially addressed (fully in some cases) by this technology over existing therapies.
19. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	The adverse events seen with Tirzepatide are gastro-intestinal (nausea, diarrhoea & vomiting) as has been reported with GLP-1 receptor agonists and appear to be reported at a similar frequency and severity.
20. Do the clinical trials on the technology reflect current UK clinical practice? • If not, how could the results be extrapolated to the UK setting? • What, in your view, are the most important outcomes, and were they measured in the trials?	I feel that the results from the SURPASS trial programme can be extrapolated to the UK (and sites from the UK were included in the programme)? The most important outcomes from the SURPASS trials were safety, reduction in HbA1c (the primary outcome in each trial) and reduction in weight. All were measured and reported.

• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes? • Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Reduction in HbA1c is a standard surrogate and predicts long term outcomes of type 2 diabetes, especially microvascular ones. No unexpected adverse events have emerged to this point.
21. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
22. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE guideline [NG28]?	The SUSTAIN FORTE trial has examined the use of high dose semaglutide (Ozempic 2mg QW); I do not believe that this was included in the update of NG28 in March 2022.
23. How do data on real-world experience compare with the trial data?	To date, there have been no publications of real-world data from health systems which have already adopted this technology (e.g. USA)
24. NICE considers whether there are any equalities issues at each stage of an evaluation. Are there any potential equality issues that should be taken into account when considering this condition and this treatment? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics. Please state if you think this evaluation could	I am not aware of any potential equality issues at this point in the evaluation.

• exclude any people for which this treatment is or will be licensed but who are protected by the equality legislation • lead to recommendations that have a different impact on people protected by the equality legislation than on the wider population • lead to recommendations that have an adverse impact on disabled people. Please consider whether these issues are different from issues with current care and why. More information on how NICE deals with equalities issues can be found in the NICE equality scheme. Find more general information about the Equality Act and equalities issues here.	

Part 2: Key messages

In up to 5 sentences, please summarise the key messages of your statement:

Tirzepatide is the first of a new class of glucose lowering therapy which, as a single peptide molecule, stimulates the GLP-1 and GIP receptors

Tirzepatide is self-administered as a once weekly subcutaneous injection

The reduction of HbA1c with Tirzepatide is better than seen with placebo and other active glucose lowering therapies

The reduction of weight (and body mass index) with Tirzepatide is better than seen with placebo and other active glucose lowering therapies

The safety and side-effect profile of Tirzepatide are similar to those seen with GLP-1 receptor agonists; results of a cardiovascular outcomes trial are awaited

Thank you for your time.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☒ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

Personal statement

- Diabetes is associated with significant morbidity and mortality and is a major contributor to vascular conditions such as retinopathy, nephropathy, diabetic foot disease and risk of amputation, myocardial infarctions and strokes
- A major risk factor for these complications is poor glycaemic (glucose) control
- Diabetes management consumes 10% of the NHS budget
- 80% or more of the costs of diabetes are consumed by the management of diabetes complications
- Given these observations, a more aggressive approach to improving glycaemic control and other risk factors to facilitate a reduction in diabetes complications is warranted. However recent evidence suggests that less than a half of individuals with type 2 diabetes achieve optimal glycaemic control and this figure is not significantly improving
- Therefore, additional glucose lowering therapies are warranted to assist individuals and populations to achieve optimal glycaemic control. From my own experience and others, coupled with study data, existing glucose lowering therapies do not always achieve the desired improvement in glycaemic control and/or are not tolerated thereby contributing to sub-optimal metabolic management and increased complication risk
- Compounding the difficulties in achieving glycaemic control, the majority of individuals with type 2 diabetes are overweight/obese which is associated with insulin resistance that worsens the ability to obtain optimized glycaemic control and is also associated with increased health issues per se
- Tirzepatide , (a dual glucose-dependent insulinotropic polypeptide and glucagon-like peptide-1 (GLP-1) receptor agonist) is a once weekly injectable glucose lowering therapy that offers a unique approach to improve glycaemic control and is associated with weight loss
- The SURPASS trials published so far indicate HbA1c drops of between 1.91 and 2.59 % (21 and 29 mmol/mol), greater drops than are seen with oral glucose lowering therapies or GLP-1 RA injections
- The SURPASS trials also demonstrate major weight loss of between 6.3 and 12.9 kg, again higher than is typically observed with oral glucose lowering therapies or GLP-1 RA injections
- The SURPASS trials did not raise any significant safety concerns
- At present there is no cardiovascular outcome data with tirzepatide but the SURPASS-CVOT trial is due to be published in 2024
- Based upon this information, I support the introduction of tirzepatide into clinical practice for treating glycaemic control in type 2 diabetes
- Where improvement in glycaemic control is required (in the absence of existing cardiovascular disease or high cardiovascular risk), I would support the use of tirzepatide as an option alongside other GLP-1 RA injections (as is the approach in the American Diabetes Association guidelines). In particular, given the magnitude of improvement in glycaemic control (and accompanying weight loss), this would be a very useful therapy for those individuals with an HbA1c greater than 2% (22 mmol/mol) above target which is not typically achieved with GLP-1 RAs
- Given insulin is typically associated with increased weight gain, higher hypoglycaemia risk than other glucose lowering therapies, requires daily injection(s) and more vigorous monitoring of glucose measurements, I would support the use of Tirzepatide before the use of insulin

Professor [redacted] MD FRCP FRCP (Edin)

Consultant in Diabetes and Endocrinology, Portsmouth Hospitals University Trust

in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]

Produced by Kleijnen Systematic Reviews (KSR) Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University Medical Center (UMC+)

Authors
Nigel Armstrong, Health Economics Manager, KSR Ltd, United Kingdom (UK)
Bram Ramaekers, Health Economist, Maastricht UMC+
Evangelos Danopoulos, Systematic Reviewer, KSR Ltd, UK
Sabine Grimm, Health Economist, Maastricht UMC+
Andrea Fernandez Coves, Health Economist, Maastricht UMC+
Mirre Scholte, Health Economist, Maastricht UMC+
Xiaoyu Tian, Health Economist, KSR Ltd, UK
Jiongyu Chen, Health Economist, KSR Ltd, UK
Lisa Stirk, Information Specialist, KSR Ltd, UK
Rachel Croft, Information Specialist, KSR Ltd, UK
Manuela Joore, Health Economist, Maastricht UMC+, the Netherlands
Robert Wolff, Managing Director, KSR Ltd, UK

Correspondence to
Nigel Armstrong, Kleijnen Systematic Reviews Ltd
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, YO19 6FD
United Kingdom

Date completed 17/03/2023

Source of funding: This report was commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme as project number NIHR135771.

Declared competing interests of the authors: None.

Acknowledgements

We gratefully acknowledge the expert clinical advice input from Dr Mohammed Kamrudeen, Honorary Senior Lecturer and Consultant Endocrinologist, Hull University Teaching Hospitals NHS Trust.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Confidential comparator prices are highlighted in green throughout the report.

Any de-personalised data are highlighted in pink throughout the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Armstrong N, Ramaekers B, Danopoulos E, Grimm S, Fernandez Coves A, Scholte M, Tian X, Chen J, Stirk L, Croft R, Joore MA, Wolff R. Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2023.

Contributions of authors

Nigel Armstrong acted as project lead on this assessment, critiqued the clinical effectiveness evidence and economic evaluation and contributed to the writing of the report. Bram Ramaekers acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Sabine Grimm, Andrea Fernandez Coves, Mirre Scholte and Manuela Joore acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Evangelos Danopoulos, Xiaoyu Tian and Jiongyu Chen acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Lisa Stirk and Rachel Croft critiqued the search methods in the submission and contributed to the writing of the report. Thomas Debray acted as JAVA expert and biostatistician, critiqued the statistical evidence, and facilitated the critique of the economic model. Robert Wolff critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report, and supervised the project.

Abbreviations

ADA	American Diabetes Association
AE	adverse event
AIC	academic in confidence
APPADL	Ability to Perform Physical Activities of Daily Living
AWMSG	All Wales Medicines Strategy Group
BG	blood glucose
BID	twice a day
BMI	body mass index
BP	blood pressure
BPM	beats per minute
BRAVO	Building, Relating, Assessing, and Validating Outcomes
CADTH	Canadian Agency for Drugs and Technologies in Health
CEA	cost effectiveness analysis
CFB	change from baseline
CI	confidence interval
CIC	commercial in confidence
CKD	chronic kidney disease
COPD	chronic obstructive pulmonary disease
CRD	Centre for Reviews and Dissemination
CrI	credible interval
CS	company submission
CSR	Clinical Study Report
CV	cardiovascular
CVD	cardiovascular disease
CVOT	Cardiovascular Outcomes Trial
DBP	diastolic blood pressure
DIC	deviance information criterion
DPP-4	dipeptidyl-peptidase 4
DPP-4i	dipeptidyl-peptidase 4 inhibitors
DSA	deterministic sensitivity analyses
EASD	European Association for the Study of Diabetes
ECG	electrocardiogram
eGFR	estimated Glomerular Filtration Rate
EPI	epidemiology
EQ-5D	European Quality of Life-5 Dimensions
EAG	Evidence Assessment Group
ESRD	end stage renal disease
EU	Europe
EUR	Erasmus University Rotterdam
FAS	full analysis set
FE	fixing errors
FPG	fasting plasma glucose
FSG	fasting serum glucose
FV	fixing violations
GI	gastrointestinal
GLP-1	glucagon-like peptide-1
Hb	haemoglobin
HbA1c	glycated haemoglobin
HDL	high density lipoprotein
HDL-C	high density lipoprotein-cholesterol
HE	Health Economic
HQO	Health Quality Ontario
HR	hazard ratio

HR	heart rate
HRQoL	health-related quality of life
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
ICER	Institute for Clinical and Economic Review
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IWQOL-LITE-CT	Impact of Weight on Quality of Life-Lite Clinical Trials Version
JAGS	Just Another Gibbs Sampler
KDIGO	Kidney Disease Improving Global Outcomes
KSR	Kleijnen Systematic Reviews Ltd
LDL	low density lipoprotein
LDL-C	low density lipoprotein-cholesterol
LIRA	liraglutide
LYs	life years
MACE	major adverse cardiovascular events
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical Subject Headings
MHRA	Medicines and Healthcare Products Regulatory Agency
mITT	modified Intent-to-Treat
MJ	matters of judgement
NASH	non-alcoholic steatohepatitis
N/A	not applicable
NCPE	National Centre for Pharmacoeconomics
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	network meta-analysis
NPH	neutral protamine Hagedorn
NR	not reported
OAD	oral antidiabetic drug
ORs	odds ratios
OUS	outside the USA
PAS	Patient Access Scheme
PBAC	Pharmaceutical Benefits Advisory Committee
PBS	Pharmaceutical Benefits System
PICO	population, intervention, comparator and outcome
PPG	postprandial glucose
PRESS	Peer Review of Electronic Search Strategies
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life year
QD	once a day
QoL	quality of life
QW	once weekly
RA	receptor agonist
RCT	randomised controlled trial
RR	relative risk; risk ratio
SAE	serious adverse event
SBP	systolic blood pressure
SBU	Swedish Agency for Technology Assessment and Assessment of Social Services
ScHARR	School of Health and Related Research

SD	standard deviation
SEMA	semaglutide
SGLT-2i	sodium-glucose co-transporter-2 inhibitor
SIGN	Scottish Intercollegiate Guidelines Network
SLR	systematic literature review
SMD	standardised mean difference
SMQ	standardised MedDRA query
SmPC	summary of product characteristics
SoC	standard of care
SOC	system organ class
SPSL	severe pressure sensation loss
STA	Single Technology Appraisal
SU	sulfonylurea
T1D	type 1 diabetes
T2D	type 2 diabetes
TEAE	treatment emergent adverse events
TZD	thiazolidinediones
TZP	tirzepatide
UK	United Kingdom
UKPDS	United Kingdom Prospective Diabetes Study
UMC+	University Medical Center+
US	United States
USA	United States of America
WBC	white blood cell count
WHO	World Health Organization

Table of Contents

Table of Tables

Table of Figures

1. EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the Evidence Assessment Group (EAG) as being potentially important for decision making. If possible, it also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 presents the key model outcomes. Section 1.3 discusses the decision problem, Section 1.4 issues relates to the clinical effectiveness, and Section 1.5 issues related to the cost effectiveness. Other key issues are discussed in Section 1.6 while a summary is presented in Section 1.7.

Background information on the condition, technology and evidence and information on key as well as non-key issues are in the main EAG report, see Sections 2 (decision problem), 3 (clinical effectiveness) and 4 (cost effectiveness) for more details.

All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1 Overview of the EAG's key issues

Table 1.1: Summary of key issues

ID1457	Summary of issue	Report Sections
1	Mismatch between scope and decision problem in terms of line of therapy and comparators might lead to a lack of evidence for the scope of interest in decision making	2
2	Mismatch between decision problem and evidence in terms of line of therapy/OAD therapy experience might lead to an overestimate of the effectiveness and cost effectiveness of tirzepatide	2, 3, and 4
3	Mismatch between the administration of tirzepatide in clinical practice by titration and the tirzepatide trial evidence, the NMA and the CEA, according to maintenance dose strata, is likely to lead to biased estimates of effectiveness and cost effectiveness in an unknown direction	2, 3, and 4
4	Lack of comparative evidence on the effect of treatments on macro- and micro-vascular complications	3
5	NMA of high risk of bias due to lack of feasibility assessment/assessment of trial comparability and insufficient sensitivity analyses	3.3, and 3.4
6	Model approach adopted by the company is not adequately justified	4.2.2
7	Selection and use of risk models to estimate complications not adequately justified	4.2.2
8	Extrapolation of treatment effectiveness: lack of justification for no treatment waning	4.2.6, and 5.1
9	Treatment discontinuation/intensification: limited reasons for discontinuation	4.2.6
10	AEs: not all incorporated for all treatments	4.2.7
11	Age-adjustment for utility values: none for older age	4.2.8
12	Discrepancies related to utility and cost values	4.2.8, 4.2.9
13	Potentially inappropriate PSA	5.1

ID1457	Summary of issue	Report Sections
14	No full deterministic one-way sensitivity analyses provided	5.2
15	Technical verification insufficient/model results not reproducible	5.3
AEs = adverse events; CEA = cost effectiveness analysis; NMA = network meta-analysis; OAD = oral antidiabetic drug; PSA = probabilistic sensitivity analysis		

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost per QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Reductions in diabetes-related complications associated with reductions in glycated haemoglobin (HbA1c) and body mass index (BMI)

Overall, the technology is modelled to affect costs by:

- Additional treatment costs associated with tirzepatide
- Reductions in diabetes-related complication costs (greatest cost savings were associated with cardiovascular events avoided)

It should be noted that one-way sensitivity analyses of all input parameters were not provided by the company. Hence, an opportunity was therefore missed to identify potentially influential parameters.

1.3 The decision problem: summary of the EAG's key issues

Table 1.2: Key issue 1: Mismatch between scope and decision problem in terms of line of therapy and comparators might lead to a lack of evidence for the scope of interest in decision making

Report Section	2
Description of issue and why the EAG has identified it as important	The population in the NICE scope is much broader than in the decision problem, which is limited to combination therapy and only a line of therapy consistent with GLP-1 RAs in response to failure of at least three OADs.
What alternative approach has the EAG suggested?	The EAG requested clarification that the company's intention was to only address the clinical and cost effectiveness of tirzepatide as a combination therapy only and in the restricted population described, which the company confirmed was the case.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Additional evidence would be required if a decision was to be made for tirzepatide as monotherapy where metformin is inappropriate. See Key issue 2 for combination therapy.
EAG = Evidence Assessment Group; GLP-1 RAs = glucagon-like peptide-1 receptor agonists; NICE = National Institute for Health and Care Excellence; OADs = oral antidiabetic drugs	

Table 1.3: Key issue 2: Mismatch between decision problem and evidence in terms of line of therapy/OAD therapy experience might lead to an overestimate of the effectiveness and cost effectiveness of tirzepatide

Report Section	2, 3, 4
Description of issue and why the EAG has identified it as important	The population in the decision problem is different to that in the SURPASS trials and the NMA in that almost no patients have experienced triple OAD therapy, most having failed on only metformin or metformin plus one other OAD. A clinical expert did suggest that GLP-1 RAs might be given at an earlier line of therapy, which is inconsistent with NICE Guideline NG28, but does seem to be consistent with the ADA/EASD consensus report, but this might also mean the other OADs, e.g., and SGLT-2i might be comparators. There is only a little evidence on whether OAD experience might be a treatment effect modifier. This is in the form of a subgroup analysis in SURPASS-4, which is the only trial where concomitant triple OAD therapy is possible, that suggests an interaction of OAD combination on the treatment effect, but the direction of effect is unclear.
What alternative approach has the EAG suggested?	The EAG suggested that the decision problem be amended to more consistent with the evidence, but the company reiterated that the line of therapy and GLP-1 RAs as comparators were how they expected tirzepatide to be given in clinical practice.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	If the line of therapy is earlier than failure of three OADs then the SURPASS trials might be more appropriate, but this needs to be recognised in the decision problem. Consideration then also needs to be given to comparison with OADs e.g., and SGLT 2i in the NMA and the CEA. Scenario analyses assuming the population characteristics from the SURPASS trials (instead of based on THIN second intensification cohort) would then also be appropriate. If, on the other hand, the decision problem does not change, then there remains uncertainty as to the appropriateness of the clinical evidence.
ADA = American Diabetes Association; CEA = cost effectiveness analysis; EAG = Evidence Assessment Group; EASD = European Association for the Study of Diabetes; GLP-1 = glucagon-like peptide-1; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; OADs = oral antidiabetic drugs; RA = receptor agonists; SGLT-2i = Sodium-glucose co-transporter-2 inhibitor	

1.4 The clinical effectiveness evidence: summary of the EAG's key issues

Table 1.4: Key issue 3: Mismatch between the administration of tirzepatide in clinical practice by titration and the tirzepatide trial evidence, the NMA and the CEA, according to maintenance dose strata, is likely to lead to biased estimates of effectiveness and cost effectiveness in an unknown direction

Report Section	2
Description of issue and why the EAG has identified it as important	The marketing authorisation for tirzepatide is for it to be administered via titration from a maintenance dose of 5 mg, through 10 mg to 15 mg as required to obtain an adequate response in HbA1c reduction. However, the comparisons between tirzepatide and the GLP-1 RAs, SURPASS trials, the NMA and the CEA are

Report Section	2
	stratified by maximum maintenance dose into 5 mg, 10 mg and 15 mg, without titration being permitted. This means that there is lack of applicability to clinical practice. Given the observation in the SURPASS trials and the NMA, which includes SURPASS-2 and SURPASS-3, of a dose response relationship for glycaemic, as well as body weight/BMI control, it is likely that efficacy would be underestimated for the 5 mg and overestimated for the 15 mg stratum. It also appears that all the comparator trials were designed in the same way.
What alternative approach has the EAG suggested?	The EAG would prefer a comparison of treatments as in clinical practice, including titration as appropriate. This would also mean that the treatment strategies in the economic model would not be restricted to within dose steps but include the possibility for individual patients to switch between treatment dosages for those treatments that are titrated. Given the current nature of the comparison, the EAG would tentatively suggest that, if the 5 mg and the 15 mg dose outcomes might be an under or overestimate respectively, then the 10 mg outcomes might be closest to titration. An equivalent analysis of the comparator outcomes, notwithstanding that some are not titrated and some available in only two dose levels, suggests that the company's chosen comparator doses for the tirzepatide 10 mg dose might also be the most appropriate.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Ideally, a comparison of treatments as they would be administered in clinical practice is required but appears that no such evidence exists. The economic model should also be updated to allow patients to switch between treatment dosages to make comparisons between treatments that are titrated as in clinical practice.
BMI = body mass index; CEA = cost effectiveness analysis; EAG = Evidence Assessment Group; GLP-1 = glucagon-like peptide-1; HbA1c = glycated haemoglobin; NMA = network meta-analysis; RA = receptor agonists	

Table 1.5: Key issue 4: lack of comparative evidence on micro and macrovascular complications

Report Section	3
Description of issue and why the EAG has identified it as important	The outcomes in the trials included in the CS are surrogates for the micro- and macrovascular complications. Therefore, it is uncertain what the treatment effect would be on these final endpoints. One tirzepatide trial, SURPASS-CVOT, was identified as reporting some of these outcomes, in particular MACE, but it was reported to be ongoing. For the comparators, other CVOTs were excluded from the NMAs, but these were only of the surrogates.
What alternative approach has the EAG suggested?	The CS does not contain the data required for this type of comparison.
What is the expected effect on the cost effectiveness estimates?	Unknown.

Report Section	3
What additional evidence or analyses might help to resolve this key issue?	The SURPASS-CVOT trial should be combined with any similar comparator trials in order to provide this comparative evidence.
CS = Company submission; CVOT = cardiovascular outcomes trial; EAG = Evidence Assessment Group; MACE = major adverse cardiovascular events; NMA = network meta-analysis	

Table 1.6: Key issue 5: NMA of high risk of bias due to lack of feasibility assessment/assessment of trial comparability and insufficient sensitivity analyses

Report Section	3.3
Description of issue and why the EAG has identified it as important	The NMA was based on a SLR not specific to the CS submitted to NICE. Trials were included without a systematic assessment of heterogeneity and with an assumption that the treatment effect is independent of concomitant background OAD therapy. Substantial heterogeneity seems to exist and have to some degree been identified by the company, but appropriate sensitivity analyses were not conducted. Also, two different estimands were used in the same NMA network, one including and the other excluding patients who required rescue therapy.
What alternative approach has the EAG suggested?	The EAG recommended a feasibility assessment/assessment of trial comparability based on potential treatment effect modification and sensitivity analyses to exclude trials to improve comparability as appropriate.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	The EAG continue to recommend a feasibility assessment/assessment of trial comparability based on potential treatment effect modification and sensitivity analyses to exclude trials to improve comparability as appropriate.
CS = company submission; EAG = Evidence Assessment Group; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; OAD = oral antidiabetic drug; SLR = systematic literature review	

1.5 The cost effectiveness evidence : summary of the EAG's key issues

The company's cost effectiveness results are presented in Section 5, the EAG's summary and detailed critique in Section 4, and the EAG's suggested amendments to the company's model are presented in Section 6. The key issues in the cost effectiveness evidence are discussed in the issue Tables below.

Table 1.7: Key issue 6: Model approach adopted by the company

Report Section	4.2.2
Description of issue and why the EAG has identified it as important	The EAG did not find compelling justification to support the company's modelling approach. This includes 1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the UKPDS OM2 model or CORE Diabetes Model that were used for (updating of the) NICE Guideline NG28 focusing on the management of T2D and 2) the selected model type, described as a "discrete time event" model instead of commonly used model types such as a DES or individual-patient state transition model. Moreover, the exact technical implementation of the model

Report Section	4.2.2
	was not clear to the EAG, this becomes even more problematic when deviating from commonly used model types.
What alternative approach has the EAG suggested?	The EAG preference would entail either using the CORE Diabetes model (consistent with NICE Guideline NG28) or providing extensive justification with supporting evidence why the company deviated from this approach. Additionally, when deviating from the NICE Guideline NG28 modelling approach, the impact of this should be assessed. Similarly, the EAG would prefer commonly used model types such as a DES or individual-patient state transition model instead of a "discrete time event" model. Moreover, deviating from commonly used model types requires substantial and compelling justification as well as detailed description of the model implementation.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	See above.
EAG = Evidence Assessment Group; NICE = National Institute for Health and Care Excellence; T2D = type 2 diabetes	

Table 1.8: Key issue 7: Selection and use of risk models to estimate complications

Report Section	4.2.2
Description of issue and why the EAG has identified it as important	For the estimation of macrovascular complications and blindness risks, the company adopted a model averaging approach, the justification for this approach was not compelling to the EAG. Moreover, the appropriateness of the selected predictive models to estimate the risk of complications in patients with T2D is not justified (in detail), nor is the applicability to the specific decision problem justified.
What alternative approach has the EAG suggested?	Provide extensive justification for the selection and use of risk models to estimate complications and scenario analyses to examine the impact of the adopted approach.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Scenario analyses selecting single predictive models based on the best match of the derivation cohort to the decision problem (as requested in clarification question B4c). Moreover, extensive justification for the model averaging approach, the selected predictive models and the applicability to the specific decision problem.
EAG = Evidence Assessment Group; T2D = type 2 diabetes	

Table 1.9: Key issue 8: Extrapolation of treatment effectiveness

Report Section	4.2.6 and 5.1
Description of issue and why the EAG has identified it as important	The QALY gains are predominantly accrued after the first year (i.e., beyond the trial time horizon) and mostly likely related to utilities for weight. Hence the extrapolation of (treatment) effectiveness is an important aspect of the model. The company made a simplifying assumption of constant risk factors (i.e., no risk factor progression) for SBP, HDL, LDL and weight (i.e., BMI) after year 1 up to treatment intensification. Moreover, the company did assume no waning of the relative treatment effect while on the initial treatment (i.e., before switching to basal insulin therapy).
What alternative approach has the EAG suggested?	The EAG would prefer assuming UKPDS OM2 risk factor progression for all risk factors (instead of assuming these being constant after the first year up to switching to basal insulin therapy). Moreover, additional justification for assuming no waning of the relative treatment effect (before switching to basal insulin therapy) is warranted.
What is the expected effect on the cost effectiveness estimates?	The alternative approach suggested by the EAG likely increases the estimated ICER.
What additional evidence or analyses might help to resolve this key issue?	See above.
BMI = body mass index; EAG = Evidence Assessment Group; HDL = high density lipoprotein; ICER = incremental cost-effectiveness ratio; LDL = low density lipoprotein; QALY = quality-adjusted life year; SBP = systolic blood pressure; UKPDS = United Kingdom Prospective Diabetes Study	

Table 1.10: Key issue 9: Treatment discontinuation/intensification

Report Section	4.2.6
Description of issue and why the EAG has identified it as important	Patients were assumed to intensify therapy, discontinuing the initial treatment and switching to basal insulin therapy, when HbA1c levels rose above 7.5%. No other reasons (e.g., drug intolerance, patient preferences) for treatment discontinuation were included in the modelling.
What alternative approach has the EAG suggested?	The EAG would prefer including other causes for treatment discontinuation (than reaching the HbA1c threshold).
What is the expected effect on the cost effectiveness estimates?	The alternative approach suggested by the EAG likely increases the estimated ICER.
What additional evidence or analyses might help to resolve this key issue?	See above.
EAG = Evidence Assessment Group; HbA1c = glycated haemoglobin; ICER = incremental cost-effectiveness ratio	

Table 1.11: Key issue 10: Adverse events: not all incorporated for all treatments

Report Section	4.2.7
Description of issue and why the EAG has identified it as important	The main concerns of the EAG relate to that only nausea is incorporated (hypoglycaemia only for basal insulin therapy) as an AE. Including hypoglycaemia only for basal insulin therapy might inflate the impact of discontinuing treatment (i.e., treatment intensification) and hereby potentially inducing bias favouring more effective treatments (i.e., tirzepatide).
What alternative approach has the EAG suggested?	As illustrated in clarification response Tables 20-22, incorporating additional AEs would potentially increase the estimated ICER (but might depend on the comparator).
What is the expected effect on the cost effectiveness estimates?	The alternative approach suggested by the EAG likely increases the estimated ICER.
What additional evidence or analyses might help to resolve this key issue?	See above.
AEs = adverse events; EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio	

Table 1.12: Key issue 11: Age-adjustment for utility values: none for older age

Report Section	4.2.8
Description of issue and why the EAG has identified it as important	The company base-case uses a relatively high utility value for patients with T2D (0.815) and does not adjust utility values for older age. Over time, this potential overestimation will likely only increase as utility values are not adjusted for age.
What alternative approach has the EAG suggested?	The EAG prefers the base-case scenario to include age-adjustment, ensuring that the utility does not exceed the age-matched general population utility.
What is the expected effect on the cost effectiveness estimates?	Using age-adjusted utility values will increase the face validity of the results and will result in a more conservative ICER estimate.
What additional evidence or analyses might help to resolve this key issue?	See above.
EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; T2D = type 2 diabetes	

Table 1.13: Key issue 12: Discrepancies related to utility and cost values

Report Section	4.2.8 and 4.2.9
Description of issue and why the EAG has identified it as important	There are discrepancies in the uncertainty measures and distributions related to utility values and costs listed in the CS and those listed in the original sources.

Report Section	4.2.8 and 4.2.9
What alternative approach has the EAG suggested?	According to the EAG, all input data should be in line with the data presented in the original sources. This includes deterministic values, measures of uncertainty and appropriate distributions.
What is the expected effect on the cost effectiveness estimates?	Uncertain. The discrepancies in the uncertainty measures and distributions related to utility values will either in- or decrease the uncertainty surrounding the model outcomes. Costs mentioned in the CS are both higher and lower than those reported in the original sources, therefore the combined effect on the ICER is difficult to determine.
What additional evidence or analyses might help to resolve this key issue?	See above.
CS = company submission; EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio	

Table 1.14: Key issue 13: Potentially inappropriate PSA

Report Section	5.1
Description of issue and why the EAG has identified it as important	The implementation of the PSA is not clear and includes bootstrapping that is not standard in PSAs. It is unclear whether all imprecision (i.e., all uncertain parameters) is taken into account the PSA, and whether stochastic uncertainty is removed from the PSA.
What alternative approach has the EAG suggested?	Implementation of the PSA according to Corro-Ramos et al 2020.[1]
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Detailed step-by-step explanation of implementation of the PSA.
EAG = Evidence Assessment Group; PSA = probabilistic sensitivity analysis	

Table 1.15: Key issue 14: No full deterministic one-way sensitivity analyses provided

Report Section	5.2
Description of issue and why the EAG has identified it as important	No full deterministic one-way sensitivity analyses (for all input parameters) were provided, and an opportunity was therefore missed to identify potentially influential parameters.
What alternative approach has the EAG suggested?	Implement deterministic one-way sensitivity analyses (for all input parameters) and present results in tornado diagrams (for all doses and in the comparison with semaglutide).

Report Section	5.2
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	See above.
EAG = Evidence Assessment Group	

Table 1.16: Key issue 15: Technical verification insufficient/model results not reproducible

Report Section	5.3
Description of issue and why the EAG has identified it as important	There remain doubts over the internal validity of the model. Model outcomes could not be reproduced by the EAG. The EAG could not find how BMI-related utilities were implemented in the model (black box character). No full overview of input parameters has been provided.
What alternative approach has the EAG suggested?	Correct the model if necessary. Provide step-by-step guide to running the model. Provide a filled in TECH-VER checklist. Provide detailed description of how the BMI-related utilities were implemented and where this can be found in the code. Provide full overview of all input parameters and how they were included in the PSA.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	See above.
BMI = body mass index; EAG = Evidence Assessment Group; PSA = probabilistic sensitivity analysis	

1.6 Other key issues: summary of the EAG's view

Not applicable.

1.7 Summary of the EAG's view

The company's cost effectiveness assessment partly complied with the NICE reference case. The deviation from the NICE reference case related to the type of economic evaluation as the incremental analyses were missing. The most prominent issues highlighted by the ERG are discussed below. These issues were listed as key issues in Section 1.5 and suggestions for analyses to (partly) examine the potential impact of these issues were provided in Sections 6.1.1 and 6.1.2 of this report.

First, the EAG did question the company's modelling approach. This includes 1) the use of the PRIME type 2 diabetes (T2D) model in general instead of commonly used available alternatives mentioned such as the CORE Diabetes Model that was used for NICE Guideline 28 (NG28) focusing on the management of T2D and 2) the selected model type, described as a "discrete time event" model instead of commonly used model types such as a DES or individual-patient state transition model. Moreover,

the exact technical implementation of the model was not clear to the EAG which is particularly problematic because of the deviation from commonly used model types. Similarly, compelling justification was missing for the company's model averaging approach as well as the appropriateness and applicability of the selected predictive models to estimate the risk of complications in patients with T2D.

Second, the population considered in the company submission (CS) was adults with T2D that is inadequately controlled with three or more antidiabetic agents, which is not aligned with the population from the SURPASS trials or the expected UK clinical use. Moreover, the company base-case included three different maintenance doses of tirzepatide: 5 mg, 10 mg, or 15 mg. Comparisons were made within each recommended maintenance dose step, and not between recommended maintenance dose steps. In addition, patients were not able to move between dose steps in the model. This does not seem to reflect clinical practice.

Third, the QALY gains are predominantly accrued after the first year and mostly likely related to utilities for weight. Hence the extrapolation of (treatment) effectiveness is an important aspect of the model. The company made a simplifying assumption of constant risk factors (i.e., no risk factor progression) for systolic blood pressure (SBP), high density lipoprotein (HDL), low density lipoprotein (LDL) and weight (i.e., BMI) after year 1 up to treatment intensification. Moreover, the company did assume no waning of the relative treatment effect while on the initial treatment (i.e., before switching to basal insulin therapy). Additionally, patients were assumed to switch to basal insulin therapy, only incase glycated haemoglobin (HbA1c) levels rose above 7.5%, i.e., no other reasons (e.g., drug intolerance, patient preferences) for treatment discontinuation were included in the modelling.

Fourth, the company base-case used a relatively high utility value for patients with T2D (0.815) and did not adjust utility values for older age, potentially resulting in utility values that are higher than expected for the age-matched general population. Moreover, the EAG highlighted discrepancies in input parameters related to utility values and costs listed in the CS and those listed in the original sources.

Fifth, the implementation of the probabilistic sensitivity analysis (PSA) was not clear and included bootstrapping, which is not standard in PSAs. It is unclear whether all imprecision (i.e., all uncertain parameters) was taken into account in the PSA, and whether stochastic uncertainty was removed from the PSA. Related to this, no full deterministic one-way sensitivity analyses (for all input parameters) were provided, and an opportunity was therefore missed to identify potentially influential parameters.

Finally, there remain doubts over the internal validity of the model. Model outcomes could not be reproduced by the EAG. The EAG could not find how BMI-related utilities were implemented in the model and no full overview of input parameters has been provided.

The CS base-case cost effectiveness results (probabilistic) indicated that tirzepatide 5 mg is both more effective and more costly than the comparators amounting to ICERs ranging between [redacted] per QALY gained (Table 5.1). Tirzepatide 10 mg was more effective in all comparisons and more costly in all comparisons but the one with liraglutide, with ICERs ranging between [redacted] per QALY gained, and tirzepatide 10 mg dominating liraglutide (Table 5.2). A similar pattern of results was projected for tirzepatide 15 mg, which was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg and had ICERs ranging between [redacted] per QALY gained versus the other comparators.

The EAG could not reproduce the company's base-case results locally with JAVA model files provided by company. Moreover, the web version of the model only had limited flexibility to make model

adjustments. Therefore, instead of implementing the EAG base-case, the EAG highlights suggested adjustments for the company to implement and produce the EAG base-case and scenario analyses. These adjustments should be implemented transparently and the EAG should be able to reproduce these analyses. It is expected that the EAG base-case would result in substantially higher ICERs compared with the company base-case as most suggested adjustments would likely increase the estimated ICER (Section 6.1.1), while for the remaining adjustments the impact is unknown.

There is large remaining uncertainty about the (long-term) effectiveness and relative effectiveness of tirzepatide, which can be at least partly resolved by the company by conducting further analyses. According to the EAG the current company's base-case is flawed and the EAG suggested adjustments could conceivably change, most likely increase, the ICER. Moreover, the current assessment does not provide an appropriate estimation of the comparators listed in the scope. Therefore, the EAG believes that the CS base-case does not represent an unbiased ICER of tirzepatide compared with relevant comparators (as would be used in clinical practice).

2. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 2.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the CS	Rationale if different from the final NICE scope	EAG comment
Population	Tirzepatide monotherapy: Adults with T2D that is inadequately controlled with diet and exercise alone and in whom the use of metformin is considered inappropriate. Tirzepatide with other antidiabetic agents: Adults with T2D that is inadequately controlled with one or more antidiabetic agents	Tirzepatide with other antidiabetic agents: Adults with T2D that is inadequately controlled with three or more antidiabetic agents	This submission positions tirzepatide for use in patients with T2D that is inadequately controlled with three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. This is the anticipated positioning of tirzepatide in UK clinical practice	There is a mismatch between the scope and the decision problem and the decision problem and the clinical effectiveness evidence (see Section 2.1).
Intervention	Tirzepatide alone or with other antidiabetic agents	Tirzepatide with other antidiabetic agents	N/A	There is a mismatch between how tirzepatide would be given in clinical practice i.e., the maintenance dose titrated from 5 mg through 10 mg to 15 mg, and how it was considered in the clinical and cost-effectiveness evidence i.e., as a fixed maintenance dose of 5, 10 or 15 mg (see Section 2.2).
Comparator(s)	The following interventions as monotherapy or in combination regimens, in accordance with NICE guidance:	The following interventions in combination regimens: GLP-1 RAs: Dulaglutide	GLP-1 RAs are considered the only relevant comparators for tirzepatide in this submission, as this aligns with the anticipated position for tirzepatide in	The company have not presented a convincing argument for the restriction to GLP 1-RAs. The problem with dosing also applies to

	Final scope issued by NICE	Decision problem addressed in the CS	Rationale if different from the final NICE scope	EAG comment
	sulfonylureas DPP-4 inhibitors pioglitazone GLP-1 mimetics SGLT-2 inhibitors insulin	Exenatide (standard and modified-release formulations) Liraglutide Lixisenatide Semaglutide (oral and injectable formulations)	the UK clinical pathway of care (see above)	the comparators that would be titrated in clinical practice (see Section 2.3).
Outcomes	The outcome measures to be considered include: HbA1c/glycaemic control complications of diabetes, including cardiovascular, renal and eye Mortality BMI Frequency and severity of hypoglycaemia Changes in cardiovascular risk factors Adverse effects of treatment HRQoL	The outcome measures to be included are: HbA1c Change in body weight BMI Frequency and severity of hypoglycaemia Adverse effects of treatment HRQoL (APPADL and IWQOL-LITE-CT)	Aligned with the final NICE scope. A CV safety meta-analysis confirming CV safety is described in Section B.2.9. Further data on CV outcomes are not yet available; they are expected to become available upon completion of the SURPASS-CVOT trial in 2025.[2] A dedicated addendum study to SURPASS-CVOT is ongoing to further investigate the impact of tirzepatide treatment on diabetic retinopathy progression	There are no comparative data on the micro- and macrovascular complications of diabetes, including CV outcomes.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be	NR	NR	Partly: the deviation from the NICE reference case related to the type of economic evaluation as the incremental analyses were missing.

	Final scope issued by NICE	Decision problem addressed in the CS	Rationale if different from the final NICE scope	EAG comment
	sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and PSS perspective. The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account.			
Subgroups to be considered	None.	NR	N/A	N/A

Based on Table 1 and pages 10 to 12 of the CS.[3]
APPADL = ability to perform physical activities of daily living; BMI = body mass index; CS = company submission; CV = cardiovascular; CVOT = Cardiovascular Outcomes Trial; DPP-4 = dipeptidyl-peptidase 4; EAG = Evidence Assessment Group; GLP-1 = glucagon-like peptide-1; HbA1c = glycated haemoglobin; HRQoL = health-related quality of life; IWOOL-LITE-CT = Impact of Weight on Quality of Life-Lite Clinical Trials Version; N/A = not applicable; NHS = National Health Service; NICE = National Institute of Health and Care Excellence; NR = not reported; PAS = Patient Access Scheme; PSS = Personal Social Services; QALY = quality-adjusted life year; RAs = receptor antagonists; SGLT-2 = Sodium-glucose co-transporter-2 inhibitor; T2D = type 2 diabetes; UK = United Kingdom

2.1 Population

The population in the NICE final scope is, for:

- tirzepatide monotherapy:
  - Adults with type 2 diabetes (T2D) that is inadequately controlled with diet and exercise alone and in whom the use of metformin is considered inappropriate
- Tirzepatide with other antidiabetic agents:
  - Adults with T2D that is inadequately controlled with one or more antidiabetic agents

The population in the decision problem is only for:

- Tirzepatide with other antidiabetic agents:
  - Adults with T2D that is inadequately controlled with three or more antidiabetic agents

EAG comment: Table 2.2 shows a comparison between the National Institute for Health and Care Excellence (NICE) scope and the decision problem, as well as the Medicines and Healthcare Products Regulatory Agency (MHRA) license, the NG28 and the main trial evidence for tirzepatide in the form of the SURPASS trials, as presented in the clarification letter,[3-7] highlights the profound mismatch between the scope and the decision problem, which is a subgroup of the scope. It also highlights the mismatch between the main SURPASS trial evidence, which is generally at an earlier line of therapy/less treatment experienced: in fact, two trials exclude triple therapy experience within the 3 months prior to Visit 1: SURPASS-2 excludes any antihyperglycemic medication except metformin, SURPASS-3 excludes any other than metformin or and sodium-glucose co-transporter-2 inhibitor (SGLT-2i), SURPASS-4 does permit triple therapy of metformin, an SGLT-2i and an sulfonylurea (SU), which applied to only about [redacted] of the trial population (Table 14, CS).[3] The network meta-analysis (NMA) studies were also chosen to: *"…align with SURPASS-2 and 3 trials and included studies conducted in patients with one to two OADs as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice"* (page 103, CS).

Table 2.2: Comparison of various population definitions (NICE scope, license, trials, NICE guideline, decision problem)

NICE scope	MHRA therapeutic indications	SURPASS trials populations	SURPASS trials treatment positioning	Treatment positioning of tirzepatide according to CS and NICE NG28	Decision problem addressed in the CS
Tirzepatide monotherapy: • Adults with T2D that is inadequately controlled with diet and exercise alone and in whom the use of metformin is considered inappropriate. Tirzepatide with other	Mounjaro is indicated for the treatment of adults with insufficiently controlled T2D mellitus as an adjunct to diet and exercise • As monotherapy when metformin is considered inappropriate due to intolerance or contraindications	SURPASS-2 Patients with T2D who had inadequate glycaemic control with metformin monotherapy (≥1500 mg/day) and had not been treated with any other OADs during the 3 months prior to the start of the study	Second-line treatment	At the same position as a GLP-1 RA i.e. when triple therapy with metformin and two other oral drugs is not effective, tolerated or contraindicated and patients: have a BMI ≥35 kg/m^2 (adjust accordingly for people from	Tirzepatide with other antidiabetic agents: • Adults with T2D that is inadequately controlled with three or more antidiabetic agents

NICE scope	MHRA therapeutic indications	SURPASS trials populations	SURPASS trials treatment positioning	Treatment positioning of tirzepatide according to CS and NICE NG28	Decision problem addressed in the CS
antidiabetic agents: • Adults with T2D that is inadequately controlled with one or more antidiabetic agents	• In addition to other medicinal products for the treatment of diabetes	SURPASS-3 Patients with T2D who had inadequate glycaemic control on stable doses of metformin with or without an SGLT-2i	Second- or third-line treatment	Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity, or have a BMI <35 kg/m^2 and: for whom insulin therapy would have significant occupational implications or when weight loss would benefit other significant obesity related comorbidities	
		SURPASS-4 Patients with T2D with high CVD risk, who had inadequate glycaemic control on stable doses of at least 1 and no more than 3 oral antidiabetic drugs (OADs), including metformin, an SGLT-2i and/or an SU	Second-, third-or later line treatment		
		SURPASS-5 Patients with T2D, with background therapy of insulin glargine with or without metformin	Unknown line given that prior treatment experience not specified.		

BMI = body mass index; CS = company submission; CVD = cardiovascular disease; GLP-1 RA = glucagon-like peptide-1; MHRA = Medicines and Healthcare Products Regulatory Agency; NICE = National Institute for Health and Care Excellence; RA = receptor agonist; SGLT-2i = sodium-glucose co-transporter-2 inhibitor; T2D = type 2 diabetes

The company were asked to confirm the decision problem population and that there is no evidence that better aligns to it. They were also asked to amend the decision problem to one with better alignment to the evidence. In response to the clarification letter, the company confirmed the decision problem i.e. as combination therapy and a replacement for a glucagon-like peptide-1 receptor antagonist (GLP-1 RA) as it might be prescribed according to NICE Guideline 28 (NG28).[4]

The EAG also noted that the indication for GLP-1 RAs was more precise than only according to treatment experience: *"If triple therapy with metformin and 2 other oral drugs is not effective, not*

tolerated or contraindicated, consider triple therapy by switching one drug for a GLP-1 mimetic for adults with type 2 diabetes who:[7]

- *have a body mass index (BMI) of 35 kg/m2 or higher (adjust accordingly for people from Black, Asian and other minority ethnic groups) and specific psychological or other medical problems associated with obesity or*
- *have a BMI lower than 35 kg/m2 and for whom insulin therapy would have significant occupational implications or weight loss would benefit other significant obesity-related comorbidities."*

It is unclear as to whether the intention would be for these criteria to also apply to tirzepatide: if so then there would be a further discrepancy with the SURPASS trials, which specify a BMI of at least 25 kg/m^2 (SURPASS-2 to 4) or at least 23 kg/m² (SURPASS-5). The company were asked to discuss the implications of this lower BMI threshold in terms of clinical effectiveness, to which they responded that they:[4] *"…consider that the likely position of tirzepatide in NHS practice will be driven by GLP-1 RA use rather than driven by the specific criteria in NG28 themselves and as such have defined the decision problem addressed in terms of GLP-1 RA use rather than the criteria listed in NG28. Nonetheless, given the NG28 restrictions apply in current practice there is no difference in the population between Lilly's definition and the NG28 GLP-1 RA population described in the question."*

In relation to the mismatch between the decision problem and the trial evidence as set out in Table 2.2, the company argued that the subgroup analysis by type of concomitant baseline therapy in SURPASS-4 provided some reassurance about the generalisability of the trial evidence. However, although they noted no significant interaction effect on weight change, they did admit a significant effect on HbA1c change, dismissing this as:[4] *"likely due to the small sample size in the 'other' category leading to high variability."* The EAG do not consider this to be the most likely explanation for the observation of a significant effect of concomitant therapy, the most likely one being that concomitant therapy does impact the treatment effect on glycated haemoglobin (HbA1c). The company confirmed that there are no other studies of tirzepatide in combination with up to three oral antidiabetic agents, as would be the case for GLP-1 RAs according to NG28 (see Table 2.2). They also refused to change the decision problem to an earlier line of therapy in order to better match the lower use of concomitant therapy found in the SURPASS trials.

The EAG did consult a clinical expert, who suggested that GLP-1 RAs might be prescribed at an earlier line of therapy: *"Most clinicians will use a GLP- RA high up in an algorithm and the closest algorithm that most of us use in practice is the ADA/EASD* [American Diabetes Association and the European Association for the Study of Diabetes] *Consensus where second line use is advocated based on clinical parameters*".[8] The EAG can confirm that GLP-1 RAs seem to be recommended as early as at second line in the ADA/EASD consensus report.[9] This might mean that there is a better match between the clinical evidence and the place in the care pathway of the comparators. However, there would still be a mismatch between the decision problem and clinical evidence populations. Therefore, the EAG consider that the mismatch between the decision problem and the trial evidence is a concern, particularly because of the lack of patients at the line of therapy consistent with GLP-1 RAs as comparators at least according to the NICE guidelines, for which there is some evidence of an effect on treatment effect in one trial, SURPASS-4, but for which there can be no further evidence because of this mismatch. This is therefore a key issue.

The EAG would like to point out that defining the population in terms of inadequate glycaemic control (failure) on a particular oral antidiabetic drug (OAD) combination is equivalent to defining the

baseline treatment characteristics in the SURPASS trials and the background (concomitant) therapy when tirzepatide is added to that combination. This has implications therefore for precise nature of the intervention and the comparator i.e., which OADs form the background treatment.

2.2 Intervention

Tirzepatide is administered via injection once weekly (QW), using a single-dose pre-filled autoinjector pen device. The dose should be injected in the abdomen, thigh or upper arm, rotating the injection site with each dose. The dose can be administered at any time of day, with or without meals. Tirzepatide is initiated at 2.5 mg QW. After 4 weeks, increase to 5 mg QW. If needed, the dose can be increased in 2.5 mg increments every 4 weeks up to 15 mg. The recommended maintenance doses are 5 mg, 10 mg and 15 mg (Table 2, CS).[3]

EAG comment: The intervention reported in the decision problem is as in the NICE scope. The license does not specify whether tirzepatide should be given in combination with background OAD therapy, or if given in combination, the nature of that combination. Indeed, consistent with the population in the decision problem, one would expect that the intervention would be in combination with at least two OADs if one OAD is switched for tirzepatide as is recommended for GLP-1 RAs in NG28.[7] However, this is not the case in all of the SURPASS trials, as already stated in Section 2.1 In fact, the only tirzepatide trial that included some patients in the population most consistent with the decision problem i.e. failed on three OAD therapies, although only a small minority, was SURPASS-4, as acknowledged by the company: *"SURPASS-4 is also relevant to this decision problem, as the population is the closest to the anticipated positioning of tirzepatide with some patients having previously received triple oral therapy (metformin, an SU and an SGLT2i)."* (p.31) However, those patients had tirzepatide added to the therapy failed on, rather than one of the OADs being switched for tirzepatide. This was also excluded from the NMA, except in a sensitivity analysis that also incuded studies with unclear proportion of metformin as background therapy (See Section 3.4.2.3). Indeed, the NMA only included SURPASS-2, where only metformin was background therapy and -3, where only a minority had SGLT-2i added to the metformin (see Sections 3.2 and 3.3). The comparators trials were also chosen for their similarity to SURPASS-2 and -3, and so the same applies to those. Therefore, as well as a mismatch in population, there is also an implicit assumption that the effectiveness of all treatments and the treatment effect between tirzepatide and comparators is additively independent i.e., the same regardless of the nature of the background therapy. Given the overlap with the issues already identified population mismatch, this has not been identified as a separate key issue, notwithstanding the key issue relating to comparability of which it is also a part (see Sections 3.3 and 3.4).

However, the cost effectiveness analysis (CEA) compares tirzepatide to comparators defined by dose according to tirzepatide dose i.e., the comparators for 5 mg, 10 mg and 15 mg are not identical. Table 46 in the company submission (CS) shows the same for the NMA. However, Table 2 in the CS suggests that all patients are treated the same and will move from a maintenance dose of 5 mg to 10 mg or 15 mg as required. In the clarification letter the company were asked to justify these three subgroups in relation to United Kingdom (UK) clinical practice. If there would be no such subgroups and all patients will move to higher doses as required in clinical practice, then they were asked to conduct the NMA and CEA without subgroups and allowing a comparison with all comparators at all doses. In response to the clarification letter, the company state that:[4] *"the CS does not specify three subgroups."* The company argued that the CS has been *"misunderstood"* when the EAG argues that comparators varying by maintenance dose imply subgroups by maintenance dose. Instead, they claim that such stratified comparisons are required because tirzepatide is titrated in dose steps equivalent to these maintenance doses i.e., 5 mg, 10 mg and 15 mg. However, this reasoning is flawed because for the very reason that

the company argue it is valid i.e., if tirzepatide is titrated then each patient needs to be able to move to a higher dose as required starting at 5 mg, progressing through 10 mg and then onto 15 mg, but this is not how tirzepatide is analysed in the CEA, the NMA or the SURPASS trials. Instead, patients are constrained to receive only one of the three maintenance doses i.e., those stratified to the lower doses may not proceed to a higher dose and those stratified to the highest dose may not remain at a lower dose.

This is therefore a key issue, for which the EAG have tentatively suggested might be mitigated by assuming that the 10 mg dose stratum will provide evidence that might be closer to a titration-based approach. This is because there does appear to be a dose-response relationship, as presented in Section 3.2.3.

2.3 Comparators

As shown in Table 2.1, the comparators in the NICE scope might be any oral antidiabetic or insulin depending on the line of therapy. However, in the decision problem, the only comparators are the GLP-1 RAs. Also, already mentioned, the NMA (see Section 3.4) and the CEA (see Section 4) stratify the comparators by the intervention dose strata (see Table 2.3).

Table 2.3: Overview of comparators and doses

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide 5 mg	Dulaglutide 1.5 mg Semaglutide 0.5 mg Oral semaglutide 7 mg Liraglutide 1.2 mg
Tirzepatide 10 mg	Dulaglutide 3.0 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
Tirzepatide 15 mg	Dulaglutide 4.5 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
Table 46, CS.[3] CS = company submission	

EAG comment: Most comparators in the scope are omitted from the decision problem. The EAG considers that this might be appropriate if the remaining comparators (GLP-1 RAs) are consistent with the population in the decision problem, which might be the case if tirzepatide would only be prescribed in place of the GLP-1 RAs. However, as stated in Section 2.1, the EAG have requested the decision problem be amended to increase consistency with the SURPASS trials i.e. earlier line and less treatment experienced for SURPASS-2 to 4 or possible later for SURPASS-5, although this is unclear given no mention of prior experience for this trial.[3] Such a new decision problem would imply the possibility of including the omitted comparators. As stated above, the company response to the clarification letter was that GLP-1 RAs are the only comparators.[4] In the response to the clarification letter, the company also reiterated that, given that GLP-1 RAs are the appropriate comparators then sulfonylureas, DDP-4 inhibitors, pioglitazone, SGLT-2 and insulin are not.[4] They stated: "*Due to the anticipated positioning of tirzepatide, sulfonylureas, DDP-4 inhibitors, pioglitazone, SGLT-2 and insulin do not represent relevant comparators as they are prescribed at a different position within the treatment pathway; it was*

therefore not relevant to provide comparative efficacy for them." (page 8[4]). The EAG would like to point out that the excluded comparators could be in the right line of therapy as comparators to tirzepatide if the decision problem was amended to more in line with the SURPASS trials, i.e., after failure on one or two OADs. As reported in Section 2.1, the EAG also elicited clinical expert opinion that GLP-1 RAs might be prescribed earlier in the care pathway.[8] This could potentially mean that some of the other OADs might be appropriate alternatives, which the clinical expert verified was the case: *"an SGLT-2i can use used as an alternative for CV risk reduction".*[8]

The other main issue is the stratification of the comparators, which is problematic for two reasons, considering that the company have rejected the notion of subgroups, thus implying that any comparator should be compared to the intervention. Of course, the EAG have already argued that lack of subgroups by dose of intervention means that the intervention should not be stratified in the first place and that this is based on tirzepatide being titrated as required up to 15 mg, as opposed to constrained to one of three maximum doses. In fact, this applies also to most of the comparator treatments, the NICE recommended doses being:[10]

- For exenatide:
    - Standard-release formulation: 5 mg twice a day (BID), increased, if necessary, after at least 1 month to a maximum dose of 10 mg BID. It should be administered within 1 hour before two main meals (at least 6 hours apart). It should not be administered after a meal
    - Modified-release formulation: 2 mg QW on the same day each week (at any time, with or without meals). The day of weekly administration can be changed, if necessary, as long as the next dose is administered at least 24 hours later

- For liraglutide:
    - 0.6 mg once a day (QD), increased after at least 1 week to 1.2 mg QD. This can be further increased, if necessary, after an interval of at least 1 week to a maximum dose of 1.8 mg QD

- For lixisenatide:
    - 10 mg QD for 14 days, to be taken within 1 hour before the first meal of the day or the evening meal, increased to 20 mg QD thereafter

- For dulaglutide:
    - 0.75 mg QW as monotherapy; 1.5 mg once weekly; increased if necessary to 3 mg once weekly after at least 4 weeks, then increased if necessary to 4.5 mg once weekly after another 4 weeks, a starting dose of 0.75 mg once weekly may be considered for potentially vulnerable patients; maximum 4.5 mg per week.11

- For semaglutide:
    - 0.25 mg QW for 4 weeks, then 0.5 mg QW for at least 4 weeks, then increased to 1 mg QW if needed.

Therefore, the intervention and comparators should be as shown in Table 2.4.

Table 2.4: Overview of comparators and doses

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide titrated up to 15 mg QW	Dulaglutide titrated up to 4.5 mg QW Semaglutide titrated up to 1 mg QW Liraglutide titrated up to 1.8 mg QD Exenatide standard-release titrated up 10 mcg BID Exenatide modified-release 2 mg QW Lixisenatide 20 mcg QD
BID = twice a day; QD = once a day; QW = once weekly	

To further illustrate the point, one might imagine that the comparison between the tirzepatide and comparator within a stratum would make sense if one could, before prescribing the treatment, know which patients would require only the lower, or only the intermediate dose, and which would require the higher dose. However, no means by which this might be achieved has been suggested: indeed, in response to the clarification letter, the company verified this by stating that: *"…the most appropriate dose for an individual should be determined by the clinician in collaboration with the patient, based on clinical characteristics and patient tolerability and cannot be determined a priori…"*.[4]

Of course, it could be argued that the stratification is necessary because this is the form of the trial evidence i.e., neither tirzepatide nor, apparently, any of the comparators have been studied in trials where there is titration. However, as explained in Section 3.4, one cannot know the result of any trade-off between efficacy and safety as a result of titration, which means that it is unclear what proportion of a patient cohort that might be titrated to a particular dose of tirzepatide would be titrated to any of the doses of semaglutide or liraglutide. Put another way, one cannot simply assume that, just because both tirzepatide and semaglutide are available in three doses, that all patients titrated to the middle dose of tirzepatide will also be titrated to the middle dose of semaglutide. However, as with the intervention, the EAG might tentatively suggest that the titrated comparator dose that is closest to clinical practice might be one that lies in the middle, although such a middle dose is only available for dulaglutide. Therefore, it might be conservative to choose the highest of the other two titrated comparator, which means that the comparators aligned with tirzepatide 10 mg in Table 2.3 might be the closest to clinical practice. This is because, as for tirzepatide, there does seem to be a dose-response relationship (see Section 3.4.2).

The inappropriate stratification has already been identified as a key issue in Section 2.2. The mismatch between the decision problem and the clinical evidence population has also already been identified as a key issue, the implications of which also extend to the choice of comparators, i.e., if GLP-1 RAs are comparators and they might be prescribed earlier than on failure of triple OAD therapy, then other OAD therapy, such as an SGLT-2i might also be comparators.

As already argued in Section 2.2, the difference in background therapy between decision problem and trials implied by the population also applies to the comparators. Therefore, as well as a mismatch in population, there is also an implicit assumption that the effectiveness of all treatments and the treatment effect between tirzepatide and comparators is additively independent i.e., the same regardless of the nature of the background therapy. This has therefore been identified as part of the key issues on population mismatch and NMA trial heterogeneity (see Sections 3.3 and 3.4).

2.4 Outcomes

See Table 2.1 for the difference between the scope and the decision problem.

EAG comment: The EAG noted that complications of diabetes, including cardiovascular, renal and eye, and mortality are outcomes defined in the NICE scope but not included in the decision problem addressed by the company. Instead the rates of these micro- and macrovascular complications were estimated using risk models in the economic model as a function of surrogates such as HbA1c, which relies on an assumption of a causal relationship between the treatment effect on change in surrogates and the downstream final outcomes. Therefore, the lack of comparative evidence on micro- and macrovascular complications has been identified as a Key Issue.

2.5 Other relevant factors

None.

3. CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

A systematic literature review (SLR) to identify clinical evidence for the treatment of T2D for the population of the Single Technology Appraisal (STA), targeting tirzepatide and GLP-1 RAs as the relevant comparators. Additional details on the SLR are reported in Appendix D of the CS[12]. The SLR was executed according to a pre-specified protocol which unfortunately was not provided in the CS.

The SLR was conducted in September 2021 and updated in October 2021 and June 2022 to capture any recently published evidence. This Section of the EAG report describes and critiques the methods of the review including searching, inclusion criteria, data extraction, quality assessment and evidence synthesis.

3.1.1 Searches

The following paragraphs contain summaries and critiques of the searches related to clinical effectiveness presented in the CS.[3] The Canadian Agency for Drugs and Technologies in Health (CADTH) evidence-based checklist for the Peer Review of Electronic Search Strategies (PRESS), was used to inform this critique.[13, 14] The CS[3] was checked against the STA specification for company/sponsor submission of evidence.[15] The EAG has presented only the major limitations of each search strategy in the report.

Appendix D of the CS details the SLR undertaken to identify relevant clinical evidence for the efficacy and safety of tirzepatide and the relevant comparators, GLP-1 RAs, for the treatment of T2D.[12] The original search was conducted in September 2021 and update searches were conducted in October 2021 and June 2022.

A summary of the sources searched is provided in Table 3.1.

Table 3.1: Data sources for the clinical effectiveness systematic review (as reported in CS)

Resource	Host/Source	Date Ranges	Date searched
Electronic databases			
MEDLINE and MEDLINE In-Process	Ovid ProQuest ProQuest	1/1/90-22/9/21 22/9/21-18/10/21 18/10/21-21/6/22	22/9/21 18/10/21 21/6/22
Embase	Ovid ProQuest	1/1/90-22/9/21 22/9/21-18/10/21 18/10/21-21/6/22	22/9/21 18/10/21 21/6/22
CENTRAL	Ovid Cochrane Library Cochrane Library	1/1/90-22/9/21 22/9/21-18/10/21 18/10/21-21/6/22	22/9/21 18/10/21 21/6/22
Conferences			
Annual Meeting of the European Association for the Study of Diabetes	Internet	Not stated	9/9/21
American Diabetes Association Scientific Sessions			
International Diabetes Federation World Diabetes Congress			

Resource	Host/Source	Date Ranges	Date searched
Annual Meeting of the American Association of Clinical Endocrinologists			
Annual European Congress of Endocrinology			
Annual Meeting of the Endocrine Society			
Annual Meeting of the Professional Society for Health Economics and Outcomes Research (ISPOR) (EU and US meetings)			
Trials registries			
ClinicalTrials.gov	Internet	2020-date	Not stated
WHO ICTRP			
CS = company submission; CENTRAL = Cochrane Central Register of Controlled Trials; EU = Europe; WHO ICTRP = World Health Organisation International Clinical Trials Registry Platform; US = United States			

EAG comment:

- Searches were undertaken in September 2021 and update searches were conducted in October 2021 and June 2022 to identify relevant clinical evidence for the efficacy and safety of tirzepatide and the relevant comparators, GLP-1 RAs, for the treatment of T2D. The CS, Appendix D and the Company's response to clarification provided sufficient details for the EAG to appraise the literature searches.[3, 4, 12]
- A good range of databases, websites, grey literature resources and trials registers were searched. Reference checking was conducted.
- Database searches were limited to a publication date of 1990 onwards but were not limited by language of publication.
- Searches were well structured, transparent and reproducible, and a good range of subject indexing terms (MeSH/EMTREE) and free text was used.
- Full details of the database host(s), dates searched and date ranges covered were not provided in the CS,[3] but were included in the Company's response to clarification.[4]
- Database search strategies contained a population facet for T2D. This facet was then combined with terms for tirzepatide and the relevant comparators.
- Study design filters to identify randomised controlled trials (RCTs) were applied to the searches of Embase and MEDLINE. The study design filters were not referenced, so it was unclear whether the filters used were published objectively-derived filters. The filters contained a combination of subject heading terms and free text terms and the EAG deemed them to be adequate.
- Separate searches for safety outcomes were not conducted. Guidance by the Centre for Reviews and Dissemination (CRD)[16] and Golder et al 2019.[17] recommend that if searches have been limited by a study design filter, additional searches should be undertaken to ensure that adverse events that are long-term, rare or unanticipated are not missed.

3.1.2 Inclusion criteria

The eligibility criteria for the SLR are presented in Table 3.2.

Table 3.2: Eligibility criteria used in search strategy for RCT and non-RCT evidence

	Description	Comparison to NICE scope/decision problem
Inclusion criteria		
Population	Adult patients (≥18 years of age) with T2D	The population is much broader and does not match the final scope or the decision problem addressed in the CS (see Table 2.2). See the EAG comment for the company's explanation
Interventions[a]	• Tirzepatide 5 mg QW • Tirzepatide 10 mg QW • Tirzepatide 15 mg QW • Albiglutide 30 mg QW • Albiglutide 50 mg QW • Dulaglutide 0.75 mg QW • Dulaglutide 1.5 mg QW • Dulaglutide 3.0 mg QW • Dulaglutide 4.5 mg QW • Exenatide 2.0 mg QW • Exenatide 5 µg BID • Exenatide 10 µg BID • Liraglutide 0.9 mg QD • Liraglutide 1.2 mg QD • Liraglutide 1.8 mg QD • Lixisenatide 20 µg QD • Loxenatide 100 µg QW • Loxenatide 200 µg QW • Semaglutide 0.5 mg QW • Semaglutide 1.0 mg QW • Semaglutide 2.0 mg QW • Semaglutide oral 3.0 mg QD • Semaglutide oral 7.0 mg QD • Semaglutide oral 14.0 mg QD	The intervention in the final scope is tirzepatide alone or with other antidiabetic agents. See the EAG comment for the company's explanation
Comparators [a]	Any of the listed interventions or one of the comparators listed below: • Basal insulin: o Insulin detemir o Insulin glargine o Insulin degludec o NPH-insulin o Bolus insulin: o Insulin lispro o Insulin aspart o Insulin glulisine	Consistent with final scope, but broader than the decision problem

	Description	Comparison to NICE scope/decision problem
	• Premixed insulin: o Biphasic NPH (50/50, 30/70, and 25/75 mixes only) (BiNPH50, BiNPH25) o Biphasic lispro (50/50 and 25/75 mixes only) (LM50, LM25) o Biphasic aspart (30/70 mix only) (BiAsp30) • DPP-4 inhibitor: o Sitagliptin o Linagliptin o Vildagliptin o Saxagliptin o Alogliptin • Sulfonylureas (SU): o Glimepiride o Glibenclamide/Glyburide (glibenclamide) o Glipizide o Gliclazide • TZD: o Pioglitazone • SGLT-2 inhibitors: o Dapagliflozin 5 mg or 10 mg QD o Canagliflozin 100 mg or 300 mg QD o Empagliflozin 10 mg or 25 mg QD o Ertugliflozin 5 mg or 15 mg QD o Ipragliflozin 25 mg or 50 mg QD o Tofogliflozin 20 mg QD • Metformin • Placebo	
Outcomes	• HbA1c • WBC • Hb • Hypoglycaemic events • Body weight • Composite <7% HbA1c, no weight gain, no hypoglycaemia • BMI • SBP	The outcomes are not aligned with the final scope

	Description	Comparison to NICE scope/decision problem
	• DBP • Heart rate • Triglycerides • Total cholesterol • LDL • HDL • eGFR • FPG • PPG • Safety outcomes	
Study design	• RCTs • ≥16 weeks on a single treatment	
Language restrictions	English language only	No justification offered
Exclusion criteria		
Population	• Patients with T1D mellitus • Patients with gestational diabetes • Children <18 years of age	Consistent with the final scope
Interventions	Liraglutide (saxenda) or any GLP-1 RAs used in treatment of obesity and not T2D	Consistent with the final scope
Comparators	NR	-
Outcomes	NR	-
Study design	• Treatment duration <16 weeks • Crossover studies that do not report pre-crossover data	No rationale offered, although will probably increase comparability
Language restrictions	Studies published in non-English language	No rationale offered

Table 7 of Appendix D of the CS[12]
BID = twice a day; BMI = body mass index; BPM = beats per minute; CS = company submission; DBP = diastolic blood pressure; DDP-4 = dipeptidyl-peptidase-4; EAG = Evidence Assessment Group; eGFR = estimated Glomerular Filtration Rate; FPG = fasting plasma glucose; GLP-1 = glucagon-like peptide-1; Hb = haemoglobin; HbA1c = glycated haemoglobin; HDL = high density lipoprotein; LDL = low density lipoprotein; NICE = National Institute for Health and Care Excellence; NPH = neutral protamine Hagedorn; NR = not reported; PPG = postprandial glucose; QW = once weekly; QD = once a day; RA = receptor antagonist; RCT = randomised controlled trial; SBP = systolic blood pressure; SLR = systematic literature review; SU = sulfonylurea; TZD = thiazolidinediones; T1D = type 1 diabetes; T2D = type 2 diabetes; WBC = white blood cell count
[a]To be eligible for inclusion in the SLR a study must include one of the interventions and at least one of the comparators listed above as treatment arms

No date restrictions were reported, although searches were limited to from 1990 (see Table 3.1). Table 10 of Appendix D of the CS[12] reports the excluded studies at full text review stage of the original search.

In this Table, 244 records are excluded with the reason 'Conference before 2020'. It is not clear why this exclusion criterion was applied.

Although the company mentioned that an update of the SLR was conducted in June 2022, it is incomplete. The company states that "*Due to time constraints, at the time of submission this update has only been completed as far as screening, with a final list of included publications determined. The results of the screening are presented in Section D.7.2. The data extraction and reporting for the update will be completed after the NICE submission*" (page 137 of Appendix D, CS[12]). In this update 15 publications met the inclusion criteria but were not included in the synthesis.

EAG comment: The EAG highlighted in the clarification letter the misalignment of the population, intervention, comparator and outcomes (PICOs) for the SLR and the NICE scope, to which the company replied that the SLR was broader and so would not have missed anything relevant to the decision problem.[4] The EAG also inquired how the list of interventions was populated, to which the company responded that comparison with GLP-1 RAs was of interest to Health Technology Assessment (HTA) agencies around the world.[4] These choices are consistent with the decision problem, notwithstanding the limitations of the tirzepatide and NMA evidence (see Sections 2 and 3.2 to 3.4) in terms of line of therapy and treatment experience/concomitant therapy. Indeed, the company also confirmed in the clarification letter that all studies were included regardless of prior/background treatment.[4]

3.1.3 Critique of study selection and data extraction

The records identified by the search strategies were screened in the software DistillerSR in two steps of title/abstract screening and full-text screening based on the pre-defined eligibility criteria. Both steps were executed in double, by two reviewers independently. Discrepancies were resolved by a third reviewer. Reasons for exclusion were recorded for both steps. Only the reasons for exclusion at full-text screening were reported. The screening process was recorded in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams (see Figures 1 and 2, Appendix D of the CS[12]).

The data were extracted in an Excel file by one reviewer and checked by a second one. The variables that were to be extracted are reported in Table 8, Appendix D of the CS[12]).

EAG comment: The EAG considers that the study selection and data extraction processes used appropriate methods to minimise error and bias. The EAG noted that the PRISMA flow diagrams, presented in Figures 1 and 2 in Appendix D, were not consistent with the summary of search results provided in the text (Section D.6.1). Figures 1 and 2 are described as flow charts for the first and up-date searches, respectively, however, Figure 1 appears to be a duplicate of Figure 2. In response to the clarification letter, the company have now provided a complete PRISMA diagram containing the results of all the searches.[4]

3.1.4 Quality assessment

According to the company, quality assessment of the included studies was conducted according to the Cochrane risk of bias assessment tool and the CRD tool. It was not reported if RoB1 or the updated and current RoB2 was used. From the results of the assessment reported in Table 46 of Appendix D[12]] it appears that six criteria were assessed: randomisation, allocation concealment, blinding, withdrawals, outcome selection and reporting and statistical methodology. These are the six of the seven criteria for assessment of risk of bias in RCTs by CRD. They do not cover the complexity of the RoB2 tool.[18]

The CS did not clarify if one or more reviewers were involved in the quality assessment process.

EAG comment: In response to the clarification letter, the company clarified that an acceptable form of standard practice of three reviewers was employed in quality assessment.[4]

3.1.5 Evidence synthesis

The SLR identified 246 publications of 205 unique studies. Among them are the SURPASS studies which examined tirzepatide as the intervention. The SURPASS trials were the base of the clinical evidence in the CS.

Trials SURPASS-1-5 are completed while SURPASS-6 and SURPASS-CVOT are ongoing. SURPASS-AP-Combo, SURPASS-J-Mono and SURPASS-J-Combo were conducted in Asian populations. The company stated that data for the first study was not available, while the other two studies were only used in the safety analysis as they were deemed not generalisable to the UK population.

In SURPASS-1 the intervention was tirzepatide monotherapy compared to placebo in T2D patients who were naïve to antihyperglycaemic injectable therapy. As such the population and treatment were judged to be not relevant to this CS and its results was not included in the clinical evidence.

From the rest of studies identified by the SLR, only those examining GLP-1 treatments were included in an NMA along with only SURPASS-2 and 3. Details on the NMA are presented in Section 3.4.

EAG comment: The EAG agree that SURPASS-1 has limited applicability to the decision problem, given the placement of tirzepatide as add-on therapy, notwithstanding the appropriateness of the decision problem to the NICE scope (see Section 2.1). The lack of applicability of the trials in an Asian population also seems reasonable.

3.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

3.2.1 Trial design

The four phase 3 SURPASS-2 to 5 trials provided the clinical efficacy evidence for tirzepatide. The company ranked the trials in terms of relevance of the comparators and patient population to the decision problem addressed in the CS: "*patients with T2D that is inadequately controlled with three or more antidiabetic agents, when the GLP-1 RA class would otherwise be considered*" (page 30 of the CS[3]).

- SURPASS-2 was found to be the most relevant to the decision problem by the company because the comparator is the GLP-1 RA semaglutide 1 mg, thus comparing tirzepatide plus metformin to semaglutide plus metformin
- SURPASS-3 was found to be relevant as the population had received 1-2 prior therapies of metformin with or without an SGLT-2i. The company acknowledged that the use of insulin degludec as the comparator in this trial is not relevant to the anticipated positioning of tirzepatide
- SURPASS-4 was found to be relevant as part of the population had previously received triple oral therapy (metformin, an SU and an SGLT-2i). The company notes that the population is narrower because all patients had high cardiovascular (CV) risk but supported that this feature would provide important CV safety data. They also acknowledge that the use of insulin glargine as the comparator in the trial is not relevant to the anticipated positioning of tirzepatide, similar to SURPASS-3

- SURPASS-5 was relevant as is compares "…*tirzepatide for the treatment of T2D with placebo in patients with background therapy of insulin glargine, with or without metformin*." (page 31 of the CS[3]). No mention on the positioning of insulin glargine was made for this trial

The company stated that dosing of tirzepatide in all SURPASS trials began at 2.5 mg and could be increase in increments of 2.5 mg every 4 weeks up to a maximum maintenance of 15 mg from week 21 onwards. However, in all SURPASS trials tirzepatide was randomised to one of three doses of tirzepatide, 5 mg, 10 mg or 15 mg maintenance or the comparator, which varied between the trials.

Eligibility included the same HbA1c threshold of ≥7.0% (≥53 mmol/mol) to ≤10.5% (≤91 mmol/mol) except SURPASS-4 where the lower limit was 7.5% (58 mmol/mol). All excluded patients with a prior history of proliferative diabetic retinopathy, diabetic maculopathy, or non-proliferative diabetic retinopathy that requires acute treatment, but SURPASS-4 also excluded patients with hepatitis and signs and symptoms of liver disease. All trials stipulated stable weight for 3 months and a BMI of at least 25 kg/m², except SURPASS-5 where it was 23 kg/m².

Concomitant medication was the same as that on entry to the trial i.e., metformin for SURPASS-2, metformin with or without an SGLT-2i for SURPASS-2, at least one and up to triple oral therapy (metformin, an SU and an SGLT-2i) for SURPASS-4 and insulin with or without metformin for SURPASS-5.

EAG comment: Given the dose related subgroups in the CEA and the NMA, the company were asked to explain why the tirzepatide trials were not stratified by eligibility for each tirzepatide maintenance dose with comparator dose chosen accordingly, as opposed to randomised to maintenance dose. This makes judging the applicability to clinical practice of the effectiveness of tirzepatide versus each of the comparators difficult since, according to the license, patients would not be constrained to any maintenance dose target. If the highest dose is the most effective (see trial results Section), the effectiveness of the lowest dose is probably an underestimate and that of the highest and overestimate of clinical practice. In fact, SURPASS-5 is the only exception to the dose-related ranking of glycaemic control with 10 mg and 15 mg being equal for change in HbA1c from baseline and a reversal in meeting the two higher targets (7% and 6.5%), which might be explained by a ceiling effect. It is unclear of the direction of any bias in the middle dose because some patients might never reach the dose and others might need to exceed it, depending on the decision rule for dose escalation, which is likely to include an HbA1c target. In SURPASS-2 and SURPASS-3 randomisation was stratified based on baseline HbA1c (≤8.5% or >8.5% [69 mmol/mol]), so the company were asked if HbA1c level be a proxy for eligibility for each tirzepatide dose. The company responded to the clarification letter by stating that;*"…the most appropriate dose for an individual should be determined by the clinician in collaboration with the patient, based on clinical characteristics and patient tolerability and cannot be determined a priori…"*.[4]

3.2.2 Baseline characteristics

A summary of the patients' baseline characteristics in the SURPASS trials that form the basis of this CS is presented in Table 3.3. The data refer to the modified intent-to-treat (mITT) population of patients with T2D included in the final analysis of the trials.

Table 3.3: Baseline characteristics of patients included in the SURPASS-2, 3, 4 and 5 trials

Intervention/comparator	TZP 5 mg				TZP 10 mg				TZP 15 mg				SEMA 1 mg	Insulin degludec	Insulin glargine	Placebo	Overall population			
SURPASS trial	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**
N	**470**	**358**	**329**	**116**	**469**	**360**	**328**	**119**	**470**	**359**	**338**	**120**	**469**	**360**	**1000**	**120**	**1,878**	**1,437**	**1,995**	**475**
Demographics																				
Age (years), mean ± SD	56.3 ± 10.0	57.2 ± 10.1	62.9 ± 8.6	61.5 ± 9.8	57.2 ± 10.5	57.4 ± 9.7	63.7 ± 8.7	60.4 ± 10.2	55.9 ± 10.4	57.5 ± 10.2	63.7 ± 8.6	60.5 ± 9.9	56.9 ± 10.8	57.5 ± 10.1	63.8 ± 8.5	60.0 ± 9.6	56.6 ± 10.4	57.4 ± 10.0	63.6 ± 8.6	60.6 ± 9.9
Female, n (%)	265 (56.4)	158 (44.1)	131 (39.8)	55 (47.4)	231 (49.3)	165 (45.8)	119 (36.3)	47 (39.5)	256 (54.5)	165 (46.0)	135 (39.9)	55 (45.8)	244 (52.0)	147 (40.8)	364 (36.4)	54 (45.0)	996 (53.0)	635 (44.2)	749 (37.5)	211 (44.4)
Race, n (%)																				
White	382 (81.3)	323 (90.2)	260 (79.3)	95 (81.9)	376 (80.2)	328 (91.1)	259 (79.0)	94 (79.0)	392 (83.4)	327 (91.1)	285 (84.6)	94 (78.3)	401 (85.5)	329 (91.4)	825 (82.7)	97 (80.8)	1551 (82.6)	1307 (91.0)	1629 (81.8)	380 (80.0)
American Indian or Alaska native	53 (11.3)	0	[redacted]	[redacted]	53 (11.3)	1 (0.3)	[redacted]	[redacted]	57 (12.1)	1 (0.3)	[redacted]	[redacted]	45 (9.6)	2 (0.6)	[redacted]	[redacted]	208 (11.1)	4 (0.3)	[redacted]	[redacted]
Asian	6 (1.3)	20 (5.6)	15 (4.6)	20 (17.2)	11 (2.3)	19 (5.3)	16 (4.9)	21 (17.6)	5 (1.1)	20 (5.6)	8 (2.4)	22 (18.3)	3 (0.6)	17 (4.7)	31 (3.1)	22 (18.3)	25 (1.3)	76 (5.3)	70 (3.5)	85 (17.9)
Black or African American	28 (6.0)	13 (3.6)	13 (4.0)	1 (0.9)	21 (4.5)	12 (3.3)	17 (5.2)	2 (1.7)	15 (3.2)	8 (2.2)	11 (3.3)	3 (2.5)	15 (3.2)	11 (3.1)	32 (3.2)	0	79 (4.2)	44 (3.1)	73 (3.7)	6 (1.3)
Multiple	1 (0.2)	1 (0.3)	[redacted]	[redacted]	8 (1.7)	0	[redacted]	[redacted]	0	1 (0.3)	[redacted]	[redacted]	3 (0.6)	0	[redacted]	[redacted]	12 (0.6)	2 (0.1)	[redacted]	[redacted]
Native Hawaiian or other Pacific Islander	0	1 (0.3)	[redacted]	-	0	0	[redacted]	-	1 (0.2)	2 (0.6)	[redacted]	-	2 (0.4)	1 (0.3)	[redacted]	-	3 (0.2)	4 (0.3)	[redacted]	-
Missing	-	-	[redacted]	-	-	-	[redacted]	-	-	-	[redacted]	-	-	-	[redacted]	-	-	-	[redacted]	-
Weight (kg), mean ± SD	92.5 ± 21.8	94.43 ± 18.86	90.3 ± 20.3	95.8 ± 19.8	94.8 ± 22.7	93.80 ± 19.81	90.6 ± 18.2	94.5 ± 22.2	93.8 ± 21.8	94.90 ± 20.98	90.0 ± 16.3	96.3 ± 22.8	93.7 ± 21.1	93.98 ± 20.59	90.2 ± 19.0	94.1 ± 21.8	93.7 ± 21.9	94.28 ± 20.06	90.3 ± 18.7	95.2 ± 21.6
BMI (kg/m²), mean ± SD	33.8 ± 6.9	33.58 ± 5.87	32.6 ± 6.1	33.6 ± 5.9	34.3 ± 6.6	33.41 ± 6.21	32.8 ± 5.5	33.4 ± 6.2	34.5 ± 7.1	33.68 ± 6.11	32.5 ± 5.0	33.4 ± 5.9	34.2 ± 7.2	33.42 ± 6.06	32.5 ± 5.5	33.2 ± 6.3	34.2 ± 6.9	33.52 ± 6.06	32.6 ± 5.5	33.4 ± 6.1

Intervention/comparator	TZP 5 mg				TZP 10 mg				TZP 15 mg				SEMA 1 mg	Insulin degludec	Insulin glargine	Placebo	Overall population			
SURPASS trial	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**
N	**470**	**358**	**329**	**116**	**469**	**360**	**328**	**119**	**470**	**359**	**338**	**120**	**469**	**360**	**1000**	**120**	**1,878**	**1,437**	**1,995**	**475**
BMI category, n (%)																				
<30	[redacted]	104 (29.1)	[redacted]	[redacted]	[redacted]	116 (32.2)	[redacted]	[redacted]	[redacted]	109 (30.4)	[redacted]	[redacted]	[redacted]	117 (32.5)	[redacted]	[redacted]	[redacted]	446 (31.0)	[redacted]	[redacted]
30 to <35	[redacted]	136 (38.0)	[redacted]	[redacted]	[redacted]	119 (33.1)	[redacted]	[redacted]	[redacted]	121 (33.7)	[redacted]	[redacted]	[redacted]	120 (33.3)	[redacted]	[redacted]	[redacted]	496 (34.5)	[redacted]	[redacted]
≥35	[redacted]	118 (33.0)	[redacted]	[redacted]	[redacted]	125 (34.7)	[redacted]	[redacted]	[redacted]	129 (35.9)	[redacted]	[redacted]	[redacted]	123 (34.2)	[redacted]	[redacted]	[redacted]	495 (34.4)	[redacted]	[redacted]
Disease Characteristics																				
Duration of diabetes (years), mean ± SD	9.1 ± 7.2	8.47 ± 5.83	11.14 ± 7.08	14.1 ± 8.1	8.4 ± 5.9	8.43 ± 6.59	11.96 ± 7.45	12.6 ± 6.2	8.7 ± 6.9	8.52 ± 6.47	11.48 ± 7.54	13.7 ± 7.5	8.3 ± 5.8	8.12 ± 6.04	12.03 ± 7.66	12.9 ± 7.4	8.6 ± 6.5	8.38 ± 6.24	11.78 ± 7.51	13.3 ± 7.3
HbA1c (%), mean ± SD	8.32 ± 1.08	8.17 ± 0.89	8.52 ± 0.84	8.30 ± 0.88	8.30 ± 1.02	8.18 ± 0.89	8.59 ± 0.91	8.36 ± 0.83	8.26 ± 1.00	8.21 ± 0.94	8.52 ± 0.98	8.23 ± 0.86	8.25 ± 1.01	8.12 ± 0.94	8.50 ± 0.85	8.37 ± 0.84	8.28 ± 1.03	8.17 ± 0.91	8.52 ± 0.88	8.31 ± 0.85
HbA1c (mmol/mol), mean ± SD	67.46 ± 1.84	65.81 ± 9.69	69.59 ± 9.21	[redacted]	67.20 ± 11.20	65.91 ± 9.76	70.43 ± 9.95	[redacted]	66.78 ± 10.97	66.18 ± 10.24	69.63 ± 10.68	[redacted]	66.69 ± 10.99	65.20 ± 10.28	69.41 ± 9.32	[redacted]	67.03 ± 11.25	65.78 ± 9.99	69.65 ± 9.65	[redacted]
HbA1c category, n (%)																				
≤8.5% (69 mmol/mol)	293 (62.3)	248 (69.3)	[redacted]	[redacted]	294 (62.7)	249 (69.2)	[redacted]	[redacted]	303 (64.5)	252 (70.2)	[redacted]	[redacted]	302 (64.4)	256 (71.1)	[redacted]	[redacted]	1192 (63.5)	1005 (69.9)	[redacted]	[redacted]
>8.5% (69 mmol/mol)	177 (37.7)	110 (30.7)	[redacted]	[redacted]	175 (37.3)	111 (30.8)	[redacted]	[redacted]	167 (35.5)	107 (29.8)	[redacted]	[redacted]	167 (35.6)	104 (28.9)	[redacted]	[redacted]	686 (36.5)	432 (30.1)	[redacted]	[redacted]
FSG (mg/dl), mean ± SD	173.8 ± 51.9	171.73 ± 47.86	172.27 ± 49.11	162.9 ± 53.9	174.2 ± 49.8	170.42 ± 47.64	175.47 ±51.93	162.3 ± 52.0	172.4 ± 54.4	168.42 ± 45.95	174.14 ± 53.84	160.3 ± 54.2	171.4 ± 49.8	166.73 ± 41.90	168.40 ± 49.72	164.1 ± 45.0	172.9 ± 51.5	169.33 ± 45.89	171.17 ± 50.75	162.4 ± 51.3
FSG (mmol/l), mean ± SD	9.7 ± 2.9	9.53 ± 2.66	[redacted]	[redacted]	9.7 ± 2.8	9.46 ± 2.64	[redacted]	[redacted]	9.6 ± 3.0	9.35 ± 2.55	[redacted]	[redacted]	9.5 ± 2.8	9.26 ± 2.33	[redacted]	[redacted]	9.6 ± 2.9	9.40 ± 2.55	[redacted]	[redacted]
SBP (mm Hg), mean ± SD	130.5 ± 14.1	130.73 ± 13.59	133.28 ± 14.18	[redacted]	131.5 ± 13.8	131.10 ± 13.12	135.08 ±16.11	[redacted]	130.5 ± 14.3	131.85 ± 12.85	134.34 ± 15.02	[redacted]	130.0 ± 13.0	132.45 ± 13.63	134.57 ± 15.67	[redacted]	130.6 ± 13.8	131.53 ± 13.30	134.40 ± 15.40	[redacted]

Intervention/comparator	TZP 5 mg				TZP 10 mg				TZP 15 mg				SEMA 1 mg	Insulin degludec	Insulin glargine	Placebo	Overall population			
SURPASS trial	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**	**-2**	**-3**	**-4**	**-5**
N	**470**	**358**	**329**	**116**	**469**	**360**	**328**	**119**	**470**	**359**	**338**	**120**	**469**	**360**	**1000**	**120**	**1,878**	**1,437**	**1,995**	**475**
DBP (mm Hg), mean ± SD	78.6 ± 8.9	78.59 ± 8.52	78.39 ± 8.75	[redacted]	80.0 ± 9.6	79.22 ± 8.69	78.60 ± 9.50	[redacted]	79.0± 9.0	79.25 ± 9.16	78.24 ± 9.16	[redacted]	79.3 ± 8.6	79.57 ± 9.18	78.41 ± 9.62	[redacted]	79.2 ± 9.0	79.16 ± 8.89	78.41 ± 9.38	[redacted]
Non-proliferative diabetes retinopathy	-	-	68 (20.7)	-	-	-	63 (19.2)	-	-	-	89 (26.3)	-	-	-	187 (18.7)	-	-	-	407 (20.4)	-
History of CV disease	[redacted]	[redacted]	275 (83.6)	[redacted]	[redacted]	[redacted]	296 (89.7)	[redacted]	[redacted]	[redacted]	293 (86.7)	[redacted]	[redacted]	[redacted]	874 (87.0)	[redacted]	[redacted]	[redacted]	1738 (86.8)	[redacted]
eGFR (CKD-EPI, ml/min per 1.73 m²), mean ± SD	96.6 ± 17.5	95.14 ± 17.22	80.28 ± 22.66	[redacted]	95.5 ± 16.6	93.65 ± 16.90	81.43 ± 20.44	[redacted]	96.3 ± 16.9	93.09 ± 17.25	81.55 ± 21.22	[redacted]	95.6 ± 17.3	94.63 ± 16.78	81.47 ± 20.78	[redacted]	96.0 ± 17.1	94.13 ± 17.04	81.28 ± 21.11	[redacted]
eGFR category, n (%)																				
<60 ml/min per 1.73 m²	19 (4.0)	16 (4.5)	62 (18.8)	[redacted]	15 (3.2)	13 (3.6)	56 (17.1)	[redacted]	11 (2.3)	12 (3.3)	58 (17.2)	[redacted]	19 (4.1)	15 (4.2)	166 (16.6)	[redacted]	[redacted]	56 (3.9)	342 (17.1)	[redacted]
≥60 ml/min per 1.73 m²	451 (96.0)	342 (95.5)	[redacted]	[redacted]	454 (96.8)	347 (96.4)	[redacted]	[redacted]	459 (97.7)	347 (96.7)	[redacted]	[redacted]	450 (95.9)	345 (95.8)	[redacted]	[redacted]	1814 (96.6)	1381 (96.1)	[redacted]	[redacted]
Macroalbuminuria	-	-	25 (7.7)	-	-	-	33 (10.3)	-	-	-	24 (7.1)	-	-	-	79 (8.1)	-	-	-	161 (8.2)	-
Microalbuminuria	-	-	76 (23.5)	-	-	-	97 (30.4)	-	-	-	103 (30.6)	-	-	-	270 (27.6)	-	-	-	546 (27.9)	-

Tables 10, 12, 14 and 16 of the CS[3]

BMI = body mass index; BPM = beats per minute; CKD-EPI = chronic kidney disease-epidemiology; CS = company submission; CV = cardiovascular; DBP = diastolic blood pressure; eGFR = estimated Glomerular Filtration Rate; FSG = fasting serum glucose; HbA1c = glycated haemoglobin; SBP = systolic blood pressure; SD = standard deviation; SEMA = semaglutide; TZP = tirzepatide; % = percentage

[a]≤8.0% (69 mmol/mol)

[b]>8.0% (69 mmol/mol)

Table 3.4: Concomitant treatments at baseline

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator	Overall population
SURPASS-2					
Metformin	100%				
SURPASS-3					
Metformin alone, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SGLT-2i, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	458 (31.9)
Metformin dose (mg/day), mean ± SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-4					
Metformin alone, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SU, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SGLT-2i, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Metformin plus SU plus SGLT-2i, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SU alone, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SGLT-2i alone, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SU + SGLT-2i, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-5					
Insulin dose mean ± SD	39.1 ± 25.4	34.7 ± 15.4	40.5 ± 29.1	36.3 ± 18.0	37.6 ± 22.7
Metformin, n (%)	99 (85.3)	99 (83.2)	97 (80.8)	99 (82.5)	394 (82.9)

Tables 12, 14 and 16 of the CS[3]
CS = company submission; SD = standard deviation; SGLT-2i = sodium-glucose co-transporter-2 inhibitor; SU = sulfonylurea; TZP = tirzepatide

EAG comment: The proportion of Asian population in the SURPASS-5 trial (17.9% overall) is higher than in the rest SURPASS-trials (1.3% - 5.3% overall). In SURPASS-4 a very high proportion of the population (86.8% overall) had a history of CV disease. The proportion of patients with CV disease in the rest SURPASS trial was comparatively very low (8% - 18.3% overall) (see Table 3.4).

Most importantly, the concomitant treatments at baseline, which patients continued during the trial as concomitant therapy, vary a lot between the SURPASS trials as presented in Table 3.3. The EAG also requested information on treatment history to which the company, in response to the clarification letter, stated that only such data recorded was on antihyperglycaemic treatments within the last 3 months, which were the background/concomitant treatments in the trial, i.e. those reported in Table 3.3.[4] They also made the point that this information would not be useful as it would not be available in any of the comparator trials included in the NMA. They also suggested that a proxy for effect of line of therapy might be the variation in outcomes between the SURPASS trials, presenting a comparison of change in HbA1c, percentage achieving the <7% HbA1c target and change in body weight, which they argued: *"demonstrated efficacy despite differences in line of therapy and background therapy."* This comparison can be observed in Section 3.2.3 below, and it does appear to show that there is little variation in outcome of tirzepatide by dose in the SURPASS trials, despite variation in background therapy. What one cannot separate out is the independent effects of line of and background therapy, although one might reasonably assume that line generally increases with number of oral therapies.[7] However, this is not the same as little variation in the treatment effect i.e. tirzepatide versus the same

comparator, and this cannot be observed because of concurrent variation in comparator. Indeed, as described in Section 3.2.5, it does appear that baseline therapy does have an impact on the treatment effect, although the direction is difficult to establish in the only trial, SURPASS-4, where OAD up to triple therapy is observed.[19]

3.2.3 Clinical effectiveness results

3.2.3.1 Glycaemic control

For all SURPASS trials, all outcomes and all doses of tirzepatide, there was a statistically significant difference versus the comparator in favour of tirzepatide. There was also a dose-response relationship whereby the higher dose was more effective, except the dose 15 mg for tirzepatide in SURPASS -5 (versus placebo) for change in HbA1c, proportion of patients achieving HbA1c <7.0% target and HbA1c ≤6.5% target at 40 weeks.

Table 3.5: HbA1c, percentage

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg)				
N	470	469	469	468
Baseline	8.33	8.31	8.25	8.24
Change from baseline to 40 weeks	−2.09*	−2.37*	−2.46*	−1.86*
Change difference from SEMA (95% CI) to 40 weeks	−0.23** (−0.36, −0.10)	−0.51** (−0.64, −0.38)	−0.60** (−0.73, −0.47)	n/a
SURPASS-3 (versus insulin degludec)				
N	358	360	358	359
Baseline	8.17	8.19	8.21	8.13
Change from baseline to 52 weeks	−1.93*	−2.20*	−2.37*	−1.34*
Change difference from insulin degludec (95% CI) at 52 weeks	−0.59** (−0.73, −0.45)	−0.86** (−1.00, −0.72)	−1.04** (−1.17, −0.90)	n/a
SURPASS-4 (versus insulin glargine)				
N	326	321	334	978
Baseline	8.52	8.60	8.52	8.51
Change from baseline to 52 weeks	−2.24*	−2.43*	−2.58*	−1.44*
Change difference from insulin glargine (95% CI) at 52 weeks	−0.80** (−0.92, −0.68)	−0.99** (−1.11, −0.87)	−1.14** (−1.26, −1.02)	n/a
SURPASS-5 (versus placebo)				
N	116	118	118	119
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	−2.23*	−2.59*	−2.59*	−0.93*

Change difference from placebo (95% CI) at 40 weeks	−1.30** (−1.52, −1.07)	−1.66** (−1.88, −1.43)	−1.65** (−1.88, −1.43)	n/a
Table 26, 32, 36, 40 of the CS[3] CI = confidence interval; CS = company submission; HbA1c = glycated haemoglobin; SEMA = semaglutide; TZP = tirzepatide *p<0.001 vs. baseline; **p<0.001 vs. comparator				

3.2.3.2 Patients achieving HbA1c below a specific threshold

Table 3.6: Patients achieving HbA1c <7.0%, percentage

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg) proportion of patients at 40 weeks	85.5*	88.9	92.2	81.1
SURPASS-3 (versus insulin degludec) proportion of patients at 52 weeks	82.4	89.7	92.6	61.3
SURPASS-4 (versus insulin glargine) proportion of patients at 52 weeks	81.0	88.2	90.7	50.7
SURPASS-5 (versus placebo) proportion of patients at 40 weeks	93.0	97.4	94.0	33.9
Figure 8, 13, 17, 21 of the CS[3] CS = company submission; HbA1c = glycated haemoglobin; TZP = tirzepatide All comparisons between tirzepatide and comparator statistically significant with p<0.001 except: *<0.05				

Table 3.7: Patients achieving HbA1c ≤6.5%, percentage

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg) proportion of patients at 40 weeks	74.0	82.	87.1	66.2
SURPASS-3 (versus insulin degludec) proportion of patients at 52 weeks	71.4	80.3	85.3	44.4
SURPASS-4 (versus insulin glargine) proportion of patients at 52 weeks	66.0	76.0	81.1	31.7
SURPASS-5 (versus placebo) proportion of patients at 40 weeks	80.0	94.7	92.3	17.0
Figure 8, 13, 17, 21 of the CS[3] CS = company submission; HbA1c = glycated haemoglobin; TZP = tirzepatide All comparisons between tirzepatide and comparator statistically significant with p<0.001				

Table 3.8: Patients achieving HbA1c <5.7%, percentage

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg) proportion of patients at 40 weeks	29.3	44.7	50.9	19.7
SURPASS-3 (versus insulin degludec) proportion of patients at 52 weeks	25.8	38.6	48.4	5.4
SURPASS-4 (versus insulin glargine) proportion of patients at 52 weeks	23.0	32.7	43.1	3.4
SURPASS-5 (versus placebo) proportion of patients at 40 weeks	26.1	47.8	62.4	2.5
Figures 8, 13, 17, 21 of the CS[3] CS = company submission; HbA1c = glycated haemoglobin; TZP = tirzepatide All comparisons between tirzepatide and comparator statistically significant with p<0.001				

3.2.3.3 Body weight and BMI

For all SURPASS trials, patients on all three tirzepatide doses showed significant weight loss and reduced BMI compared to the comparator group. There was also a dose-response relationship whereby the higher dose was more effective.

Table 3.9: Body weight change from baseline, percentage (kg)

Characteristics	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg)				
N	470	469	469	468
Baseline	92.6	94.6	93.9	93.8
Change from baseline to 40 weeks	−7.8[a]	−10.3[a]	−12.4[a]	−6.2[a]
Change difference from SEMA (95% CI) to 40 weeks	−1.7[b] (−2.6, −0.7)	−4.1[b] (−5.0, −3.2)	−6.2[b] (−7.1, −5.3)	N/A
SURPASS-3 (versus insulin degludec)				
N	358	360	358	359
Baseline	94.5	94.3	94.9	94.2
Change from baseline to 52 weeks	−7.5[a]	−10.7[a]	−12.9[a]	2.3[a]
Change difference from insulin degludec (95% CI) at 52 weeks	−9.8[c] (−10.8, −8.8)	−13.0[c] (−14.0, −11.9)	−15.2[c] (−16.2, −14.2)	N/A
SURPASS-4 (versus insulin glargine)				
N	326	321	334	978
Baseline	90.3	90.7	90.0	90.3
Change from baseline to 52 weeks	−7.1[a]	−9.5[a]	−11.7[a]	1.9
Change difference from insulin glargine (95% CI) at 52 weeks	−9.0[d] (−9.8, −8.3)	−11.4[d] (−12.1, −10.6)	−13.5[d] (−14.3, −12.8)	N/A
SURPASS-5 (versus placebo)				
N	116	118	118	119
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks	−6.2[a]	−8.2[a]	−10.9[a]	1.7[e]
Change difference from placebo (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Table 27, 33, 37, 42 of CS[3]
CI = confidence interval; CS = company submission; SEMA = semaglutide; TZP = tirzepatide; N/A = not applicable
[a]p<0.001; [b]p<0.001 versus semaglutide 1 mg; [c]p<0.001 versus insulin degludec for the mean change difference; [d]p<0.001 versus insulin glargine; [e]p<0.01 versus baseline; [f]p<0.001 versus placebo for the mean change difference

Table 3.10: BMI change from baseline

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg)				
N	470	469	469	468
Baseline	[redacted]	[redacted]	[redacted]	[redacted]

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from SEMA (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-3 (versus insulin degludec)				
N	358	360	358	359
Baseline	33.6	33.5	33.7	33.4
Change from baseline to 52 weeks	−2.7*	−3.8*	−4.6*	0.8*
Change difference from insulin degludec (95% CI) at 52 weeks	−3.6** (−3.9, −3.2)	−4.7** (−5.0, −4.3)	−5.5** (−5.8, −5.1)	N/A
SURPASS-4 (versus insulin glargine)				
N	328	326	337	998
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 52 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from insulin glargine (95% CI) at 52 weeks	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-5 (versus placebo)				
N	116	118	118	119
Baseline	33.6	33.5	33.4	33.3
Change from baseline to 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Change difference from placebo (95% CI) at 40 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Tables 28, 34, 38, 43 of CS[3]
BMI = body mass index; CI = confidence interval; CS = company submission; SEMA = semaglutide; TZP = tirzepatide; n/a = not applicable
*p<0.05, **p<0.01, ***p<0.001 versus comparator

3.2.3.4 Lipids

All treatment groups had increased high density lipoprotein-cholesterol (HDL-C) compared to the baseline. For all SURPASS trials, all outcomes and all doses of tirzepatide, there was a statistically significant difference versus the comparator in favour of tirzepatide excepting SURPASS-5 (versus placebo). There was a dose-response relationship only in SURPASS-4 among all trials whereby the higher dose was more effective.

Table 3.11: HDL-C change from baseline (mg/dl)

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg)				
Baseline (mg/dl)	42.9	42.7	42.9	42.7
Change from baseline to 40 weeks (mg/dl)	2.9	3.4	3.0	1.9
SEMA 1 mg-adjusted percent change at 40 weeks (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-3 (versus insulin degludec)				
Baseline (mg/dl)	42.8	42.1	42.4	44.4

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
Change from baseline to 52 weeks (mg/dl)	2.4	4.4	4.4	0.4
Insulin degludec adjusted change difference at 52 weeks (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-4 (versus insulin glargine)				
Baseline (mg/dl)	41.3	40.2	40.4	40.6
Change from baseline to 52 weeks (mg/dl)	2.8	4.0	4.4	1.2
Insulin glargine adjusted change difference (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-5 (versus placebo)				
Baseline (mg/dl)	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to 40 weeks (mg/dl)	[redacted]	[redacted]	[redacted]	[redacted]
Placebo adjusted change difference (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

Table 29 of the CS[3]; Table GPGL.5.11 of the SURPASS-2 CSR; Table GPGH.5.10. of the SURPASS-3 CSR[20]; Table GPGM.5.18. of the SURPASS-4 CSR[19]; Table GPGI.5.9 of the SURPASS-5 CSR[21]
CI = confidence interval; CSR = Clinical Study Report; HDL-C = high-density lipoprotein-cholesterol; N/A = not applicable; SEMA = semaglutide; TZP = tirzepatide
*p<0.05; **p<0.01; ***p<0.001 versus comparator

All treatment groups had reduced low density lipoprotein-cholesterol (LDL-C) compared to the baseline. Among all trials, there was a statistically significant difference versus the comparator in favour of tirzepatide in SURPASS-4 (versus insulin degludec) and SURPASS-5 (versus placebo) for all outcomes and all doses of tirzepatide. There was a dose-response relationship only in SURPASS-5 among all trials whereby the higher dose was more effective.

Table 3.12: LDL-C change from baseline (mg/dl)

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 (versus semaglutide 1 mg)				
Baseline (mg/dl)	88.2	88.4	86.4	88.2
Change from baseline to 40 weeks (mg/dl)	−6.7	−4.9	−4.5	−5.6
SEMA 1 mg-adjusted percent change at 40 weeks (%) (95% CI)	-1.4 (-5.59, 3.02)	-0.9 (-3.50, 5.39)	-1.3 (-3.04, 5.87)	N/A
SURPASS-3 (versus insulin degludec)				
Baseline (mg/dl)	85.4	88.5	87.6	89.7
Change from baseline to 52 weeks (mg/dl)	−5.3	−5.0	−5.7	−2.4
Insulin degludec adjusted change difference at 52 weeks (%) (95% CI)	-3.39 (-8.19, 1.66)	-3.07 (-7.96, 2.07)	-3.94 (-8.78, 1.15)	N/A
SURPASS-4 (versus insulin glargine)				

	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
Baseline (mg/dl)	77.2	72.9	74.9	75.5
Change from baseline to 52 weeks (mg/dl)	−5.1	−6.3	−6.0	0.9
Insulin glargine adjusted change difference (%) (95% CI)	-7.9*** (-12.1, -3.4)	-9.5*** (-13.7, -5.1)	-9.1*** (-13.2, -4.8)	N/A
SURPASS-5 (versus placebo)				
Baseline (mg/dl)	83.6	85.7	83.9	87.5
Change from baseline to 40 weeks (mg/dl)	−7.6	−10.9	−13.2	2.4
Placebo adjusted change difference (%) (95% CI)	-11.44** (-17.63, -4.79)	-15.23*** (-21.15, -8.87)	-17.83*** (-23.70, -11.50)	N/A

Table 29 of the CS[3]; Table GPGL.5.11. of the SURPASS-2 CSR[22]; Table GPGH.5.10 of the SURPASS-3 CSR[20]; Table GPGM.5.19. of the SURPASS-4 CSR[19]; Table GPGI.5.9 of SURPASS-5 CSR[21]
CI = confidence interval; CS = company submission; CSR = Clinical Study Report; LDL-C = low-density lipoprotein-cholesterol; N/A = not applicable; SEMA = sulfonylurea; TZP = tirzepatide; % = percentage
*p<0.05; **p<0.01; ***p<0.001 versus comparator

EAG comment: For all SURPASS trials, for nearly all HbA1c outcomes, body weight change and BMI change and all doses of tirzepatide, there was a statistically significant difference versus the comparator in favour of tirzepatide. There was also a dose-response relationship whereby the higher dose was more effective, except the dose 15 mg for tirzepatide in SURPASS-5 (versus placebo) for change in HbA1c, proportion of patients achieving HbA1c <7.0% target and HbA1c ≤6.5% target at 40 weeks, although the differences could be regarded as very small and perhaps consistent with a ceiling effect. For HDL-C change, there was also a statistically significant difference in favour of tirzepatide regardless of dose for all trials, but SURPASS-5, where the difference varied in direction depending on dose, but was much smaller and could be regarded as close to zero. There also seems to be no clear dose response relationship. For LDL-C change, the treatment effect was always in favour of tirzepatide, but not with statistical significance in SURPASS-2 and SURPASS-3, where the magnitude was also relatively small and with no clear dose response relationship. In SURPASS-4 and SURPASS-5, however, there was a statistically significant difference in favour of tirzepatide, which was much larger, especially in SURPASS-5, where there was also a dose response relationship.

Although estimated Glomerular Filtration Rate (eGFR) change was included in the NMA (see Section 3.4), the SURPASS trial results were not included in the CS. However, in response to the clarification letter, the company made it clear that the majority of patients in SURPASS-1 to 5 did not have impaired renal function i.e., eGFR >60 ml/min per 1·73 m² (83% - 97%) and in SURPASS-2 to 4 patients with eGFR <45 ml/min per 1.73 m² were excluded.[4] The EAG considers that it might therefore be reasonable to conclude that there would be little change in eGFR during the follow-up period and therefore little expectation of a treatment effect on eGFR change.

3.2.4 Safety results

This Section reports on the safety results discussed in Section B.2.10 of the CS.

The CS reports safety results of tirzepatide in patients with T2D, evaluated as an endpoint in all SURPASS trials. A total of 19 completed phase 1, phase 2, and phase 3 studies have contributed safety data with up to 106 weeks of exposure to treatment. A total of 7,769 patients received an intervention

in the phase 2 and 3 studies. Of these patients, 5,415 received tirzepatide, 312 received placebo, and 2,042 received an active comparator. The phase 3 studies were examined separately from the phase 2 because the seven phase 3 studies conducted (SURPASS-1 to 5 plus two Japanese studies, SURPASS J Mono and SURPASS J Combo) had the same tirzepatide treatment groups with the same dose-escalation schedules, which were different from the phase 2 studies. The two primary analysis sets to detect drug and dose effects, respectively, are the phase 3 placebo-controlled analysis set and the phase 3 dose effect analysis set. Data on patient deaths during the trial are presented in Appendix F of the CS.

Table 3.13: Safety analysis sets

Analysis set	Studies	Time Period	Description	Treatment groups
Phase 3 placebo-controlled analysis set (N=953)	SURPASS-1, SURPASS-5	First dose of treatment to end of safety follow-up visit or date of study withdrawal	Integrated data of TZP doses compared to placebo for studies with placebo arm and same dose-escalation schedule proposed for the label	TZP 5 mg (N=237) TZP 10 mg (N=240) TZP 15 mg (N=241) TZP all doses (N=718) Placebo (N=235)
Phase 3 Dose Effect Analysis Set (N=5,119)	SURPASS-1–5, SURPASS-J Mono, SURPASS-J Combo	First dose of treatment to end of safety follow-up visit or date of study withdrawal	Integrated data for dose comparison. Includes all studies with dose-escalation schedule proposed for the label	TZP 5 mg (N=1,701) TZP 10 mg (N=1,702) TZP 15 mg (N=1,716) TZP all doses (N=5,119)

Table 61 of the CS[3]
CS = company submission; TZP: tirzepatide

3.2.4.1 Overview of adverse events

There was an incremental increase with higher dose groups for the categories of treatment emergent adverse events (TEAEs) and discontinuation of study drug due to an adverse event (AE). The percentage of patients reporting serious adverse events (SAEs) was similar across the three tirzepatide dose groups in the dose effect analysis set and tirzepatide doses and placebo groups in the placebo-controlled analysis set. In the placebo-controlled analysis set, the percentage of discontinuations from study drug due to an AE was higher in the patients treated with tirzepatide ([redacted]%) compared to placebo ([redacted]%).

Table 3.14: Overview of adverse events

Category[a]	Dose effect analysis set n (%)				Placebo-controlled analysis set n (%)		
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)	TZP all doses (N=718)	Placebo (N=235)	TZP all doses versus placebo p-value
Deaths[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SAEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study due to AE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation from study drug due to AE[c]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TEAEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 62, 63 of the CS[3]
AE = adverse event; SAEs = serious adverse events; TEAEs = treatment-emergent adverse events; TZP = tirzepatide; % = percentage
[a]Patients may be counted in more than one category
[b]Deaths are also included as SAEs and discontinuations due to AEs
[c]Patients were to remain in the study after permanent discontinuation of study drug and initiation of an alternative antihyperglycaemic medication so additional data could be collected; such patients may have subsequently discontinued the study for the same or a different reason

3.2.4.2 Treatment emergent adverse events

The most frequently reported TEAEs were within the gastrointestinal (GI) disorders system organ class (SOC) with more patients treated with tirzepatide ([redacted]%) than patients treated with placebo ([redacted]%) in the placebo-controlled analysis set. An incremental increase with higher dose groups in the dose effect analysis set (5 mg, [redacted]%; 10 mg, [redacted]%; 15 mg, [redacted]%). A total of 336 patients (35.3%) experienced at least one TEAE in GI SOC. The TEAEs in the GI SOC were mostly mild in severity.

Table 3.15: TEAEs occurring in at least 5% of patients in any treatment group

Preferred Term	Dose effect analysis set n (%)				Placebo-controlled analysis set n (%)		
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)	TZP all doses (N=718)	Placebo (N=235)	TZP all doses versus placebo p-value
Nausea	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Nasopharyngitis	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Decreased appetite	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dyspepsia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vomiting	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Constipation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lipase increased	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hyperglycaemia					[redacted]	[redacted]	[redacted]

Table 64, 65 of the CS[3]
CS = company submission; TEAE = treatment-emergent adverse events; TZP = tirzepatide; % = percentage
*p-value denotes significantly higher levels of hyperglycaemia in the placebo group compared with the TZP groups

Table 3.16: Summary of TEAEs by maximum severity in the GI SOC (AS)

Preferred Term	Placebo-controlled analysis set n (%)					TZP all doses versus placebo p-value
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)	
Patients with ≥1 GI TEAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mild	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Severe	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 66 of the CS[3]
*Total includes one patient with a missing severity
CS = company submission; GI = gastrointestinal; SOC = system organ class; TEAE = treatment-emergent adverse events; TZP = tirzepatide

3.2.4.3 Cardiovascular risk

3.2.4.3.1 Systolic and diastolic blood pressure

There was no dose-response relationship between the tirzepatide dose groups in the percentages of patients meeting the threshold criteria for abnormal systolic blood pressure (SBP) or diastolic blood pressure (DBP) in dose effect analysis set. There were no notable differences between the tirzepatide dose groups and the placebo group in placebo-controlled analysis set.

Table 3.17: Summary of patients meeting threshold criteria for abnormal SBP and DBP at post baseline

Threshold criteria for abnormal BP (mg Hg)	Dose effect analysis set n (%)				Placebo-controlled analysis set n (%)		
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)	TZP all doses (N=718)	Placebo (N=235)	TZP all doses versus placebo p-value
SBP							
≥140 and CFB ≥20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≤90 and CFB ≤-20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DBP							
≥90 and CFB ≥10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≤50 and CFB ≤ -10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 67, 68 of the CS[3]
BP = blood pressure; CFB = change from baseline; CS = company submission; DBP = diastolic blood pressure; SBP = systolic blood pressure; TZP = tirzepatide

3.2.4.3.2 Pulse rate

Incremental increases in mean pulse rate from baseline with increasing tirzepatide dose were observed in the placebo-controlled analysis set and dose effect analysis set.

3.2.4.3.3 Heart rate

Incremental increases in mean electrocardiogram (ECG)-derived heart rate from baseline with increasing tirzepatide dose were observed in the placebo-controlled analysis set and dose effect analysis set. No clinically meaningful differences in treatment-emergent heart rate abnormalities between placebo and tirzepatide in placebo-controlled analysis set or between tirzepatide doses in the dose effect analysis set were observed.

3.2.4.3.4 CV meta-analysis

A total of 142 patients experienced the primary endpoint (adjudicated major adverse cardiovascular events-4 (MACE-4)) and contributed to the complete analysis. Overall, when comparing pooled tirzepatide to pooled comparator, the hazard ratio (HR) for the primary MACE-4 composite endpoint was 0.80 (95% confidence interval (CI): 0.57, 1.11).

3.2.4.4 Retinopathy

The results did not show increased risk of worsening retinopathy with tirzepatide treatment in the studied population, and there was not a dose-response relationship whereby the higher dose may increase risk of worsening retinopathy.

Table 3.18: Summary of potential treatment-emergent diabetic retinopathy complications

Preferred Term	Dose effect analysis set n (%)			
	TZP 5 mg (N=1,701)	TZP 10 mg (N=1,702)	TZP 15 mg (N=1,716)	TZP all doses (N=5,119)
Patients with ≥1 TEAE	[redacted]	[redacted]	[redacted]	[redacted]
Diabetic retinopathy	[redacted]	[redacted]	[redacted]	[redacted]
Macular oedema	[redacted]	[redacted]	[redacted]	[redacted]
Vision blurred	[redacted]	[redacted]	[redacted]	[redacted]
Retinal detachment	[redacted]	[redacted]	[redacted]	[redacted]
Retinal vein occlusion	[redacted]	[redacted]	[redacted]	[redacted]
Retinopathy hypertensive	[redacted]	[redacted]	[redacted]	[redacted]
Visual impairment	[redacted]	[redacted]	[redacted]	[redacted]
Amaurosis fugax	[redacted]	[redacted]	[redacted]	[redacted]
Diplopia	[redacted]	[redacted]	[redacted]	[redacted]
Maculopathy	[redacted]	[redacted]	[redacted]	[redacted]
Visual acuity reduced	[redacted]	[redacted]	[redacted]	[redacted]

Table 69 of the CS[3]
CS = company submission; TEAE = treatment-emergent adverse event; TZP = tirzepatide; % = percentage

3.2.4.5 Renal safety

The placebo-controlled analysis set is presented in Table 3.17. The CS states that *'overall, these data demonstrate that treatment with tirzepatide in patients with T2D does not significantly alter kidney function.'*. There was no indication of a dose-response relationship whereby the higher dose may increase risk of renal events.

Table 3.19: Summary of treatment-emergent renal events

SMQ Preferred Term	Placebo-controlled analysis set n (%)				
	TZP 5 mg (N=237)	TZP 10 mg (N=240)	TZP 15 mg (N=241)	TZP all doses (N=718)	Placebo (N=235)
Patients with ≥1 TEAE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Acute renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal impairment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Acute kidney injury	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Chronic kidney disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Chronic kidney disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Renal failure	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 70 of the CS[3]
CS = company submission; MedDRA = Medical Dictionary for Regulatory Activities; SMQ = standardised MedDRA query; TEAE = treatment-emergent adverse event; TZP = tirzepatide

3.2.4.6 Hypoglycaemia

The percentage of patients with episodes of severe hypoglycaemia in the phase 3 global studies, by background therapy, showed that the risk of severe hypoglycaemia with tirzepatide is low. There was no evidence that treatment with tirzepatide is associated with increased rates of severe hypoglycaemia. Also, there was not a dose-response relationship whereby the higher dose may increase risk of severe hypoglycaemia.

Table 3.20: Summary of severe hypoglycaemia postbaseline through the safety follow-up

Study (comparator)	Background therapy	Parameter	TZP 5 mg	TZP 10 mg	TZP 15 mg	Comparator
SURPASS-2 GPGL (SEMA 1 mg)	Metformin	N	470	469	470	469
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-3 GPGH (insulin degludec)	Metformin ± SGLT-2i	N	356	360	359	358
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-4 GPGM (insulin glargine)	Metformin ± SGLT-2i ± SU	N	329	328	338	1,000
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]
SURPASS-5 GPGI (placebo)	Insulin glargine ± metformin	N	116	119	120	120
		n (%); Episodes	[redacted]	[redacted]	[redacted]	[redacted]

Table 71 of the CS[3]
CS = company submission; SEMA = semaglutide; SGLT-2i = sodium-glucose co-transporter 2 inhibitor; SU = sulfonylurea; TZP = tirzepatide

The risk of hypoglycaemia with blood glucose <54 mg/dl (3.0 mmol/l) was higher when tirzepatide was used in combination with insulin glargine or SU as compared to other background glucose-lowering therapies studied, which has also been observed with the GLP-1 RA class.

The percentages of patients reporting hypoglycaemia with blood glucose <54 mg/dl (3.0 mmol/l) were similar in tirzepatide and placebo-treated patients. The percentage of patients reporting hypoglycaemia with blood glucose <54 mg/dl (3.0 mmol/l) was lower in tirzepatide-treated patients compared to basal insulin-treated patients, but higher in the tirzepatide 15 mg group compared with the semaglutide 1 mg group.

These data demonstrate that treatment with tirzepatide in patients with T2D mellitus is associated with a low risk of hypoglycaemia. Severe hypoglycaemia was uncommon with tirzepatide treatment. Overall, the risk of hypoglycaemia with tirzepatide was comparable to the GLP-1 RA class.

EAG comment: Overall, the EAG agrees with the company's statement that: "*As expected, similar to the GLP-1 RA class, the most common AEs in patients treated with tirzepatide were GI related."* (page 168).[3] However, the EAG notes the following points in relation to the safety data presented:

- The CS includes the statement: *"These results did not show increased risk of worsening retinopathy with tirzepatide treatment in the studied population."* (page 175).[3] However, the data presented (Table 71 of the CS, and Table 3.16, above) on treatment-emergent diabetic retinopathy complications are from the dose-effect analysis set, i.e. for TZP-treated patients only with no comparator data.
- The CS includes the statement that: *"Overall, these data demonstrate that treatment with tirzepatide in patients with T2D does not significantly alter kidney function."* (page 175).[3] However, the data presented (Table 72 of the CS and Table 3.17, above) appear to indicate that TZP treatment was associated with higher rates of renal AE relative to placebo.
- In response to the clarification letter, the company provided a pooled analysis of cardiac risk, for the primary endpoint (adjudicated MACE-4).[4] Whilst the EAG agrees that, overall, this analysis indicates that tirzepatide was associated with a reduced risk of adverse cardiac events than the combined comparator data set, the EAG considers that this analysis is of limited value because it does not provide any information about the relative cardiac risk of tirzepatide versus relevant individual comparators.

3.2.5 Subgroup analyses

The CS presented a list of the pre-planned subgroup analyses in the SURPASS trials 2 to 5 for change from baseline in HbA1c and body weight (at 40 weeks for SURPASS-2 and 5 and 52 weeks for SURPASS-3 and 4). The characteristics are summarised in Table 3.21.

Table 3.21: Subgroup analyses for SURPASS trials (2 to 5)

Subgroup analyses
Age (<65 versus ≥65 years, age group 1)
Age (<75 versus ≥75 years, age group 2)
Race
Ethnicity
Sex
Geographic region (US versus OUS)
Duration of diabetes (<median versus ≥median, duration of diabetes group 1)
Duration of diabetes (≤5 years versus >5 to ≤10 years versus >10 years, duration of diabetes group 2)
Baseline HbA1c (≤8.5% versus >8.5%)
Baseline eGFR (<60 ml/min/1.73 m² versus ≥60 ml/min/1.73 m²)
Baseline BMI (<27 kg/m² versus ≥27 kg/m², baseline BMI group 1)
Baseline BMI (<30 kg/m² versus ≥30 to <35 kg/m² versus ≥35 kg/m², baseline BMI group 2)
Prior use of OAD (yes versus no): SURPASS-3 only
Baseline OAD use (metformin alone, metformin plus SU, metformin plus SGLT-2i, metformin plus SU plus SGLT-2i, other) – SURPASS-4 only
Baseline use of metformin (yes versus no) – SURPASS-5 only
Table 45, CS.[3]
BMI = body mass index; CS =company submission; eGFR = estimated Glomerular Filtration Rate; HbA1c = glycated haemoglobin; OAD = oral antidiabetic drug; OUS = outside the USA; SGLT-2i = sodium-glucose c-transporter-2 inhibitor; SU = sulfonylurea; US = United States; USA = United States of America

The CS did not provide the results of these analyses, so the EAG have produced a summary of characteristics that were found to have a significant interaction with the treatment effect (p<0.1).

Table 3.22: Summary of subgroup analysis results with p values for characteristics found to have a statistically significant interaction with treatment effect versus comparator (p<0.1)

Characteristics	SURPASS-2		SURPASS-3		SURPASS-4		SURPASS-5		N trials where p<0.1
	FAS	EAS	FAS	EAS	FAS	EAS	FAS	EAS	
HbA1c									
Age Group 1 (<65 versus ≥65 years)			[redacted]		[redacted]				[redacted]
Age (<75 versus ≥75 years, age group 2)					[redacted]				[redacted]
Race	[redacted]								[redacted]
Ethnicity			[redacted]	[redacted]					[redacted]
Sex					[redacted]	[redacted]			[redacted]
Baseline HbA1c							[redacted]	[redacted]	[redacted]
Baseline eGFR								[redacted]	[redacted]
Baseline BMI group 1 (<27 versus ≥27 kg/m²)	[redacted]		[redacted]	[redacted]	[redacted]	[redacted]			[redacted]

Characteristics	SURPASS-2		SURPASS-3		SURPASS-4		SURPASS-5		N trials where p<0.1
	FAS	EAS	FAS	EAS	FAS	EAS	FAS	EAS	
Baseline BMI group 2 (<30 kg/m² versus ≥30 to <35 kg/m² versus ≥35 kg/m²)		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			[redacted]
Baseline OAD use (metformin alone, metformin plus SU, metformin plus SGLT-2i, metformin plus SU plus SGLT-2i, other)					[redacted]	[redacted]			[redacted]
Baseline use of metformin							[redacted]	[redacted]	[redacted]
Body weight									
Age group 1 (<65 versus ≥65 years)	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]			[redacted]
Age (<75 versus ≥75 years, age group 2)					[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Race					[redacted]	[redacted]			[redacted]
Ethnicity	[redacted]	[redacted]		[redacted]	[redacted]	[redacted]			[redacted]
Sex					[redacted]	[redacted]			[redacted]
Geographic region (US versus OUS)	[redacted]				[redacted]				[redacted]
Duration of diabetes group 1 (<median [7.1] years versus ≥median [7.1 years])	[redacted]	[redacted]					[redacted]	[redacted]	[redacted]
Duration of diabetes group 2 (≤5 years versus >5 to ≤10 years versus >10 years)							[redacted]		[redacted]
Baseline BMI (<27 kg/m² versus ≥27 kg/m², baseline BMI group 1)					[redacted]	[redacted]			[redacted]
Baseline BMI group 2 (<30 kg/m² versus ≥30 to <35 kg/m² versus ≥35 kg/m²)	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]			[redacted]

Tables GPGL.5.18 to GPGL.5.21,[22] Tables GPGH.5.15 to GPGH.5.18,[20] Tables GPGM.5.25 to GPGM.5.28,[19] and Tables GPGI.5.16 to GPGI.5.19.[21]
BMI = body mass index; eGFR = estimated Glomerular Filtration Rate; FAS = full analysis set; HbA1c = glycated haemoglobin; OAD = oral antidiabetic drug; OUS = outside the USA; SGLT-2i = sodium-glucose c-transporter-2 inhibitor; SU = sulfonylurea; US = United States; USA = United States of America

EAG comment: Of course, lack of statistical significance does not imply no effect and statistical significance does not inform the magnitude of any effect, so the EAG would make inferences from this table with caution. However, based on the number of trials where this is observed, it does seem to be the case that age, and BMI appear to generally be treatment effect modifiers. The effect of OAD use on HbA1c change also seems to be likely given that it was observed in the form particular to those trials in two of the three of the trials where it could be observed, i.e., in SURPASS-4 as one of four combinations and in SURPASS-5 as metformin or not. However, the direction of the effect is unfortunately not easy to determine, as observed in the HbA1c changes per OAD group in SURPASS-4.[19]

3.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

An NMA was conducted with the aim to assess the relative efficacy and safety of tirzepatide versus GLP-1 RAs, especially those available in National Health Service (NHS) practice. The network (trials included in the NMA) was stated to have been defined to align with SURPASS-2 and 3 trials as these trial designs most closely align with the anticipated positioning of tirzepatide in UK clinical practice.[3] According to the company the NMA provides results on the comparative efficacy and safety of the three examined doses of tirzepatide (5 mg, 10 mg and 15 mg) versus GLP-1 RAs at the second and third line of treatment.

Evidence from RCTs was identified in the clinical SLR referred to in Section 3.1. The company has clarified that this not-NICE-specific SLR was not designed for this CS but for a broader scope to meet the needs of multiple HTA agencies around the world.[4] As such a second set of exclusion criteria was applied to select the studies included in the NMA for this CS. Studies included in the NMA were conducted in patients with one to two OADs (partially aligned with SURPASS-2 and 3 trials). "*More specifically, the population included studies including patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin. Trials with an unknown proportion of patients on metformin background therapy and trials including patients on ≥3 OADs were excluded from the main analysis. These trials were included in the sensitivity analyses described in Section B.2.9.7.3.*" (page 106 of the CS[3]) After the request of the EAG the company clarified that "*These proportions were chosen to ensure comparability to SURPASS-2, in which 100% of patients were on only metformin, and SURPASS-3, in which 68% of patients were on metformin only and 32% were on metformin and SGLT-2i.*" (page 39[4])

Baseline characteristics of mean age, female proportion, mean BMI, mean body weight, diabetes duration, mean HbA1c and trial duration (>104 weeks) were not part of the NMA eligibility criteria. Similarly, comorbidities were not defined as an exclusion criterion. Patients with any comorbidities, including cardiovascular disease (CVD)/CV high risk, obesity and non-alcoholic fatty liver disease were include in the NMA.

The comparators were chosen to reflect the treatment options for T2D patients in real world clinical practice. A very broad list of comparators was first populated for the not-NICE-specific SLR, and then narrowed down to reflect treatments in the UK. The summary of product characteristics (SmPC) for each specified comparator was used to identify licenced doses available at the time the NMA was executed. The treatments that were defined as relevant and therefore were included in the NMA are listed below (in alphabetical order):

- Dulaglutide 0.75 mg QW
    - This dose is currently only licensed as monotherapy and as a starting dose for patients who may be considered more vulnerable, therefore only relevant to a sub-population in UK clinical practice. Results from the NMA are presented although no formal comparisons have been made.
- Dulaglutide 1.5 mg QW
- Dulaglutide 3.0 mg QW
- Dulaglutide 4.5 mg QW
- Exenatide 5 µg BID (pre-filled pen)
- Exenatide 10 µg BID (pre-filled pen)
- Exenatide 2.0 mg QW (powder and solvent for prolonged-release suspension for injection)

- Lixisenatide 20 µg once daily (QD)
- Liraglutide 1.2 mg QD
- Liraglutide 1.8 mg QD
- Semaglutide 0.5 mg QW (injectable)
- Semaglutide 1.0 mg QW (injectable)
- Semaglutide 2.0 mg QW (injectable)
    - This dose is not currently available in UK clinical practice but is included in the NMA results.
- Semaglutide 7.0 mg QD (oral)
- Semaglutide 14 mg QD (oral)
- Placebo

Non GLP-1 RA treatment arms (such as basal insulin, bolus insulin, premixed insulin, dipeptidyl-peptidase 4 inhibitors (DPP-4i), sulfonylurea (SU), thiazolidinediones (TZD), sodium-glucose cotransporter-2 inhibitors (SGLT-2i) and placebo were also considered to make connections in the network. The reference treatments were: placebo, tirzepatide 5 mg, 10 mg and 15 mg. As such the results were presented as treatment relative to tirzepatide 5 mg, 10 mg and 15 mg.

In terms of endpoints, the following efficacy endpoints were included in the main analysis of the NMA:

- Change from baseline in HbA1c (%)
- Change from baseline in weight (kg)
- Change from baseline in BMI (BMI; kg/m²)
- Change from baseline in LDL (mmol/l)
- Change from baseline in HDL (mmol/l)
- Change from baseline in eGFR (ml/min/1.73 m²)

These safety endpoints were also included:

- Change from baseline in SBP
- Proportion of patients experiencing nausea

Of the 205 studies included in the not-NICE-specific SLR, a total of 72 were eligible for inclusion in the NMA (53 in the main analysis and 19 in sensitivity analyses only). The reasons for exclusion are presented in Figure 3.1. A summary of the rationale for study exclusion is reported below.

Seven trials focusing on cardiovascular outcomes were excluded due to their design and scope. Cardiovascular outcomes trials (CVOTs) had extended trial periods with mean follow-ups of at least 14 months, they included non-T2D patients, the background therapies included injectables, while glycaemic efficacy was not assessed. Five trials were excluded due to including patients with renal impairment (stage 3 or 4 chronic kidney disease (CKD) or macroalbuminuria). This exclusion criterion was set to make the included populations as generalisable as possible. Twenty studies were excluded due to their comparators (or combination of therapy) not being available in the UK clinical practice or not relevant to the study design of the SURPASS trials. Three studies were excluded because they used flexible doses that are not approved in the UK. In addition, flexible doses hindered direct comparisons with the SURPASS trials. Twenty studies were excluded as data were not available for the time interval addressed in the NMA (see Section 3.4). One study was excluded as it was focused on the effects of the Ramadan season and one because it focused on patients who were severely insulin resistant. Seventy-six studies were excluded as the background therapies were other that one to two OADs (patients treated

with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin). It should be noted that SURPASS-J-Mono, SURPASS-J-Combo and SURPASS-AP-Combo were excluded from the efficacy evidence because they were conducted in Asian population and considered not generalisable to the UK population. Nevertheless, SURPASS-J-Mono and SURPASS-J-Combo were used in the safety analysis.

Figure 3.1: Reasons for exclusion of studies in NMA

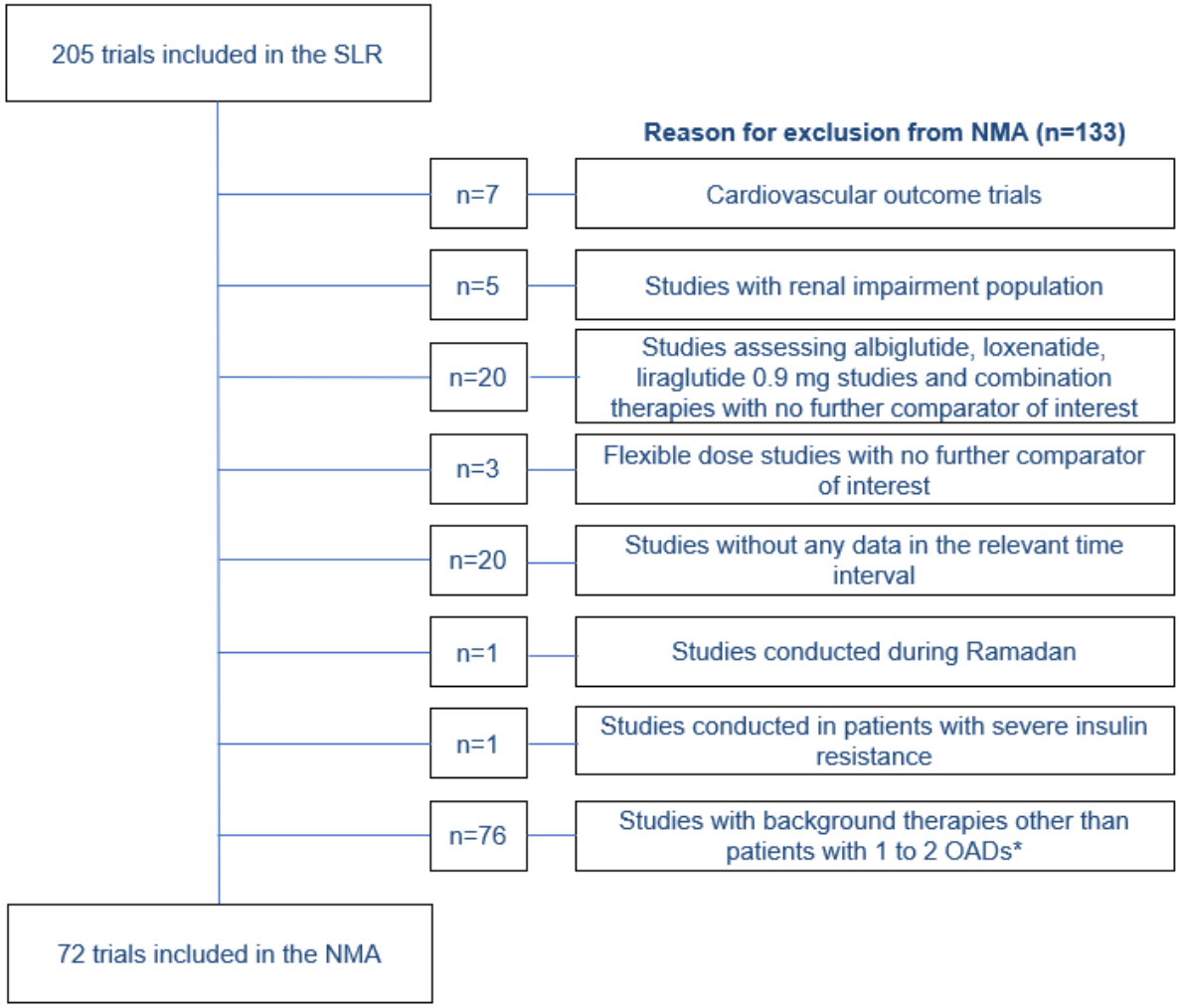


Figure 3 of the CS[3]
CS = company submission; NMA = network meta-analysis; OAD = oral antidiabetic drug; SLR = systematic literature review
*Defined as included studies including patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin.

The trial design characteristics are presented in Table 3.21 and 3.22 as well as in Figures 3.2 and 3.3. Following an EAG request,[4] the company has clarified that the summary of background therapies presented in Table 3.21 refer to treatments that were allowed within studies and were used in both intervention and comparator arms of the included studies for their full duration, with the exception of administered rescue therapies. The company also provided a more detailed list of background therapy in the files accompanying their response to clarification[4], according to which out of the 184 arms included in the NMA, 80 (43.5%) included patients that were treated with only metformin. 75 (40.8%) of the arms included patients that were treated with metformin plus only a second OAD, 21 (11.4%) of the arms included patients that were treated with metformin plus two other OADs, and 8 (4.4%) of the

arms included patients that were treated with metformin plus three other OADs. It should be noted that not all patients in each arm were treated with the same number or the same mix of background treatments. Within the arms receiving metformin plus only a second OAD, only 23 out of the 75 arms were made out of 100% of patients reiving two treatments: either metformin plus SU or metformin plus thiazolidinedione. In the arms receiving three or four OADs there was no arm where all patients were receiving three or four OADs.

Regarding the blinding status of the included trials as reported in Table 3.21, the company clarified that "*In the main analysis, 22/53 of studies were open label. In diabetes it is common to design open label studies given differences between injection devices of various comparators, as well as the distinct tolerability profile associated with GLP-1 RAs. The risk of bias thus introduced is mitigated by the objective outcomes measures used for primary and key secondary outcomes, such as HbA1c and weight, although it is acknowledged that specific safety outcomes may be more subjective and thus open to bias. In SURPASS studies, even if studies were open-label, every effort was taken to minimise the potential for biases in the study design: the study team remained blinded to the treatment assignment, within tirzepatide arms, the dose was blinded to patients, investigator and sponsor.*" (page 46[4])

Table 3.23: Summary of type of background therapy received across included studies

Background therapy received	Number of studies
Metformin monotherapy	25
Metformin alone or metformin plus SU	15
Metformin, SU, glitazones	1
Metformin alone or metformin plus glitazones	6
Metformin alone or metformin plus SGLT-2i	4
SGLT-2i alone or SGLT-2i plus metformin or SU	1
Metformin alone or metformin plus SU, DPP-4i, SGLT-2i or glinides	1
Table 47, CS.[3] CS = company submission; DPP-4i = dipeptidyl peptidase-4 inhibitors; SU = sulphonyl urea; SGLT-2i = sodium glucose cotransport-2 inhibitors	

Table 3.24: Summary of blinding status across included studies

Blinding status	Number of studies
Single-blind	2
Double-blind	26
Triple-blind	1
Open label	22
Mixed[a]	1
Not reported	1
Table 48, CS.[3] CS = company submission [a]Mixed trials included both double-blind and open-label design	

Figure 3.2: Summary of lower and upper bound for HbA1c inclusion criteria in each study included in the main analyses

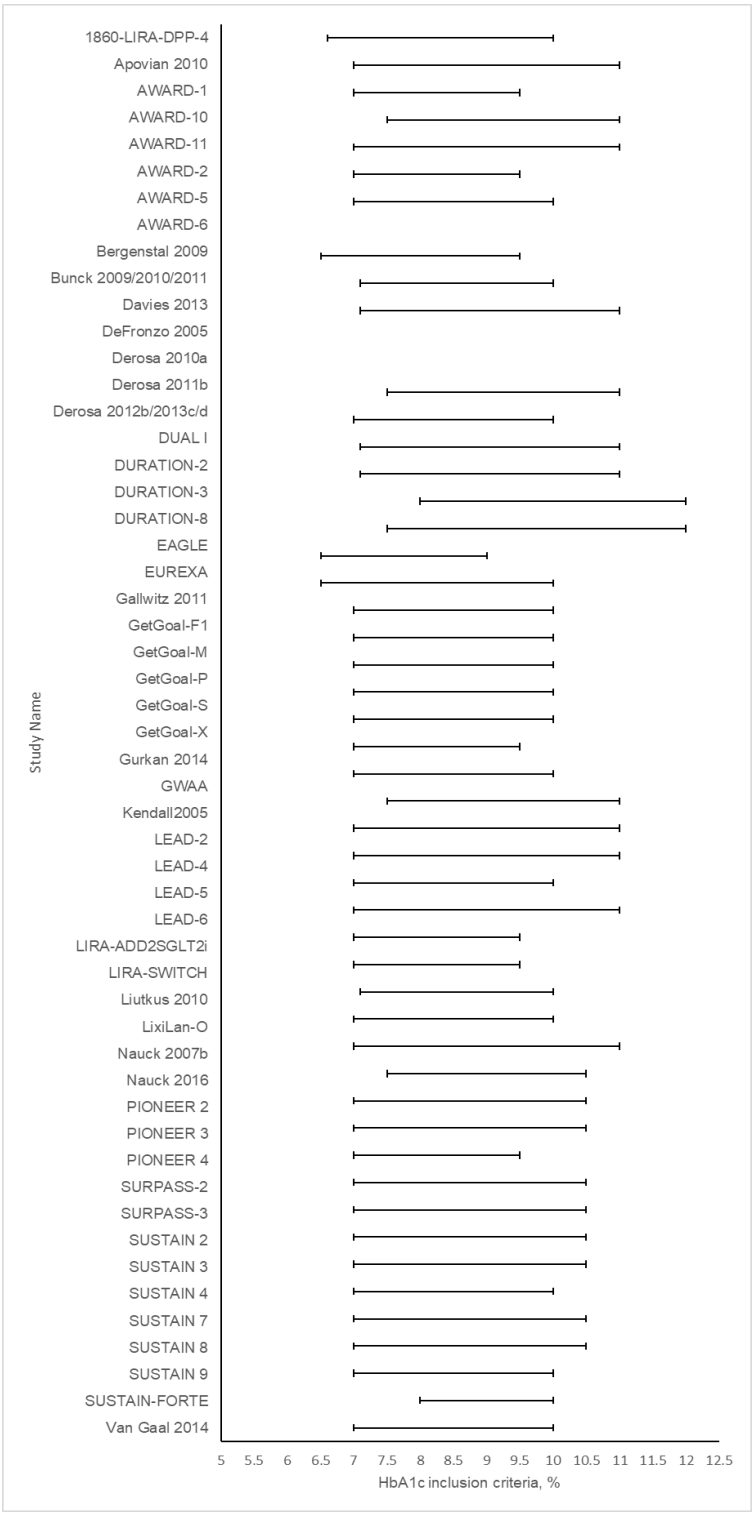


Figure 24, CS.[3]

BID = twice daily; CS = company submission; HbA1c = glycated haemoglobin; QD = once daily; QW = once weekly

Derosa 2010, Derosa 2011b and Bergenstal 2011 did not report an upper limit for HbA1c. As such, these studies are not presented on the above figure. The lower HbA1c limit for these 3 trials was 7%.

Figure 3.3: Summary of primary treatment goal (glycaemic control) and comorbidities included in each study included in the main analysis

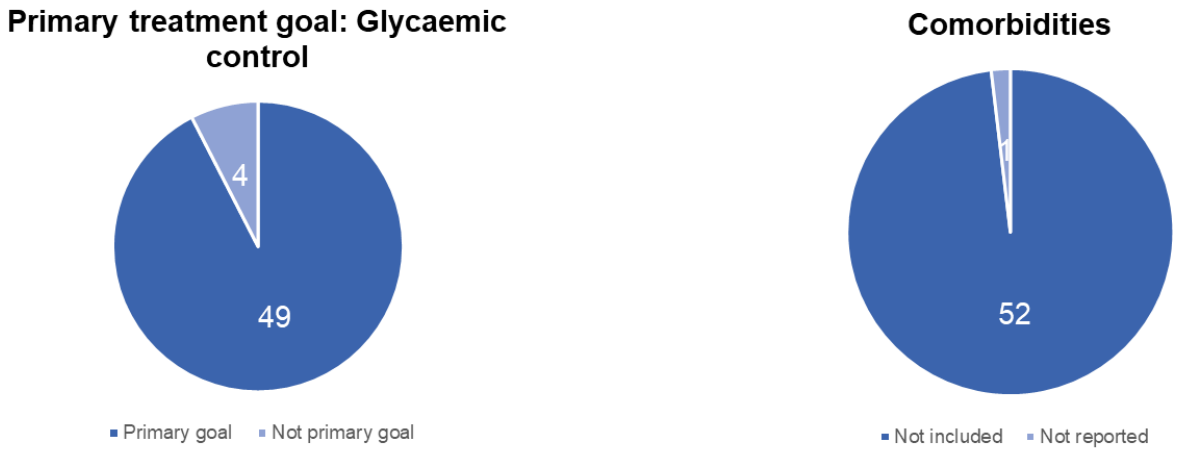


Figure 25, CS.[3]
CS = company submission
The two treatment arms that included comorbidities specified 'Obese'

The baseline characteristics of the studies included in the efficacy network are reported in Figures 30 to 37 of the CS[3] and in an Excel file provided with the company's response to request for clarification.[4] A summary of the mean and the ranges of is presented in Table 3.24. According to the company the baseline characteristics were largely consistent across the treatment arms. No further discussion was provided assessing and comparing baseline characteristics in the CS.

Table 3.25: Summary of baseline characteristics across the study arms

Baseline characteristics	Mean value	Minimum value	Maximum value
Number of patients	264.7	17.0	834.0
Proportion of female patients, %	47.4	31.0	70.0
Mean age, years	55.9	42.7	59.8
Mean baseline weight, kg	91.9	80.2	101.9
Mean baseline BMI, kg/m²	32.75	28.4	36.8
Mean baseline HbA1c, %	8.3	7.4	10.3
Mean baseline duration of diabetes, years	7.6	0.6	10.1
Mean treatment duration, weeks	46.5	24.0	156.0
Race			
Proportion of Caucasian patients, %	78.3	30.0	100.0
Proportion of Black patients, %	5.8	0.0	26.6
Proportion of Asian patients, %	9.0	0.0	45.3
Proportion of Other patients, %	4.6	0.0	21.0
Ethnicity			
Proportion of Hispanic patients, %	23.1	0.0	71.6
Proportion of non-Hispanic patients, %	66.4	0.0	96.0

Table 49 of the CS[3] and Table 8 or the response to request for clarification[4]
BMI = body mass index; CS = company submission; HbA1c = glycated haemoglobin; % = percentage

After the request of the EAG some of the baseline characteristics (female percentage, mean age, mean weight, mean HbA1c, mean duration of diabetes and treatment duration) across treatment arms were

also provided in a tabular form and are now presented in Table 3.35. Even by examining the ranges of baseline characteristics in Table 3.26 is it obvious that there are staggering variations between studies.

Table 3.26: Baseline characteristics by treatment arm

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
1860-LIRA-DPP-4	218	Liraglutide 1.8 mg	48	55 (9.1)	94.6 (18.1)	8.4 (0.7)	6.4 (5.4)	26 (78)
1860-LIRA-DPP-4	219	Sitagliptin 100 mg	45	55 (9)	93.1 (18.9)	8.5 (0.7)	6.3 (5.4)	26 (78)
1860-LIRA-DPP-4	221	Liraglutide 1.2 mg	48	55.9 (9.6)	93.7 (18.4)	8.4 (0.8)	6 (4.5)	26 (78)
Apovian 2010	96	Exenatide 10µg BID	63	54.5 (10)	94.9 (16.5)	7.7 (0.9)	5.7 (5.5)	24 (24)
Apovian 2010	98	Placebo	62	55.1 (9)	96.2 (15.6)	7.5 (0.8)	5.3 (5.1)	24 (24)
AWARD-1	141	Placebo	41.1	54.56 (10.01)	94.12 (19.28)	8.06 (1.31)	8.6 (5.78)	52 (52)
AWARD-1	280	Dulaglutide 0.75 mg	40	55.79 (9.45)	95.53 (20.56)	8.05 (1.24)	8.78 (5.47)	52 (52)
AWARD-1	279	Dulaglutide 1.50 mg	41.6	56.25 (9.72)	96.22 (19.63)	8.1 (1.34)	8.76 (5.59)	52 (52)
AWARD-1	276	Exenatide 10 µg BID	43.5	55.45 (10.15)	97.37 (18.87)	8.07 (1.34)	8.84 (5.71)	52 (52)
AWARD-10	142	Dulaglutide 1.50 mg	46	56.17 (9.26)	92.87 (19.73)	8.04 (0.65)	9.21 (5.74)	24 (36)
AWARD-10	141	Dulaglutide 0.75 mg	51	58.55 (9.14)	91.07 (20.99)	8.04 (0.61)	10.05 (6.56)	24 (36)
AWARD-10	140	Placebo	53	57.1 (9.59)	90.5 (19.47)	8.05 (0.66)	8.87 (6.13)	24 (36)
AWARD-11	616	Dulaglutide 3.0 mg	46.8	56.9 (10.2)	96.33 (20.14)	8.63 (1)	7.58 (5.52)	52 (52)
AWARD-11	614	Dulaglutide 4.5 mg	48.2	56.6 (10.2)	95.35 (20.63)	8.64 (0.91)	7.65 (5.81)	52 (52)
AWARD-11	612	Dulaglutide 1.50 mg	51.3	57.8 (9.7)	95.48 (20.17)	8.64 (0.94)	7.56 (5.78)	52 (52)
AWARD-2	262	Glargine	48.9	57.21 (9.38)	87.66 (19.62)	8.1 (0.95)	8.87 (5.98)	78 (92)
AWARD-2	272	Dulaglutide 0.75 mg	50	56.56 (9.27)	86.18 (18.15)	8.13 (0.98)	9.28 (9.53)	78 (92)
AWARD-2	273	Dulaglutide 1.50 mg	47.3	56.24 (9.76)	85.13 (17.9)	8.18 (1.03)	9.13 (6.22)	78 (92)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
AWARD-5	177	Placebo/sitagliptin 100 mg	49.2	54.91 (9.05)	87.07 (16.86)	8.1 (1.14)	6.96 (5.43)	104 (104)
AWARD-5	315	Sitagliptin 100 mg	52.1	53.75 (10.27)	85.97 (16.91)	8.09 (1.09)	7.16 (4.89)	104 (104)
AWARD-5	302	Dulaglutide 0.75 mg	55.6	54.35 (9.81)	86.22 (17.99)	8.19 (1.11)	7.34 (4.92)	104 (104)
AWARD-5	304	Dulaglutide 1.50 mg	52	53.66 (10.02)	86.67 (17.45)	8.12 (1.05)	6.95 (5.5)	104 (104)
AWARD-6	299	Dulaglutide 1.50 mg	53.8	56.49 (9.34)	93.82 (18.23)	8.06 (0.81)	7.13 (5.41)	26 (26)
AWARD-6	300	Liraglutide 1.8 mg	50.3	56.81 (9.91)	94.35 (18.96)	8.05 (0.79)	7.28 (5.41)	26 (26)
Bergenstal 2009	124	BIAsp30 BID	52.4	53.4 (9.96)	93.8 (24)	10.3 (1.92)	9.9 (5.6)	24 (24)
Bergenstal 2009	124	BIAsp30 QD	51.6	51.8 (10.9)	96.9 (25)	10.1 (1.79)	8.4 (6.3)	24 (24)
Bergenstal 2009	124	Exenatide 10 µg BID	51.6	52.5 (10.62)	96.6 (24)	10.2 (1.52)	8.6 (5.9)	24 (24)
Bunck 2009/2010/2011	36	Exenatide 10 µg BID	36.1	58.4 (NA)	90.6 (NA)	7.6 (NA)	NA (NA)	52 (56)
Bunck 2009/2010/2011	33	Glargine	33.3	58.3 (NA)	92.4 (NA)	7.4 (NA)	NA (NA)	52 (56)
Davies 2013	111	Exenatide 2 mg QW	36	59 (10)	96.7 (17)	8.37 (0.85)	8 (6)	26 (30)
Davies 2013	105	Detemir	31	58 (10)	97.9 (15.8)	8.35 (0.88)	7 (5)	26 (30)
DeFronzo 2005	110	Exenatide 5 µg BID	48.2	53 (11)	100 (22)	8.3 (1.1)	6.2 (5.9)	30 (34)
DeFronzo 2005	113	Exenatide 10 µg BID	39.8	52 (11)	101 (20)	8.2 (1)	4.9 (4.7)	30 (34)
DeFronzo 2005	113	Placebo	40.7	54 (9)	100 (19)	8.2 (1)	6.6 (6.1)	30 (34)
Derosa 2010a	65	Glibenclamide	49.23	56 (7)	82.4 (9.1)	8.9 (0.8)	NA (NA)	52 (52)
Derosa 2010a	63	Exenatide 10 µg BID	52.3	57 (8)	82 (8.3)	8.8 (0.7)	NA (NA)	52 (52)
Derosa 2011b	57	Exenatide 10 µg BID	50.87719	56 (7)	80.2 (7.5)	8.7 (0.7)	NA (NA)	52 (NA)
Derosa 2011b	54	Glimepiride	51.85185	55 (6)	81.4 (8.1)	8.8 (0.8)	NA (NA)	52 (NA)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
Derosa 2012b/2013c/d	85	Placebo	51.76	56.7 (7.3)	90.5 (10.3)	7.9 (0.6)	0.65 (NA)	52 (52)
Derosa 2012b/2013c/d	86	Exenatide 10 µg BID	50	57.3 (7.7)	89 (9.7)	8.1 (0.8)	0.63 (NA)	52 (52)
DUAL I	413	Degludec	52	54.9 (9.7)	87.4 (19.2)	8.3 (1)	7 (5.3)	52 (52)
DUAL I	834	IDegLira	48	55.1 (9.9)	87.2 (19)	8.3 (0.9)	6.6 (5.1)	52 (52)
DUAL I	414	Liraglutide 1.8 mg	50	55 (10.2)	87.4 (18)	8.3 (0.9)	7.2 (6.1)	52 (52)
DURATION-2	165	Pioglitazone	52	53 (10)	88 (20)	8.5 (1.1)	6 (5)	26 (26)
DURATION-2	166	Sitagliptin 100 mg	48	52 (11)	87 (20)	8.5 (1.2)	5 (4)	26 (26)
DURATION-2	160	Exenatide 2 mg QW	44	52 (10)	89 (20)	8.6 (1.2)	6 (5)	26 (26)
DURATION-3	223	Glargine	45	58 (9)	90.6 (16.4)	8.3 (1)	7.8 (6)	156 (194)
DURATION-3	233	Exenatide 2 mg QW	48	58 (10)	91.1 (18.6)	8.3 (1.1)	8 (5.9)	156 (194)
DURATION-8	227	Exenatide 2 mg QW	48.9	54.2 (9.6)	89.77 (20.22)	9.3 (1.06)	7.4 (5.5)	52 (53)
DURATION-8	230	Dapagliflozin	52.2	54.5 (9.2)	91.06 (19.71)	9.3 (1.03)	7.1 (5.5)	52 (53)
EAGLE	474	Glargine	47.3	57.1 (8.8)	90.8 (16.6)	9 (1)	NA (NA)	24 (24)
EAGLE	470	Liraglutide 1.8 mg	44	57.4 (8.9)	90.1 (16.7)	9.1 (1.1)	NA (NA)	24 (24)
EUREXA	487	Glimepiride	48	56 (9.1)	91.1 (14.8)	7.4 (0.7)	5.5 (4.3)	156 (156)
EUREXA	490	Exenatide 10 µg BID	44	56 (10)	92.8 (16.7)	7.5 (0.7)	5.8 (4.8)	156 (156)
Gallwitz 2011	181	BIAsp30 BID	NA	57 (9.9)	NA (NA)	7.9 (0.9)	5 (5)	26 (26)
Gallwitz 2011	182	Exenatide 10 µg BID	NA	57 (10)	NA (NA)	7.9 (0.8)	5 (4)	26 (26)
GetGoal-F1	160	Placebo	55	58.2 (9.8)	87.9 (17.3)	8 (0.8)	6.2 (4.7)	76 (79)
GetGoal-F1	161	Lixisenatide 20 µg	55	54.6 (8.9)	88 (16.8)	8.1 (0.9)	6 (4.6)	76 (79)
GetGoal-F1	161	Lixisenatide 20 µg	56	55.4 (8.9)	90.3 (19)	8 (0.9)	5.8 (3.9)	76 (79)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
GetGoal-M	170	Placebo	52.4	55 (9.4)	90.4 (20.1)	8.1 (0.9)	5.9 (4.7)	24 (27)
GetGoal-M	255	Lixisenatide 20 µg	61.6	54.5 (9.2)	90.1 (21)	8 (0.9)	6.2 (5.3)	24 (27)
GetGoal-M	255	Lixisenatide 20 µg	55.3	54.8 (10.4)	89 (20.7)	8.1 (0.9)	6.2 (5.4)	24 (27)
GetGoal-P	323	Lixisenatide 20 µg	47	56 (9.5)	92.9 (22.9)	8.1 (0.9)	8.1 (5.4)	76 (79)
GetGoal-P	161	Placebo	49	55.3 (9.5)	96.7 (25.6)	8.1 (0.8)	8.1 (5.6)	76 (79)
GetGoal-S	574	Lixisenatide 20 µg	50.5	57 (9.8)	82.6 (21.9)	8.3 (0.9)	9.1 (6)	52 (52)
GetGoal-S	285	Placebo	47.4	57.8 (10.1)	84.5 (22.8)	8.2 (0.8)	9.8 (6.2)	52 (52)
GetGoal-X	316	Exenatide 10 µg BID	40.8	57.6 (10.7)	96.1 (22.5)	8.02 (0.8)	6.8 (4.9)	24 (24)
GetGoal-X	318	Lixisenatide 20 µg	52.5	57.3 (9.2)	94 (19.6)	8.03 (0.8)	6.8 (5.5)	24 (24)
Gurkan 2014	17	Glargine	58.8	53.12 (6.99)	90.51 (14.32)	8.11 (0.76)	7.59 (4.26)	26 (26)
Gurkan 2014	17	Exenatide 10 µg BID	70	52.18 (7.26)	94.34 (11.77)	7.95 (0.81)	6.88 (3.26)	26 (26)
GWAA	282	Exenatide 10 µg BID	45	59.8 (8.8)	87.5 (16.9)	8.2 (1)	9.9 (6)	26 (26)
GWAA	267	Glargine	43.4	58 (9.5)	88.3 (17.9)	8.3 (1)	9.2 (5.7)	26 (26)
Kendall2005	241	Exenatide 10 µg BID	40.7	55 (10)	98 (21)	8.5 (1.1)	8.7 (6.4)	30 (30)
Kendall2005	245	Exenatide 5 µg BID	40.8	55 (9)	97 (19)	8.5 (1)	8.7 (5.9)	30 (30)
Kendall2005	247	Placebo	44.1	56 (10)	99 (19)	8.5 (1)	9.4 (6.2)	30 (30)
LEAD-2	242	Liraglutide 1.8 mg	41	57 (9)	NA (NA)	8.4 (1)	8 (5)	26 (104)
LEAD-2	121	Placebo	40	56 (9)	NA (NA)	8.4 (1.1)	8 (6)	26 (104)
LEAD-2	240	Liraglutide 1.2 mg	46	57 (9)	NA (NA)	8.3 (1)	7 (5)	26 (104)
LEAD-2	242	Glimepiride	43	57 (9)	NA (NA)	8.4 (1)	8 (5)	26 (104)
LEAD-4	178	Liraglutide 1.2 mg	43	55 (10)	NA (NA)	8.5 (1.2)	9 (6)	26 (26)
LEAD-4	177	Placebo	38	55 (10)	NA (NA)	8.4 (1.2)	9 (6)	26 (26)
LEAD-4	178	Liraglutide 1.8 mg	49	55 (11)	NA (NA)	8.6 (1.2)	9 (6)	26 (26)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
LEAD-5	232	Glargine	40	57.5 (10.5)	85 (17.9)	8.2 (0.9)	9.7 (6.4)	26 (26)
LEAD-5	114	Placebo	51	57.5 (9.6)	85.7 (16.7)	8.3 (0.9)	9.4 (6.2)	26 (26)
LEAD-5	230	Liraglutide 1.8 mg	43	57.6 (9.5)	85.5 (19.4)	8.3 (0.9)	9.2 (5.8)	26 (26)
LEAD-6	233	Liraglutide 1.8 mg	51	56.3 (9.8)	93.1 (20.1)	8.2 (1)	8.5 (6.2)	26 (26)
LEAD-6	231	Exenatide 10 µg BID	45	57.1 (10.8)	93 (19.5)	8.1 (1)	7.9 (5.9)	26 (26)
LIRA-ADD2SGLT-2i	100	Placebo	42	56 (9.9)	91.4 (21.4)	8 (0.6)	9.6 (6.7)	27 (29)
LIRA-ADD2SGLT-2i	203	Liraglutide 1.8 mg	38	54.7 (10.1)	91 (21)	8 (0.7)	10.1 (7.2)	27 (29)
LIRA-SWITCH	204	Sitagliptin 100 mg	39	56.5 (9.7)	91.2 (19.6)	8.2 (0.6)	7.6 (6.2)	26 (26)
LIRA-SWITCH	202	Liraglutide 1.8 mg	42	56.3 (10.6)	88.9 (19.8)	8.3 (0.6)	7.9 (5.7)	26 (26)
Liutkus 2010	111	Exenatide 10 µg BID	40	55 (8)	94.5 (17.8)	8.2 (0.9)	6.3 (4.2)	26 (26)
Liutkus 2010	54	Placebo	43	54 (9)	92.6 (18)	8.3 (0.9)	6.4 (4.6)	26 (26)
LixiLan-O	233	Lixisenatide 20 µg	43.2	58.7 (8.7)	90.8 (16.3)	8.1 (0.7)	8.9 (6.3)	30 (38)
LixiLan-O	466	Glargine	49.3	58.3 (9.4)	89.8 (16.3)	8.1 (0.7)	8.7 (5.6)	30 (38)
Nauck 2007b	253	Exenatide 10 µg BID	47	59 (9)	85.5 (15.7)	8.6 (1)	9.8 (6.3)	52 (52)
Nauck 2007b	248	BIAsp30 BID	51	58 (9)	83.4 (15.6)	8.6 (1.1)	10 (6.2)	52 (52)
Nauck 2016	202	Liraglutide 1.8 mg	35	56.3 (10.6)	101.9 (23.3)	8.4 (0.7)	6.5 (5.3)	29 (29)
Nauck 2016	202	Lixisenatide 20 µg	45	56.1 (10)	100.6 (19.9)	8.4 (0.8)	6.3 (5)	29 (29)
PIONEER 2	411	Semaglutide 14.0 mg QD	49.9	57 (10)	91.9 (20.5)	8.1 (0.9)	7.2 (5.8)	57 (59)
PIONEER 2	410	Empagliflozin 25 mg	49	58 (10)	91.3 (20.1)	8.1 (0.9)	7.7 (6.3)	57 (59)
PIONEER 3	466	Semaglutide 3.0 mg QD	45.5	58 (10)	91.6 (22)	8.3 (1)	8.4 (6.1)	83 (85)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
PIONEER 3	465	Semaglutide 14.0 mg QD	46.9	57 (10)	91.2 (21.7)	8.3 (0.9)	8.7 (6.1)	83 (85)
PIONEER 3	465	Semaglutide 7.0 mg QD	47.3	58 (10)	91.3 (20.8)	8.4 (1)	8.3 (5.8)	83 (85)
PIONEER 3	467	Sitagliptin 100 mg	49	58 (10)	90.9 (21)	8.3 (0.9)	8.8 (6)	83 (85)
PIONEER 4	285	Semaglutide 14.0 mg QD	48	56 (10)	92.9 (20.6)	8 (0.7)	7.8 (5.7)	52 (59)
PIONEER 4	284	Liraglutide 1.8 mg	48	56 (10)	95.5 (21.9)	8 (0.7)	7.3 (5.3)	52 (59)
PIONEER 4	142	Placebo	48	57 (10)	93.2 (20)	7.9 (0.7)	7.8 (5.5)	52 (59)
SURPASS-2	470	Tirzepatide 5 mg QW	56.4	56.3 (10)	92.5 (21.76)	8.32 (1.08)	9.1 (7.16)	40 (44)
SURPASS-2	469	Tirzepatide 10 mg QW	49.3	57.2 (10.5)	94.8 (22.71)	8.3 (1.02)	8.4 (5.9)	40 (44)
SURPASS-2	470	Tirzepatide 15 mg QW	54.5	55.9 (10.4)	93.8 (21.83)	8.26 (1)	8.7 (6.85)	40 (44)
SURPASS-2	469	Semaglutide 1.0 mg QW	52	56.9 (10.8)	93.7 (21.12)	8.25 (1.01)	8.3 (5.8)	40 (44)
SURPASS-3	358	Tirzepatide 5 mg QW	44	57.2 (10.1)	94.4 (18.9)	8.17 (0.89)	8.5 (5.8)	52 (59)
SURPASS-3	360	Tirzepatide 10 mg QW	46	57.4 (9.7)	93.8 (19.8)	8.18 (0.89)	8.4 (6.6)	52 (59)
SURPASS-3	359	Tirzepatide 15 mg QW	46	57.5 (10.2)	94.9 (21)	8.21 (0.94)	8.5 (6.5)	52 (59)
SURPASS-3	360	Degludec	41	57.5 (10.1)	94 (20.6)	8.12 (0.94)	8.1 (6)	52 (59)
SUSTAIN 2	409	Semaglutide 0.5 mg QW	49	54.8 (10.2)	89.9 (20.4)	8 (0.9)	6.4 (4.7)	61 (61)
SUSTAIN 2	409	Semaglutide 1.0 mg QW	50	56 (9.4)	89.2 (20.7)	8 (0.9)	6.7 (5.6)	61 (61)
SUSTAIN 2	407	Sitagliptin 100 mg	49	54.6 (10.4)	89.3 (19.7)	8.2 (0.9)	6.6 (5.1)	61 (61)
SUSTAIN 3	406	Semaglutide 1.0 mg QW	45.8	56.4 (N/A)	96.2 (N/A)	8.4 (N/A)	9 (N/A)	56 (56)
SUSTAIN 3	407	Exenatide 2 mg QW	43.7	56.7 (N/A)	95.4 (N/A)	8.3 (N/A)	9.4 (N/A)	56 (56)
SUSTAIN 4	362	Semaglutide 0.5 mg QW	46	56.5 (10.3)	93.7 (21.4)	8.1 (0.8)	7.8 (5.1)	35 (37)
SUSTAIN 4	360	Semaglutide 1.0 mg QW	49	56.7 (10.4)	94 (22.5)	8.3 (0.9)	9.3 (7.2)	35 (37)

Trial	Number of patients	Randomised treatment	Females, %	Mean age, years (SD)	Mean baseline weight, kg (SD)	Mean baseline HbA1c, % (SD)	Mean baseline duration of diabetes, years (SD)	Treatment duration, weeks (total trial duration)
SUSTAIN 4	360	Glargine	46	56.2 (10.6)	92.6 (21.5)	8.1 (0.9)	8.6 (6.3)	35 (37)
SUSTAIN 7	301	Semaglutide 0.5 mg QW	44	56 (10.9)	96.4 (24.4)	8.3 (0.9)	7.7 (5.9)	40 (40)
SUSTAIN 7	300	Dulaglutide 0.75 mg	46	55 (10.4)	95.6 (23)	8.2 (0.9)	7 (5.5)	40 (40)
SUSTAIN 7	300	Semaglutide 1.0 mg QW	46	55 (10.6)	95.5 (20.9)	8.2 (0.9)	7.3 (5.7)	40 (40)
SUSTAIN 7	300	Dulaglutide 1.50 mg	43	56 (10.6)	93.4 (21.8)	8.2 (0.9)	7.6 (5.6)	40 (40)
SUSTAIN 8	394	Semaglutide 1.0 mg QW	43	55.7 (11.1)	90.6 (22.6)	8.3 (1)	7.5 (5.9)	52 (59)
SUSTAIN 8	394	Canagliflozin 300 mg	49	57.5 (10.7)	89.8 (22.6)	8.2 (1)	7.2 (5.4)	52 (59)
SUSTAIN 9	151	Semaglutide 1.0 mg QW	41.1	57.5 (8.9)	89.6 (19.5)	8 (0.8)	9.8 (6.3)	35 (35)
SUSTAIN 9	151	Placebo	42.4	56.6 (10.1)	93.8 (22.3)	8.1 (0.8)	9.6 (5.9)	35 (35)
SUSTAIN-FORTE	481	Semaglutide 1.0 mg QW	41	58.2 (9.9)	98.6 (24.4)	8.8 (0.6)	9.8 (6.2)	40 (49)
SUSTAIN-FORTE	480	Semaglutide 2.0 mg QW	42	57.9 (10)	100.1 (22.6)	8.9 (0.6)	9.2 (6.2)	40 (49)
Van Gaal 2014	158	Lixisenatide 20 µg	65.2	42.7 (5.2)	98.5 (23.5)	8.16 (0.89)	4.4 (3.9)	24 (24)
Van Gaal 2014	161	Sitagliptin 100 mg	54.7	43.4 (4.7)	100.6 (23.8)	8.1 (1)	4.4 (3.6)	24 (24)

A24 NMA Baseline Characteristics by Treatment Arm[4]
BID = twice a day; DPP-4 = dipeptidyl-peptidase 4; HbA1c = glycated haemoglobin; LIRA = liraglutide; N/A = not applicable; SD = standard deviation; QD = once a day; QW = once weekly

EAG comment: The EAG pointed out in the clarification letter that the scope of the NMA regarding both population and comparators is not aligned to the final NICE scope nor the decision problem of the CS. Specifically, the population of the NMA is inconsistent with the decision problem i.e., inadequately controlled with three or more antidiabetic agents. This is despite the comparator trials being of GLP-1 RAs, which are recommended in the NG28 at the line of therapy specified in the decision problem. The company was asked to discuss this inconsistency to which they replied:

"*As with the design of the trial programmes for tirzepatide and for comparators, the NMA was conducted on a global level to meet the needs of multiple countries, so does not exactly match the population in the decision problem that is specific to the relatively less common position to which the NHS restricts GLP-1 RAs. However, the NMA population is aligned with the SURPASS-2 and -3 trials*

as well as other GLP-1 RA comparator trials and considered generalisable to UK clinical practice as described above in the answer to question A5d and A5e.

Results adjusted for the number of background OADs (in the meta-regression analysis) were similar to unadjusted results for all tirzepatide doses compared to all GLP-1 RAs at the same recommended maintenance dose step for HbA1c change from baseline and weight change from baseline. In addition, as described in Appendix D.9.1.3, sensitivity analysis that included studies with patients on a background therapy of three OADs (e.g., SURPASS-4) did not significantly impact the NMA results. This supports the contention that results of the NMA are generalisable to patients in the target population." (page 43[4])

The alignment to the SURPASS-2 and -3 trials does improve comparability of the network, but this does therefore imply that the whole of the NMA is inconsistent with the decision problem. If it was the case that line of therapy or treatment experience/concomitant therapy were not treatment effect modifiers, then this might not be so much of a problem. The EAG requested that the company would address all potential treatment effect modifiers, including at least: concomitant therapy, HbA1c, comorbidities e.g., CVD/CV high risk, obesity, non-alcoholic fatty liver disease, sex, age, weight, BMI, and duration of diabetes in a feasibility assessment.[23] The company maintains that the most important treatment effect modifiers were identified during the not-NICE-specific SLR and examined in a meta-regression analysis (discussed in Section 3.4). These analyses included only three characteristics: number of OADs, change from baseline in HbA1c (%) and weight (kg) (presented in Section 3.4.2.3.1). In response to the clarification letter, the company did state: "*Background therapies were reviewed and assessed to ensure that studies included were comparable to the SURPASS 2 and 3 trials and to EU guidelines. Background therapies were therefore not considered to be a key treatment-effect modifier. Information about background therapies are provided in the accompanying NMA input data file for this response."* (page 32[4]) However, no evidence was presented on how background therapies were reviewed and assessed. In response to the clarification letter, the company has provided an Excel file reporting background OADs for all trials/arms considered in all of the NMAs, but it is unclear which were part of each of the main NMAs and which the sensitivity analyses. However, it does appear that the number but also the mix of OADs treatments varied quite a lot within and between studies. Also, as with the SURPASS trials, the very fact that they the NMA is limited to patients with relatively little treatment experience or few background therapies means that assessing any treatment effect modification is seriously limited. Therefore, the presented NMA appears to have very limited value as it is not informative on the population of the decision problem. This is therefore key issue, a response to which might be to adjust the decision problem population, with a potential addition of non-GLP-1 RAs e.g., a SGLT-2i as comparators, as discussed in Section 2.

Regarding the other baseline characteristics, the company stated:

- ***"Comorbidities**: As shown in Figure 25, Document B of the CS, comorbidities were not systematically reported in most studies. In addition, the comorbidities listed below were not considered as key treatment effect modifiers for the following additional reasons:"* (page 32[4])

The fact that comorbidities were not systematically reported (52 of the studies did not include comorbidities) does not imply that these characteristics are not potential effect modifiers or that the studies in the network are comparable.

- ***"CVD and high CV risk**: CVD and high CV risk are specifically reported in CVOTs, given the specific requirements for CVOT trial design, population, objectives, and glycaemic control. As*

such, CVD and high CV risk were not considered to be a key treatment-effect modifier to be explored in this NMA which did not include CVOTs in the network." (page 32[4])

It is not clear from this response if indeed CVD/CV high risk were reported by some of the studies included in the network or not.

- ***"HbA1c**: Most studies reported consistent mean baseline HbA1c values between 8–8.5%, as shown in Figure 32 of the CS. Nevertheless, HbA1c was considered as a potential factor for consideration within the meta-regression to adjust for heterogeneity, as described in Appendix D.1.1.8 of the CS."* (page 32[4])

This statement is not supported by the presented data. The range of mean baseline HbA1c values varied from 7.4% to 10.3%. Out of the 136 arms presented, 11 arms reported values below 8% and 20% above 8.5%. (A24 NMA Baseline Characteristics by Treatment Arm[4]) The inclusion of HbA1c in meta-regression analysis is discussed in Section 3.4.

- ***"Obesity**: Across all studies, patients were consistently either overweight or obese, as shown in Figure 31 of the CS. As such, all studies included in the NMA included patients with a BMI between 30–35 kg/m². Nevertheless, similar to HbA1c, weight and/or BMI were considered as a potential factor for consideration in the meta-regression."* (page 32[4])

Figure 31 (old version: Figure 28 in the latest) of the CS refers to the proportion of female patients. The inclusion of body weight in meta-regression analysis is discussed in Section 3.4.

- ***"Baseline diabetes duration**: Baseline diabetes duration was generally similar between studies with most reporting a baseline mean between 6 and 8 years, as shown in Figure 33 of the CS; a very small number of outlier studies had lower durations."* (page 32[4])

This statement is not supported by the presented data. Baseline diabetes duration varied from 0.6 - 10.1 years. Fourteen arms reported duration less than 6 years while 55 reported more than 8 years. Eight arms did not report this information. (A24 NMA Baseline Characteristics by Treatment Arm[4])

- ***"Patients age**: As shown in Figure 29 of the CS, mean age at baseline was between 50 and 60 years in almost all studies. For the studies not excluded for other reasons, age was not considered to be a key treatment effect modifier, given feasibility and clinical judgement."* (page 32[4])

It is not clear what the company means by stating *"age was not considered to be a key treatment effect modifier, given feasibility and clinical judgement"*.

- ***"Sex**: The proportion female in each study did exhibit some variation, as shown in Figure 28 of the CS, but this was not considered as a reason to exclude studies; in principal this parameter could be added in the meta-regressions if it were considered a treatment effect modifier."* (page 32[4])

No justification is provided on why sex was not considered a treatment effect modifier.

- ***"NASH**: 6 included studies were among T2D patients with NASH – all other baseline characteristics from these studies were considered comparable to other studies, and these also had comparable study designs. Therefore, it was considered that these studies could be included in the analysis and that NASH was not a treatment effect modifier.*" (page 32[4])

Non-alcoholic steatohepatitis (NASH) data are not offered in the CS therefore this characteristic cannot be assessed by the EAG.

In the CS the company stated that "*SURPASS-J-Mono and SURPASS-J-Combo were conducted in a Japanese population and are therefore not considered generalisable to the UK population; they are not presented as part of the clinical evidence in this appraisal. Data from SURPASS-J-Mono and SURPASS-J-Combo are included in the safety analysis in Section B.2.9*." (page 30 of the CS[3]). When questioned by the EAG on whether similar, generalisability-based, exclusion criteria were applied in the SLR and when selecting studies of comparators for inclusion in the NMA the company presented a different version of why these studies were excluded from the NMA stating that "*There were no exclusions in the SLR based on generalisability criteria. SURPASS-J-Mono and SURPASS-J-Combo were both included in the SLR. However, for the NMA, SURPASS-J-Mono and SURPASS-J-Combo did not meet the inclusion criteria of oral background treatment (SURPASS-J-Mono) or include comparator of interest (SURPASS-J-Combo) and were therefore excluded from the NMA*". [4]

Initially, the company had not included the characteristics of race and ethnicity in the CS nor the NMA feasibility assessment. Following the request of the EAG in the clarification letter, the company stated that "*Trials define and collect racial and ethnicity baseline characteristics in different ways, depending on both trial design and the requirements of the different countries and locations where the studies were undertaken. As such, a formal feasibility assessment relating to these protected characteristics has not been undertaken, but the baseline data as reported by each trial according to its own definitions has been reported above; it is apparent that there is some degree of heterogeneity in the trial populations which is most likely reflective of the location of the centres recruiting for each trial.*" (page 38[4]) Therefore, these potential effect modifiers have not been assessed in the CS. In addition, the company has not provided data on all relevant characteristics per study nor per arm. Therefore, the EAG cannot assess how many studies reported these data and whether further analysis was feasible.

In conclusion, the standard methodology of an NMA is to first execute a SLR tailored to the appropriate PICOs and then execute a feasibility assessment to assess suitability of synthesis of the identified studies, according to comparability, especially of baseline characteristics that might be treatment effect modifiers. In contrast, the methodology presented in the CS has serious limitations. Apart from the mismatch between the decision problem population and those of the trials in the NMA in terms of OAD treatment experience/background therapy, there was no systematic comparison of trials, one to another, according to these characteristics or any other baseline characteristics. It also seems that, despite some considerable variation in some baseline characteristics, the effect of exclusion of trials on the basis of that variation was not tested (see Section 3.4.2.3). Therefore, the lack of a NMA feasibility assessment/assessment of comparability is a key issue.

3.4 Critique of the indirect comparison and/or multiple treatment comparison

3.4.1 NMA methods

The company states that the NMA was executed in R. The library rjags (Just Another Gibbs Sampler) was used for the Bayesian simulations. A two-stage analytical approach was used. First, a frequentist meta-analysis was conducted to assess heterogeneity and become familiar with the data and second, the NMA was conducted using Bayesian mixed treatment comparisons.

The Bayesian NMA models were computed using a Markov Chain Monte Carlo (MCMC) simulation method with 20,000-100,000 burn-in simulations and a sample of 100,000-1,500,000 (thin: 5) depending on the endpoint of the analysis. Deviance information criterion (DIC) value was used to examine consistency and choose between fixed-effects and random-effects models. Goodness of fit statistics are presented in Table 3.26. Random-effects models were used for all endpoints except for the proportion of patients reaching weight loss ≥5% and ≥10%. The company states that random effects

models were chosen over fixed-effects when DIC value were lower or similar. Goodness of fit statistics for the two endpoints where fixed-effects were used have not been provided in the CS. Continuous outcomes (change from baseline in HbA1c, body weight, BMI, etc.) are presented in terms of standardised median differences (SMD) and 95% credible interval (CrI), while binary endpoints (proportion of patients experiencing nausea, proportion of patients with at least one episode of hypoglycaemia with blood glucose (BG) <54 mg/dl (3.0 mmol/l) or severe hypoglycaemia, etc.) are presented in terms of odds ratio (OR) and 95% CrI for each tirzepatide dose versus placebo and comparators.

Meta-regression was used to explore heterogeneity, adjusting for OAD. Further meta-regression, adjusting for analysis time window and baseline covariates, such as HbA1c and weight, resulted in convergence issues. I^2 statistic and Cochrane Q values were used to assess statistical heterogeneity. Substantial heterogeneity was defined as I^2 >60%, no concerns were defined as I^2 < 40%, or I^2 between 40%–60% and Cochrane Q test p-value >0.1.

According to the CS, because tirzepatide and GLP-1 RAs exhibit dose-response relationships, for the interpretation of the NMA, comparisons were made within each recommended maintenance dose step rather than between them, according to Table 3.27.

Table 3.27: Goodness of fit statistics for all endpoints

Endpoint	Data points (N)	Residual deviance		DIC	
		Fixed-effects	Random-effects	Fixed-effects	Random-effects
HbA1c (%) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weight (kg) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BMI (kg/m²) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
LDL (mmol/l) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HLD (mmol/l) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
eGFR (ml/min/1.73 m²) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SBP (mmHg) change from baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Proportion of patients experiencing nausea (any grade permitted)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 53, CS.[3]
BMI = body mass index; CS = company submission; DIC = deviance information criterion; eGFR = estimated Glomerular Filtration Rate; HbA1c = glycated haemoglobin; HLD = high density lipoprotein; LDL = low density lipoprotein

Table 3.28: Overview of comparators and doses

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide 5 mg	Dulaglutide 1.5 mg Semaglutide 0.5 mg Oral semaglutide 7 mg Liraglutide 1.2 mg
Tirzepatide 10 mg	Dulaglutide 3.0 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
Tirzepatide 15 mg	Dulaglutide 4.5 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
Table 46, CS.[3] CS = company submission	

A key aspect of the NMA is the choice of the analysis time window. The duration of the dose escalation in the SURPASS trials is much longer than the durations of the corresponding comparators: 0-20 weeks compared to 0-12 weeks. In addition, within the SURPASS trials the time to reach the three arms/doses vary:

- Tirzepatide dose 5 mg: week 5
- Tirzepatide dose 10 mg: week 13
- Tirzepatide dose 15 mg: week 21

The highest dose of 15 mg is therefore administered for only the remaining 20 weeks of the 40-weeks SURPASS trials. Most comparator trials had a duration of 22-30 weeks, while all the comparator trials reported one outcome of interest in the time window of 20-28 weeks. The company made the decision to analyse the comparator data at 26 ± 4 (22–30) weeks while using the data from week 40 for all the tirzepatide doses. For SURPASS-4 data from week 42 was used instead since no visit was conducted at week 40. According to the company the time window of 26 ± 4 (22–30) weeks provides a balanced overall approach. The sensitivity analysis conducted by the company tested other time window scenarios (see Section 3.4.2.3). Regarding the AEs analysis, outcomes outside the 26 ± 4 weeks window were allowed.

Fifty-three studies were found eligible by the company to be included in the NMA. In eight of these studies a GLP-1 treatment was compared to a treatment that did not connect to the rest of the network and therefore these were excluded from the NMA. The company has excluded these treatments from all endpoints in the network as not treatments of interest. A full list of these treatments was not provided in the CS. In addition, pioglitazone and semaglutide 3.0 mg QD were also excluded as not treatments of interest. The exclusion of these treatments (studies/arms) resulted in a maximum of 45 studies and 23 treatments being included in the NMA, depending on the endpoint examined. Nevertheless, other treatments that were not of interest such as insulin glargine, insulin degludec, sitagliptin and glimepiride were included in the network to provide nodes to connect the network.

The company stated that substantial heterogeneity was identified in at least one of the relative comparisons for each continuous outcome. Comparisons that contributed to heterogeneity are presented

in Table 3.29. According to the company the endpoints not included in this table did not demonstrate concerns for heterogeneity. SURPASS-2 and/or SURPASS-3 contributed to the heterogeneity for change from baseline in heart rate (HR), LDL and total cholesterol. Heart rate and total cholesterol were not included in the table provided by the company. For binary outcomes, the company stated that no heterogeneity concerns were found for the majority of the endpoints except for nausea. The comparisons contributing to this heterogeneity were between tirzepatide 5 mg and 10 mg doses (based on SURPASS-2 and 3) and between liraglutide 1.8 mg and sitagliptin 100 mg. No I^2 statistic or Cochrane Q values were provided in the CS for either continuous or binary outcomes.

Table 3.29: Summary of heterogeneity in continuous variables

Outcome	Comparison contributing to heterogeneity
HbA1c	Placebo versus exenatide 10 mcg BID Placebo versus exenatide 5 mcg BID Placebo versus dulaglutide 0.75 mg Placebo versus dulaglutide 1.5 mg Glargine versus liraglutide 1.8 mg
Weight	Placebo versus exenatide 10 mcg BID Placebo versus dulaglutide 1.5 mg Liraglutide 1.8 mg versus glargine Dulaglutide 0.75 mg versus dulaglutide 1.5 mg
BMI	Placebo versus dulaglutide 1.5 mg Placebo versus liraglutide 1.8 mg Dulaglutide 1.5 mg versus dulaglutide 0.75 mg
SBP	Placebo versus exenatide 10 mcg BID Placebo versus liraglutide 1.2 mg
HDL	Placebo versus dulaglutide 0.75 mg
LDL	TZP 5 mg versus TZP 10 mg TZP 5 mg versus TZP 15 mg

Table 54, CS.[3]
BID = twice a day; BMI = body mass index; CS = company submission; HbA1c = glycated haemoglobin; HDL = high-density lipoprotein; HR = heart rate; LDL = low-density lipoprotein; TZP = tirzepatide; SBP = systolic blood pressure

3.4.2 NMA results

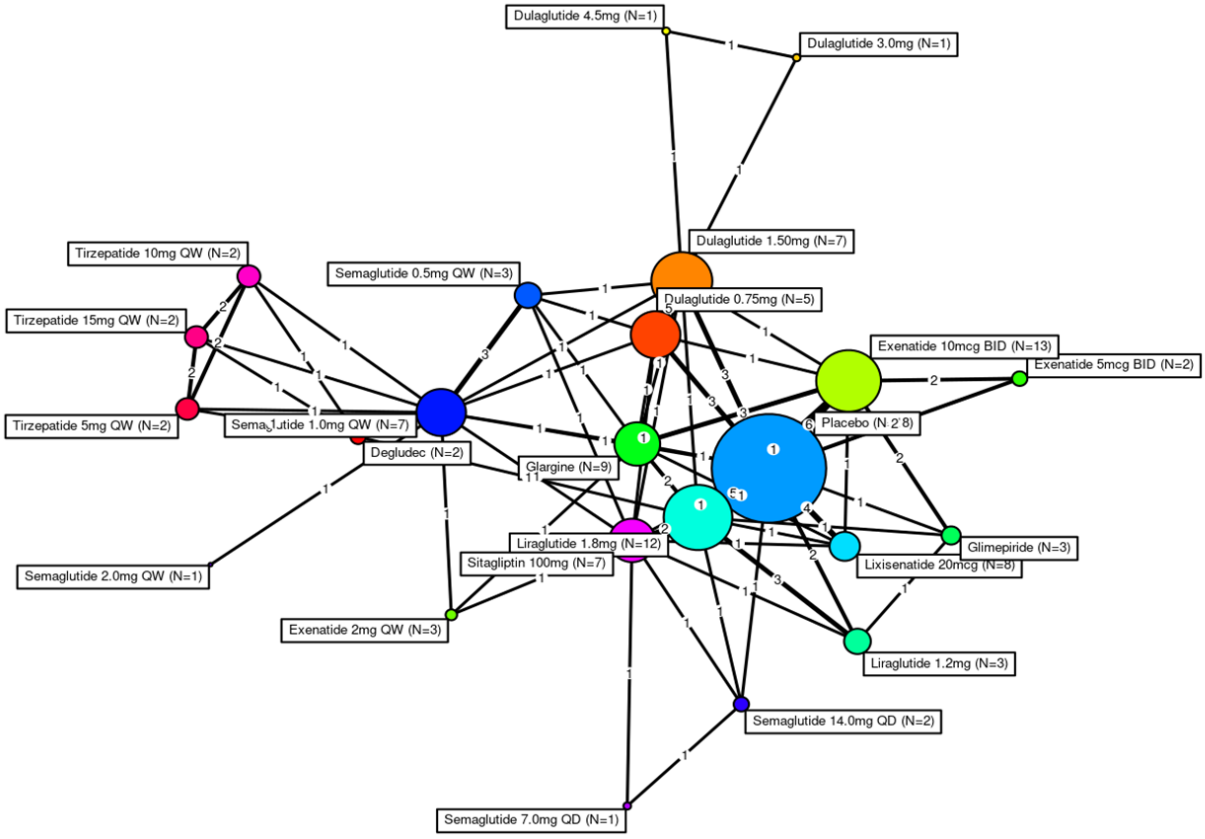
The following Sections present the base-case NMA results by outcome.

3.4.2.1 Main analysis results: Efficacy

3.4.2.1.1 HbA1c (%) change from baseline

The network diagram for HbA1c (%) change from baseline at 40 weeks (tirzepatide) and 26 ± 4 weeks (comparators) using a random-effects model is illustrated in Figure 3.4. Forty-five studies including 23 treatments were included in this network. The pairwise results are presented in Table 3.30. Tirzepatide 5 mg, 10 mg and 15 mg showed significant greater reductions in HbA1c (%) change from baseline compared with placebo and all GLP-1 RAs at the lowest, intermediate, and highest recommended maintenance dose, respectively.

Figure 3.4: Main analysis network for HbA1c (%) change from baseline



Based on Figure 40 of the CS.[3]
BID: twice a day; CS = company submission; Hb1Ac = glycated haemoglobin; QD: once a day; QW: once weekly

Table 3.30: Pairwise results (median difference [95% CrI]) for HbA1c (%) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 7.0 mg QD	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]

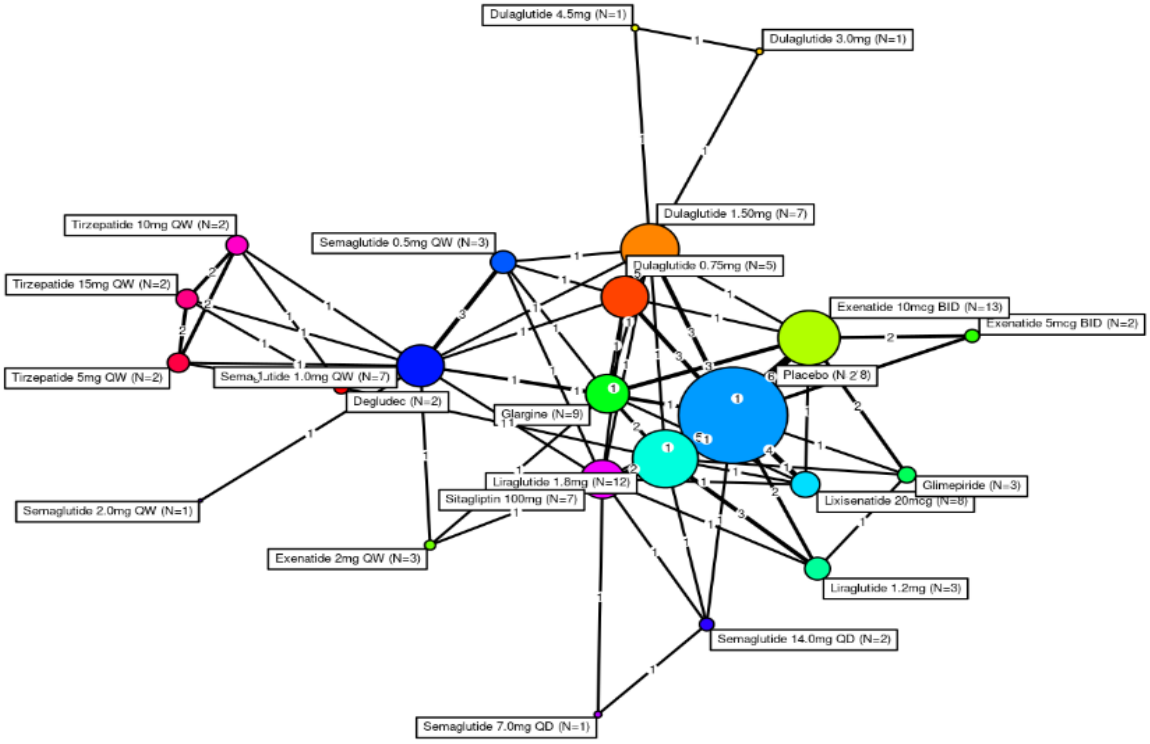
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Table 55, CS.[3]
BID = twice a day; CrI = credible interval; CS = company submission; HbA1c = glycated haemoglobin; QD = once a day; QW = once weekly; TZP = tirzepatide
NOTES: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.1.2 Weight (kg) change from baseline

All 45 studies and 23 treatments (nodes) were included in this analysis using a random-effects model. The network is illustrated in Figure 3.5 and the results presented in Table 3.31. For all doses of tirzepatide, there was a significantly favour of tirzepatide versus all comparators with a dose-response relationship whereby the higher dose was more effective.

Figure 3.5: Main analysis network for weight (kg) change from baseline



Based on Figure 44 of the CS.[3]
BID = twice a day; CS = company submission; QD = once a day; QW = once weekly

Table 3.31: Pairwise results (median difference [95% CrI]) for weight (kg) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-

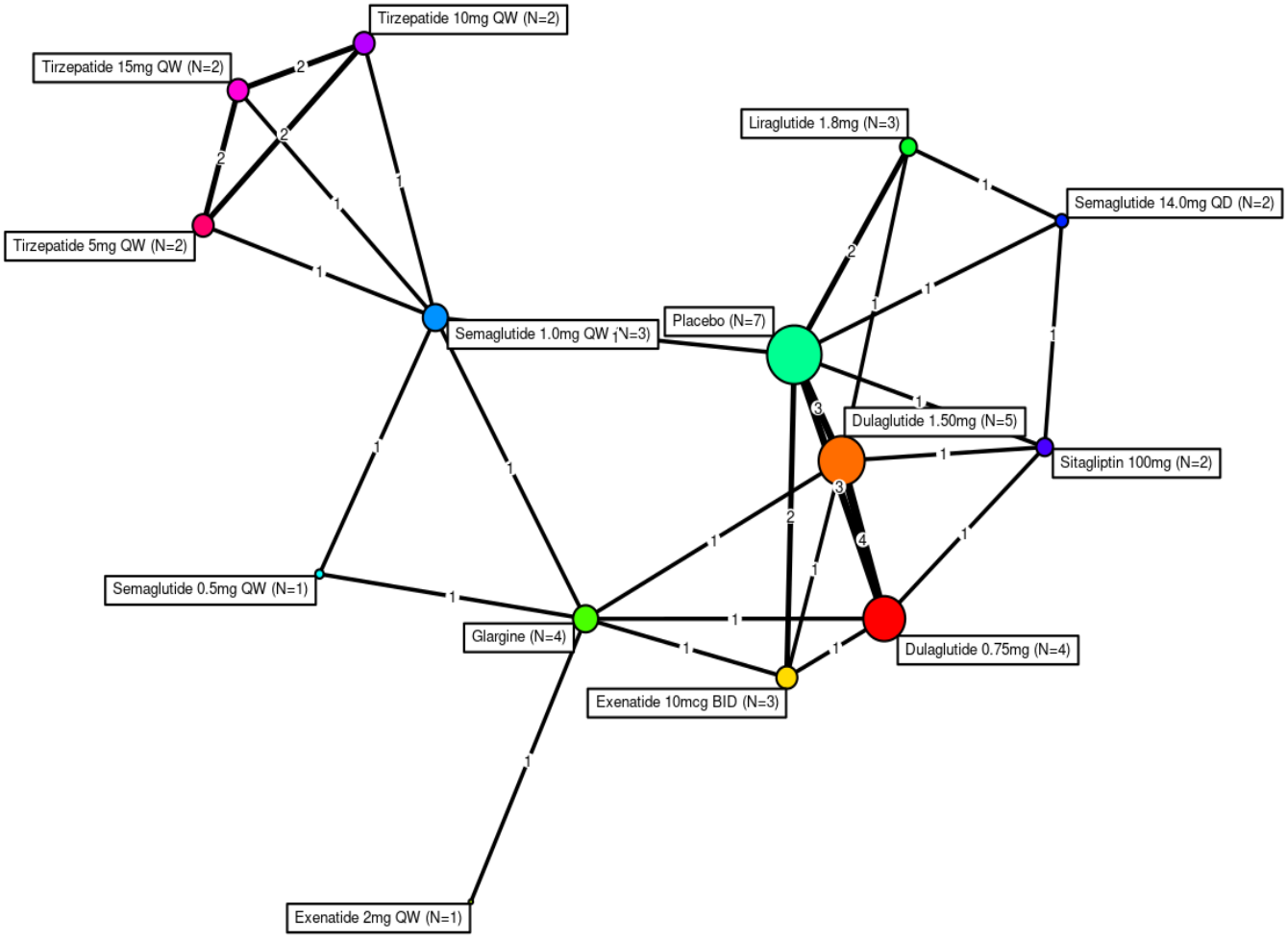
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 7.0 mg QD	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Table 56, CS.[3]
BID = twice a day; CrI = credible interval; CS = company submission; QD = once a day; QW = once weekly; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.1.3 Body mass index (kg/m²) change from baseline

In total, 15 studies and 14 treatments (nodes) were included for this endpoint using a random-effects model. The network is illustrated in Figure 3.6 and the results presented in Table 3.32. For all doses of tirzepatide, there was a significantly favour of tirzepatide versus all comparators with a dose-response relationship whereby the higher dose was more effective.

Figure 3.6: Main analysis network for BMI (kg/m²) change from baseline



Based on Figure 48 of the CS.[3]

BID = twice a day; CS = company submission; QD = once a day; QW = once weekly

Table 3.32: Pairwise results (median difference [95% CrI]) table for BMI (kg/m²) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]

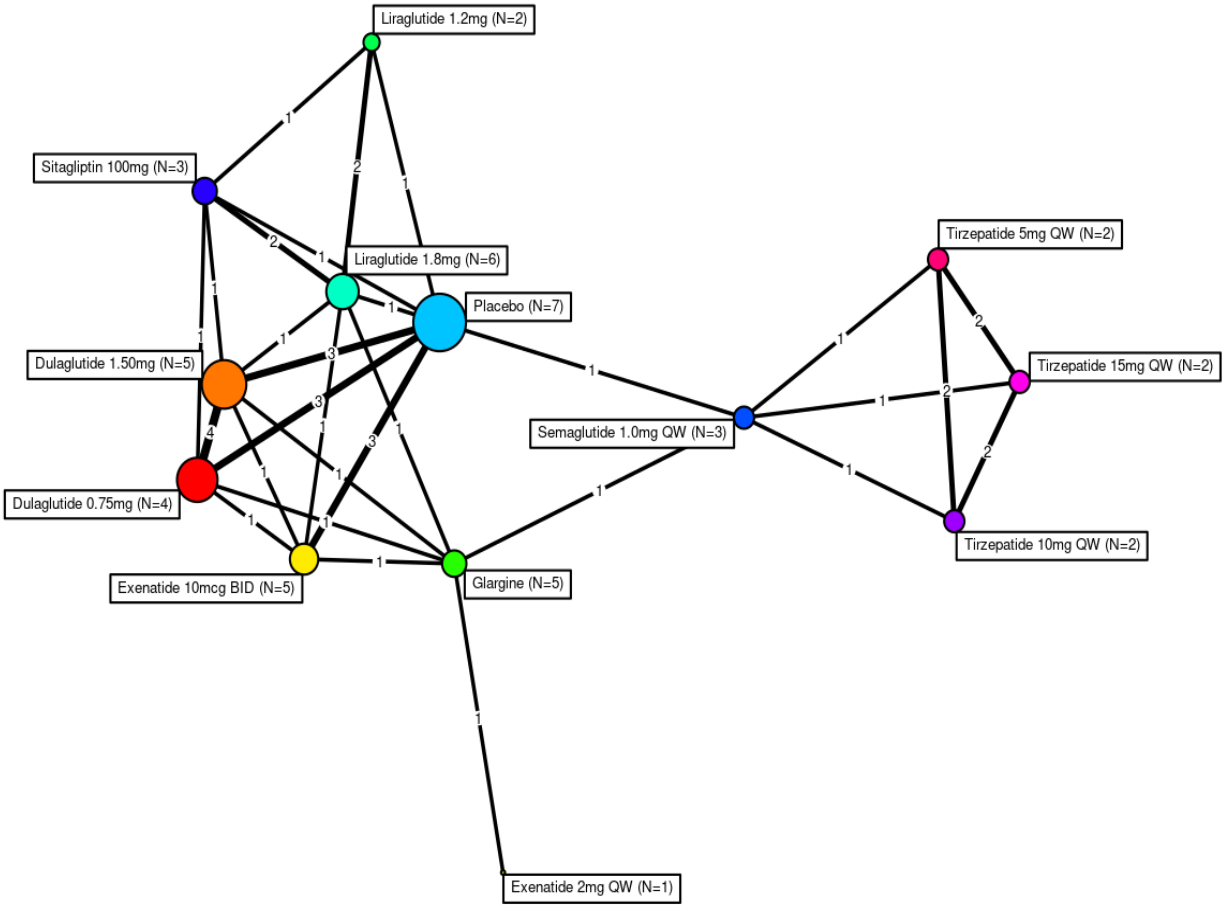
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Table 57, CS.[3]
BID = twice a day; BMI = body mass index; CrI = credible interval; CS = company submission; QD = once a day; QW = once weekly; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.1.4 Low-density lipoprotein (mmol/l) change from baseline

In total, 18 studies and 13 treatments (nodes) were included in this endpoint using a random-effects model. The network is illustrated in Figure 3.7 and the results presented in Table 3.33. Tirzepatide 5 mg showed no significant difference in LDL (mmol/l) compared to placebo and GLP-1 RAs. Tirzepatide 10 mg and 15 mg resulted in significant reductions in LDL (mmol/l) compared to placebo, but no significant differences compared to GLP-1 RAs irrespective of dose.

Figure 3.7: Main analysis network for LDL (mmol/l) change from baseline



Based on Figure 52 of the CS.[3]
BID = twice a day; CS = company submission; LDL = low density lipoprotein; QD = once a day; QW = once weekly

Table 3.33: Pairwise results (median difference [95% CrI]) table for LDL (mmol/l) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

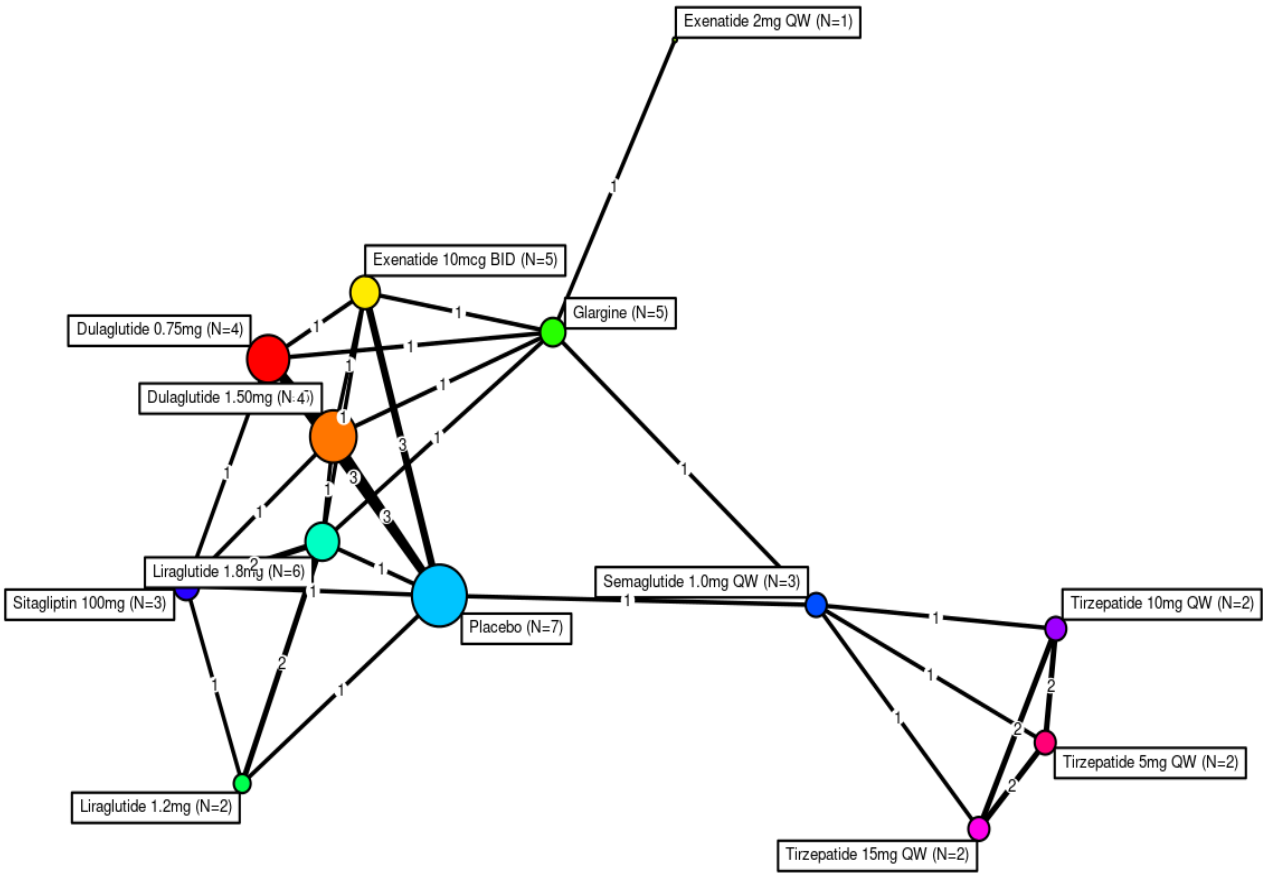
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Source: Table 58, CS.[3]
BID = twice a day; CrI = credible interval; CS = company submission; LDL = low density lipoprotein; QW = once weekly; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); cells highlighted in red show comparisons which favour other TZP doses or active treatment

3.4.2.1.5 High-density lipoprotein (mmol/l) change from baseline

In total, 18 studies and 14 treatments (nodes) were included for this endpoint using a random-effects model. The network is illustrated in Figure 3.8 and the results presented in Table 3.34. Tirzepatide 5 mg showed no statistically significant difference in HDL (mmol/l) change from baseline compared with placebo and all GLP-1 RAs. Tirzepatide 10 mg and tirzepatide 15 mg showed statistically significant increase in HDL (mmol/l) from baseline compared with semaglutide 1.0 mg. No significant difference was observed when compared with placebo and other GLP-1 RAs at any dose step.

Figure 3.8: Main analysis network for HDL (mmol/l) change from baseline



Based on Figure 56 of the CS.[3]
BID = twice a day; CS = company submission; HDL = high density lipoprotein; QD = once a day; QW = once weekly

Table 3.34: Pairwise results (median difference [95% CrI]) table for HLD (mmol/l) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Source: Table 59, CS.[3]			

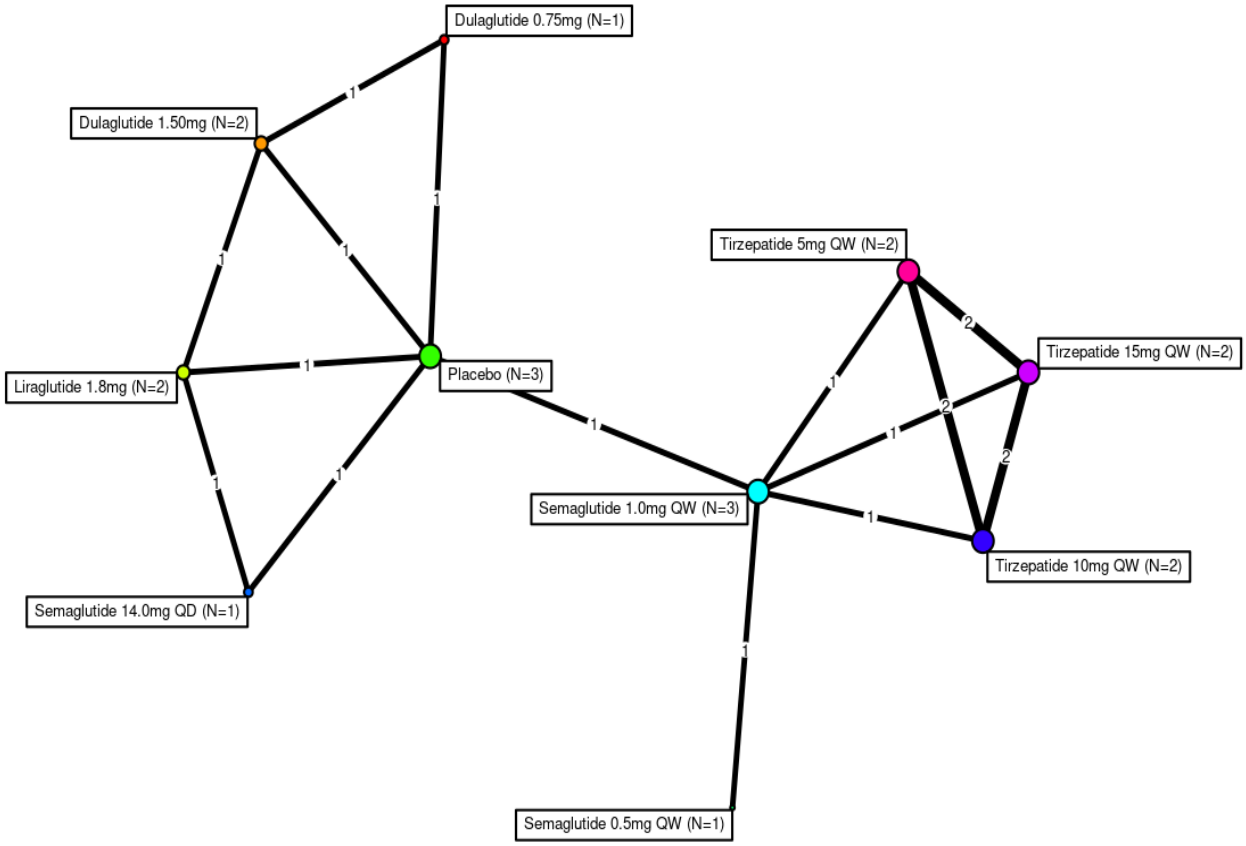
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
BID = twice a day; CrI = credible interval; CS = company submission; HDL = high density lipoprotein; QW = once weekly; TZP = tirzepatide Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); cells highlighted in red show comparisons which favour other TZP doses or active treatments			

3.4.2.1.6 eGFR (ml/min/1.73 m²) change from baseline

In total, seven studies and 10 treatments (nodes) were included for this endpoint using a random-effects model. The network is illustrated in Figure 3.9 and the results presented in Table 3.35. Tirzepatide (5 mg, 10 mg, and 15 mg) showed no statistically significant difference for change from baseline in eGFR (ml/min/1.73 m²) when compared with placebo and GLP-1 RAs at any maintenance dose. The company noted that there was high uncertainty within the network due to the limited availability of data and variability between the studies in the network in terms of the administered background treatment and the data of change from baseline for eGFR. They conclude that the results are not robust enough to draw safe conclusions.

Nevertheless, based on the similar data on the reduction of eGFR in patients treated with tirzepatide and injectable semaglutide 1 mg in SURPASS-2, the company decided to assume that tirzepatide and all comparators have an equivalent effect on renal function and used this assumption across the cost effectiveness modelling analysis.

Figure 3.9: Main analysis network for eGFR (ml/min/1.73 m²) change from baseline



Based on Figure 60 of the CS.[3]
BID = twice a day; CS = company submission; eGFR = estimated Glomerular Filtration Rate; QD = once a day; QW = once weekly

Table 3.35: Pairwise results (median difference [95% CrI]) table for eGFR (ml/min/1.73 m²) change from baseline, random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Semaglutide 14.0 mg QD	[redacted]	[redacted]	[redacted]

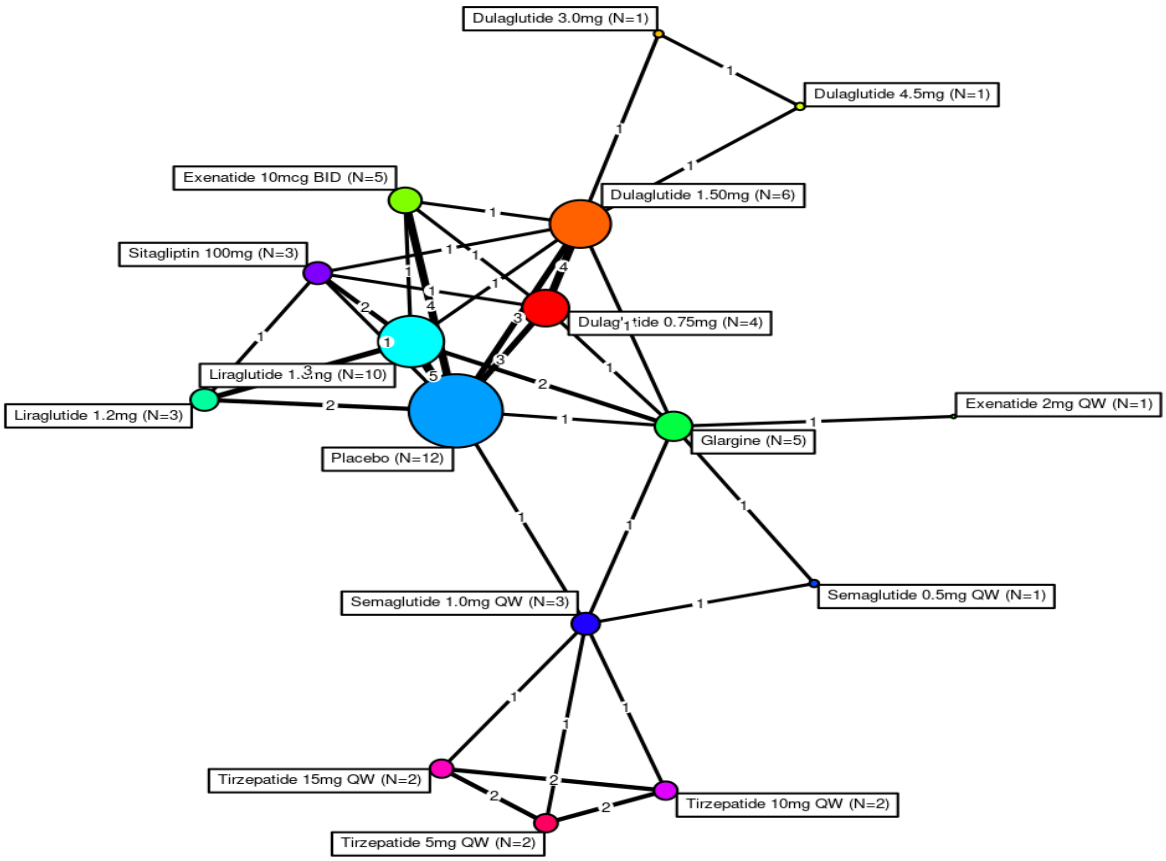
Table 60, CS.[3]
CS = company submission; eGFR = estimated Glomerular Filtration Rate; QD = once a day; QW = once weekly; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.2 Main analysis results: Safety

3.4.2.2.1 Systolic blood pressure (mmHg) change from baseline

In total, 23 studies and 16 treatments (nodes) were included for this endpoint using a random-effects model. The network is illustrated in Figure 3.10 and the results presented in Table 3.36. Tirzepatide 5 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo and dulaglutide 0.75 mg. No significant differences were observed with tirzepatide 5 mg compared with all other GLP-1 RAs at the lowest recommended maintenance dose. Tirzepatide 10 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo, liraglutide 1.8 mg, and exenatide 10.0 µg. No significant differences were observed with tirzepatide 10 mg compared with any other GLP-1 RAs at the intermediate recommended maintenance dose. Tirzepatide 15 mg showed significantly greater reductions in SBP (mmHg) from baseline compared with placebo and all GLP-1 RAs at the highest recommended maintenance dose, except dulaglutide 4.5 mg.

Figure 3.10: Main analysis network for change from baseline in SBP (mmHg)



Based on Figure 64 of the CS.[3]
BID = twice a day; CS = company submission; QD = once a day; QW = once weekly; SBP = systolic blood pressure

Table 3.36: Pairwise results (median difference [95% CrI]) for SBP (mmHg) change from baseline, fixed effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 3.0 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 4.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]

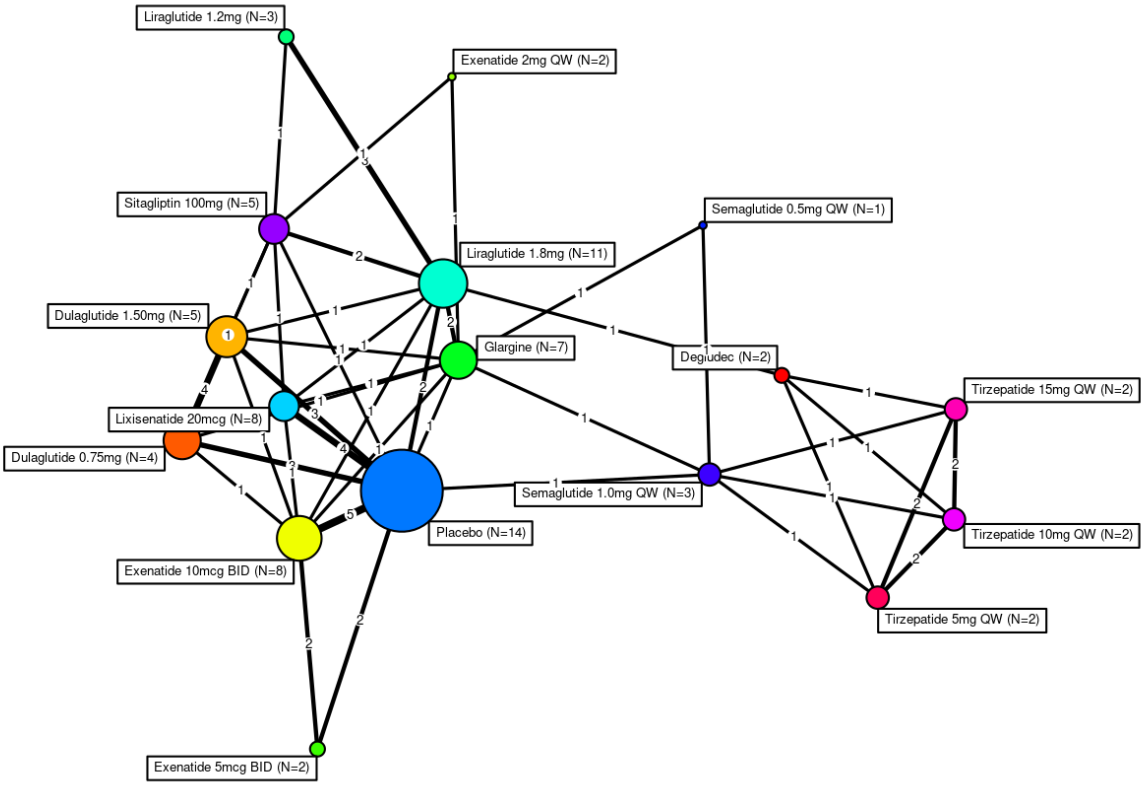
Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]

Table 61, CS.[3]
BID = twice a day; CrI = credible interval; CS = company submission; QW = once weekly; SBP = systolic blood pressure; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.2.2 Proportion of patients experiencing nausea (any grade permitted)

In total, 33 studies and 17 treatments (nodes) were included for this endpoint using a random-effects model. The network is illustrated in Figure 3.11 and the results presented in Table 3.37. A significantly higher proportion of patients receiving tirzepatide 5 mg experienced nausea (proportion of patients experiencing nausea AE) compared with placebo. Tirzepatide 5 mg showed no significant differences in nausea compared with all GLP-1 RAs at the lowest recommended maintenance dose. A significantly higher proportion of patients receiving tirzepatide 10 mg experienced nausea compared with placebo. Tirzepatide 10 mg showed no significant differences in nausea compared with all other GLP-1 RAs at the intermediate recommended maintenance dose. A significantly higher proportion of patients receiving tirzepatide 15 mg experienced nausea compared with placebo. Tirzepatide 15 mg showed no significant differences in nausea compared with all other GLP-1 RAs at the highest recommended maintenance dose.

Figure 3.11: Main analysis network for proportion of patients experiencing nausea (any grade permitted)



Based on Figure 68 of the CS.[3]
BID = twice a day; CS = company submission; QD = once a day; QW = once weekly

Table 3.37: Pairwise results (odds ratio [95% CrI]) table for proportion of patients experiencing nausea (any grade permitted), random effects model; TZP 5 mg, 10 mg or 15 mg (column) versus comparators (row)

Column versus row	TZP 5 mg	TZP 10 mg	TZP 15 mg
Placebo	[redacted]	[redacted]	[redacted]
Tirzepatide 5 mg QW	-	[redacted]	[redacted]
Tirzepatide 10 mg QW	[redacted]	-	[redacted]
Tirzepatide 15 mg QW	[redacted]	[redacted]	-
Semaglutide 0.5 mg QW	[redacted]	[redacted]	[redacted]
Semaglutide 1.0 mg QW	[redacted]	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 0.75 mg	[redacted]	[redacted]	[redacted]
Dulaglutide 1.5 mg	[redacted]	[redacted]	[redacted]
Exenatide 2 mg QW	[redacted]	[redacted]	[redacted]
Exenatide 5 mcg BID	[redacted]	[redacted]	[redacted]
Exenatide 10 mcg BID	[redacted]	[redacted]	[redacted]
Lixisenatide 20 mcg	[redacted]	[redacted]	[redacted]

Source: Table 62, CS.[3]
BID = twice a day; CrI = credible interval; CS = company submission; QW = once weekly; TZP = tirzepatide
Notes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments

3.4.2.3 Sensitivity analyses

The company reported a series of sensitivity analysis for change from baseline in HbA1c (%), weight (kg), and BMI (kg/m²) in Section D.8.1 of the appendices to the CS[12]. The sensitivity analyses included:

- Consideration of Asian population studies
- Inclusion of Phase 2 tirzepatide study
- Modification of network definition: studies including patients with unclear proportion of metformin as background therapy and studies including patients on a background therapy of three OADs (e.g., SURPASS-4) were included in this sensitivity analysis
- Exclusion of studies with insulin glargine as treatment arm
- Analysis time windows
- Different analyses timepoints and windows
- Model-based NMA for continuous outcomes
- A meta-regression adjusting for number of OADs for change from baseline in HbA1c (%) and weight (kg) was also conduced

According to Appendix D, the sensitivity analysis on the inclusion of Asian population studies included eight more studies: Araki 2015, GetGoal-M-Asia, Guo 2020, Inagaki 2012, Light-On, SUSTAIN China, Wang 2018 and Zang 2016.[12] However, they did not state whether their Asian studies, SURPASS-J-Mono and SURPASS-J-Combo, were included. There is also a mismatch to the studies reported in the file describing the studies included in this sensitivity analysis where three more studies are reported (Ji 2013; Kadowaki 2011; Li 2014) it is not clear which these studies as the full references are

not provided or whether they were included in the sensitivity analysis. For the majority of the outcomes the results were similar in this sensitivity analysis.

A sensitivity analysis was also executed including a phase 2 tirzepatide study which was excluded from the main analysis due to the difference in the dose escalation schemes (no dose escalation for 5 mg, 2-week dose escalation for 10 mg and 6-week dose escalation for 15 mg) and differences in the background therapies. It is not specified anywhere in the main CS or the appendices which Phase 2 tirzepatide study was used. In the in the excel file describing the studies included in the NMAs there a mention to the study by Frias et al. 2018.[24] The treatment effects for the majority of the comparisons regarding all three endpoints were similar to the main analysis.

A further sensitivity analysis was executed including both studies with an unclear proportion of patients receiving metformin, and trials including patients on three OADs were included. It is not clear what the rationale of this sensitivity analysis was. The results for this sensitivity analyses were similar to the main analysis.

The company states that in a further sensitivity analysis "*studies with insulin glargine were excluded from the NMA if there were only two treatment arms in the study. If more than two treatment arms were assessed in the study, the insulin glargine arm was excluded from the NMA*."(page 175[12]) . On the other hand, they also state that "*The additional studies **included** in this sensitivity analysis were AWARD-2, Bunck 2019/2010/2011, DURATION-3, EAGLE, Gurkan 2014, LEAD-5, LixiLan-O and SUSTAIN 4*." (page 175[12]) According to the company the sensitivity analysis was executed because "*In an early exploratory analysis, heterogeneity was identified in trials with insulin glargine for change from baseline in HbA1c and weight*." (page 175[12]). No further information was offered. The results of the sensitivity analyses were similar across endpoints for the majority of comparisons.

In the sensitivity analyses exploring the different time windows the company tested two scenarios. First by using an earlier timepoint for SURPASS trials at 24 weeks, closely aligned to the time point used for the comparators (26 ± 4 weeks). Second, by using a time window of 32 ± 8 weeks for all studies and the closest timepoint to 40 weeks per study where available to closely align with the 40-week time frame of the SURPASS trials. According to the company the results of both scenarios across endpoints were similar in the majority of the comparisons to the main analysis.

A model-based NMA was executed by fitting a random effects model using a time-course model for the two outcomes of HbA1c (%) and weight (kg) change from baseline. The purpose of the model was to include outcomes measured at multiple timepoints. Several candidate time-course models were tested for goodness of fit. The CS does not report which time-course model was found to fit better and was therefore used in the analysis. According to the company the results of the model-based NMA had similar median values but narrower CrIs across the two endpoints.

3.4.2.3.1 Meta-Regression

Random-effects meta-regression was used to adjust for heterogeneity. According to the company the covariates that could cause heterogeneity between the studies and treatment arms were: assessment timepoint (weeks), number of OADs (1 versus 2), baseline HbA1c and baseline weight. These covariates were identified during cross-functional workshops. No further information was offered on the nature of these workshops. An additional meta-regression was fit adjusting for baseline risk (i.e., placebo-response). The rationale for using all the above covariates was not provided.

The meta-regression heterogeneity exercise consisted of comparing the main (and unadjusted) analysis results to the meta-regression (and adjusted) analysis results based on their DIC values. Only the results

of the meta-regression analysis adjusting for number of OADs was presented in the CS (Appendix D.8.1.8[12]). The median effects for all comparisons were similar to the main analysis for both outcomes.

According to the company the meta-regression model fit adjusting for baseline risk was not significantly better to the main analysis. The data of this model were not reported. The company stated that the rest of the models adjusting for analysis time window and the baseline covariates of HbA1c and weight, exhibited convergence issues and therefore the results were also not presented.

EAG comment: The company continue to argue that the maintenance dose stratified comparisons are the most appropriate: "*When interpreting the NMA it should be considered that patients unable to tolerate higher doses of one GLP-1 RA would not be expected to tolerate higher doses of another GLP-1 RA or tirzepatide; as such the most important comparisons are within each recommended maintenance dose step, rather than between recommended maintenance dose steps".*[3] (page 163) The EAG agrees that there is generally a dose response relationship for many of the outcomes and that, within these strata, tirzepatide was often more effective than the GLP-1 RAs. In particular, for all three outcomes of HbA1c, body weight and BMI change from baseline, the three doses of tirzepatide demonstrated a statistically significantly greater reduction compared to all GLP-1 RAs within the same recommended maintenance dose step. For LDL cholesterol, there was a point estimate in favour of tirzepatide, but lack of statistical significance, except for 5 mg and 10 mg versus semaglutide 1.0 mg QW where the point estimate was in favour of the comparator and there was also a dose-response relationship. Similarly, for SBP change, there was an advantage to tirzepatide versus all comparators and a dose-response relationship. Statistical significance was reached for 15 mg versus all comparators but dulaglutide 4.5 mg, for 10 mg with all but semaglutide (any dose) and the two highest doses of dulaglutide, and for 5 mg with only the lowest dose of dulaglutide. For nausea, there was also a dose response relationship and the point estimate of the risk with tirzepatide at any dose was generally higher than for all GLP-1 RAs at any dose. The exceptions were at the lowest dose and the difference was only statistically significant versus some GLP-1 RAs and only for the higher doses of tirzepatide. The only two exceptions where there was no clear dose response relationship were HDL-C and eGFR, although for the former, there was generally an advantage to tirzepatide if rarely statistically significant and for the latter there seemed to be no clear advantage generally to either tirzepatide or comparators. These patterns were generally confirmed in the sensitivity analyses. However, as already stated in Sections 2.2 and 2.3, tirzepatide, as well as dulaglutide, semaglutide and liraglutide would not be given in clinical practice according to these fixed maintenance doses, but they would be titrated. Indeed, these dose response for nausea shows why this might need to be the case, depending on the extent of achievement of targets for efficacy such as HbA1c, body weight and LDL-C. Therefore, it is very difficult to know from any comparisons in the NMA, including within the dose strata specified by the company, precisely what the results of this trade-off would be for all of the outcomes. As stated in Sections 2.2 and 2.3 the mismatch between lack of comparison of treatments in the NMA and as would be administered in clinical practice is a key issue.

The EAG also questioned the choice of different doses of semaglutide (0.5 mg; 1 mg and oral 7 mg; 14 mg) in the NMA, to which the company replied "*The NMA was conducted on a global level and therefore includes a wide range of comparators and doses to account for various global markets. However, the discussion within the submission focuses in on relevant comparators to align with the treatments and doses available in UK.*

To clarify, injectable semaglutide (branded Ozempic) is currently available in the UK at three doses; 0.25 mg (titration dose), 0.5 mg and 1.0 mg. Semaglutide has an additional licenced dose of 2.0 mg but, as stated in Section B.2.9.2 of the CS, this was not available at the time of the clinical trials and remains

unavailable at this time in the UK. Given this, 1.0 mg injectable semaglutide is the highest available dose and is therefore the most appropriate comparator for the highest doses of tirzepatide (10 mg and 15 mg). However, as described in Table 76 of Document B in the submission, the lowest dose of tirzepatide (5 mg) was compared with the lower dose of 0.5 mg injectable semaglutide. As discussed in Section B.3.2.4, comparisons were made within each recommended maintenance dose step, rather than between.

Semaglutide is additionally available in 3 doses as an oral formulation (branded Rybelsus); 3.0 mg (titration dose), 7.0 mg and 14.0 mg. The oral formulation has a low absolute bioavailability and variable absorption. The exposure after 14.0 mg oral semaglutide is equivalent to injectable 0.5 mg semaglutide.14, 15 There is no evidence available to suggest that oral semaglutide (7.0 mg or 14.0 mg) has greater efficacy than 1.0 mg injectable semaglutide. These doses were therefore considered separately from the available doses of injectable semaglutide in the NMA and cost-effectiveness analysis."

The company states that tirzepatide at the dose of 5 mg should be compared to semaglutide 0.5 mg and the rest two doses of 10 and 15 mg to semaglutide 1 mg, just because these are the two doses currently available in the UK, at the same time acknowledging that there is indeed a further dose of injectable semaglutide of 2 mg, but not available in the UK. However, the main problem, as has already been identified as a key issue, is the lack of correspondence between any of these doses and how both tirzepatide and semaglutide would be administered in clinical practice.

According to the company "*The NMA includes reasonably homogenous trials, and the majority of key trial characteristics and baseline characteristics are consistent between treatment arms included in each network of the analysis*." (page 165[3]) and that "*Baseline characteristics were largely consistent across the included treatment arms and as such, the results are likely to be robust with minimal impact from prognostic variables. In addition, numerous sensitivity analyses were conducted to assess the robustness of the findings of the main analyses*" (page 166[3]) However, the EAG notes that a formal comparison of trials in terms of characteristics and baseline characteristics has not been executed, beyond the visual presentation of these data in tables and figures, accompanied by a statement indicating comparability (see Section 3.3). There is also a lack of transparency in the presentation of consideration of heterogeneity: I^2 statistic values or Cochrane Q values were not provided in the CS and therefore statistical heterogeneity could not be assessed by the EAG. The company also stated that there was substantial heterogeneity in all continuous outcomes and nausea, but provided no details and apparently conducted no sensitivity analysis as a result of this.

The company did present a number of sensitivity analyses and broadly the results were similar to the base-case. However, the only two that addressed the variation in baseline characteristics, including any expression of treatment experience/background therapy, were the inclusion trials with unclear proportion of patients receiving metformin, and trials including patients on three OADs, as well as the meta-regression. Both of these were limited, the first mixing two completely different expressions of treatment experience/background therapy, and the second limited to only one versus two OADs. The EAG inquired how variables were indeed identified as potential treatment effect modifiers and used in the meta-regression analyses. The company replied that "*The variables identified as potential treatment effect modifiers were pre-selected during feasibility assessment based on clinical review of the included studies; refinement of the choice of selected variables was considering during heterogeneity checks undertaken when conducting the analysis. During a series of in depth internal discussions at Lilly between medical and statistical experts, potential treatment effect modifiers were considered, including baseline characteristics (such as baseline HbA1c, weight and background therapies) and study design*

features (such as study durations, timepoints for reporting endpoints). Following these discussions, assessment timepoint (weeks), number of OADs (1 vs. 2), baseline HbA1c and baseline weight were all selected to be potentially significant sources of heterogeneity and, as such, were included in the meta-regression analysis." (page 44[4]) The company acknowledged that they identified heterogeneity in a number of outcomes coming from a few studies, but did not use these results in executing a subsequent sensitivity analysis without these studies. Thus, the company provided little justification and no evidence for the choice of these covariates.

As already mentioned in Sections 2.2 and 2.3, the EAG noted that the NMA analysis implicitly assumes that the treatment effect is independent of background therapy and inquired on this additive independence assumption. The company responded that "*This assumption does underpin the NMA analysis, as it does the individual trials for both tirzepatide and comparators and is necessary given that all treatment combinations are personalised and individual to each patient given their exhibited response at any given time. As such, all trials, other than monotherapy trials, will necessarily comprise patients on a mix of background therapies. The answers to questions A5b and A19b above do not reveal any evidence to suggest that this assumption is unreasonable but nonetheless the NMA inclusion criteria were set to reduce the heterogeneity of background therapy within the network and sensitivity analyses were undertaken around background therapy*." (page 30[4]) As with the argument for dose stratification, the company has again made a spurious argument to justify how the evidence is appropriate to inform clinical practice. With dosing, titration in clinical practice is no justification for fixed dosing in the trial evidence. With background therapy, a personalised approach is no justification for mixing patients with different background therapy. This is tantamount to arguing that treatment effect modification should not be considered at all because patients are not all the same. No response has been provided on why the effects of background therapies have not been considered to add to the observed effect. Also, the company also refers to two other clarification questions, but question A5b does not refer to this issue while question A19b does not exist.[23] In addition, the company was asked whether the variation of background therapies between some trials was taken into consideration, and how the potential impact of interaction effects on estimates of the treatment effect was examined, to which the company responded that heterogeneity between studies was examined within the feasibility assessment and that further tests of heterogeneity and consistency were executed in the NMA.[4] However, as already stated, no data from the heterogeneity assessment in the feasibility analysis were provided in the CS. In addition, the company refers to the meta-regression analysis conducted on the number of OADs (one versus two only) in the trials. They go on by stating that allowing for all differences in background therapies within the NMA would reduce the number of included studies and restrict the network, which is tantamount to admitting substantial heterogeneity. Since substantial clinical heterogeneity was identified by the company, the number of included studies should have been reduced to achieve better comparability. As already discussed, additive independence of background therapy is also inseparable from assuming no effect of variation OAD experience. Therefore, the EAG therefore remain concerned about the potential bias in the NMAs that might arise as a result of heterogeneity that has not been addressed, which is why the lack of feasibility assessment/assessment of comparability in the NMA has been identified as a key issue.

The EAG was also concerned about heterogeneity of follow-up time and so inquired at clarification why the 26-week data were not used for the intervention data as well, instead of the 40-week data, to match the comparator data and how the different time frames of the three tirzepatide doses are justified.[23] The company has provided contradictory responses to these questions. On one hand they state that "*Restricting the analysis to only include data from week 26 (the most commonly reported timepoint) would have produced results that were unfairly detrimental to tirzepatide, due to the titration*

schedule of tirzepatide.", and on the other hand that "*Change in HbA1c from baseline across timepoints in SURPASS-2 Figure 13 to Figure 16, there is a plateau for all three doses of tirzepatide from Week 24 onwards. There is therefore limited inconsistency expected using different timepoints for evaluation for each dose*."(pages 40-41[4]). Since there is a plateau in the change of HbA1c, the appropriate comparison would conceptually be to use the 26 ± 4 (22–30) weeks for comparator data and 24 weeks for the SURPASS trials (reported in the sensitivity analysis). Nevertheless, substantial differences were not identified in the results of the sensitivity analysis for the alternative time windows and so this is not regarded as a key issue.

The EAG noted that there was a lack of clarity on which effect size was used in the NMA. The company has confirmed that the efficacy estimand was used from the SURPASS trials as well as from six other trials AWARD-10, AWARD-11, LIRA-ADD2SGLT-2i, PIONEER 3, PIONEER 4, SUSTAIN-FORTE. For the rest of the studies, which make up the vast majority, the estimand was not available as this concept is relatively new and therefore the only available results were used instead.[4] The company acknowledged that the two estimands are fundamentally different in the CS and the response to clarification[4]. The company supports that the differences between the two estimands, affecting only two endpoints of the NMA, will be low as the number or patients in the SURPASS trials moving to rescue therapy was low. Nevertheless, no data were presented for the rest of the 64 studies that were included in the NMA in terms of estimand. Therefore, the potential impact of using two different estimands in the same NMA could not be assessed by the EAG.

This is a key issue, tying in with executing a NMA high risk of bias described in the previous Sections.

3.5 Additional work on clinical effectiveness undertaken by the EAG

Not applicable.

3.6 Conclusions of the clinical effectiveness section

The CS[3] and response to clarification[4] provided sufficient details for the EAG to appraise the literature searches conducted to identify relevant clinical evidence for the efficacy and safety of tirzepatide and the relevant comparators, GLP-1 RAs, for the treatment of T2D. Searches were conducted in September 2021, with updates in October 2021 and June 2022. Searches were transparent and reproducible, and comprehensive strategies were used. A good range of databases and trials registers were searched. Overall, the EAG has no major concerns about the literature searches conducted, however separate AEs searches may have retrieved additional studies.

For all SURPASS trials, for nearly all HbA1c outcomes, body weight change and BMI change and all doses of tirzepatide, there was a statistically significant difference versus the comparator in favour of tirzepatide. There was also a dose-response relationship whereby the higher dose was more effective, except the dose 15 mg for tirzepatide in SURPASS-5 (versus placebo) for change in HbA1c, proportion of patients achieving HbA1c <7.0% target and HbA1c ≤6.5% target at 40 weeks, although the differences could be regarded as very small and perhaps consistent with a ceiling effect. For HDL-C change, there was also a statistically significant difference in favour of tirzepatide regardless of dose for all trials, but SURPASS-5, where the difference varied in direction depending on dose, but was much smaller and could be regarded as close to zero. There also seems to be no clear dose response relationship. For LDL-C change, the treatment effect was always in favour of tirzepatide, but not with statistical significance in SURPASS-2 and SURPASS-3, where the magnitude was also relatively small and with no clear dose response relationship. In SURPASS-4 and SURPASS-5, however, there was a statistically significant difference in favour of tirzepatide, which was much larger, especially in SURPASS-5, where there was also a dose response relationship.

Although eGFR change was included in the NMA (see below), the SURPASS trial results were not included in the CS. Given the lack of patients with renal impairment (reduced eGFR) in the SURPASS trials, the EAG considers that it might therefore be reasonable to conclude that there would be little change in eGFR during the follow-up period and therefore little expectation of a treatment effect on eGFR change. It is therefore unclear why eGFR was included in the NMAs.

For AEs, the EAG agrees with the company that, similar to the GLP-1 RA class, the most common AEs in patients treated with tirzepatide were GI related. However, the data presented treatment-emergent diabetic retinopathy complications are from the dose-effect analysis set, i.e. for tirzepatide-treated patients only with no comparator data, the appear to indicate that tirzepatide was associated with higher rates of renal AE relative to placebo, and, for MACE, the analysis presented in the response to the clarification letter is of limited value because it does not provide any information about the relative cardiac risk of tirzepatide versus relevant individual comparators.[4] The lack of comparative evidence on these outcomes and micro- and macrovascular complication of diabetes mellitus are therefore identified as a key issue.

In the NMAs, as in the SURPASS trials, tirzepatide was compared per maintenance dose with all GLP-1 RAs per maintenance dose, and there was generally a dose response relationship for many of the outcomes and that, within these strata, tirzepatide was often more effective than the GLP-1 RAs. In particular, for all three outcomes of HbA1c, body weight and BMI change from baseline, the three doses of tirzepatide demonstrated a statistically significantly greater reduction compared to all GLP-1 RAs, including within the same recommended maintenance dose. In fact, for weight change and BMI change from baseline, all doses (5 mg, 10 mg and 15 mg) of tirzepatide were significantly more effective in reducing HbA1c (%) levels compared to placebo and all GLP-1 RAs. For HbA1c change this was also the case except for 5 mg versus semaglutide 1.0 mg QW and dulaglutide 3.0 mg and 4.5 mg, where the point estimate was in favour of tirzepatide without statistical significance. There was also a dose-response relationship of tirzepatide. For LDL-C, there was a point estimate in favour of tirzepatide, but lack of statistical significance, except for 5 mg and 10 mg versus semaglutide 1.0 mg QW where the point estimate was in favour of the comparator. There was also a dose-response relationship. For SBP change, there was an advantage to tirzepatide versus all comparators and a dose-response relationship. Statistical significance was reached for 15 mg versus all comparators but dulaglutide 4.5 mg, for 10 mg with all but semaglutide (any dose) and the two highest doses of dulaglutide, and for 5 mg with only the lowest dose of dulaglutide. For nausea, there was also a dose response relationship and the point estimate of the risk with tirzepatide at any dose was generally higher than for all GLP-1 RAs at any dose. The exceptions were at the lowest dose and the difference was only statistically significant versus some GLP-1 RAs and only for the higher doses of tirzepatide. The only two exceptions where there was no clear dose response relationship were HDL-C and eGFR, although for the former, there was generally an advantage to tirzepatide if rarely statistically significant and for the latter there seemed to be no clear advantage generally to either tirzepatide or comparators. These patterns were generally confirmed in the sensitivity analyses. However, as already stated in Sections 2.2 and 2.3, tirzepatide, as well as dulaglutide, semaglutide and liraglutide would not be given in clinical practice according to these fixed maintenance doses, but they would be titrated. Indeed, these dose response for nausea shows why this might need to be the case, depending on the extent of achievement of targets for efficacy such as HbA1c, body weight and LDL-C. Therefore, it is very difficult to know from any comparisons in the NMA, including within the dose strata specified by the company, precisely what the results of this trade-off would be for all of the outcomes. This further undermines any conclusions from the NMA as to the precise treatment effect that would be observed in clinical practice and lends further support to Key Issue already identified regarding the mismatch between dosing in clinical practice and in the NMA.

The EAG also identified some serious methodological problems with the NMA, the standard methodology of an NMA being to first execute a SLR tailored to the appropriate PICOs and then execute a feasibility assessment to assess suitability of synthesis of the identified studies, according to comparability, especially of baseline characteristics that might be treatment effect modifiers. In contrast, apart from the mismatch between the decision problem population and those of the trials in the NMA in terms of OAD treatment experience/background therapy, there was no systematic comparison of trials, one to another, according to these characteristics or any other baseline characteristics. It also seems that, despite some considerable variation in some baseline characteristics and the identification by the company of substantial heterogeneity, the effect of exclusion of trials on the basis of that variation was not tested. The company did present a number of sensitivity analyses and broadly the results were similar to the base-case. However, the only two that addressed the variation in baseline characteristics, including any expression of treatment experience/background therapy, were the inclusion trials with unclear proportion of patients receiving metformin, and trials including patients on three OADs, as well as the meta-regression. Both of these were limited, the first mixing two completely different expressions of treatment experience/background therapy, and the second limited to only one versus two OADs. The company acknowledged that they identified heterogeneity in a number of outcomes coming from a few studies but did not use these results in executing a subsequent sensitivity analysis without these studies. The EAG therefore remain concerned about trial heterogeneity, which is why the lack of feasibility assessment/assessment of comparability in the NMA has been identified as a Key Issue.

4. COST EFFECTIVENESS

4.1 EAG comment on company's review of cost effectiveness evidence

This Section pertains mainly to the review of cost effectiveness analysis studies. However, the search Section (5.1.1) also contains summaries and critiques of other searches related to cost effectiveness presented in the CS. Therefore, the following Section includes searches for the CEA review, measurement and evaluation of health effects as well as for cost and healthcare resource identification, measurement and valuation.

4.1.1 Searches performed for cost effectiveness section

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the CS.[3] The CADTH evidence-based checklist for the PRESS, was used to inform this critique.[13, 14] The CS[3] was checked against the STA specification for company/sponsor submission of evidence.[15] The EAG has presented only the major limitations of each search strategy in the report.

Appendix G of the CS provides details of a SLR conducted to identify previous health economic evaluations to inform a CEA of tirzepatide in the UK setting.[12] Searches were conducted in February 2022.

A summary of the sources searched is provided in Table 4.1.

Table 4.1: Data sources searched for economic evaluations (as reported in CS)

Resource	Host/Source	Date Ranges	Date searched
Electronic databases			
PubMed	Internet	2015-21/2/22	21/2/22
Embase	Embase.com		
Cochrane Library	Cochrane Library		
EconLit	EBSCO		
Conferences			
ISPOR/ISPOR-EU	Internet	Jan 2020-date	Not stated
Diabetes UK			
EASD Annual Congress			
HTA websites/Grey literature resources			
NICE website	Internet	2015-Jan 2022	Jan 2022
NICE Evidence Search			
SIGN website			
AWMSG website			
HTA dataset			
NCPE website			
SBU website			
HAS website			
PBS website			
PBAC website			
CADTH website			

Resource	Host/Source	Date Ranges	Date searched
HQO website			
AWMSG = All Wales Medicines Strategy Group; CADTH = Canadian Agency for Drugs and Technologies in Health; CS = company submission; EASD = European Association for the Study of Diabetes; EU = Europe; HAS = Haute Autorité de Santé; HQO = Health Quality Ontario; HTA = Health Technology Assessment; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NCPE = National Centre for Pharmacoeconomics; NICE = National Institute for Health and Care Excellence; PBAC = Pharmaceutical Benefits Advisory Committee; PBS = Pharmaceutical Benefits System; SBU = Swedish Agency for Health Technology Assessment and Assessment of Social Services; SIGN = Scottish Intercollegiate Guidelines Network; UK = United Kingdom			

Appendix H of the CS provides details of a SLR conducted to identify published HRQoL utilities to inform a CEA of tirzepatide in the UK setting.[12] Searches were conducted in February 2022. As they were intended to update a previous review[25] searches were conducted from March 2020.

A summary of the sources searched is provided in Table 4.2.

Table 4.2: Data sources searched for HRQoL studies (as reported in CS)

Resource	Host/Source	Date Ranges	Date searched
Electronic databases			
PubMed	Internet	17/3/20-31/1/22	21/2/22
Embase	Embase.com	17/3/20-31/1/22	
Cochrane Library	Cochrane Library	2020-date	
EconLit	EBSCO	01/01/15-31/1/22	
Conferences			
ISPOR	Internet	Jan 2020-date	Not stated
Diabetes UK			
EASD Annual Congress			
ADA Annual Congress			
HTA websites/Grey literature resources			
NICE website	Internet	2015-Jan 2022	Jan 2022
NICE Evidence Search			
SIGN website			
AWMSG website			
HTA dataset			
NCPE website			
SBU website			
HAS website			
PBS website			
PBAC website			
CADTH website			
HQO website			
ICER website			
ScHARR Health Utilities Database			

Resource	Host/Source	Date Ranges	Date searched
ADA = American Diabetes Association; AWMSG = All Wales Medicines Strategy Group; CADTH = Canadian Agency for Drugs and Technologies in Health; CS = company submission; EASD = European Association for the Study of Diabetes; HAS = Haute Autorité de Santé; HRQoL = health-related quality of life; HQO = Health Quality Ontario; HTA = Health Technology Assessment; ICER = Institute for Clinical and Economic Review; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NCPE = National Centre for Pharmacoeconomics; NICE = National Institute for Health and Care Excellence; PBAC = Pharmaceutical Benefits Advisory Committee; PBS = Pharmaceutical Benefits System; SBU = Swedish Agency for Health Technology Assessment and Assessment of Social Services; ScHARR = School of Health and Related Research; SIGN = Scottish Intercollegiate Guidelines Network; UK = United Kingdom			

Appendix I of the CS provides details of a SLR conducted to identify published cost data to inform a CEA of tirzepatide in the UK setting.[12] Searches were conducted in February 2022.

A summary of the sources searched is provided in Table 4.3.

Table 4.3: Data sources searched for cost/resource use studies (as reported in CS)

Resource	Host/Source	Date Ranges	Date searched
Electronic databases			
PubMed	Internet	1/1/15-31/1/22	21/2/22
Embase	Embase.com	1/1/15-31/1/22	
Cochrane Library	Cochrane Library	1/1/15-date	
EconLit	EBSCO	1/1/15-31/1/22	
Conferences			
ISPOR	Internet	Jan 2020-date	Not stated
Diabetes UK			
EASD Annual Congress			
ADA Annual Congress			
HTA websites/Grey literature resources			
NICE website	Internet	2015-Jan 2022	Jan 2022
NICE Evidence Search			
SIGN website			
National Cost Collection for the NHS website			
AWMSG website			
HTA dataset			
Research Papers in Economics website			
ADA = American Diabetes Association; AWMSG = All Wales Medicines Strategy Group; CS = company submission; EASD = European Association for the Study of Diabetes; HTA = Health Technology Assessment; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; SIGN = Scottish Intercollegiate Guidelines Network; UK = United Kingdom			

EAG comment:

- Searches were undertaken in January 2022 to identify economic, HRQoL and healthcare resource use/cost data on tirzepatide for the treatment of T2D. The CS, Appendices G, H and I and the

Company's response to clarification provided sufficient details for the EAG to appraise the literature searches.[3, 4, 12]

- A good range of databases, websites, grey literature resources and trials registers were searched. Reference checking was conducted.
- Searches were well structured, transparent and reproducible, and a good range of subject indexing terms (MeSH/EMTREE) and free text was used.
- The database searches for economic evaluations and cost/resource use had a 2015 publication date limit. As the HRQoL searches were intended to update a previous review[25] searches were conducted from March 2020.
- Searches were limited where possible to English language publications only.
- The database searches for the economic evaluation SLR contained a population facet for T2D combined with terms for tirzepatide and the relevant comparators. In the searches of PubMed, Embase and the Cochrane Library this was then limited using a cost effectiveness filter and a geographical search filter to restrict the results to economic evaluations of relevance to the UK/Ireland. The filters used were not referenced, so it was unclear whether they were published objectively-derived filters. The geographic search filter should be used with caution as it risks missing potentially relevant records, however given the large numbers of records retrieved by the searches, the EAG considers this a pragmatic approach.
- The database searches for the HRQoL utilities SLR contained a population facet for T2D. In the PubMed, Embase and Cochrane Library searches, this was then combined with a facet containing terms for diabetes-related complications, and an HRQoL filter. The HRQoL study design filters were not referenced, so it was unclear whether they were published objectively-derived filters. They contained a combination of subject heading terms and free text terms, and the EAG deemed them to be adequate.
- The database searches for the cost/resource use SLR contained a population facet for T2D. In the searches of PubMed, Embase and the Cochrane Library this was then combined with a cost filter and limited using a geographical search filter to restrict the results to cost studies of relevance to the UK/Ireland. The filters used were not referenced, so it was unclear whether they were published objectively-derived filters. The geographic search filter should be used with caution as it risks missing potentially relevant records, however given the large numbers of records retrieved by the searches, the EAG considers this a pragmatic approach.
- Conference proceedings searches were conducted for 2020-2022. The Embase searches were not limited to exclude conferences, so these will also have retrieved conference proceedings.

4.1.2 Inclusion/exclusion criteria

In- and exclusion criteria for the review on cost effectiveness studies, HRQoL studies and costs and resource use studies are presented in Table 4.4.

Table 4.4: Eligibility criteria for the SLRs

	Inclusion criteria	Exclusion criteria
Patient population	Adults aged ≥18 years with T2D, or other population samples (e.g., general population) questioned in relation to health states specifically related to T2D	T1D Gestational diabetes Mixed populations of T1D and T2D patients Patients at high risk of developing T2D Animal studies

	Inclusion criteria	Exclusion criteria
		In vitro studies
Intervention and comparators (published economic evaluations)	Tirzepatide, injectable semaglutide, oral semaglutide, dulaglutide, liraglutide, lixisenatide, exenatide, empagliflozin, canagliflozin, dapagliflozin, ertugliflozin, and insulin	Older therapies for T2D will not be reviewed
Intervention and comparators (HRQoL studies and cost/resource use studies)	All interventions treatment	N/A
Outcomes(s) 1 (published economic evaluations)	Life expectancy, quality-adjusted life expectancy, complication rates, direct costs, indirect costs, cost per life year saved, cost per event avoided and cost per QALY gained	No modelling analysis or reporting of cost effectiveness outcomes
Outcomes(s) 2 (HRQoL studies)	Health state utility values for: T2D (no complications) T2D with: Acute complications (severe hypoglycaemia, non-severe hypoglycaemia, lactic acidosis, ketoacidosis as well as fear of hypoglycaemia CV complications (angina, myocardial infarction, congestive heart failure) Stroke Peripheral neuropathy/peripheral vascular disease Foot ulcer Amputation Ophthalmic complications (diabetic retinopathy, macula oedema, severe vision loss) Renal complications (micro- and macroalbuminuria, renal transplant, haemodialysis, peritoneal dialysis) Obesity, weight loss or weight gain Treatment-related attributes (dosing frequency, mode of administration, pill burden, devices, etc.,)	Not reporting utility values (e.g., only qualitative aspects of QoL, or quantifying QoL in a manner other than utility values), or reporting utility values not relevant to the review (e.g. for patients with T2D with comorbid conditions such as arthritis or cancer, disutility for hip fracture, exacerbation of COPD etc.)
Outcomes(s) 3 (cost/resource use studies)	Annual or event costs for T2D with: Acute complications (severe and non-severe hypoglycaemia) Cardiovascular complications (ischemic heart disease, myocardial infarction, heart failure)	Not reporting per patient or per event cost estimates, or reporting costs not relevant to the modelling analysis (e.g., for patients with T2D with comorbid conditions such as arthritis or cancer)

	Inclusion criteria	Exclusion criteria
	Stroke Peripheral neuropathy/severe pressure sensation loss Foot ulcer Amputation Ophthalmic complications (macula oedema and blindness/severe vision loss) Renal complications (KDIGO CKD Stages 1-5 and ESRD)	
Study design 1 (CEA studies)	Cost effectiveness and cost-utility studies, appraisals/assessment reports from relevant HTA agencies and UK guidelines in T2D (specifically NICE and the Scottish Medicines Consortium) Guidelines	Case reports Narrative review Commentaries Letters Editorials Discussion papers Secondary sources Abstracts (published prior to 2020)[a]
Study design 2 (HRQoL studies)	RCTs Observational studies Case control studies Cross-sectional studies Guidelines	Case reports Narrative review Commentaries Letters Editorials Discussion papers Secondary sources (e.g., economic analyses sourcing utility values were sourced from the literature) Discrete choice experiment studies Abstracts (published prior to 2020)[a]
Study design 3 (cost/resource use studies)	RCTs Observational studies Case control studies Cross-sectional studies Previous HTA submissions and reviews	Case reports Narrative review Commentaries Letters Editorials Discussion papers Secondary sources (e.g., economic analyses sourcing costs were sourced from the literature) Abstracts (published prior to 2020)

	Inclusion criteria	Exclusion criteria
Time frame (published economic evaluations and cost/resource use studies)	January 2015–January 2022	Publications prior to January 2015
Time frame (HRQoL studies)	March 2020–January 2022 for database searches 2020 or 2022 for congress abstracts January 2017–January 2022 for website searches	Publications outside the time frames for inclusion
Language	English	Non-English

CS Appendices G, H, and I
CS = company submission; CEA = cost effectiveness analysis; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; CV = cardiovascular; ESRD = end stage renal disease; HRQoL = health-related quality of life; HTA = Health Technology Assessment; KDIGO = Kidney Disease Improving Global Outcomes; N/A = not applicable; NICE = National Institute for Health and Care Excellence; QALY = quality-adjusted life year; QoL = quality of life; RCT = randomised controlled trial; T1D = type 1 diabetes; T2D = type 2 diabetes; UK = United Kingdom

EAG comment: The EAG agrees that the eligibility criteria are suitable to fulfil the company's objective to identify cost effectiveness studies. The rationales for excluding cost effectiveness studies after full paper reviewing are considered appropriate given the defined in- and exclusion criteria.

4.1.3 Conclusions of the cost effectiveness review

The CS provides an overview of the included cost effectiveness, HRQoL and resource use and costs studies, but no specific conclusion was formulated.

4.2 Summary and critique of company's submitted economic evaluation by the EAG

4.2.1 NICE reference case checklist

Table 4.5: NICE reference case checklist

Element of HTA	Reference case	EAG comment on CS
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Consistent with NICE reference case
Perspective on costs	NHS and PSS	Consistent with NICE reference case
Type of economic evaluation	Cost utility analysis with fully incremental analysis	Partly consistent with NICE reference case (full incremental analysis is missing)
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Consistent with NICE reference case
Synthesis of evidence on health effects	Based on systematic review	Consistent with NICE reference case

Element of HTA	Reference case	EAG comment on CS
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults	Consistent with NICE reference case
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	Consistent with NICE reference case
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	Consistent with NICE reference case
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Consistent with NICE reference case
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Consistent with NICE reference case
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Consistent with NICE reference case

CS = company submission; HTA = Health Technology Assessment; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; PSS = Personal Social Services; QALY = quality-adjusted life year; UK = United Kingdom

4.2.2 Model structure

The company adopted the PRIME T2D model (programmed in JAVA 10), which was described as a discrete time event model. The company justified the use of the PRIME T2D model by stating that "*models developed prior to 2016, including UKPDS OM1 and OM2 and the IQVIA CORE Diabetes model, have been shown to under predict CV benefits from the GLP-1 RA class in certain situations. One reason for this could be that models developed earlier than 2016 do not fully capture the benefits of reduced body weight as they tend to be based on cohorts using traditional therapies without any weight loss benefit.*"

4.2.2.1 Patient characteristics

This model consists of a patient-level simulation with the following characteristics (according to the model interface):

Demographics

- age
- age at T2D diagnosis
- gender
- education level

Risk factors

- BMI
- HbA1c
- SBP
- total cholesterol

Complication history

- atrial fibrillation
- urinary albumin ≥50 mg/l
- peripheral vascular disease
- myocardial infarction

- race
- smoking status
- LDL
- HDL
- eGFR
- WBC
- heart rate
- haemoglobin
- stroke
- ischemic heart disease
- revascularisation
- heart failure
- ulcer
- amputation
- blindness
- renal failure
- neuropathy

4.2.2.2 Risk models to estimate health consequences

Risk models were used to estimate complications (and possibly AEs) based on the estimated patient characteristics. The company stated that the risk models (to estimate health consequences) were retrieved from the literature review as well as identified publications on existing models of T2D mellitus. Publications were selected based on the following criteria:

- Study reported endpoint data appropriate to a health economic modelling (e.g., single (not composite) endpoints, hard endpoints and/or those associated with a mortality, QoL and/or cost impact)
- Study reported or was based on longitudinal data and long-term follow up (≥4 years)
- Study enrolled a substantial number of patients (ideally >500 patients)
- Preference was given to studies with the longest duration of follow up and greatest patient numbers in cases where multiple studies reported comparable endpoint data

Selection of the most appropriate risk model was confirmed at the Advisory Board Meeting on 3 September 2019. Due to the variation between risk models in the CVD risk factors considered, no consensus could be reached on the preferred risk models. At the Advisory Board Meeting, it was agreed that for simplicity, comprehension and acceptance by health technology associations, a single risk model should be used if possible. Moreover, it was agreed that a model averaging approach could be used to combine risk models.

The model averaging approach is used in the estimation of macrovascular complication risk (myocardial infarction, stroke, ischemic heart disease and heart failure), and in the risk of blindness. In Appendix N of the CS, it is described that a weighted model averaging approach was used in which each risk model was assigned a weight based on the similarity of mean baseline characteristics between the simulated patients and the cohort used to derive the risk model (derivation cohort). The greater the similarity between the simulated patients and the derivation cohort, the larger the weight applied to the risk model. The model averaging approach was then optimized by running validation simulations to evaluate predictive performance, measured using the Chi-squared statistic, and using a genetic algorithm to minimise Chi-squared by adjusting distance coefficients for each characteristic. Notably, in each

simulation, weights are calculated using the baseline characteristics on a patient level. This means that different simulated patients will have different weighting of the risk equations in the simulation due to heterogeneity within a modelled cohort. In each year of the simulation, weighting of the risk equations is adjusted for age and duration of diabetes (but not other risk factors) for each patient, so the weighting of equations can change over time in any given simulation.

4.2.2.3 Complications and AEs considered

The following complications are considered in the model (according to Appendix N.5.3.3 of the CS):

- Myocardial infarction, first and recurrent (model averaging of Building, Relating, Assessing, and Validating Outcomes (BRAVO) and United Kingdom Prospective Diabetes Study (UKPDS) OM2)
- Stroke (model averaging of BRAVO, UKPDS OM2 and Yang et al 2007 based on the Hong Kong Diabetes Registry[26])
- Ischemic heart disease (model averaging of BRAVO and UKPDS OM2)
- Heart failure (model averaging of BRAVO, UKPDS OM2 and Yang et al based on the Hong Kong Diabetes Registry [26])
- Renal failure (model based on Zoppini et al using an Italian cohort of 1,682 patients with 10 years of follow-up [27]
- Neuropathy/severe pressure sensation loss (BRAVO model)
- Amputation (UKPDS OM2 model)
- Ulcer (UKPDS OM2 model)
- Blindness (Model averaging of BRAVO and UKPDS OM2)
- Macular oedema (model not stated in Appendix N.5 of the CS)

Adverse events considered in the model:

- Non-severe hypoglycaemic, defined as nocturnal and diurnal events (rate of hypoglycaemia was assumed to be zero for tirzepatide and all comparators)
- Severe hypoglycaemic, defined as those requiring third party medical assistance (rate of hypoglycaemia was assumed to be zero for tirzepatide and all comparators)
- Nausea (estimates were derived from the NMA, Table 81 of the CS)

4.2.2.4 Treatment intensification

After HbA1c levels rose above 7.5% (58 mmol/mol), the company assumed that treatment would be intensified. Specifically, it was assumed that the initial treatment (i.e., tirzepatide or the comparator) was discontinued and patients would switch to basal insulin therapy for the remainder of the analysis time horizon. This assumption was considered by the company to be in line with NG28 for T2D in adults which stated that if HbA1c levels are not adequately controlled by a single drug and rise to 58 mmol/mol (7.5%) or higher:

- reinforce advice about diet, lifestyle and adherence to drug treatment
- support the person to aim for an HbA1c level of 53 mmol/mol (7.0%)
- intensify drug treatment

A simplifying assumption of only one intensification step was assumed in this evaluation. The company argued that subsequent intensification steps would have very little impact on cost effectiveness, as the changes would be similar in both treatment arms.

EAG comment: The main concerns of the EAG relate to: a) the company justification to use the PRIME T2D model; b) specification of the model type; c) model averaging approach; d) selection of predictive models to estimate the risk of complications and e) approach to estimate risk of macular oedema, ulcer and revascularisation.

a) The EAG did not find compelling evidence (from the reference provided in the CS[28]) to support the company's statement (to support the use of the justification to use the PRIME T2D model): "*that models developed earlier than 2016 do not fully capture the benefits of reduced body weight as they tend to be based on cohorts using traditional therapies without any weight loss benefit*". In response to clarification question B1a, the company quoted Si et al 2020[29] reporting on the Ninth Mount Hood Diabetes Challenge: "*commonly used risk equations were generally unable to capture recent CVOT treatment effects but that calibration of the risk equations can improve predictive accuracy. Although calibration serves as a practical approach to improve predictive accuracy for CVOT outcomes, it does not extrapolate generally to other settings, time horizons, and comparators. New methods and/or new risk equations for capturing these CV benefits are needed.*" According to the company this provides evidence on the shortcomings of earlier T2D models. Nevertheless, it is unclear for the EAG that the developed de novo model, specifically the current implementation as in the CS, has a better performance to predict complications (including CV events) compared with existing diabetes models. In response to clarification question B1b, the company stated, "*it is impossible to provide definitive evidence that the PRIME T2D Model will predict outcomes more accurately than other available T2D model*" and "*suggest that comparison of the Ninth Mount Hood Diabetes Challenge results with the published validation results for the PRIME T2D Model indicate that the PRIME T2D Model may be better placed to predict cardiovascular outcomes in line with those observed from recent CVOTs*". However, no comparison of the Ninth Mount Hood Diabetes Challenge results and the current implementation of the PRIME T2D Model was provided by the company in response to this clarification question. Consequently, the justification to use the PRIME T2D model instead of commonly used available alternatives mentioned in Table 72 of the CS was not compelling according to the EAG (e.g., the CORE Diabetes Model that was used for NG28 focusing on the management of T2D, see also Table 73 of the CS). Notably, the following complications were considered in the modelling for NG28[30, 31] and were not included in Figure 69 of the CS: angina, peripheral vascular disease, diabetic retinopathy, cataract, ketoacidosis, lactic acidosis and foot ulcer (though the latter might be included in the PRIME T2D model given in Appendix N.5.6.2 of the CS elaborates on the estimation of ulcer risk)

b) The model type specified by the company is discrete time event. It is unclear to the EAG what this exactly implies. In response to clarification question B2 the company stated that "*Model nomenclature can be challenging given the conventional classifications*" and that the PRIME T2D model runs on an annual cycle length and that patient characteristics, treatments and methods of risk evaluation can change between events. The company stated that it is not a state-transition model as patients may be members of multiple "health states" simultaneously. Moreover, the company "*deliberately avoided the term "discrete event simulation (DES)" as it is synonymous with a series of 'events' that occur over time (as opposed to events occurring within an annual cycle)*". Nevertheless, it is unclear to the EAG why the company did not adopt a DES approach, also considering that DES models can include annual updates of certain model inputs (such as patient characteristics).[1]

c) For the estimation of macrovascular complications and blindness risks, the company adopted a model averaging approach. In response to clarification question B4a, the company elaborated

on this approach. However, the justification why model averaging is preferred instead of selecting a single predictive model that best matches the decision problem (with alternative models in scenario analyses) provided by the company in response to clarification question B4b is not compelling to the EAG. Unfortunately, the company did not provide scenario analyses selecting a single predictive model based on the best match of the derivation cohort to the decision problem (as requested in clarification question B4c).

d) Appendix N of the CS provides descriptions for the generic PRIME T2D Model. However, the appropriateness of the selected predictive models to estimate the risk of complications in patients with T2D is not justified (in detail). Nor is the applicability to the specific decision problem (as specified in the CS) justified. In response to clarification question B5, the company stated that "*The risk equations from the UKPDS OM2 have been widely used in the past, have been derived from a UK-specific T2D populations and are likely well-suited for patients with a low risk profile and short duration of disease. Risk equations from the BRAVO Model are better suited to patients with more advanced disease and higher risk profile (derived from the ACCORD trial population which was at high risk of cardiovascular complications). Literature review did not identify any UK-specific risk equations that could be used in a model averaging approach for patients at high risk of complications and therefore BRAVO Model equations were used. Risk equations from the Hong Kong Diabetes Registry present in the PRIME T2D Model are applicable for South East Asian populations and were not influential in the present analysis.*" Unfortunately, the company did not provide justifications (requested in clarification question B5), that the risk models used, both individually and after model averaging, are appropriate to estimate the risk of complications for the population as specified in the CS.

e) Although Figure 69 of the CS (schematic diagram of the economic model) does include macular oedema, the estimation of/risk model for macular oedema not mentioned in Appendix N.5 of the CS. Moreover, Appendix N.5.6.2 of the CS elaborates on the estimation of ulcer risk. However, ulcer is not included in Figure 69 of the CS. Similarly, revascularisation is included in Table 98 of the CS (cumulative incidence of diabetes-related complications), but it is unclear to the EAG how the revascularisation was estimated and incorporated the economic model. These issues should be clarified by the company.

4.2.3 Population

The population considered in the CS was adults with T2D that is inadequately controlled with three or more antidiabetic agents, which was narrower than the anticipated marketing authorisation for tirzepatide and the population in the final NICE scope. The MHRA therapeutic indications and NICE final scope also included tirzepatide as monotherapy for adults with T2D that is inadequately controlled with diet and exercise alone when metformin is considered inappropriate. Furthermore, the NICE final scope also included tirzepatide for adults with T2D that is inadequately controlled with one or more antidiabetic agents (instead of three or more, as stated by the company).

The modelled baseline patient characteristics were presented in Table 75 of the CS[3]. These were retrieved from the THIN second intensification cohort from NG28[32], which standard care included metformin, sulfonylurea and neutral protamine Hagedorn (NPH) insulin. When inputs were missing from the NG28 cohort the company applied the corresponding values from the SURPASS-2 clinical trial cohort (i.e., baseline total cholesterol, percentage with atrial fibrillation at baseline, percentage with peripheral vascular disease at baseline, percentage with coronary revascularization at baseline, and percentage with severe pressure sensation loss (SPSL)/neuropathy at baseline).

EAG comment: The main concerns of the EAG relate to: a) mismatch with the population considered in the NICE decision problem and clinical effectiveness evidence; and b) differences in patient characteristics between THIN cohort and SURPASS trials.

a) As discussed in Section 2.1 of this report, the decision problem addressed by the company is much narrower than the NICE scope and the MHRA marketing authorisation; with no inclusion of adults with inadequately controlled T2D receiving tirzepatide monotherapy and focusing instead on adults with inadequately controlled T2D receiving tirzepatide in combination but given at a later stage than indicated in the NICE scope (i.e., NICE scope including tirzepatide with one or more antidiabetic agents, and CS including with three or more antidiabetic agents). Furthermore, as discussed in clarification question A5[4], the company's clinical trials (SURPASS-2 -5) appear to be misaligned with the population of the company's decision problem described in the CS (see Table 2.2 of this report). The SURPASS trials were generally at an earlier line of therapy or with less treatment experienced patients (i.e., closer to the NICE final scope) and with two trials excluding patients receiving triple therapy within the three months prior to the first visit. The company justified the misalignment between the SURPASS trials and the intended placement of tirzepatide in the UK care pathway by stating that "*the SURPASS trials were designed to meet regulatory requirements of different authorities around the globe" and given that the GLP-1 Ras positioning varies globally, the trial designs do not completely align with the UK clinical practice or decision problem*". The company argues that the NMA was conducted to help this lack of direct comparison. The EAG is concerned about the mismatch between the population from the company's decision problem and the population from trial evidence, as highlighted in Section 2.1, as the decision problem is narrower than both the NICE scope and trial evidence. As the SURPASS trials appear to be a better match for the population in the NICE final scope, the EAG would have liked to explore scenario analyses with the cohort characteristics and effectiveness reported in those trials, including the list of the relevant comparators in accordance with the trials and NICE scope.

b) The baseline characteristics for the model population (provided in Table 75[3] of the CS) were mostly based on the baseline characteristics of patients in the THIN second intensification cohort, for which standard care was based on: metformin, sulfonylurea and NPH insulin. The second intensification cohort were used given that, as per the CS, population characteristics from third-line therapies (i.e., as per the company decision problem) from the THIN database were not available. The company argued that this was representative of the UK population with T2D initiating second line therapy in clinical practice, after failing diet and exercise plus metformin. When data from the THIN cohort were missing, the company used data from the SURPASS-2 cohort, since the mean duration between the two was comparable (8.6 versus 8.5 years). Notably, the SURPASS-2 population was younger (63.95 versus 56.6 years) and had higher baseline HbA1c percentage (7.50 versus 8.28), higher BMI (30.7 versus 34.5), and higher eGFR (71.37 versus 96 ml/min/1.73 m^2) than the THIN cohort. However, the EAG agrees with the choice of the THIN cohort as more representative of the UK population with T2D initiating second line therapy in clinical practice, after failing diet and exercise plus metformin. Besides, as per Table 13 from the clarification response[4], there is an erratum on the percentage of Hispanic for the SURPASS trials that should be corrected.

4.2.4 Interventions and comparators

The intervention considered in the CS was tirzepatide with other antidiabetic agents, and did not include tirzepatide alone, despite being included in the final NICE scope (see Table 1 of the CS). The company

included three different maintenance doses of tirzepatide: 5 mg, 10 mg, or 15 mg, administered via injection QW. The CS proposed an initial dose of 2.5 mg that could be increased by 2.5 mg every 4 weeks, until reaching the desired maintenance dose.[3]

The NICE scope listed the following comparators: sulfonylureas, DPP-4i, pioglitazone, GLP-1 mimetics, SGLT-2i, and insulin. Nonetheless, the comparators considered in the base-case analysis were the only following GLP-1 RAs: dulaglutide (1.5 mg, 3.0 mg, and 4.5 mg QW), liraglutide (1.2 mg and 1.8 mg QD), oral semaglutide (7 mg and 14 mg QW), and injectable semaglutide (0.5 mg and 1.0 mg QW). The company did not include the following GLP-1 RAs in its base-case (rather in the scenario analyses only) due to limited market share: lixisenatide (20 µg QD), standard exenatide (10 µg twice daily), and modified-release exenatide (2.0 mg QW). The company justified the narrower comparator group, compared with the scope, by stating that tirzepatide was anticipated to be used for patients with T2D that was inadequately controlled with three or more antidiabetic agents, and GLP-1 RA would be considered for this population,

The company asserted that patients unable to tolerate higher doses of one of the GLP-1 RAs (e.g., due to GI side-effects) would also not be expected to tolerate higher doses of a different GLP-1 RA or tirzepatide. Therefore, for the base-case analysis, comparisons were made within each recommended maintenance dose step, as opposed to between the dose having been titrated as required. The comparisons from the base-case analysis are shown in Table 4.6. The company did not consider it to be meaningful to compare the cost effectiveness between maintenance dose steps (i.e., the comparisons of a low maintenance dose with high maintenance dose).

Table 4.6: Overview of comparators per dose step (from Table 76, CS)

Tirzepatide recommended maintenance dose	Comparators
Tirzepatide 5 mg	Dulaglutide 1.5 mg Semaglutide 0.5 mg Oral semaglutide 7 mg Liraglutide 1.2 mg
Tirzepatide 10 mg	Dulaglutide 3.0 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
Tirzepatide 15 mg	Dulaglutide 4.5 mg Semaglutide 1.0 mg Oral semaglutide 14 mg Liraglutide 1.8 mg
CS, Table 76[3] CS = company submission	

EAG comment: The main concerns of the EAG relate to: a) restricting to comparisons within each recommended maintenance dose step, b) narrower intervention and comparators than in NICE final scope; c) basal insulin as the only treatment option after treatment intensification.

a) As per the CS, tirzepatide should be initiated with an initial dose of 2.5 mg that could be increased by 2.5 mg every 4 weeks up to 15 mg. In the base-case, the recommended maintenance doses of the intervention (i.e., tirzepatide) are 5 mg, 10 mg or 15 mg (see Table 4.6), which were compared to a series of GLP-1 RAs in different dosages (see Table 4.6).

Comparisons were made within each recommended maintenance dose step, and not between recommended maintenance dose steps, meaning that no comparisons were made between a low-maintenance dose with a high-maintenance dose. The EAG agrees with the company that, in clinical practice, patients would be expected to be titrated up the recommended maintenance doses as required, with the most appropriate dose being determined by the clinician based on patient's characteristics and tolerability, and aligned with the SmPC, especially given that T2D is a chronic and progressive disease, and patients may require (de)escalation through the doses. However, it is unclear that the modelling approach separately analysing dose steps, reflects clinical practice or whether, for instance, it would be possible in practice for patients to switch between these dose steps (which is currently not possible in the economic model). Furthermore, in the SURPASS trials, patients were randomised into the three maintenance doses (i.e., 5 mg, 10 mg, and 15 mg), which is not applicable to clinical practice, as patients that would fall in the 5 mg or 10 mg groups during the trials while a higher could be more suitable (and vice versa for lower doses for 10 mg or 15 mg). Therefore, the EAG would prefer that the treatment strategies incorporated in the economic model would reflect clinical practice (including the possibility for individual patients to switch between treatment dosages). Additionally, the EAG would prefer comparisons between maintenance doses, instead of restricting comparisons to within dose steps.

b) Both the intervention and comparators were narrower in the company decision problem than in the NICE final scope. For the intervention, the company's decision problem did not include tirzepatide monotherapy. For the comparators, sulfonylureas, DPP-4i, pioglitazone, SGLT-2i, and insulin were not included, as the company argued that the GLP-1 RAs were the only relevant comparator for tirzepatide in the submission. Likewise, the company did not include some GLP-1 RAs due to limited market share (e.g., lixisenatide, standard exenatide, and modified-release exenatide), despite being asked in clarification question B10.[4] The EAG agrees that including only GLP-1 RAs is consistent with the population in the company's decision problem, as discussed in Section 2.3. of this report. However, the EAG would prefer that the company's decision problem was a better match with the trial evidence (i.e., at an earlier line of therapy or in addition to insulin and with the comparators appropriate to such a line of therapy) and the NICE final scope overall. The EAG would like to highlight that the company did not provide an updated economic model, scenario analyses, and fully incremental analyses including all comparators described in the final scope as comparator despite requested in clarification question B9.[4]

c) After intensification, simulated patients in the economic model were assumed to switch to basal insulin therapy and remain there for the remainder of the simulation. However, the company did not provide supporting evidence on other possible treatments that could be used after intensification, as requested in clarification question B13.[4] The company argued that, under the assumption that the treatment effects would be equivalent in both arms after post-intensification, those treatments would have a limited effect on the cost effectiveness.

4.2.5 Perspective, time horizon and discounting

The analysis is performed from the NHS and PSS perspective. Discount rates of 3.5% are applied to both costs and benefits. The model cycle length is one year with a 50-year time horizon.

EAG comment: The approach adopted by the company in terms of perspective, time horizon and discounting is in concordance with the NICE reference case.

4.2.6 Treatment effectiveness and extrapolation

Baseline characteristics and risk factors are described in Section 4.2.3 and Table 75 of the CS. The change (treatment specific) during the first year in HbA1c (%), SBP (mmHg), HDL (mmol/l), LDL (mmol/l) and weight (kg; used to calculate BMI (kg/m²)), are based on the NMA (Section B.2.9 of the CS) and presented in Tables 77-80 of the CS. Whilst BMI was included in the NMA outputs, the change from baseline in BMI was calculated based on changes in body weight (assuming an average length of 1.68; the mean value reported for the THIN population) reported in the NMA. This was justified by stating the BMI values were not available for all comparators whereas changes in body weight were available for all comparators. For eGFR, WBC and Hb levels no change from baseline was assumed (for all treatments considered).

To handle missing inputs from the NMA, the company used a "nearest neighbour" approach. Where inputs were missing, the corresponding input from the same compound was used as a proxy, wherever possible using higher (more efficacious) doses of comparator. For example, missing changes in serum lipid levels for semaglutide 0.5 mg were taken from the corresponding values for semaglutide 1.0 mg.

4.2.6.1 Extrapolation of risk factors after the first year

Long term HbA1c progression (until treatment intensification) was based on the UKPDS Outcomes Model 2 equation.[33] For SBP, HDL, LDL and weight (i.e., BMI) no long-term changes were assumed, i.e., the risk factor values (after year 1) were assumed to remain constant until treatment intensification. Constant risk factors (SBP, HDL, LDL and weight (i.e., BMI)) after year 1 up to treatment intensification implicitly assumes no waning of relative treatment effects. The company stated that there was little long-term data on the durability of treatment effects on individual risk factors. Moreover, for HDL and LDL it was stated by the company that risk factor progression formulae show only modest changes over time. The long-term progression of risk factors (including the impact of treatment intensification) was illustrated in Figures 70-75 of the CS (for tirzepatide 10 mg and SEMA 1.0).

Long-term eGFR progression was based on data published by Zoppini et al. showing a progressive decrease in renal function over time.[27] This was preferred by the company over the UKPDS OM2 progression model for eGFR as it represents a more clinically plausible decrease over time for a range of different baseline eGFR levels (whereas the UKPDS OM2 eGFR progression formula has all patients tending towards a mean value over time).

4.2.6.2 Treatment intensification

Upon treatment intensification, i.e., switch to basal insulin therapy (primarily between years 3 and 7) the following risk factors were adjusted:

- HbA1c was assumed to decrease by a mean of 0.84%, presumably upon initiation of insulin therapy [34]
- BMI was assumed to return to baseline levels in the first year of basal insulin therapy. This approach was adopted due to the absence of data to differentiate between bodyweight responses upon initiation of insulin therapy following treatment with tirzepatide or GLP-1 RAs
- All other risk factors were assumed to return to baseline levels upon initiation of insulin therapy, as there was no evidence on durability of effect

4.2.6.3 Estimation of complications

Complications were estimated based on the risk models using patient characteristics (demographics, risk factors and complication history) as input, see Section 4.2.2.

EAG comment: The main concerns of the EAG relate to: a) assumption of constant risk factors after year 1 up to treatment intensification; b) assumptions regarding waning of relative treatment effect; c) assumptions after switching to basal insulin therapy (treatment intensification); d) treatment discontinuation assumptions; e) BMI retrieved from the NMA versus BMI calculated based on body weight; f) the "nearest neighbour" approach to handle missing inputs from the NMA for all risk factors except weight; g) assuming no change from baseline for eGFR, WBC and Hb levels.

a) The company assumed constant risk factors (i.e., no risk factor progression) for SBP, HDL, LDL and weight (i.e., BMI) after year 1 up to treatment intensification. The company argued that for most risk factor progressions other than HbA1c, only very modest changes are observed over time. As a result, the differences between the assumption of no change over time and the application of risk factor progression equations (e.g., from UKPDS OM2) is negligible whilst patients on tirzepatide or comparator therapy. According to Table 106 of the CS, assuming UKPDS OM2 risk factor progression for all risk factors would increase the ICER by roughly £2,000 (for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg). The EAG prefers assuming UKPDS OM2 risk factor progression for all risk factors.

b) The company did assume no waning of the relative treatment effect while on the initial treatment (i.e., before switching to basal insulin therapy). In response to clarification question B6, the company provided multiple scenario analyses assuming all risk factors returning to baseline after 1, 2, 3, 4 or 5 years. These analyses indicated that alternative assumptions regarding waning of relative treatment effectiveness might substantially increase the estimated ICER (see clarification response Table 14).

c) Simulated patients were assumed to switch to basal insulin therapy on intensification and remain on basal insulin therapy for the rest of the simulation. Assumptions regarding the effectiveness of basal insulin therapy were justified based on the absence of evidence. Upon initiation of insulin therapy, SBP, HDL, LDL were assumed to return to baseline levels in the company base-case. As indicated by the company in response to clarification question B6, consistently with SBP, HDL, LDL, also assuming no benefits for HbA1c and BMI (key drivers of cost effectiveness) after treatment intensification to basal insulin therapy would be the most conservative assumption. However, the company also indicated that these conservative scenarios (i.e., assuming no benefits after intensification for HbA1c or BMI respectively) would have produced ICERs between the CS base-case and the scenario analyses with no HbA1c or BMI difference which individually increased the ICER by roughly £4,000 and £7,000 respectively (for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg).

d) Patients were assumed to intensify therapy, discontinuing the initial treatment and switching to basal insulin therapy, when HbA1c levels rose above 7.5%. According to the company's response to clarification question B19, no other causes for treatment discontinuation were included. The company noted "*that changing intensification criteria had a generally modest effect on cost-effectiveness (CS Table 106) and it can be assumed that modelling discontinuation would similarly have a modest impact on cost-effectiveness provided that the a balanced approach to costs and effects was applied to the rescue medication*". The EAG disagrees with this statement given that according to Table 106 of the CS, increasing the HbA1c threshold to 8.5% and 9.5% would increase the estimated ICERs by roughly £4,000 and £5,500 (for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg) which was not considered 'modest' by the EAG. This increase in ICER might be higher if discontinuation was not restricted to patients reaching this threshold (i.e., restricted to patients on treatment with the

worst HbA1c) but also included patients discontinuing treatment for other reasons. Hence, the EAG would prefer including other causes for treatment discontinuation (than reaching the HbA1c threshold).

e) Whilst BMI was included in the NMA outputs, the company calculated change from baseline in BMI based on changes in body weight (assuming an average length of 1.68; the mean value reported for the THIN population) reported in the NMA. This was done as the NMA did not provide BMI estimates for dulaglutide 3.0 mg, dulaglutide 4.5 mg, oral semaglutide 7 mg and liraglutide 1.2 mg. According to clarification response Table 15, comparing BMI retrieved from the NMA and BMI calculated based on body weight (from the NMA), this was not a conservative assumption for all comparisons. The EAG prefers using the BMI directly retrieved from the NMA, when available, and using the BMI calculated based on body weight (from the NMA) only for dulaglutide 3.0 mg. dulaglutide 4.5 mg, oral semaglutide 7 mg and liraglutide 1.2 mg (i.e., when BMI was not available from the NMA).

f) To handle missing inputs from the NMA for all risk factors except weight (and thus BMI as used in the CS base-case), the company used a "nearest neighbour" approach. Where inputs were missing, the corresponding input from the same compound was used as a proxy, wherever possible using higher (more effective) doses of comparator. In general, the EAG considered this approach to be conservative, except for HDL and LDL 'imputed' for dulaglutide 3.0 mg and dulaglutide 4.5 mg. Here, dulaglutide 1.5 mg was used as a proxy and thus potentially underestimating the effectiveness of dulaglutide 3.0 mg and dulaglutide 4.5. However, considering the anticipated minimal impact of HDL and LDL on the results (based on Table 106 of the CS), this was considered a minor issue by the EAG. An alternative assumption that could have been explored by the company is assuming no difference in HDL and LDL between tirzepatide and dulaglutide.

g) In absence of evidence of a differential effect, the company assumed no change from baseline for eGFR, WBC and Hb levels (for all treatments). The company also indicated (response to clarification question B18), that the anticipated impact on the results would be minimal. The EAG believes the approach adopted by the company for eGFR, WBC and Hb levels is reasonable.

4.2.7 Adverse events

Averse events considered in the model:

- Non-severe hypoglycaemia (rate of hypoglycaemia was assumed to be zero for tirzepatide and all comparators)
- Severe hypoglycaemia (rate of hypoglycaemia was assumed to be zero for tirzepatide and all comparators)
- Nausea (NMA, Table 81 of the CS)

For basal insulin therapy, hypoglycaemia event rates were aligned with those used in the NICE 2022 health economic report used to inform NG28. The rates for severe and non-severe hypoglycaemia used in the modelling analysis were as follows:

- Mean annual severe hypoglycaemia rate 0.32 events per patient year
- Mean annual non-severe hypoglycaemia rate 3.84 events per patient year

EAG comment: The main concerns of the EAG relate to that only nausea is incorporated (hypoglycaemia only for basal insulin therapy) as an AE. Including hypoglycaemia only for basal insulin therapy might inflate the impact of discontinuing treatment (i.e., treatment intensification) and hereby potentially inducing bias favoring more effective treatments (i.e., tirzepatide). Moreover, in response to clarification question A37 it was stated: "*Clinically significant hypoglycaemia occurred in 10 to 14 % (0.14 to 0.16 events/patient year) of patients when tirzepatide was added to sulphonylurea and in 14 to 19 % (0.43 to 0.64 events/patient year) of patients when tirzepatide was added to basal insulin (very common)*". As illustrated in clarification response Tables 20-22, incorporating additional AEs would potentially increase the estimated ICER. Therefore, the EAG would prefer to include all relevant AEs (also including hypoglycaemia and GI AEs such as diarrhoea and vomiting).

4.2.8 Health-related quality of life

A baseline utility value of 0.815 for having T2D without complications was used. (Dis)utility values for treatment effects, complications, AEs, and administration routes were applied.

4.2.8.1 Health-related quality of life data identified in the review

Utilities used in the model were selected to be consistent with the NG28 HE analysis.[32] If not available, they were taken from other sources including the SLR. Twenty-one articles were included in the update of the Beaudet et al 2014[35] and Valentine et al 2022[25] reviews, which are described in Appendix H. If multiple utility values were available, the company stated that "*utilities from the review were prioritised based on whether they were derived using the EQ-5D instrument, with preference given to UK-specific studies where available"*.

4.2.8.2 Health state utility values

For each complication or AE, disutilities were applied using an additive approach. No age-adjustment was used in the base-case analysis. No-injection related disutility was used for all injectable formulas. A utility value was applied when a drug was administered orally or injected with the tirzepatide and dulaglutide device.

In the first year of treatment, a utility value was applied to bodyweight reductions. In each subsequent year, the impact of body weight on QoL was captured using a disutility for each unit of BMI above 25.

An overview of all (dis)utility values and sources is provided in Table 4.7. A justification was provided for utility values not derived from NG28 HE analysis.

Table 4.7: Health state utility values

Health state	Utility value	Reference	Justification
BASELINE			
T2D with no complications	0.815	NICE HE report 2022[32]	
TREATMENT EFFECTS			
Bodyweight reduction (first year on GLP-1 RA therapy)	Depending on percentage weight reduction	Linear interpolation of utility values associated with percentage weight loss from Boye et al 2022[36]	UK-specific utilities for weight change in a population with T2D and obesity

Health state	Utility value	Reference	Justification
Overweight (second year of GLP-1 RA therapy and onwards)	-0.0061 for each unit of BMI over 25	Bagust and Beale 2005[37] and NICE HE report 2022[32]	Utility associated with BMI (or body weight) state previously used in the NICE Guideline 28 modelling analysis
MACROVASCULAR EVENTS			
Myocardial infarction event (first and subsequent years)	-0.055	NICE HE report 2022[32]	
Stroke event (first and subsequent years)	-0.164	NICE HE report 2022[32]	
Ischemic heart disease (first and subsequent years)	-0.090	NICE HE report 2022[32]	
Congestive heart failure (first and subsequent years)	-0.108	NICE HE report 2022[32]	
MICROVASCULAR EVENTS			
Foot ulcer event	-0.170	NICE HE report 2022[32]	
Lower extremity amputation event	-0.280	NICE HE report 2022[32]	
Lower extremity amputation (subsequent years)	-0.122	Bagust and Beale 2005[37]	EQ-5D derived utility from a population that included UK-based patients
Blindness	-0.074	NICE HE report 2022[32]	
Macular oedema (first year)	-0.047	Mitchell et al 2012[38] assumed, corresponding to best corrected visual acuity change from 76-85 to 66-75	EQ-5D health scores were converted into utility scores using preferences from a UK population survey
Macular oedema (subsequent years)	0	Assumed	
Neuropathy	-0.066	Shao et al 2019[39]	Only utility identified matching the neuropathy/SPSL endpoint, derived using HUI-3 instrument

Health state	Utility value	Reference	Justification
RENAL COMPLICATIONS			
KDIGO CKD eGFR Stage 1	0	Assumed	Stage 1 eGFR is essentially asymptomatic
KDIGO CKD eGFR Stage 2	0	Assumed	Stage 2 eGFR is essentially asymptomatic
KDIGO CKD eGFR Stage 3	−0.004	Assumed based on Nauck et al 2019[40]	EQ-5D derived utility from a population that included UK-based patients
KDIGO CKD eGFR Stage 4	−0.004	Assumed based on Nauck et al 2019[40]	EQ-5D derived utility from a population that included UK-based patients
KDIGO CKD eGFR Stage 5	−0.164	NICE HE report 2022[32]	
ADVERSE EVENTS			
Severe hypoglycaemic event	-0.062	NICE HE report 2022[32]	
Non-severe hypoglycaemic event	-0.005	Evans et al 2013[41]	Time trade off derived utility from a population that included UK-based patients, frequently used in published cost effectiveness studies
Nausea	-0.04	Matza et al 2007[42]	Only utility estimate identified by literature review for patients on GLP-1 RAs experiencing nausea and vomiting AEs
DRUG ADMINISTRATION			
Injection	0	Assumption. Simplifying assumption of no injection-related disutility for all injectable formulas	
Oral administration	+0.004	NICE HE report 2022[32]	Utility was estimated based on the single daily injection utility of 0.029, divided by seven to compare with weekly injectables, derived from the NICE 2022 HE report for NICE Guideline 28
Injection with tirzepatide and dulaglutide device (first year on treatment)	+0.007	Boye et al 2019[43]	Only utility estimate available aligned with the observation that tirzepatide will be administered using the same pen device as dulaglutide, which has shown a utility benefit over the semaglutide administration device

BMI = body mass index; CKD = chronic kidney disease; eGFR = estimated Glomerular Filtration Rate; HE = health economic; GLP-1 = glucagon-like peptide-1; KDIGO = Kidney Disease Improving Global Outcomes;

Health state	Utility value	Reference	Justification
NICE = National Institute for Health and Care Excellence; RA = receptor agonist; SPSL = severe pressure sensation loss; T2D = type 2 diabetes; UK = United Kingdom			

EAG comment: The main concerns of the EAG relate to: a) the uncertainty measures and distributions applied to utility values; b) T2D utility value and methods for age-adjustment; c) the utility estimate associated with administration of tirzepatide and dulaglutide; d) the utility value associated with weight change in the first year; e) methods for combining disutility values; and e) how utility values were selected from the SLR.

a) In a response to clarification question B22c[4], the company provided an overview of all utility values including measures of uncertainty and distributions. The EAG found that the measures of uncertainty and chosen distributions did not match with the information provided in the sources. The EAG believes these discrepancies could have an influence on the estimated uncertainty surrounding the model outcomes and might even result in unlikely outcomes (e.g., an increase in utility when certain complications are present instead of a decrease) and should therefore be corrected.

b) The company base-case uses a relatively high utility value for patients with T2D (0.815) and does not adjust utility values for older age.[3] Considering the average age of 64 (Table 75 of the CS), the UK general population utility for that age would be 0.804 (and 0.785 for patients 65-74 years old).[44] Over time, this potential overestimation will likely only increase as utility values are not adjusted for age. The EAG prefers the base-case scenario to include age-adjustment as these estimates will provide a more conservative ICER estimate.

c) Drug administration using the tirzepatide and dulaglutide device results in a higher utility than oral administration. In response to clarification question B24[4] the company states that *"With respect to the comparison with oral semaglutide, the EAG is right to point out the potential shortcomings of this approach. Unfortunately, there is no quality of life data directly comparing administration of oral semaglutide with tirzepatide or dulaglutide to inform the analysis. We have therefore run simulations assuming that there is no device utility associated with tirzepatide in the comparison with oral semaglutide (which has an administration utility of +0.004) and the results are summarized in the table below for all three doses of tirzepatide. The findings show that removing the device related utility for tirzepatide had little impact on overall cost-effectiveness"*. Including no device utility increases the ICER by roughly £500, Table 106 of the CS. The EAG prefers no device utility associated with tirzepatide or dulaglutide in the base-case analysis.

d) The model differs from the NG28 health economic report in that it adds a utility value in the first year of treatment for changes in weight. The company stated that: *"The use of two different utilities is based on observations from the literature that there the effects of weight change versus being at a specific body weight or BMI level are different in terms of quality of life (Dennett et al. 2008). Therefore in the present analysis a weight change utility was applied in year 1 of the simulations (i.e. when the changes in body weight associated with GLP-1 RA therapy were applied in the modelling analysis)"* An alternative option would be to only incorporate a utility decrement for higher BMI values, as was done in NG28 HE report,[32] which would result in the tirzepatide strategy to gain less QALYs compared to the comparators. This scenario was explored in Table 27 response to clarification question B25a[4] and shows increases in the ICER of around £600 as compared to the company base-case. The EAG therefore prefers

this conservative scenario to be the base-case and to add the utility associated with BMI change as a scenario analysis.

e) The company base-case uses an additive method for combining multiple (dis)utility values, for example when a patient experiences both a myocardial infarction and a foot ulcer at the same time, the associated disutilities are added. Although the best method to combine multiple disutility values is still debated, the multiplicative method is considered to be the best approach overall and more conservative than the additive method.[45] The impact this method was explored in sensitivity analysis (increasing the ICER by roughly £3,000, Table 106 of the CS). Given the abovementioned, the EAG prefers the base-case scenario to include the multiplicative method of combining utility values.

f) The company did not state in the CS whether any of the identified studies of the SLR could be used to inform utility values in the economic model, and, if multiple values were available, how the appropriate study was selected. Instead, they stated "*Health-related quality-of-life utility data for the modelling analysis were principally selected to be consistent with the 2022 health economic analysis to support NG28, as it was assumed that NICE would consider these estimates to be appropriate for T2D patients in clinical practice in England (and given that the literature review did not provide more suitable estimates for use in the present analysis).*" as response to clarification question B22a.[4] Pooling of utility values was not performed because of the heterogeneous nature of QoL data.

4.2.9 Resources and costs

The cost categories included in the model were treatment costs (including treatment initiation and administration costs), and complication and AE costs.

Unit costs were based on NHS reference prices 2019/2020, NHS Electronic Drug Tariff costs 2022, and Personal Social Services Research Unit (PSSRU) Unit Costs Database of Health and Social Care Professionals 2020/2021. A healthcare payer perspective was adopted. No indirect costs were included.

4.2.9.1 Resource use and costs data identified in the review

An SLR was conducted (Appendix I of the CS) to identify relevant studies evaluating health state unit costs and resource use.[12] The SLR identified five studies reporting UK relevant resource use and cost information. The company did not summarise in the CS whether any of the identified studies could be used to inform cost and resource use in the economic model. The costs for neuropathy and renal complications (Kidney Disease Improving Global Outcomes (KDIGO) chronic kidney disease (CKD) eGFR Stage 4) were not identified via the SLR.

4.2.9.2 Treatment costs

The model considered three, QW doses of tirzepatide (5 mg, 10 mg and 15 mg in pre-filled pens). The treatment cost associated with all doses of tirzepatide is given as [redacted], [redacted] and [redacted] per week, giving an annual cost of [redacted], [redacted] and [redacted] per patient for tirzepatide 5, 10, and 15 mg respectively.

Annual costs for comparators were £955.52 for QW dulaglutide (1.5 mg, 3.0 mg and 4.5 mg in pre-filled disposable injections), £955.52 for QW semaglutide (0.5 mg and 1.0 mg in pre-filled disposable injections), £955.00 for oral semaglutide (7 mg and 14 mg), £955.49 for 1.2 liraglutide (pre-filled disposable injection), £1,433.24 for 1.8 liraglutide (pre-filled disposable injection), and £185.84 for

basal insulin (based on 40 insulin units per day according to the World Health Organisation (WHO) daily dose estimate of insulin glargine).

For all treatments, initiation costs and cost of needles were included, as well as initiation costs in the first year of therapy. Metformin was included as background therapy for all treatments (£40.18 per year). Costs of self-monitoring of blood glucose were included for basal insulin therapy.

4.2.9.3 Health state costs

No costs are associated with the T2D without complications health state.

4.2.9.4 Event costs

Table 4.8 outlines the annual costs included in the model for each complication or AE.

Table 4.8: Health state costs

Complication/adverse event	Costs (£, 2021 value)	Reference
Myocardial infarction event (first year)	8,678	NICE HE report 2022[32]
Myocardial infarction event (subsequent years)	2,157	NICE HE report 2022[32]
Stroke event (first year)	9,333	NICE HE report 2022[32]
Stroke event (subsequent years)	2,223	NICE HE report 2022[32]
Ischemic heart disease (first year)	12,565	NICE HE report 2022[32]
Ischemic heart disease (subsequent years)	2,209	NICE HE report 2022[32]
Congestive heart failure (first year)	4,929	NICE HE report 2022[32]
Congestive heart failure (subsequent years)	2,891	NICE HE report 2022[32]
Foot ulcer (first year)	3,731	NICE HE report 2022[32]
Foot ulcer (subsequent years)	0	Assumption
Amputation (first year)	14,473	NICE HE report 2022[32]
Amputation (subsequent years)	4,022	NICE HE report 2022[32]
Blindness (first year)	3,717	NICE HE report 2022[32]
Blindness (subsequent years)	1,408	NICE HE report 2022[32]
Macular oedema	681	NHS reference costs 2019/2020
Neuropathy/SPSL	293	Hunt et al 2017[46]
KDIGO CKD eGFR Stage 1	0	Assumption
KDIGO CKD eGFR Stage 2	0	Assumption
KDIGO CKD eGFR Stage 3	0	Assumption
KDIGO CKD eGFR Stage 4	465	Kent et al 2015[47]
KDIGO CKD eGFR Stage 5	21,541	NICE HE report 2022[32]
Nausea	0	Assumption
Non-severe hypoglycaemic event	0	NICE HE report 2022[32]
Severe hypoglycaemic event	384	NICE HE report 2022[32]

CKD = chronic kidney disease; eGFR = estimated Glomerular Filtration Rate; HE = health economic; KDIGO = Kidney Disease Improving Global Outcomes; NICE = National Institute for Health and Care Excellence; SPSL = severe pressure sensation loss

EAG comment: The main concerns of the EAG relate to: a) inflation of costs to present day values, b) discrepancies between costs mentioned in the source and the CS, c) costs associated with nausea, d) T2D health state costs, and e) how cost values were selected from the SLR.

a) The company did not state how costs were inflated to present day values. Furthermore, costs were inflated to different years. Treatment costs were 2022 values while complication costs were 2021 values. The EAG suggests inflating all costs to 2022 values.

b) There were discrepancies between the costs stated in the CS and the original publications for cost associated with foot ulcer and neuropathy. For foot ulcer in year 1, the CS states a cost of £3,620[3], while the original publication lists a cost of £,3520 (both 2020 values).[32] Costs associated with neuropathy are listed as £293 per year (2021 value),[3] while the original publication states a cost of £968 per year for neuropathy.[46] According to the EAG, these discrepancies should be resolved.

c) An assumption was made of £0 costs related to nausea; however, nausea can potentially be related with increased healthcare costs. Examples of nausea costs are used in previous NICE STAs, e.g., focusing on cancer treatments, although these are likely not applicable to T2D treatments. The EAG would like to suggest exploring the effect of costs associated with nausea in a scenario analysis.

d) No specific T2D health state costs were included. The company's reasoning for this choice is that *"The approach to cost estimation is aligned with previous publications in this area and approach recently used by NICE in the preparation of NG28. No specific health state costs for T2D, or other costs that would be the same across all treatment arms were included. The rationale for this was simply that inclusion of any such costs would not have a significant impact on the treatment decision, given that this would apply across all treatment arms and could only have an impact if there were substantial differences between treatments in the time spent living in the model, which is not the case in the present analysis. In line with this logic, no additional cost scenario simulations have been run as changing the annual costs associated with T2D management would not impact cost-effectiveness."* The EAG would like a scenario analysis where the impact of including T2D health state costs is explored.

e) The company did not state in the CS whether any of the identified studies of the SLR could be used to inform costs and resource use in the economic model, and, if multiple sources were available, how the appropriate study was selected.

4.2.10 Severity

The company stated that "*no severity weights were used in the evaluation of quality-adjusted life expectancy in the present analysis*".

EAG comment: No comment.

4.2.11 Uncertainty

According to the company sensitivity analyses showed that the ICERs are robust to changes in the modelling parameters.

EAG comment: The company did not explore the impact of all input parameters on the estimated ICERs, hence, the EAG does not agree with this statement that the ICERs are robust to changes in the modelling parameters. Moreover, results provided by the company in Table 106 of the CS as well as the clarification responses indicated that changes in input parameters can have a substantial impact on the estimated ICERs.

5. COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The CS base-case cost effectiveness results (probabilistic) indicated that tirzepatide 5 mg is both more effective and more costly than the comparators amounting to ICERs ranging between [redacted] per QALY gained (see Table 5.1). Tirzepatide 10 mg was more effective in all comparisons and more costly in all comparisons but the one with liraglutide, with ICERs ranging between [redacted] per QALY gained, and tirzepatide 10 mg dominating liraglutide (see Table 5.2). A similar pattern of results was projected for tirzepatide 15 mg, which was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg and had ICERs ranging between [redacted] per QALY gained versus the other comparators.

Table 5.1: Company's base-case cost effectiveness results for tirzepatide 5 mg

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)*	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 5 mg	[redacted]	13.132	[redacted]	[redacted]	--	[redacted]	--
Dulaglutide 1.5 mg	[redacted]	13.053	[redacted]	[redacted]	+0.078	[redacted]	[redacted]
Semaglutide 0.5 mg	[redacted]	13.074	[redacted]	[redacted]	+0.057	[redacted]	[redacted]
Oral semaglutide 7 mg	[redacted]	13.030	[redacted]	[redacted]	+0.101	[redacted]	[redacted]
Liraglutide 1.2 mg	[redacted]	13.022	[redacted]	[redacted]	+0.110	[redacted]	[redacted]

ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year
*For tirzepatide versus comparator

Table 5.2: Company's base-case cost effectiveness results for tirzepatide 10 mg

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)*	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 10 mg	[redacted]	13.138	[redacted]	--	--	--	--
Dulaglutide 3.0 mg	[redacted]	13.063	[redacted]	[redacted]	+0.075	[redacted]	[redacted]
Semaglutide 1.0 mg	[redacted]	13.092	[redacted]	[redacted]	+0.046	[redacted]	[redacted]
Oral semaglutide 14 mg	[redacted]	13.078	[redacted]	[redacted]	+0.060	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	13.025	[redacted]	[redacted]	+0.113	[redacted]	[redacted]

ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year
*For tirzepatide versus comparator

Table 5.3: Company's base-case cost effectiveness results for tirzepatide 15 mg

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)*	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)
Tirzepatide 15 mg	[redacted]	13.165	[redacted]	--	--	--	--
Dulaglutide 4.5 mg	[redacted]	13.087	[redacted]	[redacted]	+0.078	[redacted]	[redacted]
Semaglutide 1.0 mg	[redacted]	13.092	[redacted]	[redacted]	+0.073	[redacted]	[redacted]
Oral semaglutide 14 mg	[redacted]	13.078	[redacted]	[redacted]	+0.087	[redacted]	[redacted]
Liraglutide 1.8 mg	[redacted]	13.025	[redacted]	[redacted]	+0.141	[redacted]	[redacted]

ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year
*For tirzepatide versus comparator

The 95% CrIs around the improvement in quality-adjusted life expectancy associated with tirzepatide did not cross zero for any of the dosages or comparisons. Based on pairwise comparisons, there was a [redacted] probability (in the different comparisons) that tirzepatide 5 mg would be cost-effective, assuming a willingness to pay threshold of £20,000 per QALY gained versus the four comparators evaluated individually. For the 10 mg dose, the probability ranged between [redacted] and for the 15 mg dose between [redacted]

Fully incremental analyses were not provided.

Overall, the technology is modelled to affect QALYs by:

- Reductions in diabetes-related complications associated with reductions in HbA1c and BMI

Overall, the technology is modelled to affect costs by:

- Additional treatment costs associated with tirzepatide
- Reductions in diabetes-related complication costs (greatest cost savings were associated with CV events avoided)

EAG comment: The main concerns of the EAG relate to: a) disaggregated outcomes, b) likely inappropriate PSA, c) no fully incremental analysis was provided, d) presentation of disaggregated costs likely does not allow calculation of costs with any confidential comparator prices, and e) most benefits accrued in the modelled versus observed.

a) Disaggregated outcomes were not comprehensively presented. Fully disaggregated life years (LYs) are not available from the model, which is acceptable. The company upon request provided disaggregated QALY decrements (i.e., the difference in QALYs per patient from the end of the model horizon and the beginning of the model) in response to the clarification letter in Table 37. These QALY decrements were disaggregated to show where they occurred: CV complications, renal disease, neuropathy and diabetic foot complications, ocular complications, hypoglycaemia and a last category that was called "treatment-related". This latter category included QALY decrements experienced by patients caused by *"utilities for weight year in year 1, BMI state in years 2+, utilities associated with administration, and disutilities associated with nausea and vomiting"*. It made up the largest QALY loss prevented by tirzepatide when compared to semaglutide in the respective dose bands (see Table 5.4, printed in bold). This means that tirzepatide reduced QALY loss the most in the treatment-related category, which is probably caused by BMI reduction. It is noteworthy that reductions in diabetes-related complications are not contributing to the majority of QALY differences between tirzepatide and semaglutide, as the sum of QALY loss saved due to all complications is still smaller than the QALY loss saved by tirzepatide in the treatment-related category. Also noteworthy is that hypoglycaemia was the complication where tirzepatide achieved the largest QALY saving, whilst hypoglycaemia is only included as a complication after discontinuation, i.e., on basal insulin therapy.

Table 5.4: QALY decrements over patient lifetime per treatment arm

Intervention	Treatment-related*	Cardiovascular complications	Renal disease	Neuropathy and diabetic foot complications	Ocular complications	Hypoglycaemia
QALY decrements for tirzepatide 5 mg versus comparators						
Tirzepatide 5 mg	-0.384	-0.362	-0.018	-0.369	-0.046	-0.358
Dulaglutide 1.5 mg	-0.430	-0.368	-0.018	-0.375	-0.047	-0.376
Semaglutide 0.5 mg	-0.426	-0.367	-0.018	-0.374	-0.047	-0.373
Oral semaglutide 7 mg	-0.429	-0.368	-0.018	-0.377	-0.047	-0.388
Liraglutide 1.2 mg	-0.435	-0.368	-0.018	-0.376	-0.047	-0.384
Incremental QALYs tirzepatide 5 mg versus semaglutide 0.5 mg	0.042	0.005	0.000	0.005	0.001	0.015
QALY decrements for tirzepatide 10 mg versus comparators						
Tirzepatide 10 mg	-0.360	-0.357	-0.018	-0.366	-0.045	-0.350
Dulaglutide 3.0 mg	-0.421	-0.366	-0.018	-0.373	-0.046	-0.368
Semaglutide 1.0 mg	-0.409	-0.364	-0.017	-0.370	-0.046	-0.364
Oral semaglutide 14 mg	-0.413	-0.366	-0.018	-0.374	-0.047	-0.376
Liraglutide 1.8 mg	-0.431	-0.368	-0.018	-0.376	-0.047	-0.38
Incremental QALYs tirzepatide 10 mg versus semaglutide 1.0 mg	0.049	0.007	-0.001	0.004	0.001	0.014
QALY decrements for tirzepatide 15 mg versus comparators						
Tirzepatide 15 mg	-0.343	-0.354	-0.017	-0.364	-0.045	-0.347
Dulaglutide 4.5 mg	-0.416	-0.366	-0.018	-0.372	-0.046	-0.364
Semaglutide 1.0 mg	-0.409	-0.364	-0.017	-0.370	-0.046	-0.364
Oral semaglutide 14 mg	-0.413	-0.366	-0.018	-0.374	-0.047	-0.376
Liraglutide 1.8 mg	-0.431	-0.368	-0.018	-0.376	-0.047	-0.380
Incremental QALYs tirzepatide 15 mg versus semaglutide 1.0 mg	0.066	0.010	0.000	0.006	0.001	0.017

Intervention	Treatment-related*	Cardiovascular complications	Renal disease	Neuropathy and diabetic foot complications	Ocular complications	Hypoglycaemia
Based upon Table 37 in the clarification response[4] BMI = body mass index; QALY = quality-adjusted life year *Treatment-related utility decrements include utilities for weight year in year 1, BMI state in years 2+, utilities associated with administration, and disutilities associated with nausea and vomiting						

b) From descriptions in the CS, the PSA appeared to be inappropriate. The EAG requested further detailed clarification on the PSA methods including a step-by-step description of the implementation of the PSA, but the company did not provide this. The company did however, clarify, that patient characteristics, treatment effects, costs, and utilities are all sampled, in addition to model coefficients, when the PSA mode is enabled. The company also stated that in the PSA, bootstrap samples are drawn from the whole cohort. This seems to contradict the statement in the CS "*For PSA, the model reported results based on a nonparametric bootstrapping approach, in which samples from 1% of the simulated population (in this case comprising 3,000 patients) were drawn 1,000 times from the full patient data set at the end of the simulation."* (Section B.3.11.2 of the CS) It therefore remains unclear whether the PSA is appropriate, as bootstrapping is not standard part of a PSA.[1] Furthermore, response to clarification question B2c suggests that stochastic first-order uncertainty was potentially included in the PSA and therefore mixed with imprecision. Doubts therefore remain about the appropriateness of the PSA.

c) No fully incremental analysis was provided. Upon request, the company provided results from what they termed a fully incremental analysis in Tables 31-33 of the clarification letter response. However, this is, in fact, not a fully incremental analysis where interventions and comparators are sorted from cheapest/most QALY providing to most expensive/least QALY providing and fully incremental ICERs are provided alongside of indications of dominated and extendedly dominated. Cost effectiveness frontiers were provided in the CS (Figures 82 and 83) and incremental net health benefits were calculated upon request – however the company did not provide net health benefits for each intervention/comparator.

d) The EAG also requested a presentation of results with cost split by treatment costs and other costs, to facilitate potential analysis using any confidential prices for comparator treatments. The company provided disaggregated costs, however, the treatment costs also included, apart from the intervention/comparator costs, the costs of background therapy and the cost of basal insulin (after intensification), which means that the analysis of potential comparator confidential prices likely cannot be performed.

e) Upon request, the company provided a Table overview showing the proportion of observed versus extrapolated LYs and QALYs in the model (Tables 39-41 of the clarification response). This showed that the vast majority of QALYs was accrued in the extrapolated (not observed) period of the model: ~94% for all tirzepatide comparisons at all dosages.

5.2 Company's sensitivity analyses

The company performed and presented the results of PSA, deterministic sensitivity analyses (DSA) as well as scenario analyses. The PSA sampled 1,000 times with replacement from a sub-sample (3,000) of simulation patients.

No full one-way sensitivity analyses, examining the impact all input parameters individually, were provided. Apart from scenarios on the time horizon and the discount rate, scenario analyses were only provided for the tirzepatide 10 mg dose in comparison with semaglutide 1.0 mg. The most impactful scenarios were:

- Assuming only a HbA1c difference between treatments (increased ICER)
- Assuming only a BMI difference between treatments (increased ICER)
- Assuming no BMI difference (increased ICER)

- No weight/BMI utilities (increased ICER)
- Intensification at HbA1c 9.5% threshold instead of 7.5% (increased ICER)
- Cause-subtracted life tables for mortality risk estimation only (decreased ICER)

EAG comment: The main concerns of the EAG relate to: a) no full one-way sensitivity analysis was conducted, and b) scenario analyses were provided only for the semaglutide comparison and also only for the 10 mg tirzepatide dose.

a) No full one-way sensitivity analysis was provided by the company where all parameters values were varied according to a specified range, but instead the company performed scenario analyses on parameters and assumptions that were pre-specified. This is a deviation from the NICE reference case, which highlights that deterministic sensitivity analyses may be useful for identifying parameters that the decision is sensitive to. This results in a missed opportunity for identify potentially influential parameters.

b) Sensitivity and scenario analyses were only provided for the semaglutide comparison and only for some selected input parameters but should be provided for all comparisons and all input parameters. In response to the clarification question B30, the company stated that this request was impracticable in the short timeframe, and that sensitivity analysis for all comparators would provide little or no additional data that would help answer the decision question. However, the EAG considers that you can only be sure that no or little additional information was provided once the analysis was done. Furthermore, sensitivity and scenario analysis were only provided for the tirzepatide 10 mg analysis (for the semaglutide comparison). The company stated that these were considered generalisable to the other dosages as similar patterns of results would be observed with respect to cost effectiveness for analogous simulations with other tirzepatide doses. However, this remains uncertain.

5.3 Model validation and face validity check

5.3.1 Face validity assessment

The face validity of the original PRIME T2D model was assessed via review by clinical and diabetes modelling experts at an Advisory Board Meeting.[48] It was unclear whether further face validity checks were undertaken for the current implementation of the model.

5.3.2 Technical verification

Internal validation of the model code was performed to ensure the model was coded correctly and could accurately reproduce the results of the studies used to develop the model.

5.3.3 Comparisons with other technology appraisals

No cross validation was provided.

5.3.4 Comparison with external data used to develop the economic model

External validation was performed where the PRIME T2D Model was used to reproduce the outcomes of published studies in T2D, including long-term outcomes and outcomes from CVOT studies. It was unclear whether further external validity checks were undertaken for the current implementation of the model.

5.3.5 Comparison with external data not used to develop the economic model

[48, 49]It was unclear whether external validity checks were undertaken for the current implementation of the model.

EAG comment:

The main concerns of the EAG relate to: a) technical verification and reproducibility not demonstrated, b) face validity checks likely not conducted for the current NICE model, c) cross-validation hampered by lack of clarity on impact of differing assumptions, and d) external validation incomplete.

a) Regarding the technical verification, the company did not convince the EAG that sufficient internal validity checks were carried out on this application of the PRIME T2D model. The company did not provide a filled in TECH-VER checklist, stating that it *"would add little additional value in the current circumstances. Most of the key areas around model verification and validation have already been addressed elsewhere in the submission (see the external code audit and validation as described in Appendix N3)"* (clarification response B32). The EAG disagrees with this statement as the EAG identified an error with only superficial testing (see response to clarification question B31). It is therefore possible that further internal validity checks would bring to light further errors or issues. In addition, and more concerning, the EAG observed major issues with reproducibility. When attempting to reproduce the company base-case using the JSON files (and settings) provided by the company, the estimated HRQoL appeared to be on a different scale and was approximately 1% of that reported in the CS and obtained in the web-based dashboard for tirzepatide (e.g., [redacted] QALYs for tirzepatide 5 mg in the EAG simulation compared to [redacted] QALYs in the CS). In addition, HRQoL for the control treatment (both semaglutide and dulaglutide were tested) was much lower in the EAG's simulations than the estimated HRQoL presented in the report (e.g., in one simulation [redacted] QALYs for dulaglutide 1.5 mg in the EAG simulation compared with [redacted] QALYs in the CS), leading to a very different incremental QALY than reported in the CS. Costs could also not be reproduced but the error was, with a difference of about 30% of the costs reported in the CS. It is unclear what causes these discrepancies, as most of the inputs appear to be aligned. However, the company have not provided a full overview of inputs and the EAG found inputs in the JSON files that could not be found anywhere in the CS. Furthermore, the EAG was unable to find how the BMI-related utility was implemented in the model, which is one of the most influential model inputs and therefore especially concerning. This demonstrates that to a certain extent, there is still a black box character to this model.

b) The company undertook face validity checks using literature in the development of the model, three Advisory Boards to get expert clinical and health economic input for the development of the model on the statistical approach for handling uncertainty, modelling complications and risk factors, and approaches to risk estimation, and the PRIMA 2022 review. It is still unclear to the EAG whether the Advisory Board (in 2019) were held to inform this particular application of the model and thus the face validity of the current implementation remains unclear.

c) Upon request the company provided an overview of other relevant NICE appraisals, including NG28 (2022). The company also explained how differences in approaches led to different results. The company considered the risk equations to be the main defining feature of the different model structures (all appraisals used different models). Inputs, including intensification thresholds, were broadly similar between appraisals and estimation of progression of risk factors and AEs were broadly aligned as well. In terms of complication risk

estimation, the present submission differs in that it uses UKPDS OM2 risk formulae with BRAVO model risk formulae (model averaging), while the 2022 NICE evaluation used UKPDS OM2 in combination with HRs from CVOTs. The company stated, however, that it was difficult to comment on how the present approach would compare directly with a CVOT-calibrated UKPDS OM2 modelling approach. Whilst the PRIME T2D model has been validated against several CVOTs without calibration, the 2022 NICE HE evaluation relied on a calibration approach using data from different CVOTs in combination. The company stated: *"The use of unadjusted hazard ratios from multiple CVOTs in a long-term cost-effectiveness analysis has considerable potential to skew the outcomes if these [heterogeneity between studies] challenges are not appropriately addressed."* (clarification response B.34.b).

d) External validation of the modelled outcomes against SURPASS were not provided and the company justified this stating that given that model inputs were derived from SURPASS, model estimates at one year would be similar to observations in SURPASS. The EAG would have liked to see this verified. Regarding external validation with alternative data, the company highlighted the paucity of suitable long-term study data for model validation. A literature review to identify new studies for additional validation analyses was considered by the company as impracticable based on short timelines.

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

Based on all considerations in the preceding Sections of this EAG report, the EAG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments suggested by the EAG form the EAG base-case and were subdivided into three categories (derived from Kaltenthaler et al 2016)[50]:

- Fixing errors (FE) (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (FV) (correcting the model where the EAG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (MJ) (amending the model where the EAG considers that reasonable alternative assumptions are preferred)

6.1.1 EAG base-case

Adjustments suggested by the EAG, to derive the EAG base-case (using the CS base-case as starting point) are listed below. The EAG could not reproduce the company's base-case results locally (see Section 5.3 of this report) with JAVA model files provided by company. Moreover, the web version of the model only had limited flexibility to make model adjustments. Therefore, instead of implementing the EAG base-case, the EAG highlights suggested adjustments for the company to implement and produce the EAG base-case and scenario analyses. These analyses should show how individual adjustments impact the results plus the combined effect of all adjustments simultaneously, resulting in the EAG base-case. Additionally, the issue of the EAG being unable to reproduce the company's base-case results locally should be resolved.

6.1.1.1 Fixing errors (FE)

1. Resolve discrepancies in the uncertainty measures and distributions related to utility values and costs listed in the CS and those listed in the original sources. Utility discrepancies are listed in 4.2.8 EAG comment a, and 4.2.9 EAG comment b, as well as key issue 12.
   The impact of this adjustment on the cost effectiveness is unknown.

6.1.1.2 Fixing violations

2. Incorporate treatment strategies in the economic model that reflect clinical practice (including the possibility for individual patients to switch between treatment dosages). See Section 4.2.4 EAG comment a, as well as key issue 3.
   The impact of this adjustment on the cost effectiveness is unknown.

3. Incorporate all comparators described in the final scope. See Section 4.2.4 EAG comment b.
   The impact of this adjustment on the cost effectiveness is unknown.

4. Incorporate all relevant AEs (also including hypoglycaemia and GI AEs such as diarrhoea and vomiting). See Section 4.2.7 EAG comment as well as key issue 10.
   This adjustment will likely increase the estimated ICER (might depend on the comparator). The magnitude of the impact is unclear as no analyses were provided incorporating all abovementioned AEs simultaneously (see also clarification response Tables 20-22).

5. Incorporate age-adjustment for utility values, ensuring that the utility does not exceed the age-matched general population utility. See Section 4.2.8 EAG comment b.
   This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as the exact implementation of the "QALY age-adjustment based on Ara and Brazier" analyses performed by the company (CS Table 89) is unclear (amongst others whether the T2D utility exceeds the age-matched general population utility).

6. Inflating all costs to the same price year, preferably 2022 values. See Section 4.2.9 EAG comment a.
   The impact of this adjustment on the cost effectiveness is unknown.

6.1.1.3 Matters of judgement

7. Assuming UKPDS OM2 risk factor progression for all risk factors. See Section 4.2.6 EAG comment a as well as key issue 8.
   This adjustment will likely increase the estimated ICER.

8. Assuming additional causes for treatment discontinuation (than reaching the HbA1c threshold). See Section 4.2.6 EAG comment d, as well as key issue 9.
   This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as including additional causes for treatment discontinuation (than reaching the HbA1c threshold) was not explored by the company.

9. Using the BMI directly retrieved from the NMA, when available, and using the BMI calculated based on body weight (from the NMA) only for dulaglutide 3.0 mg. dulaglutide 4.5 mg, oral semaglutide 7 mg and liraglutide 1.2 mg (i.e., when BMI was not available from the NMA). See Section 4.2.6 EAG comment e.
   This adjustment will likely increase the estimated ICER (might depend on the comparator). Using change in BMI values directly from the NMA would increase the estimated ICER by roughly £1,300 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).

10. Assume no device utility associated with tirzepatide or dulaglutide in the base-case analysis. See Section 4.2.8 EAG comment c.
    This adjustment will likely increase the estimated ICER (might depend on the comparator). Assuming no device utility would increase the estimated ICER by roughly £500 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).

11. Assume a multiplicative approach for utility values. See Section 4.2.8 EAG comment e.
    This adjustment will likely increase the estimated ICER. Assuming a multiplicative approach to combine utility values would increase the estimated ICER by roughly £3,000 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).

6.1.2 EAG exploratory scenario analyses

The EAG suggested the following exploratory scenario analyses, conditional on the EAG base-case, to explore the impact of alternative assumptions.

12. Using the CORE Diabetes model (consistent with NG28). See Section 4.2.2 EAG comments a and b, as well as key issue 6.

The impact of this adjustment on the cost effectiveness is unknown.

13. Selecting single predictive models based on the best match of the derivation cohort to the decision problem (as requested in clarification question B4c). See Section 4.2.2 EAG comment c, as well as key issue 7.
    The impact of this adjustment on the cost effectiveness is unknown.

14. Assuming the population characteristics from the SURPASS-2 trial (instead of based on THIN second intensification cohort). See Section 4.2.3 EAG comment a.
    This adjustment will likely decrease the estimated ICER. Assuming SURPASS-2 inputs (cohort characteristics and treatment effects) would decrease the estimated ICER by roughly £2,000 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 107).

15. Assuming waning of the relative treatment effect while on the initial treatment. See Section 4.2.6 EAG comment b, as well as key issue 8.
    This adjustment will likely increase the estimated ICER. Assuming clinical benefits for 1 year only would increase the estimated ICER by roughly £21,000 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (clarification response Table 14).

16. Assuming no difference in HDL and LDL between tirzepatide and dulaglutide. See Section 4.2.6 EAG comment f.
    This adjustment will likely increase the estimated ICER (might depend on the comparator). Assuming no serum lipids difference would increase the estimated ICER by roughly £200 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).

17. Only assume a utility decrement for higher BMI values, as was done in NG28 HE report. See Section 4.2.8 EAG comment d.
    This adjustment will likely increase the estimated ICER. Assuming no body weight change utility would increase the estimated ICER by roughly £600 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).

18. Exploring the impact of including costs associated with nausea in a scenario analysis. See Section 4.2.9 EAG comment c.
    This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as including costs associated with nausea was not explored by the company.

19. Exploring the impact of including T2D health state costs. See Section 4.2.9 EAG comment d.
    This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as including T2D health state costs was not explored by the company.

6.2 Impact on the ICER of additional clinical and economic analyses undertaken by the EAG

The EAG could not reproduce the company's base-case results locally (see Section 5.3 of this report) with JAVA model files provided by company. Moreover, the web version of the model only had limited flexibility to make model adjustments. Therefore, instead of implementing the EAG base-case, the EAG highlights suggested adjustments for the company to implement and produce the EAG base-case and scenario analyses. Consequently, the EAG was unable to assess the impact of EAG suggested analyses (other than the descriptions provided in Section 6.1 of this report).

6.3 EAG's preferred assumptions

The EAG preferred base-case is described in Section 6.1 of this report.

6.4 Conclusions of the cost effectiveness section

The CS and response to clarification provided sufficient details for the EAG to appraise the literature searches conducted to identify economic, HRQoL and cost data on tirzepatide for the treatment of T2D. Searches were conducted in January-February 2022. Searches were transparent and reproducible, and comprehensive strategies were used. A good range of databases were searched. Overall, the EAG has no major concerns about the literature searches conducted.

The company's cost effectiveness assessment partly complied with the NICE reference case. The deviation from the NICE reference case related to the type of economic evaluation as the incremental analyses were missing. The most prominent issues highlighted by the ERG are discussed below. These issues were listed as key issues in Section 1.5 and suggestions for analyses to (partly) examine the potential impact of these issues were provided in Sections 6.1.1 and 6.1.2 of this report.

Firstly, the EAG did question the company's modelling approach. This includes 1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the UKPDS OM2 model or CORE Diabetes Model that were used for (updating of the) NG28 focusing on the management of T2D and 2) the selected model type, described as a "discrete time event" model instead of commonly used model types such as a DES or individual-patient state transition model. Moreover, the exact technical implementation of the model was not clear to the EAG, which is particularly problematic because of the deviation from commonly used model types. Similarly, compelling justification was missing for the company's model averaging approach as well as the appropriateness and applicability of the selected predictive models to estimate the risk of complications in patients with T2D.

Secondly, the population considered in the CS was adults with T2D that is inadequately controlled with three or more antidiabetic agents, which is not aligned with the population from the SURPASS trials or the expected UK clinical use. Moreover, the company base-case included three different maintenance doses of tirzepatide: 5 mg, 10 mg, or 15 mg. Comparisons were made within each recommended maintenance dose step, and not between recommended maintenance dose steps. In addition, patients were not able to move between dose steps in the model. This does not seem to reflect clinical practice.

Thirdly, the QALY gains are predominantly accrued after the first year and mostly likely related to utilities for weight. Hence the extrapolation of (treatment) effectiveness is an important aspect of the model. The company made a simplifying assumption of constant risk factors (i.e., no risk factor progression) for SBP, HDL, LDL and weight (i.e., BMI) after year 1 up to treatment intensification. Moreover, the company did assume no waning of the relative treatment effect while on the initial treatment (i.e., before switching to basal insulin therapy). Additionally, patients were assumed to switch to basal insulin therapy only in case HbA1c levels rose above 7.5%, i.e., no other reasons (e.g. drug intolerance, patient preferences) for treatment discontinuation were included in the modelling.

Fourthly, the company base-case used a relatively high utility value for patients with T2D (0.815) and did not adjust utility values for older age, potentially resulting in utility values that are higher than expected for the age-matched general population. Moreover, the EAG highlighted discrepancies in input parameters related to utility values and costs listed in the CS and those listed in the original sources.

Fifthly, the implementation of the PSA was not clear and included bootstrapping, which is not standard in PSAs. It is unclear whether all imprecision (i.e., all uncertain parameters) was taken into account in

the PSA, and whether stochastic uncertainty was removed from the PSA. Related to this, no full deterministic one-way sensitivity analyses (for all input parameters) were provided, and an opportunity was therefore missed to identify potentially influential parameters.

Finally, there remain doubts over the internal validity of the model. Model outcomes could not be reproduced by the EAG. The EAG could not find how BMI-related utilities were implemented in the model and no full overview of input parameters has been provided.

The CS base-case cost effectiveness results (probabilistic) indicated that tirzepatide 5 mg is both more effective and more costly than the comparators amounting to ICERs ranging between [redacted] per QALY gained (see Table 5.1). Tirzepatide 10 mg was more effective in all comparisons and more costly in all comparisons but the one with liraglutide, with ICERs ranging between [redacted] per QALY gained, and tirzepatide 10 mg dominating liraglutide (Table 5.2). A similar pattern of results was projected for tirzepatide 15 mg, which was projected to be cost saving (and therefore dominant) versus liraglutide 1.8 mg and had ICERs ranging between [redacted] per QALY gained versus the other comparators.

The EAG could not reproduce the company's base-case results locally with JAVA model files provided by company. Moreover, the web version of the model only had limited flexibility to make model adjustments. Therefore, instead of implementing the EAG base-case, the EAG highlights suggested adjustments for the company to implement and produce the EAG base-case and scenario analyses. These adjustments should be implemented transparently and the EAG should be able to reproduce these analyses. It is expected that the EAG base-case would result in substantially higher ICERs compared with the company base-case as most suggested adjustments would likely increase the estimated ICER (see Section 6.1.1), while for the remaining adjustments the impact is unknown.

There is large remaining uncertainty about the (long-term) effectiveness and relative effectiveness of tirzepatide, which can be at least partly resolved by the company by conducting further analyses. According to the EAG the current company's base-case is flawed and the EAG suggested adjustments could conceivably change, and most likely increase, the ICER. Moreover, the current assessment does not provide an appropriate estimation of the comparators listed in the scope. Therefore, the EAG believes that the CS base-case does not represent an unbiased ICER of tirzepatide compared with relevant comparators, as would be used in clinical practice.

7. REFERENCES

[1] Corro Ramos I, Hoogendoorn M, Rutten-van Molken MPMH. How to address uncertainty in health economic discrete-event simulation models: an illustration for chronic obstructive pulmonary disease. Med Decis Making 2020; 40(5):619-632

[2] Eli Lilly and Company. A study of tirzepatide (LY3298176) compared with dulaglutide on major cardiovascular events in participants with type 2 diabetes (SURPASS-CVOT). NCT04255433. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). [accessed 25.4.22]. Available from: https://clinicaltrials.gov/ct2/show/NCT04255433/

[3] Eli Lilly and Company. *Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA): Document B - Company evidence submission*: Eli Lilly and Company, 2022. 221p.

[4] National Institute for Health and Care Excellence. *Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: Response to request for clarification from the ERG*. London: NICE, 2023. 151p.

[5] National Institute for Health and Care Excellence. *Tirzepatide for treating type 2 diabetes: Final scope [Internet]*. London: NICE, 2022 [accessed 9.1.23]. 4p. Available from: https://www.nice.org.uk/guidance/gid-ta10835/documents/final-scope

[6] Eli Lilly and Company. *Summary of product characteristics: Mounjaro 15 mg solution for injection in pre-filled pen [Internet]*, 2022 [accessed 9.1.23]. 26p. Available from: https://mhraproducts4853.blob.core.windows.net/docs/4a1f72182b7ea0a07ff89549cc1e4c36b3b3a233

[7] National Institute of Health and Care Excellence. *Type 2 diabetes in adults: management [NG28] [Internet]*. London: NICE, 2022 [accessed 18.2.22] Available from: https://www.nice.org.uk/guidance/ng28/

[8] Mohammed K. Personal communication: Email to EAG, 2 February 2023.

[9] Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2022; 45(11):2753-2786

[10] National Institute for Health and Care Excellence. *Diabetes - type 2: GLP-1 receptor agonists [Internet]*. London: NICE, 2023 [accessed 6.4.23] Available from: https://cks.nice.org.uk/topics/diabetes-type-2/prescribing-information/glp-1-receptor-agonists/

[11] British National Formulary (BNF): Dulaglutide [Internet]. London: BNF Publications, [accessed 19.4.23]. Available from: https://bnf.nice.org.uk/drugs/dulaglutide/

[12] Eli Lilly and Company. *Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA): Appendices*: Eli Lilly and Company, 2022. 514p.

[13] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline statement. J Clin Epidemiol 2016; 75:40-6

[14] Canadian Agency for Drugs and Technologies in Health. *PRESS - Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E) [Internet]*. Ottawa: CADTH, 2016 [accessed 18.01.23] Available from: https://www.cadth.ca/resources/finding-evidence/press

[15] National Institute for Health and Care Excellence. *Single Technology Appraisal: company evidence submission template [Internet]*. London: NICE, 2015 [accessed 18.01.23]. 28p. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/specification-for-company-submission-of-evidence-2015-version.docx

[16] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 18.01.23] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[17] Golder S, Peryer G, Loke YK. Overview: comprehensive and carefully constructed strategies are required when conducting searches for adverse effects data. J Clin Epidemiol 2019; 113:36-43

[18] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019; 366:l4898

[19] Eli Lilly and Company. *Efficacy and safety of LY3298176 once weekly versus insulin glargine in patients with type 2 diabetes and increased cardiovascular risk. SURPASS-4. (Clinical Study Report) [PDF provided with the company's submission]*, 2021. 5827p.

[20] Eli Lilly and Company. *A randomized, phase 3, open-label trial comparing the effect of LY3298176 versus titrated insulin degludec on glycemic control in patients with type 2 diabetes. SURPASS-3. (Clinical Study Report) [PDF provided with the company's submission]*, 2021. 3104p.

[21] Eli Lilly and Company. A study of tirzepatide in participants with type 2 diabetes mellitus (T2DM). NCT03951753. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). [accessed 27.4.22]. Available from: https://clinicaltrials.gov/ct2/show/NCT03951753/

[22] Eli Lilly and Company. *A phase 3, randomized, open-label trial comparing efficacy and safety of tirzepatide versus semaglutide once weekly as add-on therapy to metformin in patients with type 2 diabetes. SURPASS-2. (Clinical Study Report) [PDF provided with the company's submission]*, 2021. 3625p.

[23] National Institute for Health and Care Excellence. *Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: Clarification questions* London: NICE, 2023. 46p.

[24] Frias JP, Nauck MA, Van J, Kutner ME, Cui X, Benson C, et al. Efficacy and safety of LY3298176, a novel dual GIP and GLP-1 receptor agonist, in patients with type 2 diabetes: a randomised, placebo-controlled and active comparator-controlled phase 2 trial. Lancet 2018; 392(10160):2180-2193

[25] Valentine W, Norrbacka K, Boye KS. Evaluating the impact of therapy on quality of life in type 2 diabetes: a literature review of utilities associated with treatment-related attributes. Patient Relat Outcome Meas 2022; 13:97-111

[26] Yang X, So WY, Kong AP, Ho CS, Lam CW, Stevens RJ, et al. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. Diabetes Care 2007; 30(1):65-70

[27] Zoppini G, Targher G, Chonchol M, Ortalda V, Negri C, Stoico V, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. Clin J Am Soc Nephrol 2012; 7(3):401-8

[28] Shao H, Shi L, Fonseca VA. Using the BRAVO risk engine to predict cardiovascular outcomes in clinical trials with sodium-glucose transporter 2 inhibitors. Diabetes Care 2020; 43(7):1530-1536

[29] Si L, Willis MS, Asseburg C, Nilsson A, Tew M, Clarke PM, et al. Evaluating the ability of economic models of diabetes to simulate new cardiovascular outcomes trials: a report on the Ninth Mount Hood Diabetes Challenge. Value Health 2020; 23(9):1163-1170

[30] National Institute for Health and Care Excellence. *Type 1 and 2 diabetes in adults: diagnosis and management. Economic modelling for periodontal treatment in adults with type 1 and type 2 diabetes. NICE guideline NG17, NG28. Economic model report [Internet]*. London: NICE, 2022 [accessed 10.3.23]. 33p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-on-periodontal-treatment-in-adults-with-type-1-and-type-2-diabetes-pdf-11131191037

[31] National Institute for Health and Care Excellence. *Type 2 diabetes in adults: management. Economic modelling for continuous glucose monitoring in adults with type 2 diabetes. NICE guideline NG28. Economic model report [Internet]*. London: NICE, 2022 [accessed 10.3.23]. 28p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-pdf-11013295213

[32] National Institute of Health and Care Excellence. *Type 2 diabetes in adults: management (update). Health economic model report [NG28] [Internet]*. London: NICE, 2022 [accessed 10.4.22]. 78p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845/

[33] Leal J, Alva M, Gregory V, Hayes A, Mihaylova B, Gray AM, et al. Estimating risk factor progression equations for the UKPDS Outcomes Model 2 (UKPDS 90). Diabet Med 2021; 38(10):e14656

[34] Willis M, Asseburg C, Nilsson A, Johnsson K, Kartman B. Multivariate prediction equations for HbA(1c) lowering, weight change, and hypoglycemic events associated with insulin rescue medication in type 2 diabetes mellitus: informing economic modeling. Value Health 2017; 20(3):357-371

[35] Beaudet A, Clegg J, Thuresson PO, Lloyd A, McEwan P. Review of utility values for economic modeling in type 2 diabetes. Value Health 2014; 17(4):462-70

[36] Boye KS, Matza LS, Stewart KD, Andrews H, Howell TA, Stefanski A, et al. Health state utilities associated with weight loss in type 2 diabetes and obesity. J Med Econ 2022; 25(1):14-25

[37] Bagust A, Beale S. Modelling EuroQol health-related utility values for diabetic complications from CODE-2 data. Health Econ 2005; 14(3):217-30

[38] Mitchell P, Annemans L, Gallagher M, Hasan R, Thomas S, Gairy K, et al. Cost-effectiveness of ranibizumab in treatment of diabetic macular oedema (DME) causing visual impairment: evidence from the RESTORE trial. Br J Ophthalmol 2012; 96(5):688-93

[39] Shao H, Yang S, Fonseca V, Stoecker C, Shi L. Estimating quality of life decrements due to diabetes complications in the United States: the Health Utility Index (HUI) diabetes complication equation. Pharmacoeconomics 2019; 37(7):921-929

[40] Nauck MA, Buse JB, Mann JFE, Pocock S, Bosch-Traberg H, Frimer-Larsen H, et al. Health-related quality of life in people with type 2 diabetes participating in the LEADER trial. Diabetes Obes Metab 2019; 21(3):525-532

[41] Evans M, Khunti K, Mamdani M, Galbo-Jørgensen CB, Gundgaard J, Bøgelund M, et al. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. Health Qual Life Outcomes 2013; 11:90

[42] Matza LS, Boye KS, Yurgin N, Brewster-Jordan J, Mannix S, Shorr JM, et al. Utilities and disutilities for type 2 diabetes treatment-related attributes. Qual Life Res 2007; 16(7):1251-65

[43] Boye KS, Matza LS, Stewart KD, Jordan J, Biricolti G, Del Santo S, et al. Patient preferences and health state utilities associated with dulaglutide and semaglutide injection devices among patients with type 2 diabetes in Italy. J Med Econ 2019; 22(8):806-813

[44] Janssen B, Szende A. Population norms for the EQ5D. In: Szende A, Janssen B, Cabases J, editors. *Self-reported population health: an international perspective based on the EQ5D*. Dordrecht: Springer, 2014.

[45] Ara R, Brazier J. *Comparing EQ-5D scores for comorbid health conditions estimated using five different methods (Unpublished)*, 2010. 1-20p. Available from: https://eprints.whiterose.ac.uk/11048/2/HSE_combining_data.pdf

[46] Hunt B, Ye Q, Valentine WJ, Ashley D. Evaluating the long-term cost-effectiveness of daily administered GLP-1 receptor agonists for the treatment of type 2 diabetes in the United Kingdom. Diabetes Ther 2017; 8(1):129-147

[47] Kent S, Schlackow I, Lozano-Kuhne J, Reith C, Emberson J, Haynes R, et al. What is the impact of chronic kidney disease stage and cardiovascular disease on the annual cost of hospital care in moderate-to-severe kidney disease? BMC Nephrol 2015; 16:65

[48] Pollock RF, Norrbacka K, Boye KS, Osumili B, Valentine WJ. The PRIME type 2 diabetes model: a novel, patient-level model for estimating long-term clinical and cost outcomes in patients with type 2 diabetes mellitus. J Med Econ 2022; 25(1):393-402

[49] Gæde P, Oellgaard J, Carstensen B, Rossing P, Lund-Andersen H, Parving HH, et al. Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial. Diabetologia 2016; 59(11):2298-2307

[50] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. Health Technol Assess 2016; 20(26):1-48

Single Technology Appraisal

Tirzepatide for treating type 2 diabetes [ID3938]

EAG report – factual accuracy check and confidential information check

"Data owners may be asked to check that confidential information is correctly marked in documents created by others in the evaluation before release." (Section 5.4.9, NICE health technology evaluations: the manual).

You are asked to check the EAG report to ensure there are no factual inaccuracies or errors in the marking of confidential information contained within it. The document should act as a method of detailing any inaccuracies found and how they should be corrected.

If you do identify any factual inaccuracies or errors in the marking of confidential information, you must inform NICE by **5pm on Wednesday 29 March 2023** using the below comments table.

All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

Please underline all confidential information, and separately highlight information that is submitted as 'commercial in confidence' in turquoise, all information submitted as 'academic in confidence' in yellow, and all information submitted as 'depersonalised data' in pink.

Issue 1 Major Issues

Description of problem	Description of proposed amendment	Justification for amendment	EAG response
Table numbers and page numbers of the CS referred to throughout the report seem to align with V1 of the CS rather than V2.	Please review against V2 of the CS. The correct table numbers and page numbers have also been listed in the typographical errors below.	These errors suggest that the EAG are referring to an outdated version of the CS, updated following requests from the NICE team.	Amended accordingly – for clarity, the reason for the discrepancy appears to be because of additional spacing and extra lines in Tables 85 and 86 for tirzepatide dosing in V2.
Table 1.7 (Key Issue 6) on Page 16 states: *"1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the CORE Diabetes Model that was used for NICE Guideline NG28 focusing on the management of T2D"* This statement is also mentioned throughout the report.	This issue should be revised to state: *"1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the **UK PDS model** that was used for NICE Guideline NG28 focusing on the management of T2D"*	As per the Health Economic model report associated with the published guideline, the UKPDS OM2 model was used for NICE Guideline NG28, rather than the CORE Diabetes model as suggested by the EAG.[1] The Core Diabetes Model was used, however, in the NICE update to diabetes guidelines on continuous glucose monitoring which modified both NG17 and NG28.	Amended – changed: "1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the CORE Diabetes Model that was used for NICE Guideline NG28 focusing on the management of T2D" Into: "1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such

			as the UKPDS OM2 model or CORE Diabetes Model that were used for (updating of the) NICE Guideline NG28 focusing on the management of T2D"
Page 27, Table 2.2. The treatment positioning of tirzepatide according to the CS is described as: *"When triple therapy with metformin and two other oral drugs, one of which is a GLP-1 RA, is not effective, tolerated or contraindicated… then change the GLP-1 RA to tirzepatide"*	The intended positioning is as follows and should be updated in this instance and throughout the report: *"For use in patients with T2D that is inadequately controlled with* ***three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered****"*	This statement in the EAG report is inaccurate and suggests a misunderstanding of the intended positioning of the treatment. • Tirzepatide is proposed for use in the same position as the GLP-1 RA class. Tirzepatide is proposed as an alternative to starting a GLP-1 RA and not as a replacement of a GLP-1 RA following intensification • As per the positioning of GLP-1 RA's in the pathway, the proposed positioning of tirzepatide is as a replacement of one of three oral agents (at	This has now been amended to be clearer and to acknowledge the possibility of switching on OAD to tirzepatide, as would be the case for GLP-1 RAs according to NG28.

		failure), therefore concomitant use in the pathway would be in addition to 2 oral agents. (Note: it is not proposed that one of the oral agents would be a GLP-1 RA). Please refer to Figure 2 in Document B of the CS for a diagram of the proposed position of tirzepatide in the current treatment pathway	
Page 30 states *"one would expect that the intervention would be in combination with at least three OADs, but this is not the case in the SURPASS trials, as already stated in Section 2.1: in fact, the only tirzepatide trial that included some patients and only a small minority was SURPASS-4 and this was excluded from the NMA"*	This statement should be revised or deleted to reflect that the proposed positioning of tirzepatide is: *"For use in patients with T2D that is inadequately controlled with* ***three or more antidiabetic agents, as a more efficacious option whenever GLP-1 RAs would otherwise be considered****"*	This statement in the EAG report is inaccurate and suggests a misunderstanding of the intended positioning of the treatment, as described in the row above. The NMA is in line with the positioning of tirzepatide for patients with 2 background antidiabetic agents.	This has now been amended to be clearer and to acknowledge the possibility of switching an OAD to tirzepatide, as would be the case for GLP-1 RAs according to NG28.

Page 32 of the EAG report states that NICE recommends the following doses of dulaglutide: "*0.75 mg QW as monotherapy; 1.5 mg QW as add-on therapy"*	This should be corrected to: *"0.75 mg QW as monotherapy; 1.5 mg, **3 mg or 4.5 mg** QW as add-on therapy"*	Further doses of dulaglutide are recommended by NICE as per the published information on the BNF.[2]	Corrected – there appears to be an inconsistency between different parts of the NICE website, but the BNF is probably the most reliable source. Note also that other corrections have been made to indicate that dulaglutide is to be titrated to one of three possible doses.
Page 33 of the EAG report states: *"However, as with the intervention, the EAG might tentatively suggest that the titrated comparator dose that is closest to clinical practice might be one that lies in the middle, although such a middle dose is only available for **liraglutide**."*	This should be updated to: *"However, as with the intervention, the EAG might tentatively suggest that the titrated comparator dose that is closest to clinical practice might be one that lies in the middle, although such a middle dose is only available for **liraglutide and dulaglutide**."*	As per the comment above, further doses of dulaglutide are recommended by NICE.	Not a factual inaccuracy, according to NICE recommendation (see reference mentioned above).
Page 100 of the EAG report states: *"Furthermore, injectable semaglutide 0.5 mg and oral semaglutide 7 mg as*	This statement should be removed or substantially revised as these doses were not equated to have the same	As noted by the EAG on Page 100 of the report, the clarification question response details that *"The exposure*	Removed.

well as injectable semaglutide 1.0 mg and oral semaglutide 14 mg are ***equated as having the same effect on patients with no evidence to support these statements****."*	effect. This statement from the EAG report was not made in the CS.	*after 14.0 mg oral semaglutide is equivalent to injectable 0.5 mg semaglutide.*[14, 15] *There is no evidence available to suggest that oral semaglutide (7.0 mg or 14.0 mg) has greater efficacy than 1.0 mg injectable semaglutide. These doses were therefore considered separately from the available doses of injectable semaglutide in the NMA and cost-effectiveness analysis."* This details that the doses were not equated to have the same effect. Equally, references were provided to support the statements made in the CS.[3, 4]	
Page 140 states: *"Incorporate all comparators described in the final scope."* under the 'Fixing Violations' subheading	This issue should be removed from this subheading and instead stated elsewhere within the report.	As the scope does not require the company to position exactly as the scope suggests, this is not a violation of the reference case and instead an	Not a factual inaccuracy. Not all comparators described in the final scope were incorporated in the economic analyses. Notably, not all GLP-1 RAs were incorporated in the economic analyses

		opinion of the EAG. To state it as such is incorrect.	despite being requested (clarification question B10). This is considered a violation i.e. where "the EAG considered that the NICE reference case, scope or best practice had not been adhered to" (as described in the EAG report).

Issue 2 Minor Comments

Description of problem	Description of proposed amendment	Justification for amendment	EAG response
Table 1.4; Page 15 states: *"However, the comparisons between tirzepatide and the GLP-1 RAs, SURPASS trials, the NMA and the CEA are stratified by maximum maintenance dose into 5 mg, 10 mg and 15 mg,* ***without titration being permitted****."*	Please update to: *"However, the comparisons between tirzepatide and the GLP-1 RAs, SURPASS trials, the NMA and the CEA are stratified by maximum maintenance dose into 5 mg, 10 mg and 15 mg."*	This is a misunderstanding as the SURPASS trials did permit titration, as explained in Section B.2.3.1 of Document B of the CS.	Not a factual inaccuracy. The trials required dose escalation: this is not the same as titration according to response.
Table 1.6; Page 16 states: *"Trials were included without a systematic*	Please update to: *"Trials were included without a systematic*	This statement does not note that the NMA did have requirements about the background OAD therapy,	Not a factual inaccuracy. There is a difference between one and two

assessment of heterogeneity and with an assumption that the treatment effect is independent of concomitant background OAD therapy"	*assessment of heterogeneity and with an assumption that the treatment effect is independent of concomitant background OAD therapy;* ***however studies included patients treated with an add-on to one OAD, defined as >90% of patients on metformin monotherapy, or add-on to one to two OADs with >50% of patients on metformin."***	as noted in the CS and response to clarification question A26.	OADs and variation in the second OAD.
Page 30 states: *"in fact, the only tirzepatide trial that included some patients and only a small minority was SURPASS-4 and this was excluded from the NMA , which only included SURPASS-2, where only metformin was background therapy and 3, where only a minority had SGLT-2i added to the*	Please update to: *"in fact, the only tirzepatide trial that included some patients and only a small minority was SURPASS-4 and this was excluded from the* ***main analysis of*** *NMA , which only included SURPASS-2, where only metformin was background therapy and 3, where only a minority had SGLT-2i added*	As described in Section D.8.1.3 of the CS, studies including patients on a background therapy of three OADs (e.g., SURPASS-4), were included in the sensitivity analysis of the NMA.	Amended.

metformin (see Sections 3.2 and 3.3).”	*to the metformin (see Sections 3.2 and 3.3).”*		
Page 38 states: “No date restrictions were reported.” In reference to the clinical SLR.	This statement is incorrect and should be removed.	As detailed within the search strategy tables (Table 1,2 and 3 of the Appendix D of the CS), the SLR was date limited from 1990-current.	Amended.
Page 48–52; Tables 3.6–3.8 and 3.10–3.12. The asterisk footnotes detailing significance are not correct for all trials included in the tables.	Please update as per the CS and source material referenced, in line with Table 3.9 where letters are used to denote the different significance levels for different trials.	The significance of the statistical tests has been misrepresented.	Not a factual inaccuracy – we have rechecked these and the significance levels are indicated by asterisks, the number of which we have correctly reported.
Page 65 of the EAG report states: *“The CS did not provide the results of these analyses, so the EAG have produced a summary of characteristics that were found to have a significant interaction with the treatment effect (p<0.1).”*	Please update to: *“**Document B** did not provide the results of these analyses, so the EAG have produced a summary of characteristics that were found to have a significant interaction with the treatment effect (p<0.1).”*	This statement should be revised as the results were provided as part of the CS within the reference pack.	Corrected.

Page 67 of the EAG report states: *"The exclusion of further studies did not precede but was rather executed within a **so-called** feasibility assessment aimed to create a network that was consistent in terms of both study design and study population."*	Please update to: *"The exclusion of further studies did not precede but was rather executed within a feasibility assessment aimed to create a network that was consistent in terms of both study design and study population."*	This language is unprofessional on the part of the EAG and should be removed.	Not a factual inaccuracy. However, the text has been amended to improve clarity.
Page 70 of the EAG report states: *"The company also provided a more detailed list of background therapy in the files accompanying their response to clarification[4], according to which out of the 184 arms included in the NMA, 80 (43.5%) were treated with only one OAD, 75 (40.8%) were treated with two OADs, 21 (11.4%) with three and 8 (4.4%) with four OADs. It should be noted that not all patients in each arm were treated with the same number or the*	This section should be removed or substantially revised as it is incorrect.	This section indicates that the EAG have misunderstood and mixed up the files provided as part of the responses to clarification questions A23 and A24. For further clarification on interpretation of background therapy, the EAG should refer to column G of the dataset provided as part of the response to A23. For further information on the which studies were included as part of the main	Amended to improve clarity.

same mix of background treatments. Within the arms receiving two OADs, only 23 out of the 75 arms were made out of 100% of patients reiving two treatments: either metformin plus SU or metformin plus thiazolidinedione; which are both not a relevant comparators – not a GLP-1 RA treatments. In the arms receiving three or four OADs there was no arm where all patients were receiving three or four OADs"		analysis and which were included as part of the sensitivity analysis, the EAG should refer to column L of the dataset provided as part of the response to A24.	
Page 70 of the EAG report states: *"only 23 out of the 75 arms were made out of 100% of patients reiving two treatments: either metformin plus SU or metformin plus thiazolidinedione**; which are both not a relevant comparators – not a GLP-1 RA treatments**"*	Please update to: *"only 23 out of the 75 arms were made out of 100% of patients reiving two treatments: either metformin plus SU or metformin plus thiazolidinedione"*	This statement is incorrect as metformin plus SU or metformin plus thiazolidinedione are background therapies from trials rather than comparators.	Corrected.

Page 81 of the EAG report states: *"Baseline diabetes duration varied from 0.6 - 10.1 years. Fourteen arms reported duration less than 6 years while 55 reported more than 8 years. Eight arms did not report this information."*	Please update to specify that these data are from the main analyses, as this is not currently mentioned.	These data are from the main analyses and this should be specified here.	Not a factual inaccuracy. Also, there is clearly an implication that these figures are from the main analyses, given the lack of any statement to the contrary.
Page 82 of the EAG report states: *"In addition, the company has not provided data per study nor per arm. Therefore, the EAG cannot assess how many studies reported these data and whether further analysis was feasible."*	Please remove this statement.	This information was provided as part of the response to Clarification Question A25 and as such this statement should be removed.	Amended for clarity.
Page 82 of the EAG report states: "The libraries and code used to create and execute the network models was not reported."	Please remove this statement.	This information was provided as part of the response to Clarification Question A34 and as such this statement should be removed.	Corrected.

Page 100 of the EAG report states: *"I^2 statistic values or Cochrane Q values were not provided in the CS"*	Please remove this statement.	These values were provided as part of the clarification question responses.	The EAG cannot locate this information.
Page 101 of the EAG report states: *"They go on by stating that allowing for all differences in background therapies within the NMA would reduce the number of included studies and restrict the network, which is tantamount to admitting substantial heterogeneity."*	Please update to: *"They go on by stating that allowing for all differences in background therapies within the NMA would reduce the number of included studies and restrict the network,* ***as this was not considered a treatment effect modifier this is unlikely to increase heterogeneity****."*	As mentioned in the CS, background therapy was not considered a treatment effect modifier and as such there is no need to restrict the network to account for these differences. It is therefore not correct to state that this would be associated with 'substantial heterogeneity'.	Not a factual inaccuracy. In fact, the EAG consider that it has not been demonstrated that background therapy is not a treatment effect modifier.
Page 102 of the EAG report states: *"For the rest of the studies, which make up the vast majority, the estimand was not available as this concept is relatively new and therefore the only available results based on the treatment-regimen estimand were used instead."*	The statement should be updated to: *"For the rest of the studies, which make up the vast majority, the estimand was not available as this concept is relatively new and therefore* ***the available reported results were used; the concept of estimand was only relevant for weight and HbA1c analyses****."*	As the concept of estimands is fairly new, some older studies included in the NMA did not specify whether the results were based on the efficacy estimand or the treatment-regimen estimand. Therefore, it is not correct to say that when details of the estimand were not available, results from	Amended.

		the treatment-regimen estimand were used.	
Page 102 of the EAG report states: *"Nevertheless, no data were presented for the rest of the 64 studies that were included in the NMA and did not provide the efficacy estimands."*	Please remove this statement.	All input data and baseline characteristics for the NMA were provided within the reference pack of the clarification questions, and the concept of estimands were not defined for these studies.	Amended for clarity.
Page 120, Section 4.2.6.2, the EAG report states that: *"HbA1c was assumed to decrease by a mean of 0.84%, presumably upon initiation of insulin therapy (though the latter is not clearly specified in the CS)[32]"*	This statement should be revised or deleted as it is inaccurate.	The text in the original submission has been overlooked by the EAG. Section B 3.3.2 of the CS states that: *"On intensification to basal insulin in the base case analysis:* *HbA1c was assumed to decrease by a mean of 0.84% based on the formula for "all" input parameters published by Willis et al. in 2017.[161]" 5*	Amended: changed: "HbA1c was assumed to decrease by a mean of 0.84%, presumably upon initiation of insulin therapy (though the latter is not clearly specified in the CS)[32]" Into: "HbA1c was assumed to decrease by a mean of 0.84% upon initiation of insulin therapy"

Page 121, Section 4.2.6.3, point c), the EAG report states that: *"However, the company also indicated that these conservative scenarios (i.e., assuming no benefits after intensification for HbA1c or BMI respectively) would have produced ICERs between the CS base-case and the scenario analyses with no HbA1c or BMI difference which individually increased the ICER by roughly £4,000 and £7,000 respectively (for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg)."*	This statement should be revised or deleted as it is inaccurate.	The description in the original submission has been overlooked by the EAG. The sensitivity analysis simulations described here assume no HbA1c benefit or no BMI benefit associated with tirzepatide **throughout the entire simulation** (i.e. during treatment and after intensification to basal insulin therapy). This is outlined in Table 89 of the CS, where the simulations were described as follows: *"No HbA1c difference between treatments (tirzepatide HbA1c changed matched to SEMA)"* *"No BMI difference treatments (tirzepatide BMI changed matched to SEMA)"*	Not a factual inaccuracy. This is in line with the company's response to clarification question B6 and CS Table 106

Page 127, Section 4.2.8.2, point e), where the EAG report uses reference 43 (Ara and Brazier, Value Health 2011; 14(5):740-5) to support the contention that: *"the multiplicative method is considered to be the best approach overall and more conservative than the additive method."*	The suggested superiority of the multiplicative approach is not clearly supported by the reference, and as such, this statement should be revised to align with the reference or should be deleted.	The reference does not support this statement in the context of the present analysis. The publications describes an analysis of SF-6D derived health-state utilities over a limited range. It shows a linear regression approach to combining utilities to be the best fit to the data collected and cites limitations with all of the methods used (including multiplicative and additive). The results show the five methods investigated were ranked differently in various sub-analyses. We do not know how generalisable the findings from this paper are likely to be for the combination of utilities derived using a different instrument, in a different population (i.e. a type 2 diabetes population) and with different baseline utility scores.	An incorrect publication of Ara and Brazier was referenced. This has been changed to: Ara, R. and Brazier, J. (2010) Comparing EQ-5D scores for comorbid health conditions estimated using five different methods. https://eprints.whiterose.ac.uk/11048/ This study used EQ-5D data from the Health Survey for England and stated that: "Comparing the three original non parametric methods in terms of average errors and proportions of estimated HSUVs accurate to within a given magnitude, when using a baseline of perfect health, we found the additive method was the least accurate and the multiplicative method was the most accurate. When using an age adjusted baseline, the accuracy for both the additive and multiplicative methods increased and the minimum method was the least

			accurate while the multiplicative method remained the most accurate." "Although the simple linear model produced more accurate results than the non parametric estimators in our data, none of the coefficients in the model were significant and the model requires validating in external data. The trend to under estimate higher HSUVs and over estimate lower HSUVs suggests that a different model specification may be warranted and additional research exploring alternatives would be beneficial. A limitation of using regressions to explore relationships between HSUVs is that models are unlikely to be valid for HSUVs obtained using different preference-based measures thus each measure would require an individualised model." Moreover, the company's quote is a fragment of the following sentence: "Although the best method to combine multiple disutility values is still debated, the multiplicative method is

			considered to be the best approach overall and more conservative than the additive method.”
Page 138, Section 5.3.5, where the EAG report states: *“External validation was performed also against data not used to develop the economic model,[46] e.g. the Steno-2 RCT.” [6]*	The text should be revised or deleted as it does not accurately reflect the validation analysis published on the PRIME T2D Model.	The PRIME T2D Diabetes Model was not validated against the Steno-2 trial. In the publication referenced in the EAG report, an example cost-effectiveness analysis was presented that was based on data published from the Steno-2 trial as it showed the impact of improvements in multiple risk factors. No validation against Steno-2 was performed or described in the publication referenced.	Amended.
Page 138 in the EAG report states: *“the Advisory Boards (in 2014, 2015 and 2019)”*	Please update to reflect that there was one advisory board in 2019.	Only one advisory board (2019) was carried out and is highlighted throughout in Appendix N.	Amended.

Issue 3 Minor Typographical and Grammatical Errors

Description of problem	Description of proposed amendment	Justification for amendment	EAG response
Page 41 states: *"…a BMI of at least 25 kg/m², except SURPASS-5 where it was 23 mg/m²."*	Please update to *"…a BMI of at least 25 kg/m², except SURPASS-5 where it was 23* ***kg****/m²."*	Minor typographical error	Corrected.
Page 44; Table 3.3. TZP 10 mg SURPASS 4 eGFR mean ± SD cell states *"81.43 ± 0.44"*	Please update the SD to "***20.44*****"**	Minor typographical error	Corrected.
Page 46; Table 3.4. For all SURPASS-4 and the SURPASS-5 Metformin rows, the units are not specified.	Please update to specify *"****n (%)****"* in the SURPASS-4 rows and in the SURPASS-5 Metformin row	Minor typographical error	Corrected.
Page 46; Table 3.4. Footnote states that the data are sourced from *"Tables 12 of the CS"*	Please update to *"Tables 12****, 14 and 16*** *of the CS"*	Minor typographical error	Corrected.
Page 48; Table 3.5. Footnote states that the data are sourced from	Please update to *"Table 26, 32,* ***36, 40*** *of the CS"*	Minor typographical error	Corrected.

"Table 26, 32, 37, 42 of the CS"			
Page 48; Table 3.5. Footnote states *"*p<0.001; **p<0.001"*	Please update to *"*p<0.001* ***versus baseline****; **p<0.001* ***versus comparator for superiority****"*	Minor typographical error	Amended.
Page 48; Tables 3.6–3.8. Footnotes state that the data are sourced from *"Table 8, 13, 18, 23 of the CS"*	Please update to *"Table 8, 13,* ***17, 21*** *of the CS"*	Minor typographical error	Corrected.
Page 49 states: *"There was a dose-response relationship only in SURPASS-4 among all* ***trails*** *whereby the higher dose was more effective."*	Please update to: *"There was a dose-response relationship only in SURPASS-4 among all* ***trials*** *whereby the higher dose was more effective."*	Minor typographical error	Corrected.
Page 49; Table 3.9. Footnote states that the data are sourced from *"Table 27, 33, 38, 44 of the CS"*	Please update to *"Table 27, 33,* ***37, 42*** *of the CS"*	Minor typographical error	Corrected.
Page 50 states: *"Among all* ***trails****, there was a statistically significant difference versus the*	Please update to: *"Among all* ***trials****, there was a statistically significant difference versus the comparator in favour of tirzepatide in SURPASS-4*	Minor typographical error	Corrected.

*comparator in favour of tirzepatide in SURPASS-4 (versus insulin degludec) and SURPASS-5 (versus placebo) for all outcomes and all doses of tirzepatide. There was a dose-response relationship only in SURPASS-5 among all **trails** whereby the higher dose was more effective."*	*(versus insulin degludec) and SURPASS-5 (versus placebo) for all outcomes and all doses of tirzepatide. There was a dose-response relationship only in SURPASS-5 among all **trials** whereby the higher dose was more effective."*		
Page 50; Table 3.10. Footnote states that the data are sourced from *"Table 28, 34, 39, 45 of the CS"*	Please update to *"Table 28, 34, **38, 43** of the CS"*	Minor typographical error	Corrected.
Page 51; Table 3.11. Footnote states that the data are sourced from *"Table GPGL.5.11 of the SURPASS-2 CSR; Table GPGL.5.10. of the SURPASS-3 CSR; Table GPGH.5.18. of the SURPASS-4 CSR17; SURPASS-5 CSR"*	Please update to *"Table GPGL.5.11 of the SURPASS-2 CSR; Table **GPGH**.5.10. of the SURPASS-3 CSR; Table **GPGM**.5.18. of the SURPASS-4 CSR; **Table GPGI.5.9.** of the SURPASS-5 CSR"*	Minor typographical error	Corrected.

Page 52; Table 3.12. Footnote states that the data are sourced from *"Table GPGL.5.11. of the SURPASS-2 CSR; Table GPGL.5.10 of the SURPASS-3 CSR; Table GPGH.5.19. of the SURPASS-4 CSR; SURPASS-5 CSR"*	Please update to *"Table GPGL.5.11. of the SURPASS-2 CSR; Table **GPGH**.5.10 of the SURPASS-3 CSR; Table **GPGM**.5.19. of the SURPASS-4 CSR; **Table GPGI.5.9.** SURPASS-5 CSR"*	Minor typographical error	Corrected.
Page 53; Table 3.13. Footnote states that the data are sourced from *"Table 63 of the CS"*	Please update to *"Table **61** of the CS"*	Minor typographical error	Corrected.
Page 54; Table 3.14. Footnote states that the data are sourced from *"Table 66, 64 of the CS"*	Please update to *"Table **62, 63** of the CS"*	Minor typographical error	Corrected.
Page 56; Table 3.15. Footnote states that the data are sourced from *"Table 66, 67 of the CS"*	Please update to *"Table **64, 65** of the CS"*	Minor typographical error	Corrected.
Page 56; Table 3.16. Footnote states that the	Please update to *"Table **66** of the CS"*	Minor typographical error	Corrected.

data are sourced from *“Table 68 of the CS”*			
Page 56; Table 3.16. Asterisk footnote missing	Please update to *“*Total includes one patient with a missing severity”*	Minor typographical error	Corrected.
Page 58; Table 3.17. Footnote states that the data are sourced from *“Table 69, 70 of the CS”*	Please update to *“Table **67, 68** of the CS”*	Minor typographical error	Corrected.
Page 60; Table 3.18. Footnote states that the data are sourced from “*Table 71 of the CS”*	Please update to *“Table **69** of the CS”*	Minor typographical error	Corrected.
Page 62; Table 3.19. Footnote states that the data are sourced from *“Table 72 of the CS”*	Please update to *“Table **70** of the CS”*	Minor typographical error	Corrected.
Page 64; Table 3.20. Footnote states that the data are sourced from *“Table 73 of the CS”*	Please update to *“Table **71** of the CS”*	Minor typographical error	Corrected.
Page 66; Table 3.21. Footnote states that the data are sourced from *“Table 48 of the CS”*	Please update to *“Table **45** of the CS”*	Minor typographical error	Corrected.

Page 67; Table 3.22. The Body Weight, Age Group 1 row for SURPASS-4 EAS states *"[redacted]"*	Please update to *"[redacted]"*	Minor typographical error	Corrected.
Page 67; Table 3.22. The Body Weight, Age Group 2 row for SURPASS-4 FAS and EAS are empty	Please update to FAS: *"**<0.0001**"*; EAS: *"**0.001**"*	Minor typographical error	Amended.
Page 71; Tables 3.23 and 3.24 state Table 49 and Table 50 as sources	Please update to *"Table **47**"* and *"Table **48**"*, respectively.	Minor typographical error	Corrected.
Page 72–73; Figures 3.2 and 3.3 state Figure 27 and Figure 28 of the CS as sources	Please update to *"Figure **24**"* and *"Figure **25**"*, respectively.	Minor typographical error	Corrected.
Page 73; Table 3.25 states Table 51 of the CS as a source	Please update to *"Table **49**".*	Minor typographical error	Corrected.
Page 80 states: *"The fact that comorbidities were not systematically reported (52 of the studies did not include comorbidities) does not **infare** that these characteristics are not*	Please update to: *"The fact that comorbidities were not systematically reported (52 of the studies did not include comorbidities) does not **infer** that these characteristics are not potential effect modifiers or that the*	Minor typographical error	Amended.

potential effect modifiers or that the studies in the network are comparable."	*studies in the network are comparable."*		
Page 81 states: *"Figure 31 of the CS refers to the proportion of female patients."*	Please update to: *"Figure **28** of the CS refers to the proportion of female patients."*	Minor typographical error	This cannot be changed given that it is
Page 81 states: *"Out of the 136 arms presented, 11 arms reported values below 8% and 20% above 8.5%."*	Please update to: *"Out of the 136 arms presented, 11 arms reported values below 8% and 20 **arms** above 8.5%."*	Minor typographical error	Amended.
Page 84 states: *"The company stated that substantial heterogeneity was identified **in at least in at least** one of the relative comparisons for each continuous outcome."*	Please update to: *"The company stated that substantial heterogeneity was identified **in at least one** of the relative comparisons for each continuous outcome."*	Minor typographical error	Corrected.
Page 84 states: *"Both of these were limited, the first mixing **to** completely different expressions of treatment*	Please update to: *"Both of these were limited, the first mixing **two** completely different expressions of treatment experience/background therapy, and*	Minor typographical error	Corrected.

experience/background therapy, and the second limited to only one versus two OADs."	*the second limited to only one versus two OADs."*		
Page 87; Table 3.30. The 'Dulaglutide 4.5 g' reports [redacted]	Please update the row title to *'Dulaglutide 4.5* ***mg'*** and the value to [redacted].	Minor typographical error	Corrected.
Page 88; Table 3.30. The Lixisenatide 20 mcg row states [redacted]" and [redacted]" in the TZP 5 mg and TZP 10 mg rows respectively.	Please update to [redacted]" and [redacted]" in the TZP 5 mg and TZP 10 mg rows respectively.	Minor typographical error	Amended.
Page 127, Section 4.2.9.2, paragraph 2, where the EAG report quotes annual for QW dulaglutide and QW semaglutide as *"£955,52"*	Please update to *"****£955.52****"*	Typographical error replacing decimal points with commas. Table 85 of the CS reports both prices as £955.52	Corrected.

Location of incorrect marking	Description of incorrect marking	Amended marking	EAG response
Page 51; Table 3.12. SURPASS-5 LDL-C data are highlighted AIC.	These data are published in the Dahl 2022 supplementary content so do not need to be AIC highlighted.	The AIC highlighting can be removed from these data.	Amended.
Page 56; Table 3.16. Unpublished TEAE data are not highlighted.	As per the CS, these data are unpublished and therefore should have AIC highlighting.	All data in this table should be AIC highlighted.	Amended.
Page 66; Table 3.22. Unpublished subgroup analysis results from the trial CSRs are not highlighted.	As per the CS, these data are unpublished and therefore should have AIC highlighting.	All data in this table should be AIC highlighted.	Amended.
[redacted]	[redacted]	This should be updated to: *"The treatment cost associated with all doses of tirzepatide is given as [redacted], [redacted] and [redacted] per week, giving an annual cost of [redacted], [redacted] and [redacted] per patient for tirzepatide 5, 10, and 15 mg respectively."*	Amended.

The cost-effectiveness results within Section 5 of the EAG report are all unredacted.		Please redact all tirzepatide prices and cost-effectiveness results including: Fixed and discounted prices, direct costs, incremental costs or differences in cost, ICERs, probabilities of cost-effectiveness, or dominance results.	Amended.

References

1. National Institute of Health and Care Excellence (NICE). NG28. Type 2 diabetes in adults: management. Health Economic Model Report. Available at: https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845/ [Accessed 10 April 2022].
2. British National Formulary (BNF). Dulaglutide. Medicinal Forms. Available at: https://bnf.nice.org.uk/drugs/dulaglutide/medicinal-forms/ [Accessed 28 March 2023].
3. National Institute of Health and Care Excellence (NICE). Semaglutide. British National Formulary. Available at: https://bnf.nice.org.uk/drugs/semaglutide/. [Accessed 14 February 2023].
4. Chubb B, Gupta P, Gupta J, et al. Once-Daily Oral Semaglutide Versus Injectable GLP-1 RAs in People with Type 2 Diabetes Inadequately Controlled on Basal Insulin: Systematic Review and Network Meta-analysis. Diabetes Ther 2021;12:1325-1339.
5. Willis M, Asseburg C, Nilsson A, et al. Multivariate Prediction Equations for HbA(1c) Lowering, Weight Change, and Hypoglycemic Events Associated with Insulin Rescue Medication in Type 2 Diabetes Mellitus: Informing Economic Modeling. Value Health 2017;20:357-371.
6. Pollock RF, Norrbacka K, Boye KS, et al. The PRIME type 2 diabetes model: a novel, patient-level model for estimating long-term clinical and cost outcomes in patients with type 2 diabetes mellitus. J Med Econ 2022;25:393-402.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Tirzepatide for the treatment of patients with type 2 diabetes

[ID3938]

Response to EAG Report

May 2023

File name	Version	Contains confidential information	Date
ID3938_Eli Lilly_Tirzepatide_Response to EAG Report_16May [REDACTED]	1	Yes	16th May 2023

Executive summary

In response to the EAG report, this addendum intends to address the EAGs concerns particularly in relation to three key issues: the decision problem (Key Issues 1-3), NMA feasibility assessment (Key Issue 5) and the cost effectiveness (Key Issues 6-15).

Decision problem

There has been a misunderstanding by the EAG of the intended positioning of tirzepatide. Firstly, the eligible population would be in line with T2D NG28 where GLP-1 RAs are used if triple therapy with metformin and two OADs are not effective, contraindicated or tolerated. One OAD is then swapped for tirzepatide/GLP-1 RA which would mean tirzepatide would be used in combination with two OADs. The decision problem and clinical effectiveness evidence is therefore is in line with UK clinical practice in positioning tirzepatide for patients with T2D on two background OADs, exactly as GLP-1 RAs are currently used.

NMA feasibility assessment

In response to the EAG concerns around lack of feasibility assessment, the company have provided a more detailed description of the feasibility assessment conducted for the NMA. Additionally, in order to address concerns about potential heterogeneity within the trials, an additional sensitivity analysis has been conducted in which studies that contributed to the increased heterogeneity were removed. For the HbA1c and body weight endpoints, the results of this sensitivity analysis were in line with those of the main analysis. For the BMI endpoint, the results were mostly aligned however there were some differences due to the smaller network and potential influence of sigma convergence issues.

Cost-effectiveness results

In order to address the issues the EAG experienced with the model, the company has worked with NICE to facilitate a series of calls between the EAG and the model developers, which the company hope has alleviated any concerns about the model. Further clarification has also been provided in the document below on reasoning for the model approach used as well as the technical validation steps employed and further detail supporting the appropriateness of the PSA. Additionally, as requested by the EAG, further analyses have been run with results presented for comparisons of all doses, rather than of doses in the same recommended maintenance dose step, as well as the results of the fully incremental analysis.

The cost -effectiveness analysis also includes [redacted] for tirzepatide (please see the EAG Preferred Base Case Simulations section below).

A.1.1 The Decision Problem

Key Issue 1: Mismatch between scope and decision problem in terms of line of therapy and comparators might lead to a lack of evidence for the scope of interest in decision making

Report Section	2
Description of issue and why the EAG has identified it as important	The population in the NICE scope is much broader than in the decision problem, which is limited to combination therapy and only a line of therapy consistent with GLP-1 RAs in response to failure of at least three OADs.
What alternative approach has the EAG suggested?	The EAG requested clarification that the company's intention was to only address the clinical and cost effectiveness of tirzepatide as a combination therapy only and in the restricted population described, which the company confirmed was the case.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Additional evidence would be required if a decision was to be made for tirzepatide as monotherapy where metformin is inappropriate. See Key issue 2 for combination therapy.
EAG = Evidence Assessment Group; GLP-1 RAs = glucagon-like peptide-1 receptor agonists; NICE = National Institute for Health and Care Excellence; OADs = oral antidiabetic drugs	

The proposed treatment positioning for tirzepatide is in line with glucagon-like peptide-1 receptor agonist (GLP-1 RA) therapy, as it is expected that tirzepatide will be used as an alternative to GLP-1 RAs. The NG28 clinical guidelines recommend that if triple therapy with metformin and two oral antidiabetic agents (OADs) is not effective, not tolerated or contraindicated, then clinicians should consider *switching* one of these drugs to a GLP-1 RA.[1] It is therefore expected that clinicians in UK clinical practice would use tirzepatide in patients with type 2 diabetes (T2D) that is inadequately controlled with metformin and two OADs, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. More specifically, if triple therapy with metformin and two OADs is not effective, not tolerated or contraindicated, clinicians would consider switching one of these drugs to a GLP-1 RA or tirzepatide. This would then result in tirzepatide being taken *in combination with two OADs*. Please see Figure 1 for the NG28 clinical guideline diagram depicting this GLP-1 RA positioning, with the additional inclusion of glucose-dependent insulinotropic polypeptide (GIP)/GLP-1 RA therapy (i.e. tirzepatide) in line with GLP-1 RA therapy in the bottom right textbox, as was presented within the original submission.

The expected eligible population for tirzepatide is therefore the same as the NG28 GLP-1 RA-eligible population, representing a narrower population than that specified in the marketing authorisation wording for tirzepatide and the NICE final scope for this evaluation. This anticipated positioning aligns with current UK clinical practice and reflects the highest unmet need for a more effective treatment option for patients for whom the alternative is a GLP-1 RA, which may not sufficiently control their HbA1c level and/or provide sufficient weight loss. The positioning of tirzepatide therefore dictates that the most relevant comparators for this submission are GLP-1 RAs only.

The SURPASS clinical trial program was designed to meet regulatory requirements of different authorities around the globe and to provide clinically meaningful data on the use of tirzepatide at

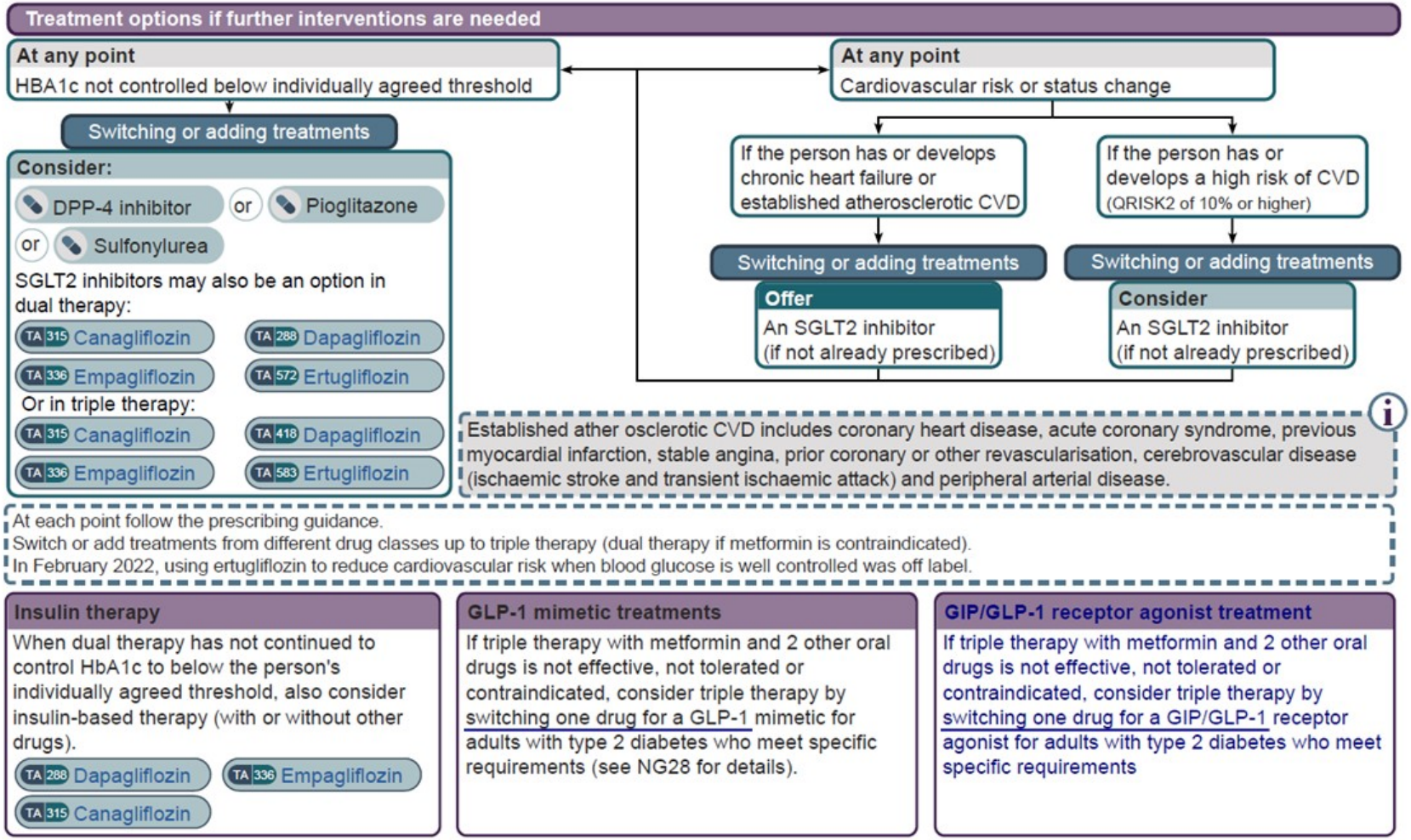
different stages of T2D and its treatment continuum from monotherapy to the failure of basal insulin treatment. As part of this, EMA guidance on conducting a T2D clinical development programme was followed.[2] The background treatments received by patients alongside tirzepatide therefore reflect the global nature of the trials. It is acknowledged that the trial designs of these global clinical trials are not entirely in alignment with UK clinical practice, or the decision problem addressed within this submission, in terms of line of therapy, comparators and background therapies. However, such a misalignment is not uncommon in NICE evaluations and those of other HTA bodies, especially of GLP-1 RAs. The liraglutide, dulaglutide and semaglutide clinical trial programmes all provide evidence which reflects the continuum of global treatment pathways and are not solely focussed on the population after failure of two OADs. For example, the Phase 3 trial designs listed in the semaglutide summary of product characteristics (SmPC) are not aligned to how the GLP-1 RA is used in NHS practice.[2, 3]

Nonetheless, the SURPASS clinical trials provide robust evidence for the clinical efficacy of tirzepatide in T2D and their relevance to the decision problem is further discussed in Section B.2.2 of the submission. The SURPASS clinical trial program supports the positioning of tirzepatide as an alternative to GLP-1 RA therapies and, within the framework of the current NG28 algorithm, there is no intention to position tirzepatide beyond the specified population presented within this submission with the current data available.

Whilst none of the SURPASS clinical trials exactly match the population as proposed for positioning in UK clinical practice, a network meta-analysis (NMA) was conducted to establish comparative efficacy for tirzepatide generalisable to the relevant population. Through the NMA, comparative efficacy is demonstrated across multiple relevant efficacy outcomes, including the most critical clinical endpoints in the management of diabetes, such as change from baseline in glycated haemoglobin (HbA1c), weight and body mass index (BMI). These endpoints are the key drivers of clinical- and cost-effectiveness analyses of a treatment in T2D.

Overall, despite the slight misalignment between the trial population and patient population expected in UK clinical practice, the company maintain that there is a robust and comprehensive evidence base on which to base decision making which is consistent with previous UK assessments of GLP-1s semaglutide and dulaglutide.[4-6]

Figure 1: Anticipated positioning of tirzepatide (GIP/GLP-1 receptor agonist treatment) alongside NICE T2D NG28 clinical guidelines for patients following insufficient control of HbA1c levels on first-line therapy



Abbreviations: BMI: body mass index; CVD: cardiovascular disease; DPP-4: dipeptidyl peptidase-4; GIP: glucose-dependent insulinotropic polypeptide; GLP-1: glucagon-like peptide-1; HbA1c: glycated haemoglobin; OAD: oral antidiabetic drug; SGLT2i: sodium glucose cotransporter-2 inhibitor.
Source: NICE guidelines on management type 2 diabetes.[1]

Key Issue 2: Mismatch between decision problem and evidence in terms of line of therapy/OAD therapy experience might lead to an overestimate of the effectiveness and cost effectiveness of tirzepatide

Report Section	**2, 3, 4**
Description of issue and why the EAG has identified it as important	The population in the decision problem is different to that in the SURPASS trials and the NMA in that almost no patients have experienced triple OAD therapy, most having failed on only metformin or metformin plus one other OAD. A clinical expert did suggest that GLP-1 RAs might be given at an earlier line of therapy, which is inconsistent with NICE Guideline NG28, but does seem to be consistent with the ADA/EASD consensus report, but this might also mean the other OADs, e.g., and SGLT-2i might be comparators. There is only a little evidence on whether OAD experience might be a treatment effect modifier. This is in the form of a subgroup analysis in SURPASS-4, which is the only trial where concomitant triple OAD therapy is possible, that suggests an interaction of OAD combination on the treatment effect, but the direction of effect is unclear.
What alternative approach has the EAG suggested?	The EAG suggested that the decision problem be amended to more consistent with the evidence, but the company reiterated that the line of therapy and GLP-1 RAs as comparators were how they expected tirzepatide to be given in clinical practice.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	If the line of therapy is earlier than failure of three OADs then the SURPASS trials might be more appropriate, but this needs to be recognised in the decision problem. Consideration then also needs to be given to comparison with OADs e.g., and SGLT 2i in the NMA and the CEA. Scenario analyses assuming the population characteristics from the SURPASS trials (instead of based on THIN second intensification cohort) would then also be appropriate. If, on the other hand, the decision problem does not change, then there remains uncertainty as to the appropriateness of the clinical evidence.
ADA = American Diabetes Association; CEA = cost effectiveness analysis; EAG = Evidence Assessment Group; EASD = European Association for the Study of Diabetes; GLP-1 = glucagon-like peptide-1; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; OADs = oral antidiabetic drugs; RA = receptor agonists; SGLT-2i = Sodium-glucose co-transporter-2 inhibitor	

Please refer to the response to Key Issue 1 for justification for the differences between the decision problem and clinical evidence in terms of line of therapy. Regarding any differences in background therapy, the results of the SURPASS-4 subgroup analysis of baseline OADs for the endpoints change from baseline in HbA1c (%), weight (kg) and BMI (kg/m^2) were in line with those of the main analysis and therefore support the generalisability of the tirzepatide results irrespective of baseline therapy. Similarly, in the NMA meta-regression analysis for HbA1c change from baseline and weight change from baseline, results adjusted for the number of background OADs were similar to the unadjusted results for all tirzepatide doses compared to all GLP-1 RAs. In addition, as part of the NMA, a sensitivity analysis (described in Appendix D.8.1.3 of the main submission) that included studies with patients on background therapy comprising

three OADs (e.g., SURPASS-4) produced results in line with the main analysis. Note, this sensitivity analysis is not in line with the positioning of tirzepatide in UK clinical practice, i.e. two OADs plus GLP-1 RA/tirzepatide, as described under Issue 1 above. Taken together, these analyses support the assertion that results of the NMA are generalisable to patients in the target population and supports that any mismatch between OAD therapy experience between the trial populations and clinical practice is unlikely to have a significant impact on the cost effectiveness of tirzepatide.

A.1.2 The Clinical Effectiveness Evidence

Key Issue 3: Mismatch between the administration of tirzepatide in clinical practice by titration and the tirzepatide trial evidence, the NMA and the CEA, according to maintenance dose strata, is likely to lead to biased estimates of effectiveness and cost effectiveness in an unknown direction

Report Section	**2**
Description of issue and why the EAG has identified it as important	The marketing authorisation for tirzepatide is for it to be administered via titration from a maintenance dose of 5 mg, through 10 mg to 15 mg as required to obtain an adequate response in HbA1c reduction. However, the comparisons between tirzepatide and the GLP-1 RAs, SURPASS trials, the NMA and the CEA are stratified by maximum maintenance dose into 5 mg, 10 mg and 15 mg, without titration being permitted. This means that there is lack of applicability to clinical practice. Given the observation in the SURPASS trials and the NMA, which includes SURPASS-2 and SURPASS-3, of a dose response relationship for glycaemic, as well as body weight/BMI control, it is likely that efficacy would be underestimated for the 5 mg and overestimated for the 15 mg stratum. It also appears that all the comparator trials were designed in the same way.
What alternative approach has the EAG suggested?	The EAG would prefer a comparison of treatments as in clinical practice, including titration as appropriate. This would also mean that the treatment strategies in the economic model would not be restricted to within dose steps but include the possibility for individual patients to switch between treatment dosages for those treatments that are titrated. Given the current nature of the comparison, the EAG would tentatively suggest that, if the 5 mg and the 15 mg dose outcomes might be an under or overestimate respectively, then the 10 mg outcomes might be closest to titration. An equivalent analysis of the comparator outcomes, notwithstanding that some are not titrated and some available in only two dose levels, suggests that the company's chosen comparator doses for the tirzepatide 10 mg dose might also be the most appropriate.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Ideally, a comparison of treatments as they would be administered in clinical practice is required but appears that no such evidence exists. The economic model should also be updated to allow patients to switch between treatment dosages to make comparisons between treatments that are titrated as in clinical practice.

Report Section	2
BMI = body mass index; CEA = cost effectiveness analysis; EAG = Evidence Assessment Group; GLP-1 = glucagon-like peptide-1; HbA1c = glycated haemoglobin; NMA = network meta-analysis; RA = receptor agonists	

The company acknowledges that, given the SURPASS trial programme was based on pre-specified maintenance doses but in clinical practice, tirzepatide will be administered via titration between maintenance doses, there is a mismatch in terms of administration here. It is important to note that this is a recurring issue in diabetes trials (including other GLP-1 RA trials) with dosing based on pre-specified maintenance dose strata rather than titration.

In an attempt to address this issue, in the original submission, the company suggested that, when *interpreting* the results, comparisons should be made within recommended maintenance dose steps rather than between dose steps. In clinical practice, patients titrate to maintenance doses of GLP-1 RAs with few patients de-escalating to lower doses,[7] which supports the interpretation of the results in this way. Nevertheless, comparisons between all dosages were made within the NMA, as presented in the original submission, and the economic model has the capability to run these comparisons. Comparisons made between all doses have therefore been conducted for the cost effectiveness analysis and are provided in the EAG Preferred Base Case Simulations section below.

Additionally, as this is an issue that applies to all relevant comparator trials, the company agree with the EAG that any attempt to structure a model around projected titration would likely lead to an increase in uncertainty due to lack of evidence.

Key Issue 4: Lack of comparative evidence on micro and macrovascular complications

Report Section	3
Description of issue and why the EAG has identified it as important	The outcomes in the trials included in the CS are surrogates for the micro- and macrovascular complications. Therefore, it is uncertain what the treatment effect would be on these final endpoints. One tirzepatide trial, SURPASS-CVOT, was identified as reporting some of these outcomes, in particular MACE, but it was reported to be ongoing. For the comparators, other CVOTs were excluded from the NMAs, but these were only of the surrogates.
What alternative approach has the EAG suggested?	The CS does not contain the data required for this type of comparison.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	The SURPASS-CVOT trial should be combined with any similar comparator trials in order to provide this comparative evidence.
CS = Company submission; CVOT = cardiovascular outcomes trial; EAG = Evidence Assessment Group; MACE = major adverse cardiovascular events; NMA = network meta-analysis	

Limited data on micro and macrovascular complications were collected in the trials for which results are available and therefore it is not possible to provide comparative evidence on micro and macrovascular complications. The SURPASS-CVOT trial is expected to collect data on macro- and micro-vascular complications.[8] Unfortunately, as the trial is still ongoing, these data are not yet available and as such, no additional information can be provided here. In lieu of this, a meta-analysis of positively adjudicated major adverse cardiovascular events (MACE) in the

SURPASS trials was performed, as described in the main submission, and found that treatment with tirzepatide was not associated with excess risk for CV events in patients with T2D.

Key Issue 5: NMA of high risk of bias due to lack of feasibility assessment/assessment of trial comparability and insufficient sensitivity analyses

Report Section	3.3
Description of issue and why the EAG has identified it as important	The NMA was based on a SLR not specific to the CS submitted to NICE. Trials were included without a systematic assessment of heterogeneity and with an assumption that the treatment effect is independent of concomitant background OAD therapy. Substantial heterogeneity seems to exist and have to some degree been identified by the company, but appropriate sensitivity analyses were not conducted. Also, two different estimands were used in the same NMA network, one including and the other excluding patients who required rescue therapy.
What alternative approach has the EAG suggested?	The EAG recommended a feasibility assessment/assessment of trial comparability based on potential treatment effect modification and sensitivity analyses to exclude trials to improve comparability as appropriate.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	The EAG continue to recommend a feasibility assessment/assessment of trial comparability based on potential treatment effect modification and sensitivity analyses to exclude trials to improve comparability as appropriate.
CS = company submission; EAG = Evidence Assessment Group; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; OAD = oral antidiabetic drug; SLR = systematic literature review	

The EAG stated that the NMA was associated with a high risk of bias due to the lack of feasibility assessment and insufficient sensitivity analyses. However, the company maintain that an adequate feasibility assessment was conducted with multiple sensitivity analyses performed in order to address any potential sources of uncertainty. In order to alleviate any concerns, further information has been provided below.

Feasibility assessment

Studies were identified in a systematic way to align with the PICO statement set out initially. Alongside the visual presentation of these data in tables and figures, detailed assessment of RCT bias was employed to reduce the uncertainty of the estimated treatment effect. Similar results to the base case were obtained in the meta-regression analysis on treatment effect modifiers, providing justification for the statement that treatment effect is not influenced by the differences observed between baseline characteristics.

Comorbidities

During the feasibility assessment, studies identified by the SLR were assessed in terms of patient comorbidities and baseline characteristics.

The following specific populations were identified as being assessed in some studies included in the SLR: patients with renal impairment, CVD/high CV risk, obesity, NAFLD and other

comorbidities. Of the studies identified, only one was eligible for inclusion in the network due to number of background therapies patients were on or had data at the relevant timepoints and most were excluded based on this. Additionally some studies were conducted in an Asian population and therefore were not included in the main analysis. As the majority of studies including patients with comorbidities were not included in the network due to the network definition, it was not feasible to assess the impact of comorbidities on the treatment effect estimates.

More details on the review and assessment of baseline characteristics and comorbidities within identified studies are listed below.

Comorbidities	Studies excluded from the main analysis and reason for exclusion	Studies included in the main analysis
Renal Impairment: Studies where a specific population with renal impairment could be identified were reviewed to ensure alignment to the SURPASS trials in terms of inclusion and exclusion criteria. With this in mind, studies only including patients with moderate or severe impairment or macroalbuminuria were excluded from the analysis.	The following studies were specifically designed for patients with CKD and studied the effect of treatments in this setting. They included patients with moderate or severe impairment or macroalbuminuria and were therefore excluded. • **AWARD 7** • **LIRA-RENAL** • **PIONEER 5** • **Hiramatsu** • **Wang 2020a**	n/a
CVD/CV high risk (not CVOT)	• **Nystrom 2017**: as this study had a trial duration of less than 22 weeks, no data available in the relevant time interval so it was not included • **Wang 2020:** as this study was conducted in an Asian population it was not included in main analysis The following studies were excluded as the background therapy received by patients was not aligned with the network definition, none were eligible for inclusion within the main analysis • **Arturi 2017:** as this study had an unclear proportion of patients on metformin + SU, • **Ikomnomodis 2020**: as this study had 32% of patients with no OAD and 68% with metformin only • **HEELA**: as patients within the trial had up to three OADs (metformin [100% of patients], SU [85%], TZP [42%]) this study was not included in the main analysis due to the network	n/a

	definition but was investigated within a sensitivity analysis	
Obesity Six studies were identified with a specific population of obese T2DM patients but only one fitted with the network definition. However the meta-regression sensitivity analysis (as presented B.2.9.5.3 of the company submission) included baseline weight to account for any differences at baseline.	• **Yin 2018:** The trial duration of this study was less than 22 weeks so there no data were available at the timepoint of interest and the study was excluded The following studies were excluded as the background therapy received by patients was not aligned with the network definition, none were eligible for inclusion within the main analysis • **Gravitas**: in this study, 23% of patients did not have background therapy, 26% had insulin, 7% received metformin, 8% SU and 11% SGLT2. • **Scale**: In this study, 11% of patients had no background therapy, 85% had metformin, 27% had SU, 8% had glitazone. • **LYDIA**: This study had an unclear proportion of patients on metformin and/or SU • **Kind-LM:** Patients in this study had no background therapy	• **Van Gaal 2014:** 100% of patients within the study received metformin so this study was aligned with the network definition and was eligible for main analysis
Non-alcoholic fatty liver disease (NAFLD) Five studies were identified as specifically including patients with NAFLD ; however none of them were eligible for the network either due to background therapy received by the patients or because they were conducted in an Asian population.	Patients in the following studies did not receive any background therapy so the studies were not aligned with the network definition • **Feng 2019** • **Liu 2020** • **Zhang2020b** The following studies were conducted in an Asian population so were not included in the main analysis • **Guo 2020** • **Light-on**	n/a
Other comorbidities	• **Jaiswal 2015**: included patients with mild-to-moderate peripheral neuropathy • **Lou 2020:** included patients with metabolic syndrome Neither study was eligible for inclusion in the network due to background therapy (background therapy was not detailed in the studies. In addition Lou 2020 was conducted in an Asian population.	n/a

Baseline Characteristics

Studies identified by the SLR were also assessed in terms of patient baseline characteristics to identify potential sources of heterogeneity within the included studies. As shown in the table below, only two studies identified as possible sources of heterogeneity were included in the main analysis. Therefore, the impact of these characteristics could not be investigated. However heterogeneity was also checked during the analysis and is discussed further below.

Baseline characteristics	Studies excluded from the main analysis and reason for exclusion	Studies included in the main analysis
Age: studies with mean baseline age <50 or >70 years were identified and assessed for potential heterogeneity. Most studies had a mean baseline age within the 50–70 year range so studies falling outside of this range were assessed to identify whether differences were likely to impact the results of the NMA analysis. Age <50 : Six studies were identified with a population of baseline age <50 years, but as the younger population of patients with T2D is growing in the UK, these studies were still considered relevant. Therefore, no reason was identified to exclude these studies from main analysis based on the age only. However, there were other reasons for exclusion detailed in the next column. Age > 70 as this is relatively older population also under represented in studies. Three studies were identified with population of baseline age >70 years. None of them were eligible for the main analysis for reasons other than the age.	**Age < 50** • ***LYDIA*** (mean age 44 years): In this study unclear proportions of patients received metformin and/or SU so study was not eligible for inclusion within the main analysis. However, this study was included in a sensitivity analysis • **Feng 2019** (mean age 47 years), **Liu 2020** (mean age 49 years) patients had no background therapy so those studies did not fit with the network definition • **Light-on** (mean age 45 years)*:* This study was conducted in an Asian population and so was not included in main analysis, but was included in a sensitivity analysis • **SIMPLE** (mean age 47 years): Patients in this study received basal insulin as background therapy, so the study did not align to the network definition **Age > 70** • **GetGoal-O**: Patients in this study received basal insulin +/- OAD as background therapy so the study did not align with the network definition • **PIONEER 5 and Hiramatsu 2018**: both studies contained patients with moderate renal impairment (eGFR	**Age < 50** **Van Gaal 2014** (mean age 43)**:** 100% of patients within the study received metformin so this study was aligned with the network definition and was eligible for main analysis **Age > 70** n/a

	30-60) so neither were eligible for inclusion within the network	
Gender Three studies with a high (>70%) proportion of females were identified. All three studies were excluded because they were not eligible for inclusion in the network for reasons detailed in the next column. 24 studies with less than 30% female were identified but none could be included in the main analysis primarily due to background therapy or because they were conducted in an Asian population.	**%Female > 70%** • **SIMPLE**: Patients in this study received background therapy of basal insulin so the study did not align with the network definition • **Rosenstock 2009b and Wu 2012**: these studies had a trial duration of less than 22 weeks so no data available were available at the relevant timepoints **%Female <30%** • 24 studies were identified with a low (<30%) proportion of females, amongst which three aligned with the network definition (**Araki 2015, Light-On, Zhang 2012**), but those studies were conducted in an Asian population.	**%Female > 70%** n/a **%Female <30%** n/a
Baseline BMI and weight Studies with high BMI (BMI > 40 kg/m²) or weight (weight >120 kg) at baseline were identified but none of these were eligible for inclusion in the network due to background therapy.	**Weight > 120 kg** 7 studies were identified, however due to background therapy received by patients, none were eligible for inclusion within the main analysis • **SCALE** • **GRAVITAS** • **Jaiswal 2015** • **SIMPLE,** • **ELEGANT** • **DURATION-1** • **Vanderheiden 2016** **BMI > 40 kg/M²** • **Vanderheiden 2016** This study could not be included in the analysis due to background therapies received by patients	n/a
Diabetes duration Studies with mean diabetes duration of >15 years were identified, but	• **Rosenstock 2020**: this study was investigated albiglutide which was not	

none of these were eligible for inclusion in the network mostly due to background therapy.	a relevant comparator within the NMA so was out of scope The following studies did not align with the network definition due to the number or variety of background therapies received by patients: • **GRAVITAS**: 23% of patients did not have background therapy, 26% had insulin,, 7% metformin, 8% SU, 11% SGLT2 • **MDI Liraglutide**: 100% of patients had basal and mealtime insulins (+ OADs) • **PIONEER 8:** 100% of patients had insulins as background therapy (+ OAds) • **Joubert 2021:** 100% of patients had insulins as background therapy (+ OAds) • **Van Eyk 2019**: 77% of patients had insulins as background therapy so this study did not fit with network definition The following studies were conducted in an Asian population so were excluded from the main analysis and investigated in a sensitivity analysis • **Nakaguchi2020** • **Ishii2019** The following studies contained patients with renal impairments and as such could not be included within the main analysis • **AWARD 7** • **LIRA-RENAL**	
HbA1c at baseline Amongst the studies identified with either very high (>10% in at least 1 arm) or very low (<7% in at least 1 arm) HbA1c at baseline, only Bergenstal 2009 was included in the network. Other studies from the SLR were considered consistent in terms of	**Hba1c > 10%** • **SIMPLE**: patients in this study received basal insulin as background therapy so the study was not aligned with the network definition	**Hba1c > 10%** • **Bergenstal2009**: 100% of patients had both metformin and SU as background therapy so the study aligned with the network definition and was included in the main

HbA1c. In addition, a meta-regression with adjustment for baseline HbA1c was to be performed to evaluate any potential impact of HbA1c differences at baseline.	**Hba1c <7%** • **LIBRA**: ~30% of patients received no background therapy so the study was not aligned with the network definition • **Yamamoto 2018**: patients received no background therapy so the study was not aligned with the network definition • **Hiramatsu2018**: patients with moderate renal impairment were included so the study was excluded from the analysis • **Matikainen 2018**: the trial duration of this study was less than 22 weeks so no data were available at the relevant timepoint	analysis
Race/ethnicity	Studies were identified as either Asian or non-Asian studies. 56 studies were identified as having a mainly Asian population; 52 included only Asian countries, and 4 included a >50% proportion of patients of Asian ethnicity. All were excluded from the main analysis and included in a sensitivity analysis.	

Sensitivity analysis for unclear metformin use and/or 3 OADs

The network definition meant that trials were included with:

- Add-on 1 OAD as defined as >90% of patients on metformin monotherapy*
- Add-on 1–2 OADs with >50% of patients of metformin*

*Trials with an unknown proportion of patients on metformin background therapy were excluded from the main analysis as well as trials including patients on 0 or ≥3 OADs. Trials with an unknown proportion of patients on metformin and trials with ≥3 OADs were included in the sensitivity analyses. This is to ensure alignment with SURPASS-2 and -3 trials and with current international guidelines.

A sensitivity analysis investigating the effect of modifying the network definition was performed, which included studies including patients with unclear proportion of metformin as background therapy and studies including patients on a background therapy of three OADs, this network

definition aligned with SURPASS-4 . Seven studies were included in this sensitivity analysis: DURATION-6, Ferdinand 2014, HARMONY7, HEELA, LYDIA, SURPASS-4 and SUSTAIN 10. Amongst these studies, six included patients on a background therapy of three OADs and one (LYDIA) included patients with unclear proportion of metformin as background therapy. Of note, a meta-regression adjusting for the number of OADs was also performed on the main analysis.

To note, the limitation of this network definition is also an issue for NMAs investigating other GLP-1 RAs. For example, an independently reviewed NMA investigating the efficacy of semaglutide was only able to conduct an NMA in patients whose diabetes was uncontrolled on one or two OADs.[4]

Heterogeneity

The EAG raised concerns that potential heterogeneity between the trials had not been adequately tested within the NMA. The company maintain that heterogeneity was thoroughly tested and have provided further information on this to alleviate any concerns.

The EAG noted that no I^2 statistics had been provided within the company submission, these statistics have now been provided in **Appendix A**.

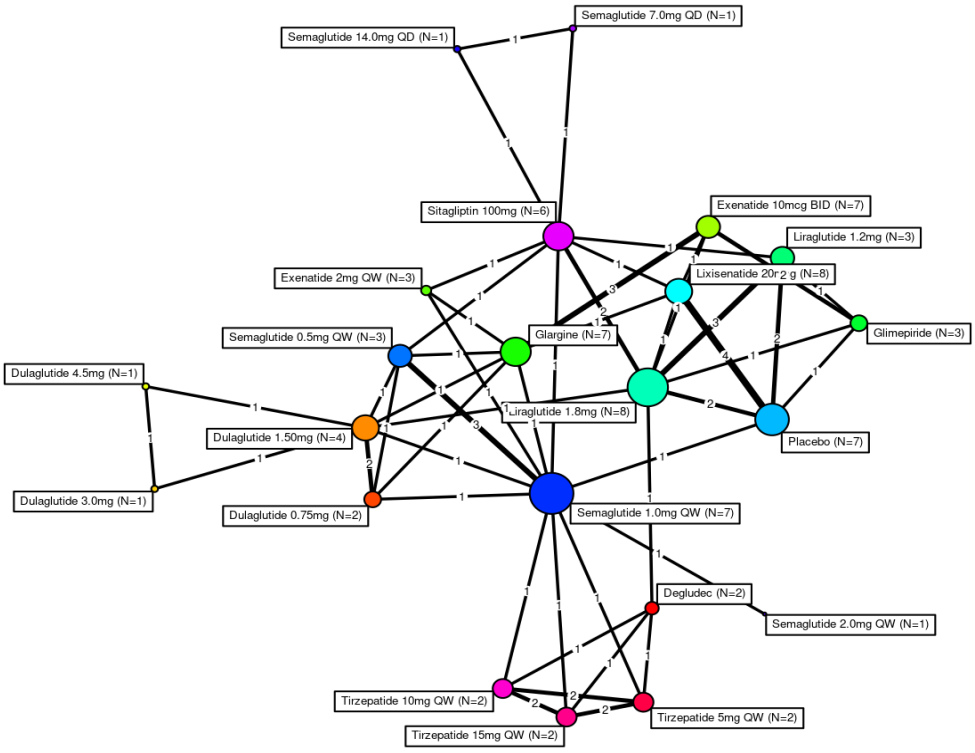
Within the EAG report, it was noted that '*the company also stated that there was substantial heterogeneity in all continuous outcomes and nausea, but provided no details and apparently conducted no sensitivity analysis as a result of this'.* The company would like to highlight that there appears to be a misunderstanding of the definition of substantial heterogeneity by the EAG. When referring to substantial heterogeneity in the CS, the company is referring to specific treatment comparisons that were classified as having substantial heterogeneity as per advice published by the Cochrane Consumers and Communication Group.[9] This does not mean the whole network has substantial heterogeneity, as may have been understood by the EAG. As shown in **Appendix A**, out of all the treatment comparisons that were assessed for heterogeneity only a small portion had substantial heterogeneity. Due to this small portion of studies/treatment comparisons yielding increased heterogeneity, sensitivity analyses for the removal of these studies were not performed. However, the company assessed and reviewed the studies that added to heterogeneity and it was concluded that all were aligned with the inclusion/exclusion criteria of the NMA as well as the network definition.

For transparency, the company have also conducted a sensitivity analyses on HbA1c, body weight and BMI in which the following studies that contributed to the increased heterogeneity were removed: Apovian 2010, AWARD-1, DeFronzo 2005, Derosa 2012b/2013c/s, Kendall 2005, Liutkus 2010, AWARD-10, AWARD-5, EAGLE, LEAD-5, LIRA-ADD2SGLT2i, PIONEER 4. The same model specifications were used as the main analysis. To note, the BMI network had issues with sigma convergence, alternative priors were tested but did not result in better fit and did not resolve the issue with sigma.

In summary, the results from this sensitivity analysis were similar to the results of the main analysis for the HbA1c and body weight models. However, there were some differences in the BMI model following the removal of these studies as the network has fewer studies and thus removal of the studies is more impactful. Additionally, the results of the heterogeneity sensitivity analysis for BMI should be interpreted with caution due to the sigma convergence issues noted above.

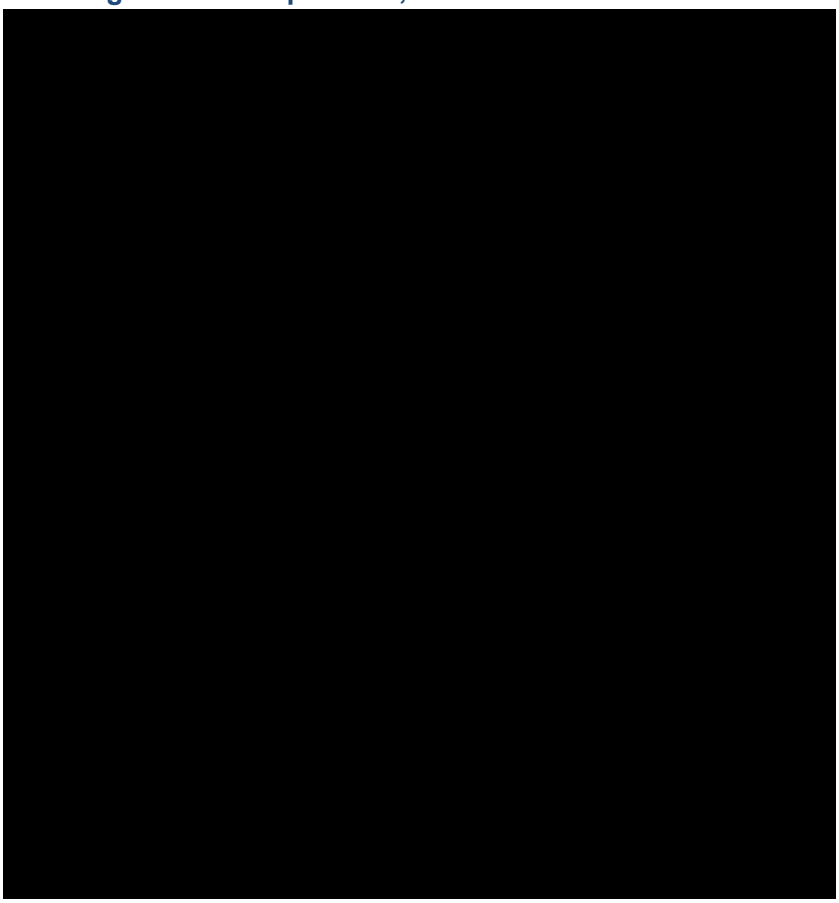
HbA1c

Figure 2: Sensitivity analysis network diagram for HbA1c (%) change from baseline



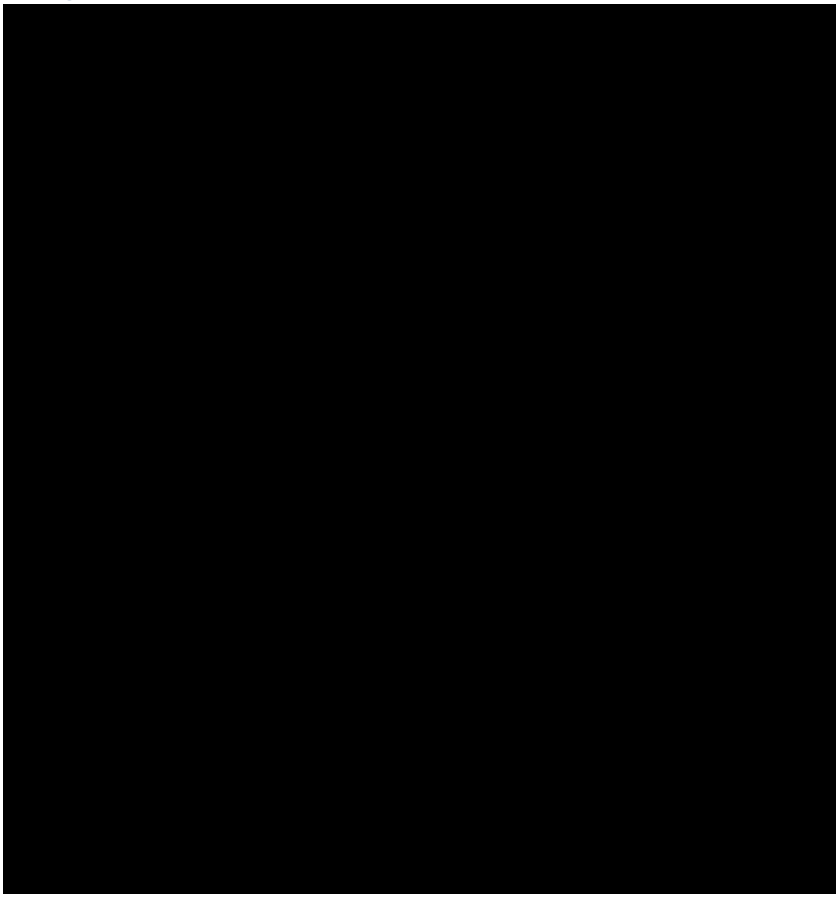
Abbreviations: BID: twice daily; QD: every day; QW: every week.

Figure 3: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 5 mg versus comparators, random effects model



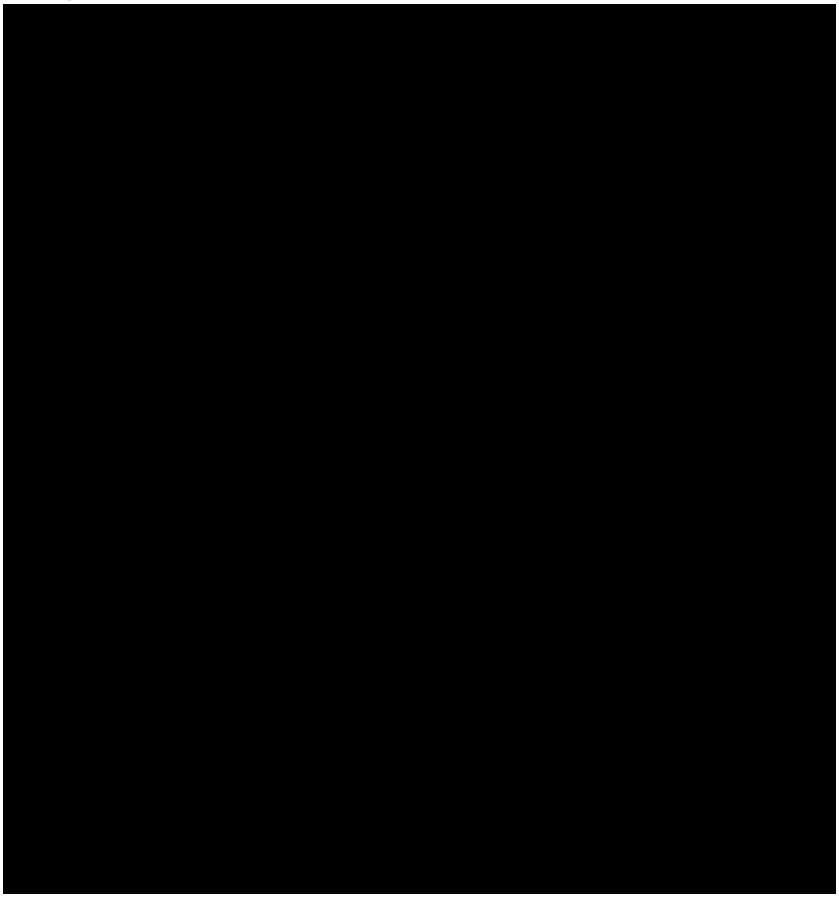
Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 4: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 10 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 5: Forest plot (median difference [95% CrI]) for HbA1c (%) change from baseline, TZP 15 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: CrI: credible interval; BID: twice a day; HbA1c: haemoglobin A1c; QD: once a day; QW: once a week; TZP: tirzepatide.

Table 1: Pairwise results (median difference [95% CrI]) table for HbA1c (%) change from baseline, random effects model; TZP 5, 10 or 15 mg (column) vs comparators (row)

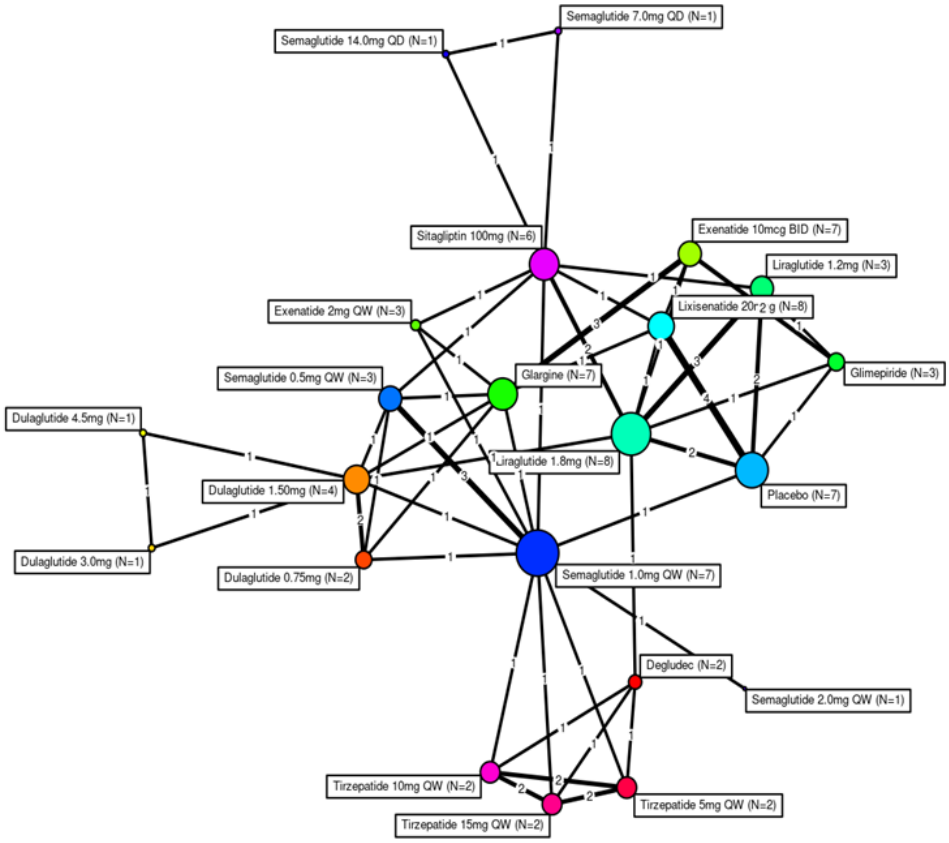
Column vs. row	TZP 5 mg		TZP 10 mg		TZP 15 mg	
	Main	Sensitivity	Main	Sensitivity	Main	Sensitivity
Placebo						
Tirzepatide 5 mg QW						
Tirzepatide 10 mg QW						
Tirzepatide 15 mg QW						
Semaglutide 0.5 mg QW						
Semaglutide 1.0 mg QW						
Liraglutide 1.2 mg						
Liraglutide 1.8 mg						
Dulaglutide 0.75 mg						
Dulaglutide 1.50 mg						
Dulaglutide 3.0 mg						
Dulaglutide 4.5 mg						

Column vs. row	TZP 5 mg		TZP 10 mg		TZP 15 mg	
	Main	Sensitivity	Main	Sensitivity	Main	Sensitivity
Semaglutide 7.0 mg QD						
Semaglutide 14.0 mg QD						
Exenatide 2 mg QW						
Exenatide 5 mcg BID						
Exenatide 10 mcg BID						
Lixisenatide 20 mcg						

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

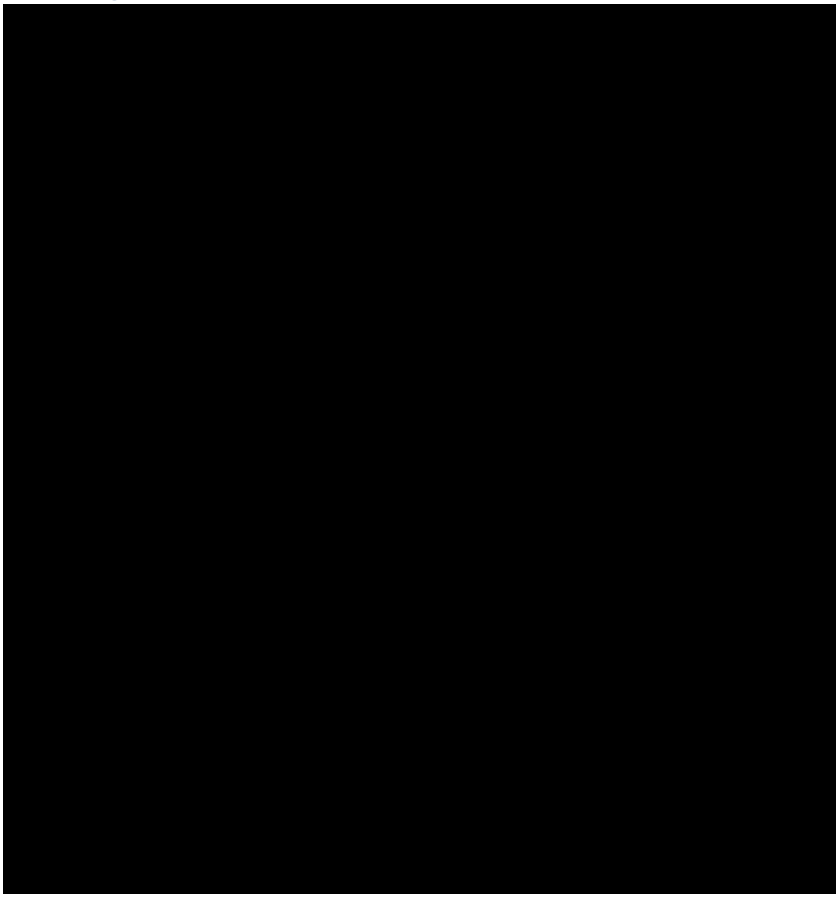
Body Weight

Figure 6: Sensitivity analysis network for weight (kg) change from baseline



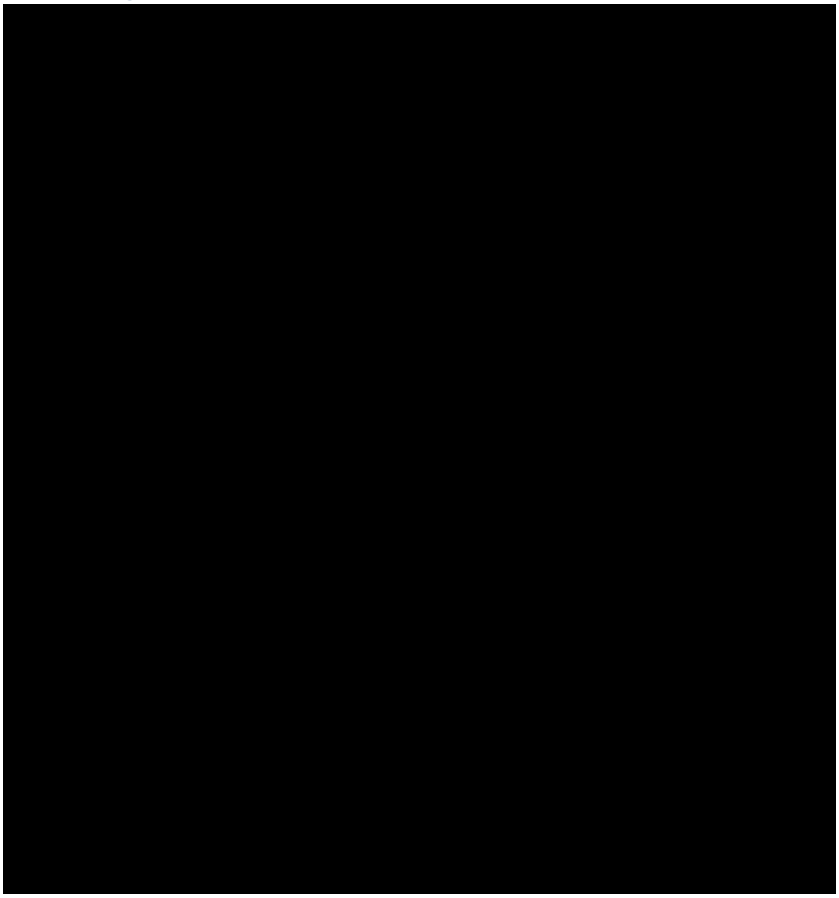
Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Figure 7: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 5 mg versus comparators, random effects model



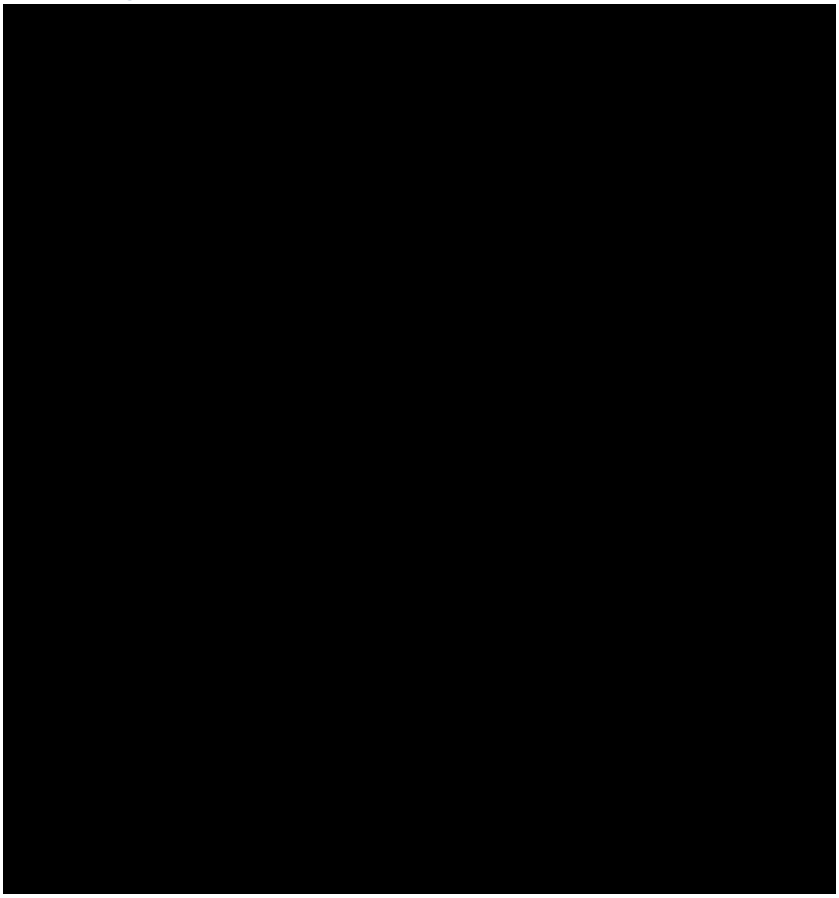
Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 8: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 10 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 9: Forest plot (median difference [95% CrI]) for weight (kg) change from baseline, TZP 15 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Table 2: Pairwise results (median difference [95% CrI]) table for weight (kg) change from baseline, random effects model; tirzepatide 5/10/15 mg (column) vs comparators (row)

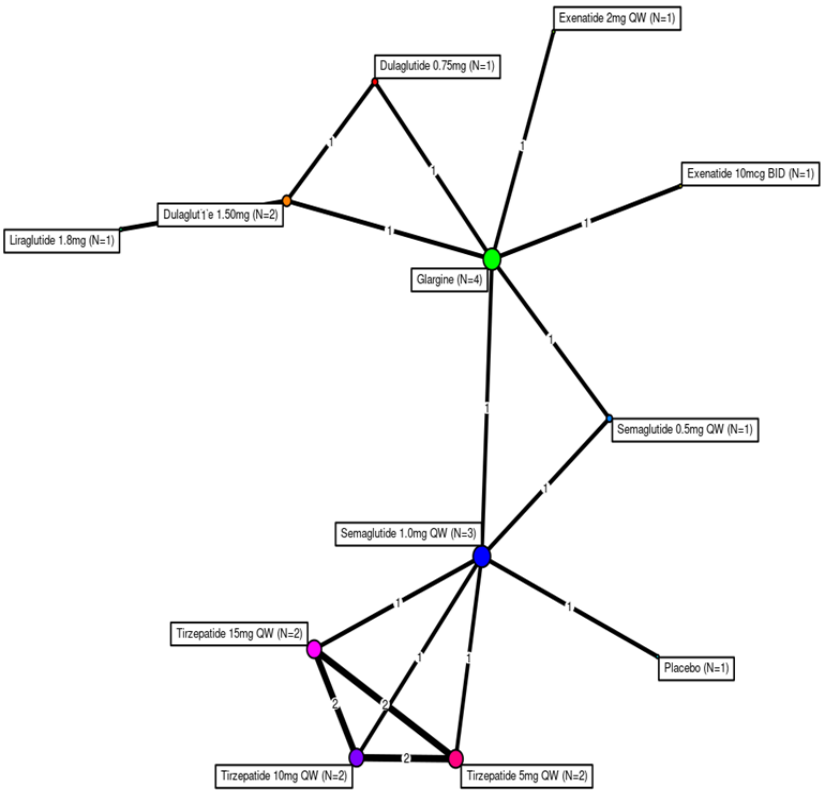
Column vs. row	TZP 5 mg		TZP 10 mg		TZP 15 mg	
	Main	Sensitivity	Main	Sensitivity	Main	Sensitivity
Placebo						
Tirzepatide 5 mg QW						
Tirzepatide 10 mg QW						
Tirzepatide 15 mg QW						
Semaglutide 0.5 mg QW						
Semaglutide 1.0 mg QW						
Liraglutide 1.2 mg						
Liraglutide 1.8 mg						
Dulaglutide 0.75 mg						
Dulaglutide 1.50 mg						
Dulaglutide 3.0 mg						
Dulaglutide 4.5 mg						

Column vs. row	TZP 5 mg		TZP 10 mg		TZP 15 mg	
	Main	Sensitivity	Main	Sensitivity	Main	Sensitivity
Semaglutide 7.0 mg QD						
Semaglutide 14.0 mg QD						
Exenatide 2 mg QW						
Exenatide 5 mcg BID						
Exenatide 10 mcg BID						
Lixisenatide 20 mcg						

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

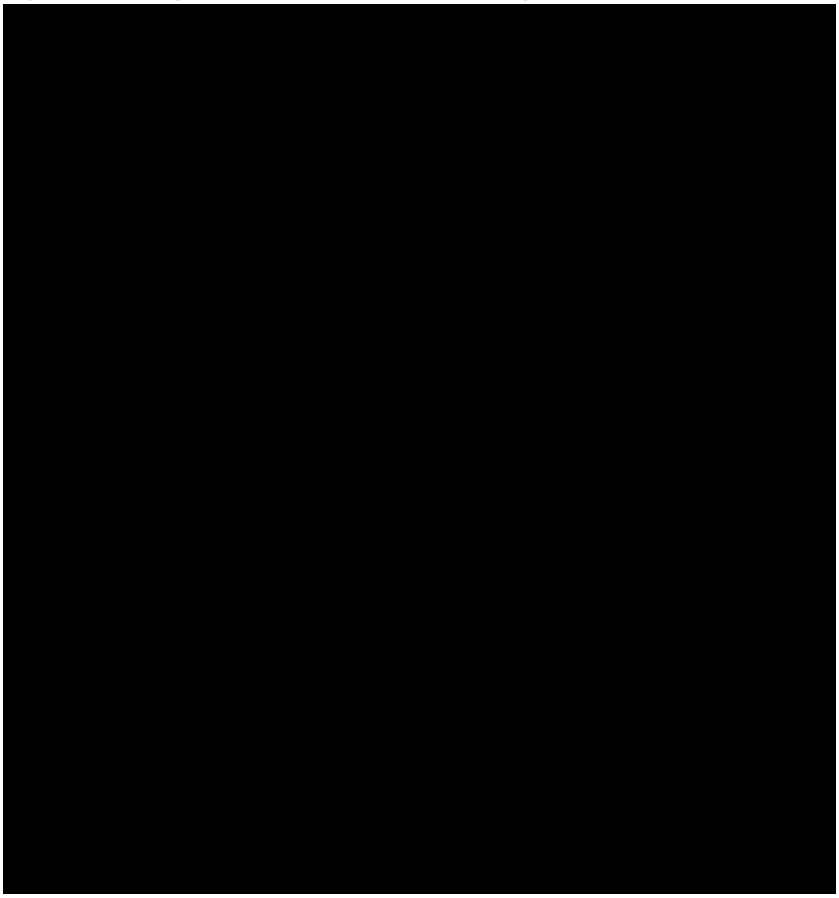
BMI

Figure 10: Sensitivity analysis network for body mass index (kg/m2) change from baseline



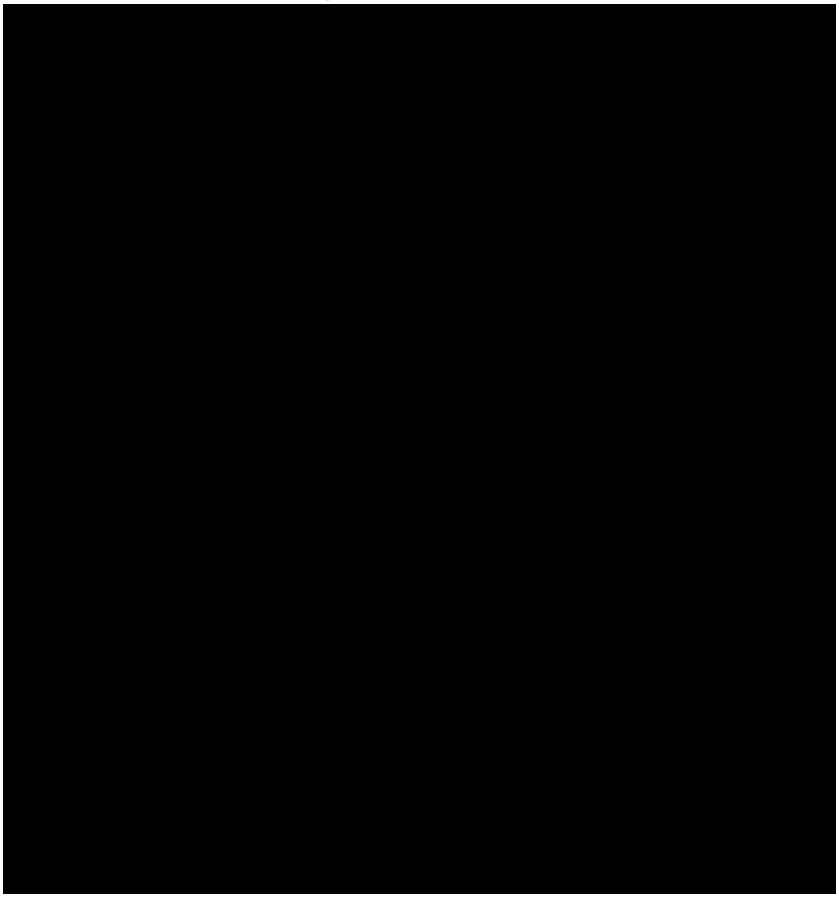
Abbreviations: BID: twice a day; QD: once a day; QW: once a week.

Figure 11: Figure 46: Forest plot (median difference [95% CrI]) for body mass index (kg/m2) change from baseline, TZP 5 mg versus comparators, random effects model



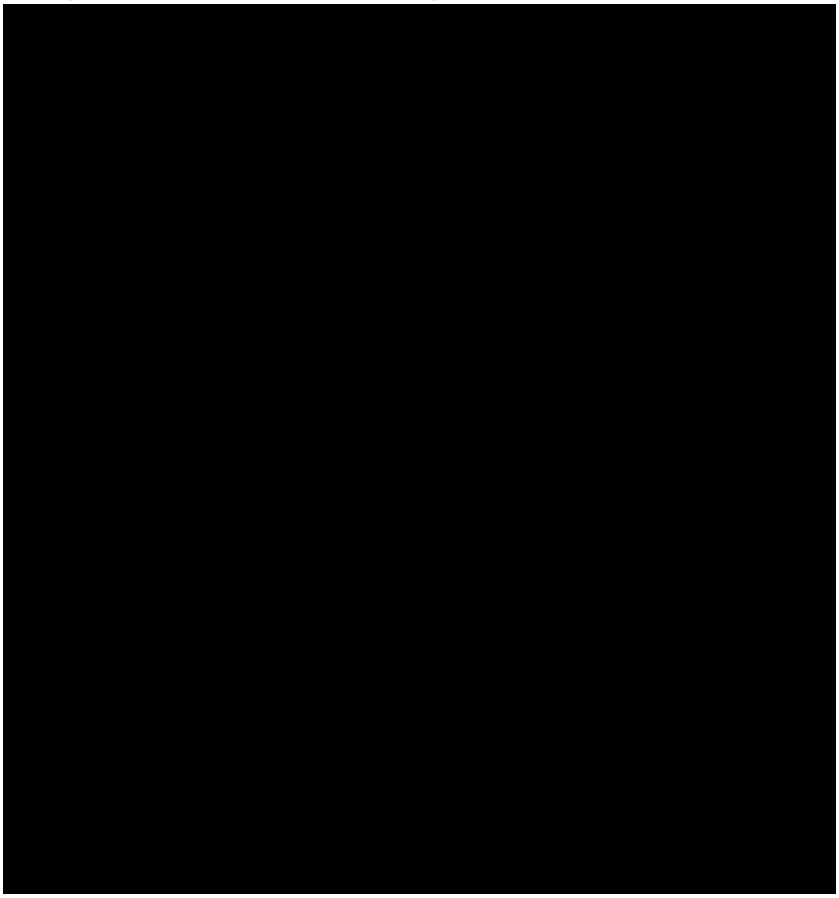
Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 12: Forest plot (median difference [95% CrI]) for body mass index (kg/m2) change from baseline, TZP 10 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Figure 13: Figure 48: Forest plot (median difference [95% CrI]) for body mass index (kg/m2) change from baseline, TZP 15 mg versus comparators, random effects model



Footnotes: Semaglutide 2.0 mg is not currently available in NHS practice but is included in the NMA for completeness. Degludec, glargine, sitagliptin 100 mg and glimepiride were used as nodes to connect other treatments and were not included in the PICOTS. Hence the results for these treatments shown in the figures should not be interpreted.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

Table 3: Pairwise results (median difference [95% CrI]) table for body mass index (kg/m^2) change from baseline, random effects model; tirzepatide 5/10/15 mg (column) vs comparators (row)

Column vs. row	TZP 5 mg		TZP 10 mg		TZP 15 mg	
	Main	Sensitivity	Main	Sensitivity	Main	Sensitivity
Placebo						
Tirzepatide 5 mg QW						
Tirzepatide 10 mg QW						
Tirzepatide 15 mg QW						
Semaglutide 0.5 mg QW						
Semaglutide 1.0 mg QW						
Liraglutide 1.8 mg						
Dulaglutide 0.75 mg						
Dulaglutide 1.50 mg						
Semaglutide 14.0 mg QD						
Exenatide 2 mg QW						
Exenatide 10 mcg BID						

Footnotes: Cells highlighted in green show comparisons which significantly favour TZP (dose according to column header); Cells highlighted in red show comparisons which favour other TZP doses or active treatments.
Abbreviations: BID: twice a day; QD: once a day; QW: once a week; TZP: tirzepatide.

The company would also like to bring attention to the sensitivity analyses and decisions made within the main analysis to mitigate bias, study variation and test assumptions held, including choosing a random effects model for the main analyses and running meta-regression adjusting for treatment effect modifiers as well as baseline risk. All of these additional analyses yielded similar results to the main analyses and did not result in better model fit, therefore suggesting that the main analysis of the NMA produced results robust to the observed variation in study design, patient population, background therapy and baseline characteristics. The regression coefficient for most of the meta-regression models was not significant. Significant regression coefficients were only seen for HbA1c and body weight both adjusting for HbA1c, in both cases the regression coefficient was close to zero; these results are supported by the similar pairwise results produced by the model.

Additional NMA issues

The company would also like to respond to some additional issues raised by the EAG in regards to the NMA.

- The EAG noted that the model fit statistics for the meta-regression model adjusting for baseline risk were not reported. For transparency, these statistics are now presented in Table 4. To note, as mentioned in Section B.2.9.5.4 of the company submission, meta-regression models adjusting for analysis time window and baseline covariates resulted in convergence issues and were therefore not run.

Table 4: Model fit statistics for the meta-regression model adjusting for baseline risk

Endpoint	Analysis	Residual deviance	DIC	Number of data points	Regression coefficient (95% CI)	Sigma (95% CI)
HbA1c	Main analysis	113.02	209.04	114	-	0.13 (0.09, 0.17)
	Adjusted for baseline risk	110.25	207.63	114	−0.50 (−0.77, −0.21)	0.12 (0.09, 0.16)
BW	Main analysis	103.87	190.74	114	-	0.43 (0.26, 0.62)
	Adjusted for baseline risk	104.58	192.84	114	−0.18 (−0.42, 0.09)	0.43 (0.27, 0.62)
BMI	Main analysis	41.09	76.88	41	-	0.18 (0.06, 0.35)
	Adjusted for baseline risk	40.67	76.53	41	−0.34 (−0.67, 0.03)	0.16 (0.03, 0.31)

Abbreviations: BMI: body mass index; BW: body weight; CI: confidence interval; DIC: deviance information criterion; HbA1c: glycated haemoglobin

- The EAG stated that '*heart rate and total cholesterol were not included*' in comparisons presented between studies that contributed to heterogeneity. The company would like to clarify that heart rate and total cholesterol were not included as these endpoints were not part of the company submission and reference to this was mentioned in error.

- In reference to the sensitivity analysis exploring the inclusion of studies with an Asian population, the EAG noted a discrepancy between studies reported in the file provided and the write up. Namely, that reference was made to the following studies within the file:

Ji 2013, Kadowaki 2011, Li 2014. The company can confirm that these studies were not included in any sensitivity analyses as although they were conducted in an Asian population, some patients within these studies received three OADs as background therapy, meaning that the studies were not aligned with the network definition. Additionally, the studies were not eligible for inclusion within the sensitivity analysis investigating studies with three OADs as background therapy as they were conducted in an Asian population.

- The company submission noted that *'in an early exploratory analysis, heterogeneity was identified in trials with insulin glargine for change from baseline in HbA1c and weight'*, the EAG requested further information on this. The company can confirm that during the feasibility assessment heterogeneity was identified in two studies (LEAD-5 and EAGLE) due to difference in insulin titrations that could impact HbA1c and weight results. Considering there were other studies with insulin glargine, the company decided to broaden the review to all studies with an insulin glargine arm. From this review it was identified that end total insulin dose could vary between studies and this could impact results for HbA1c and weight. The decision was made to run a sensitivity analysis excluding studies with insulin glargine arms, as described in Section B.2.9.5.3 in the company submission. Results from this sensitivity analysis were consistent with the main analysis, supporting that removal of these studies did not have a material impact on results.

- The EAG stated that '*in response to the clarification letter, the company has provided an Excel file reporting background OADs for all trials/arms considered in all of the NMAs, but it is unclear which were part of each of the main NMAs and which the sensitivity analyses.'* The company would like to clarify that information on which studies were included as part of the main analysis and which were included as part of the sensitivity analysis is provided in column L of the dataset provided as part of the response to A24 with reference made to the data provided in response to A23 on background OADs.

- Among the studies found to be eligible for inclusion within the NMA, the EAG noted that '*in eight of these studies a GLP-1 treatment was compared to a treatment that did not connect to the rest of the network and therefore these were excluded from the NMA. The company has excluded these treatments from all endpoints in the network as not treatments of interest'.* The company note that these studies were connected but were branching out and were therefore excluded as they did not add additional information to the network.

- In reference to the model-based NMA, the EAG report notes that *'the CS does not report which time-course model was found to fit better and was therefore used in the analysis.*' The company can clarify that in the model-based approach, various candidate time-course models were fitted with data up to 1 year for a fixed model. Candidate time-course models included linear, Emax, quadratic, exponential, equally spaced piecewise linear (knots at week 13, 26 and 39) and unequally spaced piecewise linear (knots at week 6.5, 13 and 26 were a better match of the shape of curves in the first 6 months). Studies were included in this analysis when at least three timepoints within 52 weeks were reported for the endpoint. Model fit statistics were compared, and the candidate model fit was also examined graphically, and the best time-course model was selected for the fixed model. For endpoints below, the chosen time-course models were the unequally spaced piecewise linear:

- Change from baseline in HbA1c
- Change from baseline in weight

A random effect model was then fitted using the more appropriate time-course model. The final chosen model was decided by comparing the DICs between fixed effect and random effect. The model fit of the final chosen model was also examined by comparing the posterior mean residual deviance to the total number of data points, and the DICs between inconsistency model and the consistency model. For both endpoints the final chosen model was the random effect model with unequally spaced piecewise linear time-course model.

- As noted in the factual accuracy check, the EAG incorrectly stated in regard to estimands within the NMA that, apart from the SURPASS trials and six other studies, in most studies *'the estimand was not available as this concept is relatively new and therefore the only available results based on the treatment-regimen estimand were used instead.'* To provide further clarity around this point, a detailed review was performed for the studies included in the network, investigating the language used in the publications regarding the handling of data after initiation of rescue therapy. Among the 45 studies included in the main analysis, the estimand was clearly defined in 10 studies: AWARD-10, AWARD-11, AWARD-2, AWARD-6, LIRA-ADD2SGLT2i, PIONEER 3, PIONEER 4, SURPASS-2, SURPASS-3, SUSTAIN-FORTE. For these studies data from the efficacy estimand was used within the NMA. In 11 studies, the estimand was not defined but there was clear language indicating data was excluded after initiating rescue therapy, thus deemed similar to the efficacy estimand. These studies were: AWARD-1, GetGoal-P, GetGoal-S, LIRA-SWITCH, Nauck 2016, SUSTAIN 2, SUSTAIN 3, SUSTAIN 4, SUSTAIN 7, SUSTAIN 9, Van Gaal 2014. For the remaining 24 studies, the estimand was not defined and no clear language was included regarding the handling of data after initiation of rescue therapy, in these studies only the available data was included.

- The company would like to note that the meta-regression model was only run adjusting for OADs as other models had convergence and auto-correlation issues. These issues were seen in models adjusted for assessment time window, baseline HbA1c and body weight and are detailed further in a file included in the reference pack entitled 'Auto-correlation issues'.[10] However, in order to ensure the assessment time window was appropriate, two sensitivity analyses were conducted as described in the Document B (Section B.9.5.3) of the company submission.

A.1.3 Cost-effectiveness issues

Key issue 6: Model approach adopted by the company

Description of issue and why the EAG has identified it as important	The EAG did not find compelling justification to support the company's modelling approach. This includes 1) the use of the PRIME T2D model in general instead of commonly used available alternatives mentioned such as the CORE Diabetes Model that was used for NICE Guideline NG28 focusing on the management of T2D and 2) the selected model type, described as a "discrete time event" model instead of commonly used model types such as a DES or individual-patient state transition model. Moreover, the exact technical implementation of the model was not clear to the EAG, this becomes even more problematic when deviating from commonly used model types.
What alternative approach has the EAG suggested?	The EAG preference would entail either using the CORE Diabetes model (consistent with NICE Guideline NG28) or providing extensive justification with supporting evidence why the company deviated from this approach. Additionally, when deviating from the NICE Guideline NG28 modelling approach, the impact of this should be assessed. Similarly, the EAG would prefer commonly used model types such as a DES or individual-patient state transition model instead of a "discrete time event" model. Moreover, deviating from commonly used model types requires substantial and compelling justification as well as detailed description of the model implementation.

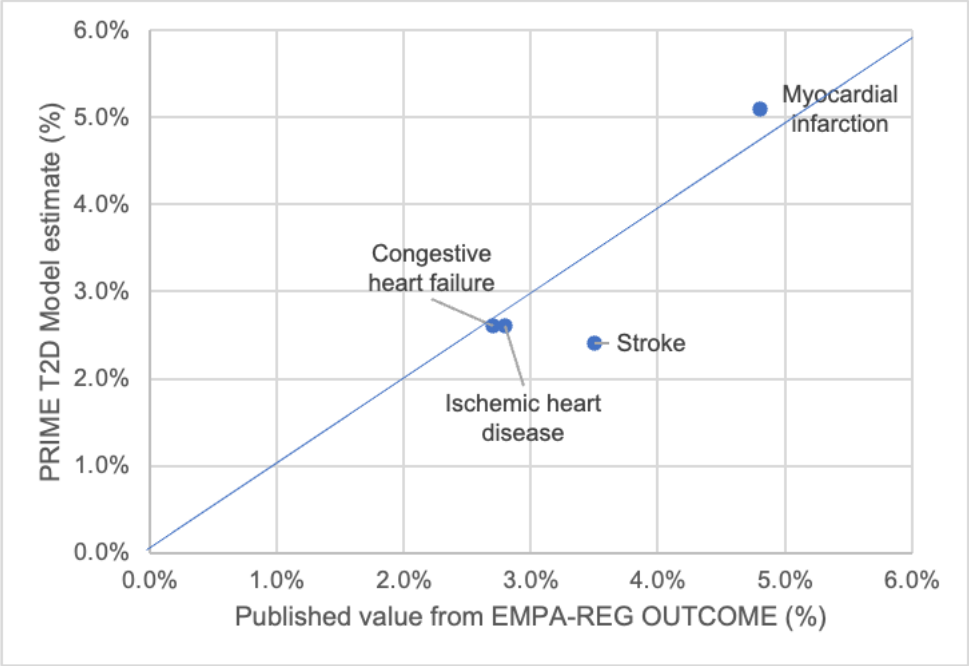
Choice of model: The PRIME T2D Model was chosen for the analysis as, based on published evidence, it may be better suited to predicting long-term outcomes for type 2 diabetes patients receiving GLP-1 receptor agonists than the CORE Diabetes Model or the UKPDS OM2 (details below).[11-13] It should be noted that the CORE Diabetes Model was not used to support the preparation of NICE Guidelines NG28 as intimated by the EAG (a calibrated version of the UKPDS OM2 was used), but the CORE Diabetes Model was used to provide supporting information in economic modelling efforts on periodontal treatment in type 1 and type 2 diabetes (NG17 and NG28) and on continuous glucose monitoring in adults with type 2 diabetes (NG28).

As outlined previously in the original submission and in the response to clarification questions, the CORE Diabetes Model and UKPDS OM2 performed poorly in validations against cardiovascular outcomes trials at the Ninth Mount Hood Challenge Meeting published in 2020.[13] For example, prior to calibration the CORE Diabetes Model underpredicted the risk of stroke by around 54% and the UKPDS OM2 overpredicted the risk of myocardial infarction by 27% in the active treatment arm of EMPA-REG (see Appendix 2 in Si *et al.*).[13] Without appropriate calibration, there is a risk that these models may under/overestimate the risk of diabetes-related complications in a cost-effectiveness evaluation, particularly when agents such as GLP-1 receptor agonists are involved that may alter cardiovascular risk profiles. Crucially, the calibration of existing type 2 diabetes model with hazard ratios from CVOTs is a complex challenge with considerable potential to provide misleading results when comparing multiple interventions as recently summarized by Evans *et al.* (2023).[11] Main concerns focus on the heterogeneity of the trials, with different study durations, inclusion criteria, rescue medication protocols and endpoint definitions, which results in significant uncertainty when comparing two or more interventions

evaluated in separate CVOTs, as robust adjustment for these differences is very challenging. This is compounded by differences in endpoint definitions in a given model (which need to match those in the CVOT to be suitable for calibration) and the challenge of double-counting treatment effects (the hazard ratios from CVOTs are typically not adjusted for improvements in conventional risk factors such as HbA1c). The use of unadjusted hazard ratios from multiple CVOTs in a long-term cost-effectiveness analysis has considerable potential to skew the outcomes if these challenges are not appropriately addressed. As outlined by Evans *et al.* it is likely that these challenges can only be overcome by combining patient-level data from CVOTs to prepare novel risk equations that can better model modern therapies for type 2 diabetes.[11] At this moment in time, there are no published data that would allow the appropriate calibration of the UKPDS OM2 or CORE Diabetes Model for the present analysis.

In contrast, the published validation results for the PRIME T2D Model indicate that the PRIME T2D Model may be better placed to predict cardiovascular outcomes in line with those observed from recent CVOTs without calibration.[12, 13] For example with the EMPA-REG OUTCOMES trial (c.f. CORE Diabetes Model and UKPDS OM2 results above), the root mean squared difference for four endpoints in the active treatment arm was 0.7%, with the PRIME T2D Model generally matching published outcomes well, although slightly underestimating the risk of stroke (see figure below and further details were provided in the PRIME T2D Model Technical Report as part of the original submission).

Figure 14: PRIME T2D Model validation scatterplot for the EMPA-REG OUTCOME study



The PRIME T2D Model has also been shown to validate well against the UK-based Lipids in Diabetes Study.[12] Shortcomings associated with the UKPDS OM2's ability to predict cardiovascular risk in a modern UK population over 10 years of follow up (ASCEND study) were recently highlighted by Keng et al. (2022), where the authors outlined, in particular, the lack of a revascularization endpoint and poor performance in older patients as key challenges with the UKPDS OM2.[14] The authors also cite earlier diagnosis and improved risk factor control in modern diabetes care as potential reasons for poorly predicted outcomes (UKPDS data were collected between 1977 and 2007). The approach used with the PRIME T2D Model allows for

inclusion of a revascularization endpoint and integration of more recent data for risk evaluation (via model averaging), which are designed to address these shortcomings. During the development of the PRIME Type 1 Diabetes Model, we were able to show that a model averaging approach, when used to evaluate the risk of cardiovascular endpoints, was superior to any individual risk equations alone.[15] The evidence indicated that risk equations performed well in validations against the derivation populations (or similar populations) but poorly in populations with different characteristics or risk profiles. This is the essential tenet of the model averaging approach: risk equations are weighted to match the risk profile of individual patients to avoid the situations where risk equations from low risk populations (e.g. UKPDS) are applied to high risk patients (e.g. patients in a simulation with long duration of diabetes, advanced disease, history of complications and elevated risk factors). Importantly, validation results to date with the PRIME T2D Model support a model averaging approach in type 2 diabetes.[12]

Unclear technical implementation: In addition to the publication and technical report on the PRIME T2D Model, the EAG were provided full access to the model source code as well as online access via the model interface (it is perhaps noteworthy that the source code would not be available for review with the CORE Diabetes Model, the model suggested by the EAG). Technical support was available throughout the review period and all clarification questions were answered in full. The company also made the model developers directly available to the EAG to assist with understanding of the model and its functionality.

Deviation from the NICE Guideline NG28 modelling approach: Conceptually, the PRIME T2D Model is not a deviation from the modelling approaches cited by the EAG (specifically the UKPDS OM2 and the CORE Diabetes Model). All three models are patient-level simulations that rely, primarily, on publish risk equations to evaluate the risk of complications and mortality, and are capable of integrating country- and/or population specific costs and utility data. This approach is consistent across almost all the models of type 2 diabetes in the Mount Hood Registry (https://www.mthooddiabeteschallenge.com/registry)[16] and presented at Mount Hood Challenge meetings. The EAG statement that the model is a deviation from previous approaches used by NICE or diabetes modelling in general is factually incorrect. The term “discrete time event model” used to describe the PRIME T2D Model has not been correctly understood by the EAG. This description is analogous to the term “discrete-time illness-death“ model used by Clarke *et al.* to describe the UKPDS Outcomes Model; this type of model was used by NICE to support the recent preparation of NG28 guidelines.[17] We deliberately avoided the term “discrete event simulation (DES)” as it is synonymous with a series of ‘events’ that occur over time (as opposed to events occurring within an annual cycle) and, perhaps more crucially, assumes no change in the system between events.[18] Similarly, the PRIME T2D Model was not described as a state-transition model (STM) because it is not aligned with the conventional definition of an STM as simulated patients may be members of multiple "states" simultaneously (e.g. CKD stage 3, heart failure, and history of MI) without the model explicitly incorporating the notion of any composite states.[19] Further, every effort was made during the cost-effectiveness analysis to use data inputs consistent with the NICE modelling approach for Guidelines NG 28 (including costs, utilities and treatment intensification assumptions) as outlined in the original submission. We would therefore question the EAG use of the term “deviation.”

Key issue 7: Selection and use of risk models to estimate complications

Description of issue and why the EAG has	For the estimation of macrovascular complications and blindness risks, the company adopted a model averaging approach, the justification for this approach was not compelling to the EAG. Moreover, the appropriateness of the selected predictive models to

identified it as important	estimate the risk of complications in patients with T2D is not justified (in detail), nor is the applicability to the specific decision problem justified.
What alternative approach has the EAG suggested?	Provide extensive justification for the selection and use of risk models to estimate complications and scenario analyses to examine the impact of the adopted approach.
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Scenario analyses selecting single predictive models based on the best match of the derivation cohort to the decision problem (as requested in clarification question B4c). Moreover, extensive justification for the model averaging approach, the selected predictive models and the applicability to the specific decision problem.

As outlined in the response to key issue 6 (above), the justification for the approach to estimation of macrovascular complication risk is primarily based on the published validation work using the model averaging approach with the PRIME T2D Model and the shortcomings of alternative approaches, which are also documented in the published literature. The approach to (and justification for) model averaging in the PRIME T2D Model was previously detailed in the response to clarification questions (question B.4) but has also been summarized below.

The PRIME T2D Model has been validated against cardiovascular outcomes trials, including EMPA-REG OUTCOME (empagliflozin), REWIND (dulaglutide), LEADER (liraglutide), and DEVOTE (insulin degludec), using the model averaging approach, and been shown to compare well to published outcomes.[12]

Several other models of type 2 diabetes have not performed as well in published validations. For example, at the ninth Mount Hood Challenge, the CORE Diabetes Model and UKPDS OM2 did not reliably reproduce trial outcomes without calibration (details above).[13] Given the heterogeneous nature of the cardiovascular outcomes trials, methodological challenges with calibration (e.g. assumptions of non-changing hazards over time) and the fact that cardiovascular outcomes trial data is not yet available for tirzepatide, appropriate calibration was not possible for the present health economic evaluation.[11]

The importance of a revascularization endpoint for successful prediction cardiovascular risk in a modern UK population was recently outlined by Keng et al. (2022). Specifically, Keng et al. noted that: Patients are now more likely to receive preventive coronary revascularization possibly contributing to the decline in the risk of MI and the apparent increase in the risk of other IHD (which includes coronary revascularization). Excluding coronary revascularization from the other IHD endpoint leads to a similar pattern of overprediction (see Appendix Fig. 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.09.005). Furthermore, the observed cumulative incidence of a combined endpoint of MI or coronary revascularization (assuming that coronary revascularizations prevented some MIs) was also lower than the cumulative incidence of MI predicted by the UKPDS-OM2. These exploratory analyses suggest that the UKPDS-OM2 is overpredicting the risk of IHD in general and that other IHD seems well predicted partly

because of the higher rate of coronary revascularizations in ASCEND. The PRIME T2D Model includes a revascularization endpoint in contrast to the CORE Diabetes Model and the UKPDS OM2, neither of which include revascularization.[14]

The model averaging approach in the PRIME T2D Model includes the BRAVO, UKPDS OM2, and Yang et al. macrovascular risk models, which can be parameterised with cohort, risk factor, and treatment effect data from the cohort and trial results of interest (e.g. SURPASS-2). The product and trial-agnostic nature of the PRIME T2D Model necessitates this approach, and model averaging allows the model to derive weights on a per-patient basis to tailor the overall modelling approach to the cohort; in the absence of risk equations derived directly from the trial or trials in question, we consider this approach to be preferable to the selection of a single risk model parameterised from a different population than that under investigation (an approach commonly employed elsewhere in diabetes modelling efforts). In addition to addressing concerns around the structural uncertainty inherent in using a single specific risk model, the approach allows the model to adapt risk estimation to difference populations at different stages of disease progression. Establishing which model is the "best match" to the decision problem is challenging, hence the default approach of allowing PRIME to weight the use of risk models automatically. The most prominent diabetes risk models (e.g. UKPDS OM1, UKPDS OM2, the IQVIA Core Diabetes Model, and the Cardiff Model) are all based — at least in part — on the UKPDS population, which was a population with newly-diagnosed type 2 diabetes, with the first patients enrolled in 1977, prior to the existence of statins, insulin analogs, SGLT-2 inhibitors, or GLP-1 receptor agonists. The incorporation, through a model averaging framework, of risk models derived from more modern populations of patients such as ACCORD (in the BRAVO model) and the Hong Kong Diabetes Registry (in the Yang et al. risk equations) allow the model to tailor the weighting of each model to each simulated patient. We believe this approach to be better suited to the decision problem than selecting a single model as the basis of the analysis and validation analysis indicates that the approach may be better suited to predicting long-term clinical outcomes in a modern type 2 diabetes population.

Key issue 8: Extrapolation of treatment effectiveness

Description of issue and why the EAG has identified it as important	The QALY gains are predominantly accrued after the first year (i.e., beyond the trial time horizon) and mostly likely related to utilities for weight. Hence the extrapolation of (treatment) effectiveness is an important aspect of the model. The company made a simplifying assumption of constant risk factors (i.e., no risk factor progression) for SBP, HDL, LDL and weight (i.e., BMI) after year 1 up to treatment intensification. Moreover, the company did assume no waning of the relative treatment effect while on the initial treatment (i.e., before switching to basal insulin therapy).
What alternative approach has the EAG suggested?	The EAG would prefer assuming UKPDS OM2 risk factor progression for all risk factors (instead of assuming these being constant after the first year up to switching to basal insulin therapy). Moreover, additional justification for assuming no waning of the relative treatment effect (before switching to basal insulin therapy) is warranted.
What is the expected effect on the cost	The alternative approach suggested by the EAG likely increases the estimated ICER.

effectiveness estimates?	

As detailed in the text below, there are considerations around individuals risk factors and the equations recommended by the EAG that should be factored into assumptions on risk factor progression:

- HbA1c, shown to be one of the main drivers of cost-effectiveness, follows UKPDS progression and has a treatment waning effect in the submitted modelling analysis
- An assumption of treatment waning with respect to the effects on body weight (or BMI) or SBP for GLP-1 receptor agonists is at odds with the published literature and could bias a cost-effectiveness evaluation (details below)[20-22]
- UKPDS risk factor progression has been incorporated into the EAG preferred base case analysis in all cases where it is not contradicted by the available clinical evidence

The comments made by the EAG are not wholly correct and do not take into account the differences between expected progressions for different risk factors. The statement that "the company did assume no waning of the relative treatment effect while on the initial treatment" does not apply to HbA1c, shown to be one of the main drivers of cost-effectiveness, which followed a progression based on UKPDS data that led to a relative reduction in treatment benefit whilst on treatment. Changes in body weight (or BMI) were found to be the other key driver of cost-effectiveness and, in line with published data on GLP-1 receptor agonists, weight loss was assumed to be maintained during therapy (with BMI remaining constant) but returned to baseline levels on in the first year after switching to basal insulin (conservative assumption).[23] Similar assumptions were applied to the progression of SBP over time. It is important to note that the EAG assertion of a waning treatment effect on body weight (or BMI) or SBP is at odds with the clinical evidence currently available (and introducing a spurious assumption here would have an impact on cost-effectiveness). Long-term data from the CVOTs for dulaglutide, semaglutide and liraglutide show that body weight and SBP remain stable whilst on GLP-1 receptor agonist therapy, which contrasts to the rapid return to a population mean observed with the UKPDS risk factor progression equations.[20, 22, 24] Due to copyright, the exact figures can not be copied here, however, evidence of this point can be seen in Figure 4 of Gerstein *et al.* 2019 (REWIND), Figure S5 of Marso *et al.* 2016 (LEADER) and Figure 2 and Figure S6 of Marso et al. 2016 (SUSTAIN-6).[20-22] These data indicate that an assumption of no changes in body weight (or BMI) or SBP whilst on therapy with a GLP-1 receptor agonist is better matched to the available clinical data. The UKPDS OM2 risk factor progression equations for SBP and BMI are shown below:

- SBP increases to a level approximately 4 mmHg above baseline in the first 5 years of the simulation
- BMI increases gradually over time without having a notable waning effect (on between treatment differences) over the first 5 years of the simulation

Figure 15: Example UKPDS OM2 risk factor progression for SBP based on semaglutide 1.0 mg and tirzepatide 10 mg treatments show SBP going increasing in the first 5 years of a modelling simulation

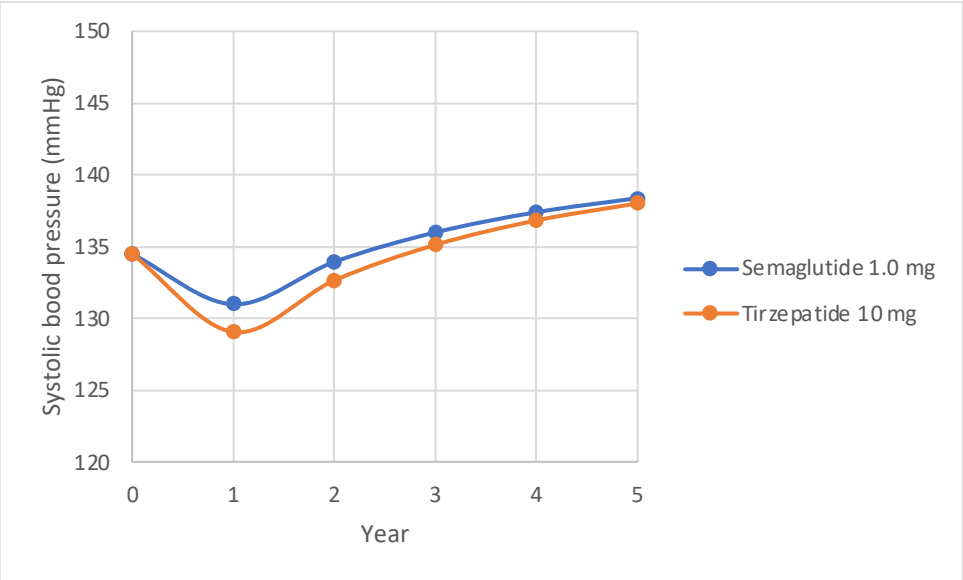
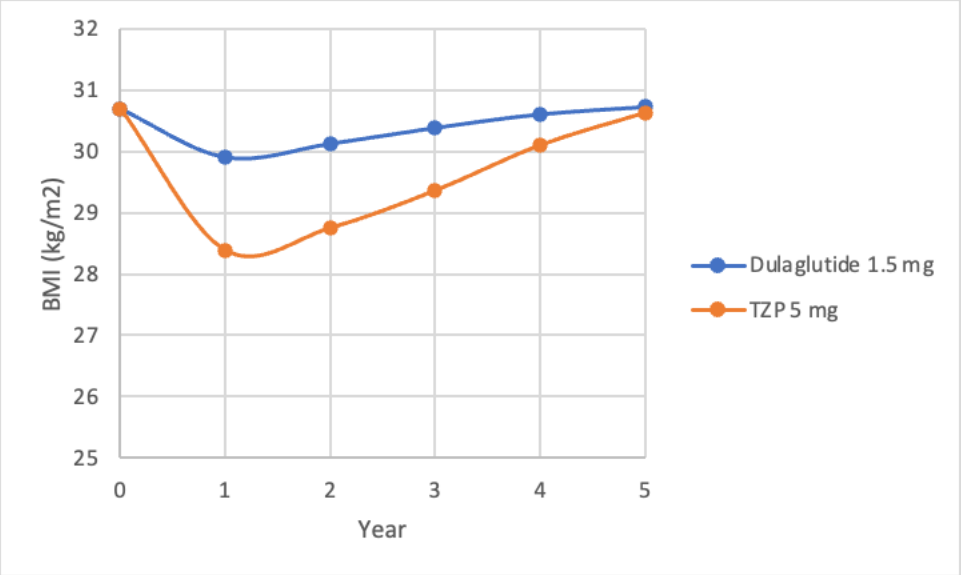


Figure 16: Example UKPDS OM2 risk factor progression for BMI based on dulaglutide 1.5 mg and tirzepatide 5 mg treatments show BMI returning to baseline levels in the first 5 years of a modelling simulation



In light of these observations, and the mismatch of UKPDS OM2 risk factor progression for SBP and BMI relative to the published data from CVOTs, the following assumptions were made for the EAG preferred base case analysis (see EAG Preferred Base Case Simulations section for more details):

- UKPDS OM2 risk factor progression will be assumed all risk factors whilst on insulin therapy and for HbA1c, LDL, HDL, eGFR, white blood cells count, heart rate and haemoglobin levels whilst on tirzepatide or comparator treatments in line with EAG recommendations

- For SBP and BMI, no change will be assumed whilst simulated patients remain on tirzepatide or comparator therapy, but will return to baseline levels and follow UKPDS OM2 risk factor progression after switching to basal insulin therapy

Key issue 9: Treatment discontinuation/intensification

Description of issue and why the EAG has identified it as important	Patients were assumed to intensify therapy, discontinuing the initial treatment and switching to basal insulin therapy, when HbA1c levels rose above 7.5%. No other reasons (e.g., drug intolerance, patient preferences) for treatment discontinuation were included in the modelling.
What alternative approach has the EAG suggested?	The EAG would prefer including other causes for treatment discontinuation (than reaching the HbA1c threshold).
What is the expected effect on the cost effectiveness estimates?	The alternative approach suggested by the EAG likely increases the estimated ICER.

As previously outlined in the response to clarification questions, assumptions around treatment discontinuation have considerable potential to bias a cost-effectiveness analysis and therefore were not included in the base case for the present evaluation.

In diabetes modelling, the assumption of treatment discontinuation is linked to a rescue therapy (as no treatment is not an option) and the nature of this rescue therapy (in terms of assumed cost and effectiveness) can have a marked impact on cost-effectiveness. For example, an expensive new intervention can have a low ICER in a modelling scenario where there is a high discontinuation rate and a lower cost rescue medication, particularly if that medication lowers HbA1c (as simulated patients can accrue some of the clinical benefits on the new intervention but at lower costs as they switch to the rescue medication). Crucially, there is a paucity of published clinical evidence on the nature of the rescue medication in such scenarios and they are inevitably based on modeller's assumptions. Further, multiple assumptions on treatment discontinuation can lead to complex treatment algorithms (where some patients intensify therapy [to a given treatment] and others switch to rescue medication [to a different treatment]) for the intervention and comparators, making it challenging to ascertain cost-effectiveness due to the interaction of costs and effects from the different therapies involved. This obfuscates the research question of whether the new intervention is cost-effective versus relevant comparators (instead providing information on which assumption-based rescue medications represent the best value for money).

To avoid the potential for rescue medication influencing the outcomes of the modelling analysis, treatment intensification was only modelled using an HbA1c threshold. This is aligned with an assumption that patients who do not tolerate the interventions well are likely to miss doses, leading to poorer glycaemic control and meeting the criterion for intensification. It should be noted that changing intensification criteria had a generally modest effect on cost-effectiveness in sensitivity analysis (CS Table 106) and it can be assumed that modelling discontinuation would similarly have a modest impact on cost-effectiveness provided that a balanced approach to costs and effects was applied to the rescue medication.

Key issue 10: Adverse events: not all incorporated for all treatments

Description of issue and why the EAG has identified it as important	The main concerns of the EAG relate to that only nausea is incorporated (hypoglycaemia only for basal insulin therapy) as an AE. Including hypoglycaemia only for basal insulin therapy might inflate the impact of discontinuing treatment (i.e., treatment intensification) and hereby potentially inducing bias favouring more effective treatments (i.e., tirzepatide).
What alternative approach has the EAG suggested?	As illustrated in clarification response Tables 20-22, incorporating additional AEs would potentially increase the estimated ICER (but might depend on the comparator).
What is the expected effect on the cost effectiveness estimates?	The alternative approach suggested by the EAG likely increases the estimated ICER.

As previously described in the response to clarification questions and detailed below, the approach to modelling adverse events was consistent with previous submissions on GLP-1 receptor agonists given the available data and was in some respects conservative:

Conservatively, a utility for the more severe health state of nausea and vomiting was applied to rates of vomiting in the present modelling analysis (other GI adverse events were not included in previous evaluations of GLP-1 receptor agonists)

As outlined in Section B.2.9 of the CS, GLP-1 RAs are known to be associated with GI AEs, including nausea and vomiting, in the early months of treatment (titration phase). Nausea rates for tirzepatide and all comparators were derived from the NMA and were assumed to negatively impact quality of life in year 1 of the simulation (CS Section B.3.4.5 and Table 81) in the analysis as this aspect of tolerability may have been a differentiator between different GLP-1 RA agonists in the cost-effectiveness evaluation. The NMA provided separate rates of nausea and vomiting with no information on the combined “nausea and vomiting” endpoint. For the base case analysis, it was assumed that: 1) the rate of nausea reported from the NMA would represent the rate of the combined nausea and vomiting endpoint, and 2) a disutility representing the more severe health state of nausea and vomiting would be applied to nausea rates in the analysis (conservative assumption). Sensitivity analysis showed that the impact of nausea utilities on cost-effectiveness outcomes was minimal (CS Table 106). Including both the nausea and vomiting rates from the NMA in the same simulation would have created a risk of double-counting events and biasing the cost-effectiveness evaluation.

As outlined in CS Section B.3.4.4, rates of hypoglycaemia were not reported in the NMA due to many studies reporting zero events; therefore rates of hypoglycaemia were set to zero for tirzepatide and all comparators in the base case analysis. This assumption is likely to be a reasonable approximation for the interventions included in the present analysis based on the very low hypoglycaemia rates observed in the SURPASS trial programme and clinical studies of other T2D medications such as GLP-1 RAs. For basal insulin therapy, hypoglycaemic event rates were aligned with those used in the NICE 2022 health economic report used to inform NG28.

Hypoglycaemia rates were not included in the NMA due to the heterogeneous nature of the published data and therefore could not be included in the base case analysis without the risk of introducing bias. With respect to severe hypoglycaemia (the endpoint “proportion of patients experiencing at least one severe hypoglycaemic event” many studies reported zero events and

was therefore not analysed in the NMA. For non-severe hypoglycaemia (the endpoint "proportion of patients with at least one episode of hypoglycaemia with BG <54mg/dL") there was limited availability of data to connect the network. For the combined hypoglycaemia endpoint ("proportion of patients with at least one episode of hypoglycaemia with BG <54 mg/dL or severe hypoglycaemia), it was not feasible to analyse this endpoint despite the existence of a connected network as studies reported variability due to different background therapies.

Key issue 11: Age-adjustment for utility values: none for older age

Description of issue and why the EAG has identified it as important	The company base-case uses a relatively high utility value for patients with T2D (0.815) and does not adjust utility values for older age. Over time, this potential overestimation will likely only increase as utility values are not adjusted for age.
What alternative approach has the EAG suggested?	The EAG prefers the base-case scenario to include age-adjustment, ensuring that the utility does not exceed the age-matched general population utility.
What is the expected effect on the cost effectiveness estimates?	Using age-adjusted utility values will increase the face validity of the results and will result in a more conservative ICER estimate.

In the submission, the approach to the base case analyses was to be consistent wherever appropriate with the NICE NG28 health economic analysis. With specific reference to the use of age-adjusted utilities, this was not included in the NICE health economic analysis and therefore did not form part of the submitted base case (but it was explored in sensitivity analysis). In response to the EAG preference on age-adjusted utility, this approach has been used for the EAG preferred base case simulations as described later in this response document.

Key issue 12: Discrepancies related to utility and cost values

Description of issue and why the EAG has identified it as important	There are discrepancies in the uncertainty measures and distributions related to utility values and costs listed in the CS and those listed in the original sources.
What alternative approach has the EAG suggested?	According to the EAG, all input data should be in line with the data presented in the original sources. This includes deterministic values, measures of uncertainty and appropriate distributions.
What is the expected effect on the cost effectiveness estimates?	Uncertain. The discrepancies in the uncertainty measures and distributions related to utility values will either in- or decrease the uncertainty surrounding the model outcomes. Costs mentioned in the CS are both higher and lower than those reported in the original sources, therefore the combined effect on the ICER is difficult to determine.

It should be noted that the changes requested by the EAG with respect to this point will not influence the base case cost-effectiveness outcomes reported, but only the PSA simulations and, in particular, the variance around those results. For the EAG preferred base case simulations (detailed later in this response document), every effort has been made to match the uncertainty measures and distributions with the source data although this is very challenging given the sporadic nature of reporting variance (particularly around costs) and almost none of the data sources for costs or utilities describing distribution shapes or forms.

Key issue 13: Potentially inappropriate PSA

Description of issue and why the EAG has identified it as important	The implementation of the PSA is not clear and includes bootstrapping that is not standard in PSAs. It is unclear whether all imprecision (i.e., all uncertain parameters) is taken into account the PSA, and whether stochastic uncertainty is removed from the PSA.
What alternative approach has the EAG suggested?	Implementation of the PSA according to Corro-Ramos et al 2020.[25]
What is the expected effect on the cost effectiveness estimates?	Unknown.
What additional evidence or analyses might help to resolve this key issue?	Detailed step-by-step explanation of implementation of the PSA.

Probabilistic sensitivity analysis (PSA) in the PRIME T2D Model was implemented on the principle that PSA should capture uncertainty around all aspects of the simulation, thereby characterising the full extent of the uncertainty around the modelled answer to the decision problem, rather than only a subset of that uncertainty arising from probability distributions around model parameters/coefficients alone. We understand this approach to align with the high-level objective of PSA as specified in "NICE health technology evaluations: the manual", in which NICE note that the economic evaluation should "present an overall assessment of uncertainty to committees to inform decision making".

The aspects of randomness, heterogeneity, and uncertainty that are captured in the PRIME Diabetes Model cover patient heterogeneity, "random walk" through the model (i.e. stochastic or first-order uncertainty as described by Briggs et al.),[26] intra-cycle ordering clinical event risk exposure, model coefficients (e.g. beta coefficients of regression equations), and other user-specified model inputs (treatment effects, costs, and utilities). Note that the NICE glossary defines "parameter uncertainty" as "Uncertainty about the mean values of parameters (for example, health outcomes, utilities and resource use) included in the model". Health outcomes in the PRIME T2D Model are governed by patient characteristics, treatment effects, regression equation coefficients, and uniform distribution sampling; all of the latter sources of uncertainty in the above list (model coefficients and user-specified model inputs) could therefore fall under this definition.

When all of these random and distributional phenomena are captured simultaneously (as is the case in a "default" PSA), the PRIME T2D Model thereby generates an analysis that characterizes the full extent of uncertainty around the modelled answer to the decision problem in the target population, rather than the uncertainty around model parameters alone; most notably, it captures the effects of the interactions between patient heterogeneity and all other parameter uncertainty in the model. However, it should be noted that, in the interests of enabling PSA to be run to suit the expectations and requirements of different stakeholders, every class of uncertainty can easily be "gated" either in the code and/or the model parameters, allowing, e.g. patient heterogeneity or model parameter input sampling to be disabled independently. It is unclear if the EAG have explored these possibilities by modifying the Java source code, but our current understanding is that no such modifications or exploratory analyses have been conducted or even attempted by the EAG.

When PSA is active, the following parameters and functions in the model are sampled or otherwise randomised:

- Patient characteristics (i.e. patient heterogeneity), based on user input mean and standard deviation values.
- Sub-model execution order to reduce any systematic bias introduced by sub-models running consistently in the same order. The EAG's apparent ignorance of the importance of random sub-model ordering in reducing bias is profoundly concerning.
- Sub-model coefficients, based on published standard errors around the coefficient values from the respective publications on which the sub-models are based.
- Treatment effects, based on user input mean and standard deviation values.
- Costs, based on user input mean and standard deviation values.
- Utilities, based on user input mean and standard deviation values.

Once a simulation capturing all of the above randomness, heterogeneity, and uncertainty is complete, the uncertainty around the economic and quality of life outcomes is evaluated using a non-parametric bootstrap, which is performed on the in-memory simulated cohort, implemented using parallel Java streams in the ResultsController Java class. The non-parametric bootstrap treats the in-memory patient population as a large-scale clinical trial or registry analysis cohort, with the bootstrap iterations performed to yield the expected outcomes from a cost-utility analysis, namely the incremental cost and QALY estimates required to generate a cost-effectiveness scatterplot and acceptability curve.

With sufficient patients, this approach of simultaneously sampling from all sources of heterogeneity and uncertainty in the model gives stable results in PRIME T2D Model; Monte Carlo error is greatly reduced by the substantial number of patients that can be simulated in PRIME T2D Model in relatively short amounts of time (600,000 per simulation in the company submission base case). This computational efficiency is attributable to the statically-typed and bytecode-compiled nature of the Java programming language, in addition to the threaded implementation of the core simulation and the highly-parallelised implementation of the non-parametric bootstrap.

We hope it is apparent that the efforts that have gone into architecting and implementing PSA in the PRIME T2D Model are not inconsiderable, with the intention of capturing and accurately characterising as much uncertainty in the modelled outcomes as possible. Indeed, we believe that the extent of random sampling that occurs in a PRIME T2D Model PSA may be unprecedented in NICE diabetes technology appraisals; for instance, running the company submission base case (with 300,000 patients per simulation arm) with PSA active results in over 2.8 billion samples being drawn from the random number generator (RNG).

Probabilistic Sensitivity Analysis Implementation

Regarding the lack of clarity noted by the EAG around the implementation, we would note that all model source code was provided to the EAG and all of the source code for the supporting statistical library (the Apache Commons Mathematics Library) is open source. The latter code is directly navigable from the PRIME T2D Model source code when utilizing any modern integrated development environment (IDE) with support for the Maven build automation tool. As such, even the exact implementation of the RNG (a Mersenne twister) and the probability distribution sampling has been available to the EAG for the duration of their evaluation of the model, a

degree of transparency that is not possible in models developed in Microsoft Excel, where the RNG and probability distribution sampling implementations are closed source and have demonstrable errors in their implementations.

We would also note that the points in the code at which decisions whether values should be drawn from parameter distributions are made are not obscure or in any way obfuscated, for example:

- Whether sampling of costs is active is governed by a Boolean value named sampleCosts, which is referenced in the EconomicsController Java class.
- Whether sampling of utilities is active is governed by a Boolean value named sampleUtilities, which is referenced in the QualityOfLifeController Java class.
- Whether sampling of treatment effects is active is governed by a Boolean value named sampleTreatmentEffects, which is referenced in the TreatmentController Java class.
- Whether sampling of model coefficients is active is governed by a single line of code in the PatientController.java superclass from which all complication-evaluating Java classes inherit.
- The simulated cohort of patients is generated (based on the user-defined cohort characteristics) in the CohortController Java class. Patient heterogeneity is thereby introduced in this class, which comprises just 250 lines of code (LOC), of which ~180 LOC are responsible for generating the cohort.
- Random walk (stochastic uncertainty) through the model is governed by sampling from uniform distributions in the processPatient() methods of each Java class responsible for modeling a given complication.

Furthermore, we received confirmation from the EAG on April 24, 2023 that they had successfully replicated the results from the web version of the PRIME T2D Model locally, in a compiled version provided to them alongside the source code, confirming that the EAG's local environment and the web environment were identical with regard to both random number generation and simulation trajectories. Given all of the above, we are unable to comprehend how the implementation of the probabilistic sensitivity analysis could reasonably be described as "not clear"; every aspect of the PSA implementation is visible in the source code, rationally named, modifiable by the EAG, and executable in their local simulation environment.

Refutation of the EAG Proposal to Implement Probabilistic Sensitivity Analysis using Techniques Described by Corro-Ramos et al.

Regarding the alternative approach to PSA proposed by the EAG, we would firstly note that it is unclear why a study with only nine citations in CrossRef — five of which are self-citations by authors of the original manuscript — is being cited as the recommended approach for implementing PSA. The code repository cited in the Corro-Ramos et al. 2020 manuscript includes a single 992-line R source code file, which is dependent on 29 unique comma-separated variable files, none of which is provided in the code repository. This precludes the ability to run the model at all, let alone replicating the results presented in the manuscript. The only basis we can see for citing this as an exemplary approach to the implementation of probabilistic sensitivity analysis is the relationship between the authors of the Corro-Ramos et al. 2020 manuscript and members of the EAG itself, who we note have previously authored studies together.

Regardless, certain aspects of the approach outlined by Corro-Ramos et al. would not be feasible in the PRIME T2D Model; for instance, the PRIME T2D Model relies on coefficients and regression outputs from multiple previously-published models, none of which has publicly available data on the correlation or covariance matrices for the model coefficients, nor are the patient-level data publicly available. We would also note that the computational inefficiency arising from the dynamically-typed, interpreted nature of the R programming language necessitates techniques such as those described by Corro-Ramos et al. where we note that the primary PSA analysis simulated just 100 patients over 300 PSA iterations, versus 300,000 patients over 1,000 bootstrap iterations in the PRIME T2D Model as used in the company submission base case. Without the techniques described by Corro-Ramos et al., the Monte Carlo error in an R model would likely be impractical to mitigate by increasing the number of patients or PSA iterations; we believe that no such limitations affect PRIME T2D Model.

Utilisation of Bootstrapping in the Probabilistic Sensitivity Analysis

As detailed above, we consider the approach of capturing all available sources of randomness, heterogeneity, and uncertainty in the PRIME T2D Model — followed by a non-parametric bootstrap analysis — to be the most appropriate and robust method of providing a holistic quantification of uncertainty around the modelled solution to the decision problem. Regardless of the exact methodological approach utilised in the PRIME T2D Model, we note the apparently directly contradictory predilections of the EAG with regard to the use of bootstrapping in PSA; on the one hand, the EAG recommend the analyses be run using the closed-source, proprietary CORE Diabetes Model, the documentation for which unequivocally describes that non-parametric bootstrapping is the foundation of its PSA implementation (see below). On the other hand, the EAG characterise the use of non-parametric bootstrapping in the PRIME T2D Model as "not standard".

The use of non-parametric bootstrapping in the CORE Diabetes Model is very clearly and explicitly documented in the seminal publication on the CORE Diabetes Model by Palmer et al., who note (emphasis added):

"Nonparametric bootstrap methods are used to evaluate uncertainty in cost-effectiveness outcomes measured. Each probability in the model is simulated using a first-order Monte Carlo approach to represent sampling uncertainty. After 1,000 simulations of 1,000 non-identical patients, 1,000 bootstrap samples are drawn and the joint distribution of mean incremental costs and mean effectiveness gained is evaluated."

Lack of Feedback from PRIMA Review

Finally, while this does not pertain directly to the present technology appraisal, we would also question why the EAGs comments on the PSA implementation were not echoed in the preceding NICE Preliminary Independent Model Advice (PRIMA) review of the PRIME T2D Model, which was conducted, at least in part, by the same Java expert as presently subcontracted by the EAG. Such comments could have enabled a constructive dialog on the preferred PSA implementation ahead of time.

Key issue 14: No full deterministic one-way sensitivity analyses provided

Description of issue and why the EAG has	No full deterministic one-way sensitivity analyses (for all input parameters) were provided, and an opportunity was therefore missed to identify potentially influential parameters.

identified it as important	
What alternative approach has the EAG suggested?	Implement deterministic one-way sensitivity analyses (for all input parameters) and present results in tornado diagrams (for all doses and in the comparison with semaglutide).

As outlined in the response to clarification questions, the request to provide sensitivity analysis for all input parameters is impracticable. A standard simulation has over 185 input parameters (not including life tables). To do this for the comparators suggested by the EAG would be 1,110 simulations (assuming high and low estimates for each input parameter and three doses of tirzepatide versus semaglutide) and produce three tornado diagrams, each with 185 variables. This would not provide useful information with respect to the decision question.

All key model inputs that have an influence on cost-effectiveness were explored in sensitivity analysis in the original submission (CS, Table 106), where 12 PSA simulations and 48 one-way sensitivity analyses were described. Tornado diagrams summarized these data were provided to the EAG in the response to clarifications document in February, 2023.

Key issue 15: Technical verification insufficient/model results not reproducible

Description of issue and why the EAG has identified it as important	There remain doubts over the internal validity of the model. Model outcomes could not be reproduced by the EAG. The EAG could not find how BMI-related utilities were implemented in the model (black box character). No full overview of input parameters has been provided.
What alternative approach has the EAG suggested?	Correct the model if necessary. Provide step-by-step guide to running the model. Provide a filled in TECH-VER checklist. Provide detailed description of how the BMI-related utilities were implemented and where this can be found in the code. Provide full overview of all input parameters and how they were included in the PSA.

The EAG confirmed that model outcomes could indeed be reproduced without issue using the model interface (for which they had access since August 2022). The technical challenges faced by the EAG in terms of reproducing the base case analysis locally (without using the online version of the model) have been resolved further to the call on 14 April 2023, during which it was demonstrated that the model results could be reproduced using the JAVA code and the JSON files provided. Further technical support was provided to ensure the EAG could reproduce the submitted model outcomes.

The evaluation of BMI-related utilities and how they were entered into the model were provided in the response to clarification questions in February 2023 (this involved entering an annual utility score associated with BMI or other treatment related aspects that influence quality of life) into a field entitled “treatment related utility” in the model interface. We believe the term “black box character” is inappropriate in the present context given that the EAG was provided with a detailed technical report describing model functionality, had access the model since August 2022, were provided with the full source code and JSON files required to the model off-line (and independently) and all model inputs for the cost-effectiveness evaluation were detailed in the original submission. The offer of technical support with the model has been open since August, 2022. In comparison with any previous submissions in type 2 diabetes, this represents an unprecedented level of transparency as the EAG have had access to exactly how the model works (source code) since August 2022.

We dispute the EAG claim that no full overview of input parameters has been provided. The EAG were given full access to the base case simulations and settings (model inputs) via the model

interface in August 2022. All model inputs were detailed in the submission documentation and could be verified in the model interface. Further, all model inputs for the base case analysis were provided in JSON files to the EAG so that the model could be run off-line. The model inputs have, therefore, been provided in triplicate.

With respect to the alternative approaches suggested by the EAG:

Correct the model if necessary: No corrections are/were necessary and the EAG have overcome the technical issues they faced when trying to run the model locally.

Provide step-by-step guide to running the model: A model user guide was submitted in August 2022 with step-by-step instructions for running a simulation with the PRIME T2D Model. The EAG confirmed there were no issues using the online version of the model in March 2023 and the technical challenges the EAG faced running a local version of the model off-line have been resolved.

Provide a filled in TECH-VER checklist: A completed TECH-VER checklist has not been provided as 1) this is not mandated or recommended by NICE, and 2) completion of the checklist did not form part of the model development process (as much of the checklist is not directly relevant to the development of a patient-level simulation of type 2 diabetes). Instead, the following information was provided (we do not believe the TECH-VER checklist would add significantly to the model verification/validation steps already taken):

A technical report detailed the model verification and validation steps throughout model development, including external (third party) verification of the model code, advisory board meetings and adherence with ISPOR good modelling practice guidelines

Details of the NICE PRIMA review of the PRIME T2D Model, including the NICE PRIMA review groups recommendations and response

Provide detailed description of how the BMI-related utilities were implemented and where this can be found in the code: This was provided in the response to clarification questions where the following information was provided:
With respect to nausea, weight loss, BMI and device utilities, the model features a treatment-related utility function that is editable by the user and can be used to define separate utilities to be applied in year 1 and years 2+ of any given simulation. The treatment related utilities are added to the annual utility score for each patient as calculated based on the inputs in the Utilities element. In the current set of simulations, the treatment-related utility function was used to capture the following utilities:
- Year 1: body weight change utility (no separate BMI utility), device utility and the nausea and vomiting utility
- Years 2+: BMI utility only (no body weight change utility)

In the source code, the variable associated with a treatment-related utility is named `QoLCategory.RX` and can be found in the `QualityOfLifeController.java` controller.

Provide full overview of all input parameters and how they were included in the PSA: Details of the approach to PSA are provided in the response to key issue 13 (above) and a complete list of input parameters for the EAG preferred base case simulations are provided later in this response document.

Other Points Made by the EAG Requiring Clarification

Section 4.2.2 Model structure	Page 115-16
EAG comment: The main concerns of the EAG relate to: a) the company justification to use the PRIME T2D model; b) specification of the model type; c) model averaging approach; d) selection of predictive models to estimate the risk of complications and e) approach to estimate risk of macular oedema, ulcer and revascularisation. … it is unclear for the EAG that the developed de novo model, specifically the current implementation as in the CS, has a better performance to predict complications (including CV events) compared with existing diabetes models. Notably, the following complications were considered in the modelling for NG28[27, 28] and were not included in Figure 69 of the CS: angina, peripheral vascular disease, diabetic retinopathy, cataract, ketoacidosis, lactic acidosis and foot ulcer (though the latter might be included in the PRIME T2D model given in Appendix N.5.6.2 of the CS elaborates on the estimation of ulcer risk) Unfortunately, the company did not provide justifications (requested in clarification question B5), that the risk models used, both individually and after model averaging, are appropriate to estimate the risk of complications for the population as specified in the CS.	

a), b) and c) With respect to the EAG comments on the justification for using the PRIME T2D Model, the specific type of model used, and the use of model averaging, these have all been addressed in response to key issues 6 and 7 (above).

d) The theme of the selection of predictive models has largely been addressed in responses to key issues 6 and 7 (above). Specifically, the validation analysis using the PRIME T2D Model indicates that it is better placed to predict long-term clinical outcomes than either of the other models mentioned by the EAG (CORE Diabetes Model and the UKPDS OM2) when compared with the results presented at the Ninth Mount Hood Challenge meeting for those models, and given the challenges around calibration of risk equations as discussed by Evans *et al.* (2022). [11, 13] The ability of model averaging to adapt to as patients become older and risk profiles change is, intuitively, also an advantage over a fixed approach with a selected predictive model. In addition, the inclusion of a revascularization endpoint is considered a key feature of a modern diabetes-modelling analysis according to Keng *et al.* (2022) which would not be available with the CORE Diabetes Model or the UKPDS OM2, but is possible using the current methodology with the PRIME T2D Model.[14]

e) With respect to complications included in NG28 and in the present modelling analysis, some clarity is required around the EAG comments. In the 2022 NICE update for the management of type 2 diabetes in adults (EAG report reference 30), a calibrated version of the UKPDS OM2 was used. The other evaluations references by the EAG include *Economic modelling for continuous glucose monitoring in adults with type 2 diabetes* and *Economic modelling for periodontal treatment in adults with type 1 and type 2 diabetes* (references 28 and 29 in the EAG report), where the CORE Diabetes Model was used. The complications modelled by these two models as well as the PRIME T2D Model are summarized in the following table.

UKPDS OM2	CORE Diabetes Model	PRIME T2D Model
Myocardial infraction	Myocardial infarction	Myocardial infarction
Ischaemic heart disease	--	Ischaemic heart disease
--	Angina	--
Heart failure	Heart failure	Heart failure
Stroke	Stroke	Stroke
--	--	Revascularization
--	Peripheral vascular disease	--
--	Micro-, macroalbuminuria	KDIGO CKD Stages 1-4
Renal failure	Renal failure	Renal failure
--	Retinopathy	--
--	Macular edema	Macular edema
Blindness	Blindness	Blindness
--	Neuropathy	Neuropathy
Ulcer	Ulcer	Ulcer
Amputation	Amputation	Amputation

It is perhaps noteworthy that the PRIME T2D Model is the only one of the three models to include an revascularization endpoint as recommended by Keng *et al.* (2022).[14] For the complications listed by the EAG as missing from the present analysis, the following considerations may be relevant:

Angina: The CORE Diabetes Model uses risk equations from UKPDS OM2 by default to evaluate the risk of macrovascular endpoints. It is notable that there is no UKPDS OM2 risk equation for angina and therefore another approach must be being used to evaluate the risk of angina and, perhaps, adjust for the competing risk of the ischaemic heart disease endpoint (that is not modelled in the CORE Diabetes Model). As neither of the CORE Diabetes Model evaluations states which risk equations were used, it is impossible to know how this was done. There is significant overlap between the angina and ischemic heart disease endpoints and it would be very unusual for any model to include both.

Peripheral vascular disease: In the CORE Diabetes Model, the approach to modelling peripheral vascular disease is the same in both type 1 and type 2 diabetes and is based on data from the Framingham study with risk-adjustment based on the UKPDS.[24] Given how old this data is, the minimal role of PVD in a cost-effectiveness evaluation (modest impact on costs-and effects) and

recommendations from the advisory board meetings during model development, it was decided (conservatively) not to include this endpoint in PRIME.

Diabetic retinopathy: Diabetic retinopathy in the CORE Diabetes Model is modelled based on data from a type 1 diabetes population in the Diabetes Control and Complications Trial (DCCT).[24] Similarly, given how old this data is and the fact is not from a population with type 2 diabetes, the minimal role of retinopathy in a cost-effectiveness evaluation (modest impact on costs-and effects) and recommendations from the advisory board meetings during model development, it was decided (conservatively) not to include these endpoints in PRIME.

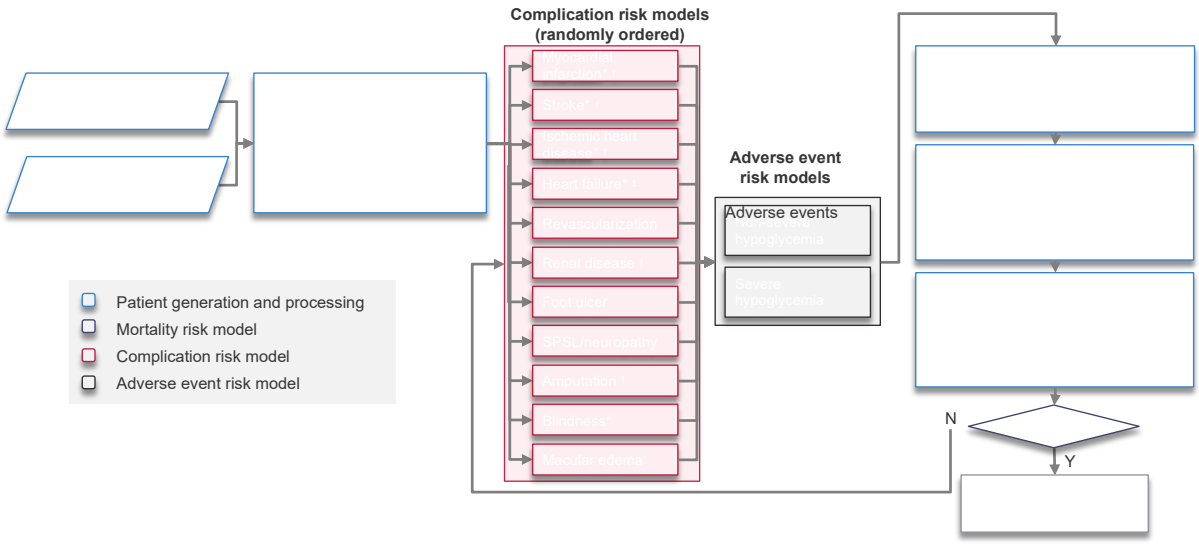
Cataract: Again, the risk of cataracts in the CORE Diabetes Model is based, in part, on data from a type 1 diabetes populations. For the same reasons outlined above for diabetic retinopathy, this endpoint was not included in the PRIME T2D Model.

Ketoacidosis: Ketoacidosis is an adverse event that occurs in primarily patients with type 1 diabetes but very rarely occurs in the type 2 diabetes population. For this reason, it was not included in the PRIME T2D Model.

Lactic acidosis: Lactic acidosis is rare in type 2 diabetes populations and there is little or no evidence suggesting that lactic acidosis occurs at differential frequencies with different treatments or interventions.[29] As a result, it would have almost no impact on a cost-effectiveness evaluation and, in line with recommendations from the advisory board meetings during model development, it was decided not to include this endpoint in PRIME.

We apologise for any confusion the EAG encountered around which complications were included in the PRIME T2D Model. In addition to the technical report, the EAG received access to the model interface in August 2022 (which shows all model inputs and outputs), the source code (showing all model calculations), and JSON files (with all base case inputs), which could all have provided additional clarity on the endpoints included in the present modelling analysis. We have included a schematic of the PRIME T2D Model to confirm the endpoints included in the modelling evaluation.

Figure 17: Schematic diagram of the PRIME T2D Model



* Model averaging is used in this controller; † denotes complications with an increased risk of mortality in the year of complication onset and in subsequent years; ‡ denotes complications with an increased risk of mortality associated with a history of this complication; RNG, random number generator; SPSL, severe pressure sensation loss

Section 4.2.3 Population	**Page 117**
EAG comment: The main concerns of the EAG relate to: a) mismatch with the population considered in the NICE decision problem and clinical effectiveness evidence; and b) differences in patient characteristics between THIN cohort and SURPASS trials. …the EAG agrees with the choice of the THIN cohort as more representative of the UK population with T2D initiating second line therapy in clinical practice, after failing diet and exercise plus metformin. Besides, as per Table 13 from the clarification response[30], there is an erratum on the percentage of Hispanic for the SURPASS trials that should be corrected.	

a) Questions on the target population and the decision problem have been addressed in The Decision Problem section of the response

b) In line with the EAG recommendations, a scenario analysis of the EAG preferred base case has been performed using the SURPASS-2 cohort (with the proportion of Hispanic patients cross-checked with the source literature) to document the (modest) impact on cost-effectiveness of changes to the cohort characteristics (see EAG Preferred Base Case Simulations section for details).

Section 4.2.4 Interventions and Comparators	**Page 117-119**
EAG comment: The main concerns of the EAG relate to: a) restricting to comparisons within each recommended maintenance dose step, b) narrower intervention and comparators than in NICE final scope; c) basal insulin as the only treatment option after treatment intensification. … the EAG would prefer that the treatment strategies incorporated in the economic model would reflect clinical practice (including the possibility for individual patients to switch between treatment dosages). Additionally, the EAG would prefer comparisons between maintenance doses, instead of restricting comparisons to within dose steps.	

a) Fully incremental analysis has been provided in the EAG Preferred Base Case Simulations section below . With respect to the comments on “reflecting clinical practice”, almost any plausible sequence of treatments is possible within the modelling environment. However, any such approach needs to be supported by clinical evidence on the effects of switching treatments to be valid and, currently, the effectiveness data required to create the sorts of scenario suggested by the EAG do not exist. As a result, any such analysis would need to be assumption based and, as previously outlined in the response to key issue 9, assumptions around treatment switching and rescue medications can have an important influence on cost-effectiveness (as well as obfuscating the cost-effectiveness profile of individual

agents/doses). For these reasons, no additional modelling was performed to address the EAG comment.

b) Questions on the comparators relevant to the final scope have been addressed elsewhere in the response (see The Decision Problem section)

c) With respect to the EAG comments on basal insulin as the only treatment option after treatment intensification, the original submission included a sensitivity analysis where a further treatment intensification step was included to basal-bolus therapy, which decreased the ICER for tirzepatide. For the base case simulations, it was conservatively assumed that treatment effects should be equivalent in both arms after post-intensification, such that any subsequent treatments would have a limited effect on cost-effectiveness. Moreover, assumptions around subsequent treatment intensification steps (as there is little evidence to document the changes in risk factors for multiple treatment steps following GLP-1 receptor agonist treatment in patients with type 2 diabetes) are subject to the same issues described in the response to *key issue 9* in terms of influencing and obfuscating cost-effectiveness results based purely on modelling assumptions.

Section 4.2.6 Treatment effectiveness and extrapolation	**Page 120-122**
EAG comment: The main concerns of the EAG relate to: a) assumption of constant risk factors after year 1 up to treatment intensification; b) assumptions regarding waning of relative treatment effect; c) assumptions after switching to basal insulin therapy (treatment intensification); d) treatment discontinuation assumptions; e) BMI retrieved from the NMA versus BMI calculated based on body weight; f) the “nearest neighbour” approach to handle missing inputs from the NMA for all risk factors except weight; g) assuming no change from baseline for eGFR, WBC and Hb levels.	

a) The assumption of constant risk factors after year 1 up to treatment intensification has been addressed in detail, citing relevant clinical evidence, in the response to *key issue 8*.

b) Similarly, the use of UKPDS OM2 risk factor progressions and assumptions around the waning of treatment effects has been addressed in the response to *key issue 8*. It should be noted that for the key drivers of cost-effectiveness (HbA1c and BMI), HbA1c has been associated with a waning of treatment effect throughout the submission and switching to use UKPDS OM2 risk factor progression will do little to introduce a waning affect associated with BMI differences. Moreover, it could be argued that artificially decreasing clinical benefits (without any clinical evidence of waning) without any corresponding impact on costs has considerable potential to bias a cost-effectiveness evaluation. It should also be noted that, at the population level, there is a “waning” of treatment effects as the cohort gradually switches to insulin therapy and risk factors (such as BMI and SBP) return to baseline levels. See the risk factor progression curves presented in the *EAG Preferred Base Case Simulations* section for details.

c) With respect to the assumptions around treatment effects following intensification to basal insulin therapy, the EAG noted that: *… company also indicated that these conservative scenarios (i.e., assuming no benefits after intensification for HbA1c or BMI respectively) would have produced ICERs between the CS base-case and the scenario analyses with no HbA1c or BMI difference which individually increased the ICER by roughly £4,000 and £7,000 respectively (for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg).* This

represents a misunderstanding of the sensitivity analysis performed. The scenarios referred to by the EAG assumed that there was no HbA1c difference between treatments or there was no BMI difference between treatments at any stage in the simulation (i.e. during treatment with tirzepatide or semaglutide and during treatment with basal insulin therapy). These scenarios were run as part of an effort to identify the key drivers of cost-effectiveness and not (as assumed by the EAG) to explore different assumptions following intensification to basal insulin therapy.

d) The topic of intensification of therapy, discontinuing the initial treatment and switching to basal insulin therapy, the challenges associated with modelling treatment pathways and assumptions around rescue medication has been addressed in the response to *key issue 9*.

e) The EAG recommendation to use change from baseline in BMI directly from the NMA has been adopted in the *EAG Preferred Base Case Simulations* sections (see below).

f) and g) No response required to EAG comments.

Section 4.2.7 Adverse events	**Page 122-123**
EAG comment: The main concerns of the EAG relate to that only nausea is incorporated (hypoglycaemia only for basal insulin therapy) as an AE. Including hypoglycaemia only for basal insulin therapy might inflate the impact of discontinuing treatment (i.e., treatment intensification) and hereby potentially inducing bias favouring more effective treatments (i.e., tirzepatide). Moreover, in response to clarification question A37 it was stated: "*Clinically significant hypoglycaemia occurred in 10 to 14 % (0.14 to 0.16 events/patient year) of patients when tirzepatide was added to sulphonylurea and in 14 to 19 % (0.43 to 0.64 events/patient year) of patients when tirzepatide was added to basal insulin (very common)*". As illustrated in clarification response Tables 20-22, incorporating additional AEs would potentially increase the estimated ICER. Therefore, the EAG would prefer to include all relevant AEs (also including hypoglycaemia and GI AEs such as diarrhoea and vomiting).	

The comments on modelling adverse events have been largely addressed in the response to *key issue 10*. The impact of hypoglycaemia (cited by the EAG as potentially benefitting more efficacious treatments due to later intensification) would be very modest in the cost-effectiveness evaluation (as sensitivity analysis on hypoglycaemia indicated in the original submission). It may also be a fair assumption that the interventions that improve glycaemic control most, thereby delaying insulin therapy, may be associated with a small benefit in this regard. It should also be noted that hypoglycaemia rates quoted by the EAG refer to concomitant use of sulfonylurea and insulin (both of which are known to be associated with increased hypoglycaemia risk), and do not correspond to the treatment regimens investigated in the modelling analysis (where metformin was assumed to be the only concomitant therapy).

Section 4.2.8 Health-related quality of life	**Page 123-127**
EAG comment: The main concerns of the EAG relate to: a) the uncertainty measures and distributions applied to utility values; b) T2D utility value and methods for age-adjustment; c) the utility estimate associated with administration of tirzepatide and dulaglutide; d) the utility value associated with weight change in the first year; e) methods for combining disutility values; and e) how utility values were selected from the SLR.	

a) The EAG comments on uncertainty measures applied to utility values are addressed in the response to *key issue 12*.

b) The EAG comments on the T2D utility value and age-adjustment are addressed in the response to *key issue 11*.

c) The utility associated with the administration of tirzepatide and dulaglutide based on the study by Boye *et al.* (2019) has been removed from the EAG preferred base case simulations (see *EAG Preferred Base Case Simulations* below).[31]

d) The utility associated with weight loss in year 1 of the simulation, based on the study by Boye *et al.* (2021), has been removed from the EAG preferred base case simulations (see *EAG Preferred Base Case Simulations* below).[32]

e) An additive approach was adopted for the present analysis as this is best aligned with previous health economic evaluations, including those performed by NICE, in type 2 diabetes. This includes:

Additive approach to combining utilities: National Institute for Health and Care Excellence. Type 1 and 2 diabetes in adults: diagnosis and management. Economic modelling for periodontal treatment in adults with type 1 and type 2 diabetes. NICE guideline NG17, NG28. Economic model report [Internet]. London: NICE, 2022 [accessed 10.3.23]. 33p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-on-periodontal-treatment-in-adults-with-type-1-and-type-2-diabetes-pdf-11131191037

Additive approach to combining utilities: National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. Economic modelling for continuous glucose monitoring in adults with type 2 diabetes. NICE guideline NG28. Economic model report [Internet]. London: NICE, 2022 [accessed 10.3.23]. 28p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-pdf-11013295213

Additive approach to combining utilities: National Institute of Health and Care Excellence. Type 2 diabetes in adults: management (update). Health economic model report [NG28] [Internet]. London: NICE, 2022 [accessed 10.4.22]. 78p. Available from: https://www.nice.org.uk/guidance/ng28/evidence/health-economic-model-report-pdf-10959500845/

Cost minimisation approach (no utilities involved): Ertugliflozin with metformin and a dipeptidyl peptidase-4 inhibitor for treating type 2 diabetes. Technology appraisal guidance (TA583), 5 June 2019. https://www.nice.org.uk/guidance/ta583

Cost minimisation approach (no utilities involved): Ertugliflozin as monotherapy or with metformin for treating type 2 diabetes. Technology appraisal guidance [TA572], 27 March 2019. https://www.nice.org.uk/guidance/ta572

Additive approach to combining utilities: Dapagliflozin in combination therapy for treating type 2 diabetes. Technology appraisal guidance [TA288], 26 June 2013. https://www.nice.org.uk/guidance/ta288

Additive approach to combining utilities: Dapagliflozin in triple therapy for treating type 2 diabetes. Technology appraisal guidance [TA418], 23 November 2016. https://www.nice.org.uk/guidance/ta418

Approach not reported (assumed additive based on data presented): Canagliflozin, dapagliflozin and empagliflozin as monotherapies for treating type 2 diabetes. Technology appraisal guidance [TA390], 25 May 2016. https://www.nice.org.uk/guidance/ta390

Approach not reported (assumed additive based on data presented): Empagliflozin in combination therapy for treating type 2 diabetes. Technology appraisal guidance [TA336], 25 March 2015. https://www.nice.org.uk/guidance/ta336

None of the health economic analyses in type 2 diabetes available on the NICE website used a multiplicative approach to combine quality of life utilities. The EAG commented that: *Although the best method to combine multiple disutility values is still debated, the multiplicative method is considered to be the best approach overall and more conservative than the additive method.*[33] The reference cited (Ara R, Brazier J. Estimating health state utility values for comorbid health conditions using SF-6D data. Value Health 2011; 14(5):740-5) does not support this statement in the context of the present analysis. The publication describes an analysis of SF-6D derived health-state utilities over a limited range. Ara and Brazier showed that a linear regression approach to combining individual utilities provided the best fit to the data collected (on combination health states) and cited limitations with all of the methods they investigated (including multiplicative and additive approaches). The results ranked five methods of combining utilities and the respective rankings changed across various sub-analyses. In the present analysis, utilities were mainly derived from the EQ-5D instrument. There is no information in the Ara and Brazier publication on how generalisable the findings are likely to be for the combination of utilities derived using a different instrument, in a different population (i.e. a type 2 diabetes population) and with different baseline utility scores. It therefore seemed appropriate to use an additive approach for the present modelling analysis, given the consistency with previous NICE evaluations and in the absence of evidence to the contrary.

f) No response required.

Section 4.2.9.2 Treatment costs	**Page 127-129**
EAG comment: The main concerns of the EAG relate to: a) inflation of costs to present day values, b) discrepancies between costs mentioned in the source and the CS, c) costs associated with nausea, d) T2D health state costs, and e) how cost values were selected from the SLR.	

Response:

a) In line with the EAG comment, all complication costs have been inflated to 2022 value for the EAG preferred base case simulations (see *EAG Preferred Base Case Simulations* section)

b) In response to the EAG comment, all costs have been checked against the source material and amended if necessary for the EAG preferred base case simulations (see *EAG Preferred Base Case Simulations* section)

c) As suggested in the EAG comment, a scenario analysis has been performed to evaluate the impact of including a cost associated with nausea (see *EAG Preferred Base Case Simulations* section)

d) As suggested in the EAG comment, the impact of including an annual T2D health state cost has been investigated in a scenario analysis (see *EAG Preferred Base Case Simulations* section)

e) Following the literature review, costs and resource use estimates were selected to best align with the NICE health economic evaluation used to support the development of NG28 wherever possible. Only two complication costs were derived from other sources (neuropathy and CKD stage 4) for which only one estimate was identified in the published literature (see *EAG Preferred Base Case Simulations* section for details of the costs used)

Section 4.2.11 Uncertainty	**Page 129-**
EAG comment: The company did not explore the impact of all input parameters on the estimated ICERs, hence, the EAG does not agree with this statement that the ICERs are robust to changes in the modelling parameters. Moreover, results provided by the company in Table 106 of the CS as well as the clarification responses indicated that changes in input parameters can have a substantial impact on the estimated ICERs.	

The EAG comment has been addressed in the response to *key issue 14* (above).

Section 5.1 Company's cost-effectiveness results	**Page 130-136**
EAG comment: The main concerns of the EAG relate to: a) disaggregated outcomes, b) likely inappropriate PSA, c) no fully incremental analysis was provided, d) presentation of disaggregated costs likely does not allow calculation of costs with any confidential comparator prices, and e) most benefits accrued in the modelled versus observed.	

a) With respect to the EAG comments on disaggregated QALY decrements, whilst these were not tabulated in the original submission report they were presented for the base case simulations via the model interface (access provided to the EAG in August 2022). These estimates were subsequently reproduced in tabular form in response to the clarification letter from the EAG in February 2023.

b) The comments made by the EAG on PSA are addressed in the response to key issue 13 (above).

c) Fully incremental outcomes and estimates of net health benefits were provided to the EAG in the response to the clarification letter in February 2023. In addition, fully incremental outcomes and estimates of net health benefits have been provided for the EAG preferred base case analysis (see EAG Preferred Base Case Simulations below).

d) An explanation on how treatment costs are entered into the model was provided to the EAG in response to the clarification letter in February, 2023. The relevant input for the model is an annual cost associated with treatment and the calculation for this was detailed in the original submission and in the response to the clarification questions. It takes a simple line of arithmetic to adjust pharmacy costs for the modelling analysis, which would in turn allow

analysis of potential comparator confidential prices. These analyses would have been easily run via the model interface using the EAG accounts given the access provided in August, 2022. We would reject the EAG assertion that: *the analysis of potential comparator confidential prices likely cannot be performed.*

e) The concept of most benefits being accrued during the modelled (not observed) period is central to health economic analysis in type 2 diabetes and is entirely consistent with previous submissions and NICE health economic evaluations in this area (see response to *Section 4.2.8 Health-related quality of life* above for a list of recent submissions/evaluations). Indeed, this approach is endorsed by published guidelines on diabetes modelling based on the knowledge that improvements in risk factors, such as HbA1c, BMI, SBP and serum lipid levels, can reduce the risk of diabetes-related complications over a long-term time horizon.[34] We do not see this comment as a valid criticism of the analysis.

Section 5.2 Company's sensitivity analysis	**Page 136-137**
EAG comment: The main concerns of the EAG relate to: a) no full one-way sensitivity analysis was conducted, and b) scenario analyses were provided only for the semaglutide comparison and also only for the 10 mg tirzepatide dose.	

a) The EAG comment on full one-way sensitivity analysis is addressed in the response to *key issue 14* (above).

b) In the original submission, the base case analysis was accompanied by 12 PSA simulations and 48 one-way sensitivity analysis simulations that were sufficient to identify key drivers of cost-effectiveness in the submission. A submission addendum was provided to the EAG in March 2022 with sensitivity analyses for all three doses of tirzepatide (the original submission focused on sensitivity analyses for the comparison of tirzepatide 10 mg with semaglutide 1.0 mg), which provided evidence of the generalizability of one-way sensitivity analysis results across tirzepatide doses.

Section 5.3 Model validation and face validity check	**Page 137-139**
EAG comment: The main concerns of the EAG relate to: a) technical verification and reproducibility not demonstrated, b) face validity checks likely not conducted for the current NICE model, c) cross-validation hampered by lack of clarity on impact of differing assumptions, and d) external validation incomplete.	

a) The response to *key issue 15* addresses many of the main aspects of the EAG commentary on validation and face validity, but the following points may also be relevant:

The EAG confirmed that model outcomes could indeed be reproduced without issue using the model interface (for which they had access since August 2022). The technical challenges faced by the EAG in terms of reproducing the base case analysis locally (without using the online version of the model) have been resolved further to the call on 14 April, during which it was demonstrated that the model results could be reproduced using the JAVA code and the JSON files provided. Further technical support was provided to ensure the EAG could reproduce the submitted model outcomes.

The EAG stated that: *the company did not convince the EAG that sufficient internal validity checks were carried out on this application of the PRIME T2D model.* In terms of addressing this comment, we would point out the following:

- There are not multiple versions or applications of the PRIME Diabetes Model as suggested by the EAG. The EAG were provided online access to the model which runs the model source code via an online interface. This source code is the same code that was verified by an independent third-party during model development.
- This version of the model was the same one that was reviewed through the PRIMA process. No changes to the model calculations were made following PRIMA review (only interface changes to allow users model flexibility and additional functionality was added – the PRIMA review report and response was made available as part of the submission).
- The source code and all necessary JSON files to run the model were provided to the EAG in August, 2022. At the same time online access was provided (moved to a new server in October, 2022 after an EAG request following slow response times with the online model). The online model results have been reproducible in our hands when running the model off-line throughout the review process. The JSON files required to run the offline model code were provided again in February 2023 (response to clarification questions). And after recently resolving the EAG's technical issues, the online model results are now reproducible in the EAG's hands (as confirmed by email).
- A technical report describing the model development process (following good practice guidelines), verification and validation for this version of the model was provided as part of the original submission.
- This level of verification, review and transparency is unprecedented in previous submissions on type 2 diabetes to NICE, and we do not believe the EAG claims on internal validity withstand scrutiny.

The suggested use of the TECH_VER checklist has been addressed in the response to key issue 15.

The EAG claimed to have: identified an error with only superficial testing (see response to clarification question B31). The issue raised by the EAG was not a calculation error but a bug in the graphical representation of a PSA scatterplot (the numerical simulation results were correct). This type of bug would not have been identified by the EAG-suggested TECH-VER checklist, was not detected during PRIMA review and has since been amended (note: this was not a change to the model calculations or to the source code).

b) With respect to the role of Advisory Board meetings in the development of the model, the earlier meetings (2014 and 2015) were focused on developing an effective model framework/structure, including the approach to PSA (which applies to present version of the PRIME T2D Model), and the 2019 Advisory Board was focused exclusively on the current (and only) version of the PRIME T2D Model. As summarized in the model Technical Report Section 4.3.1 (provided as part of the original submission), recommendations from this Advisory Board meeting were used to influence the choice of complications to be modelled, the approaches to risk evaluation (including model averaging) and the selection of input data for incorporation into the model.

c) No response required

d) No response required

A.1.4 EAG Preferred Base Case Simulations

Methodology

The base case simulations were re-run incorporating recommendations from the EAG and new pack prices for tirzepatide. Full incremental base case results are presented in following section (see *Base Case Results*) along with accompanying PSA (see *Probabilistic Sensitivity Analysis Results*) and scenario analyses in line with EAG feedback (see *Scenario Analysis Results*).

EAG Suggestions for the Preferred Base Case

The following table summarizes EAG suggestions for the base case analysis with the corresponding actions taken and/or rationale for non-inclusion in the preferred base case simulations.

Table 5: Summary of EAG suggestions for the preferred base case

EAG Comment	Action taken or response
Section 6.1.1.1 Fixing errors (FE) 1. Resolve discrepancies in the uncertainty measures and distributions related to utility values and costs listed in the CS and those listed in the original sources. Utility discrepancies are listed in *4.2.8 EAG comment a*, and *4.2.9 EAG comment b*, as well as *key issue 12*.	Measures of variance for all complication costs and utilities used were derived from the original published sources wherever possible and incorporated in the preferred base case simulations (see summary tables in the *Revised Model Inputs* section below). Note: this change will only affect the results of PSA simulations.
Section 6.1.1.2 Fixing violations 2. Incorporate treatment strategies in the economic model that reflect clinical practice (including the possibility for individual patients to switch between treatment dosages). See *Section 4.2.4 EAG comment a*, as well as key issue 3.	No changes were made to the simulation inputs in response to this comment. See response to *Section 4.2.4 Interventions and Comparators* for details and rationale.
3. Incorporate all comparators described in the final scope. See *Section 4.2.4 EAG comment b*.	No changes were made to the simulation inputs in response to this comment. See response to *Section 4.2.4 Interventions and Comparators* for details and rationale.
4. Incorporate all relevant AEs (also including hypoglycaemia and GI AEs such as diarrhoea and vomiting). See *Section 4.2.7 EAG comment* as well as *key issue 10*. This adjustment will likely increase the estimated ICER (might depend on the comparator). The magnitude of the impact is unclear as no analyses were provided incorporating all abovementioned AEs simultaneously (see also clarification response Tables 20–22).	No changes were made to the base case simulation inputs in response to this comment. See response to *Section 4.2.7 Adverse events* for details and rationale. Scenario analysis was performed to explore the effects of including other AEs in the cost-effectiveness analysis (see *Scenario Analysis Results* for details).
5. Incorporate age-adjustment for utility values, ensuring that the utility does not exceed the age-matched general population utility. See Section 4.2.8 EAG comment b.	Age-adjustment for utility values was included in the preferred base case simulation inputs (see summary tables in the *Revised Model Inputs* section below).

EAG Comment	Action taken or response
This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as the exact implementation of the "QALY age-adjustment based on Ara and Brazier" analyses performed by the company (CS Table 89) is unclear (amongst others whether the T2D utility exceeds the age-matched general population utility).	
6. Inflating all costs to the same price year, preferably 2022 values. See *Section 4.2.9 EAG comment a*.	All costs were inflated to 2022 values for the preferred base case simulation inputs (see summary tables in the *Revised Model Inputs* section below).
Section 6.1.1.3 Matters of judgement 7. Assuming UKPDS OM2 risk factor progression for all risk factors. See *Section 4.2.6 EAG comment a* as well as *key issue 8*.	UKPDS OM2 risk factor progressions were used throughout the analysis except in cases where clinical evidence provided contrary evidence (see response to *key issue 8*): • UKPDS OM2 risk factor progression was assumed all risk factors whilst on insulin therapy and for HbA1c, LDL, HDL, eGFR, white blood cells count, heart rate and haemoglobin levels whilst on tirzepatide or comparator treatments in line with EAG recommendations • For SBP and BMI, no change was assumed whilst simulated patients remain on tirzepatide or comparator therapy, but returned to baseline levels and followed UKPDS OM2 risk factor progression after switching to basal insulin therapy
8. Assuming additional causes for treatment discontinuation (than reaching the HbA1c threshold). See *Section 4.2.6 EAG comment d*, as well as *key issue 9*.	No changes were made to the simulation inputs in response to this comment. See response to *Section 4.2.6 Treatment effectiveness and extrapolation* for details and rationale.
9. Using the BMI directly retrieved from the NMA, when available, and using the BMI calculated based on body weight (from the NMA) only for dulaglutide 3.0 mg. dulaglutide 4.5 mg, oral semaglutide 7 mg and liraglutide 1.2 mg (i.e., when BMI was not available from the NMA). See Section 4.2.6 EAG comment e.	BMI values were directly retrieved from the NMA wherever possible for the preferred base case analysis simulation inputs.
10. Assume no device utility associated with tirzepatide or dulaglutide in the base-case analysis. See *Section 4.2.8 EAG comment c*.	No device utility was used for tirzepatide or dulaglutide for the preferred base case analysis simulation inputs.
11. Assume a multiplicative approach for utility values. See *Section 4.2.8 EAG comment e*.	No changes were made to the simulation inputs in response to this comment. See response to *Section 4.2.8 Health-related quality of life* for details and rationale.

Abbreviations: AE: adverse event; BMI: body mass index; EAG: evidence assessment group; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein; ICER: incremental cost-effectiveness ratio; LDL: low density lipoprotein; NMA: network meta-analysis; QALY: quality-adjusted life year; SBP: systolic blood pressure; UKPDS OM2: United Kingdom Prospective Diabetes Study Outcomes Model 2.

Revised Model Inputs

A summary of changes to the model inputs for the preferred base case simulations is provided in the following table (Table 6).

Table 6: Overview of revised inputs for the modelling analysis

Simulation element	Change(s) from submitted base case
Cohort	No changes made
Treatment effects and risk factor progressions	Change from baseline in BMI was taken (where available) directly from the NMA results UKPDS risk factor progressions were used for all risk factors with the exceptions of SBP and BMI during treatment with tirzepatide or comparators
Treatment costs	Pack prices for tirzepatide were as follows: • Tirzepatide 5 mg [redacted] (28 days) • Tirzepatide 10 mg [redacted] (28 days) • Tirzepatide 15 mg [redacted] (28 days)
Complication costs	All complication costs were inflated to 2022 values Costs queried by the EAG were checked against source data and amended if necessary Variance estimates were extracted from source data wherever possible and included in the model inputs
Health-related quality of life utilities	An age-adjusted additive approach to utility estimation was used based on Ara and Brazier 2010[33] Variance estimates were extracted from source data wherever possible and included in the model inputs No weight loss utility (Boye *et al.* 2021) was used in the preferred base case analysis[32] No device utilities for tirzepatide or dulaglutide were used in the preferred base case analysis
Other settings	No other changes made

Abbreviations: BMI: body mass index; EAG: evidence assessment group; NMA: network meta-analysis; SBP: systolic blood pressure; UKPDS: United Kingdom Prospective Diabetes Study.

The revised treatment effects used in the preferred base case analysis are summarized in Table 7 with differences from the original submission (BMI changes only) marked in italics and underlined.

Table 7: Treatment effects applied in the first year of the simulation for tirzepatide and comparators

	TZP 5 mg mean (SD)	TZP 10 mg mean (SD)	TZP 15 mg mean (SD)	DULA 1.5 mg mean (SD)	DULA 3.0 mg mean (SD)	DULA 4.5 mg mean (SD)	SEMA 0.5 mg mean (SD)	SEMA 1.0 mg mean (SD)	ORAL SEMA 7 mg mean (SD)	ORAL SEMA 14 mg mean (SD)	LIRA 1.2 mg mean (SD)	LIRA 1.8 mg mean (SD)
HbA1c change from baseline (%)	[redacted]											
SBP change from baseline (mmHg)	[redacted]											
BMI change from baseline (kg/m2)	[redacted]											
HDL change from baseline (mmol/L)	[redacted]											
LDL change from baseline (mmol/L)	[redacted]											

Values in italics and underlined are changed from the original submission

Abbreviations: BMI: body mass index; DULA: dulaglutide; eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; LIRA: liraglutide; SBP: systolic blood pressure; SD: standard deviation; SEMA: semaglutide; TZP, tirzepatide.

A summary of the complication costs used in the preferred base case analysis (inflated to 2022 values) is provided in Table 8.

Table 8: Summary of direct costs associated with diabetes-related complications used in the modelling analysis

	Mean, 2022 value (£)	Standard error (£)	Original source
Macrovascular complications			
Myocardial infarction, year 1	8,862	1,322	Alva *et al.* Diabet Med. 2015;32(4):459-66
Myocardial infarction, year 2	2,203	250	Alva *et al.* Diabet Med. 2015;32(4):459-66
Stroke, year 1	9,530	2,164	Alva *et al.* Diabet Med. 2015;32(4):459-66
Stroke, year 2	2,270	379	Alva *et al.* Diabet Med. 2015;32(4):459-66
Ischemic heart disease, year 1	12,831	1,799	Alva *et al.* Diabet Med. 2015;32(4):459-66
Ischemic heart disease, year 2	2,256	248	Alva *et al.* Diabet Med. 2015;32(4):459-66
Revascularization, year 1	3,593	359	Shao et al. Pharmacoeconomics. 2019; 37(7): 921-929
Revascularization, year 2	0	0	Assumed
Congestive heart failure, year 1	5,033	1,127	Alva *et al.* Diabet Med. 2015;32(4):459-66
Congestive heart failure, year 2	2,952	510	Alva *et al.* Diabet Med. 2015;32(4):459-66
Microvascular complications			
Foot ulcer, year 1	3,705	371	Kerr *et al.* Diabet. Med. 2019;36: 995-1002, no variance reported, 10% assumed
Foot ulcer, year 2	0	0	Assumed
Amputation, year 1	14,779	2,962	Alva et al. Diabet Med. 2015;32(4):459-66
Amputation, year 2	4,107	837	Alva et al. Diabet Med. 2015;32(4):459-66
Blindness, year 1	3,796	1,409	NI Alva et al. Diabet Med. 2015;32(4):459-66
Blindness, year 2	1,438	229	Alva et al. Diabet Med. 2015;32(4):459-66
Macular oedema	696	70	NHS reference costs 2019/2020*, no variance reported, 10% assumed
Neuropathy/SPSL, all years	1,098	110	Hunt et al. Diabetes Ther. 2017;8(1):129-147, no variance reported, 10% assumed
Renal complications			
KDIGO CKD eGFR stage	0	0	Assumed
KDIGO CKD eGFR stage 2	0	0	Assumed
KDIGO CKD eGFR stage 3	0	0	Assumed
KDIGO CKD eGFR stage 4	472	31	Kent et al. BMC Nephrol. 2015;16:65.

	Mean, 2022 value (£)	Standard error (£)	Original source
KDIGO CKD eGFR stage 5	21,996	2,200	Alva et al. Diabet Med. 2015;32(4):459-66

Abbreviations: CKD: chronic kidney disease; eGFR: estimated glomerular filtration rate; SPSL: severe pressure sensation loss; KDIGO: Kidney Disease Improving Global Outcomes. *Day Case, BZ87A, Minor vitreous retinal procedures, 19 years and over.[35]

A summary of the utilities associated with diabetes-related complication and associated variance estimates used in the preferred base case analysis is provided in Table 9.

Table 9: Utilities and disutilities used in the modelling analysis for diabetes-related complications and hypoglycaemic events

Baseline	Utility	Standard error	Original source
T2D with no complications	+0.815	+0.040	Alva *et al.* Health Econ. 2014;23(4):487-500
Complication/adverse event	**Disutility**	**Standard error**	**Original source**
Macrovascular complications			
Myocardial infarction event	−0.055	0.006	Alva *et al.* Health Econ. 2014;23(4):487-500
History of myocardial infraction	−0.055	0.006	Alva *et al.* Health Econ. 2014;23(4):487-500
Stroke event	−0.164	0.030	Alva *et al.* Health Econ. 2014;23(4):487-500
History of stroke	−0.164	0.030	Alva *et al.* Health Econ. 2014;23(4):487-500
Ischemic heart disease (each year)	−0.090	0.018	Alva *et al.* Health Econ. 2014;23(4):487-500
Revascularization	−0.038	0.011	Shao et al. Pharmacoeconomics. 2019; 37(7): 921-929
History of revascularization	−0.016	0.005	Shao et al. Pharmacoeconomics. 2019; 37(7): 921-929
Congestive heart failure (each year)	−0.108	0.031	Alva *et al.* Health Econ. 2014;23(4):487-500
Microvascular complications			
Foot ulcer (year of event)	−0.170	0.019	Beaudet *et al.* Value Health. 2014;17(4):462-470.
Lower extremity amputation (year of event)	−0.280	0.056	Alva *et al.* Health Econ. 2014;23(4):487-500
Lower extremity amputation (subsequent years)	−0.122	0.025	Hayes *et al.* Value Health. 2016;19:36-41
Blindness (each year)	−0.074	0.025	Alva *et al.* Health Econ. 2014;23(4):487-500
Macular oedema (first year)	−0.047	0.005	Mitchell *et al.* Br J Ophthalmol 2012;96:688-693
Macular oedema (subsequent years)	0	0	Assumed

Baseline	Utility	Standard error	Original source
Neuropathy/SPSL (each years)	−0.066	0.007	Shao *et al.* Pharmacoeconomics. 2019; 37(7): 921-929
Renal complications			
KDIGO CKD eGFR stage 1	0	0	Assumed
KDIGO CKD eGFR stage 2	0	0	Assumed
KDIGO CKD eGFR stage 3	−0.004	0.010	Nauck *et al.* Diabetes Obes Metab. 2019;21:525–532.
KDIGO CKD eGFR stage 4	−0.004	0.010	Nauck *et al.* Diabetes Obes Metab. 2019;21:525–532.
KDIGO CKD eGFR stage 5	−0.164	0.016	Alva *et al.* Health Econ. 2014;23(4):487-500
Adverse events			
Severe hypoglycaemic event	−0.062	0.004	Evans et al. Health Qual Life Outcomes. 2013; 11: 90
Non-severe hypoglycaemic event	−0.005	0.001	Evans et al. Health Qual Life Outcomes. 2013; 11: 90

Abbreviations: CKD: chronic kidney disease; eGFR: estimated glomerular filtration rate; KDIGO: Kidney Disease Improving Global Outcomes; SPSL: severe pressure sensation loss; T2D: type 2 diabetes.

Base Case Results

In line with the revised methodology outlined above, long-term projections with the PRIME T2D Model showed that all three doses of tirzepatide were associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 10, Table 11 and Table 12). Tirzepatide 5 mg was associated with greater lifetime direct costs than eight of the nine comparators, with incremental costs ranging between −£409 and £742. Incremental cost-effectiveness ratios (ICERs) ranged from dominant to £16,817 per QALY gained (Table 10). Tirzepatide 10 mg was associated with higher direct costs than nine comparators, with ICERs for tirzepatide 10 mg ranging between £3,625 and £18,115 per QALY gained (Table 11). A similar pattern of results was projected for tirzepatide 15 mg, with ICERs ranging between £4,498 and £15,209 per QALY gained versus comparators (Table 12). Incremental results between relevant comparators as well as estimates of net health benefit (NHB) for each comparison are provided for the present analysis in Table 10, Table 11 and Table 12). Progression curves for the key risk factors of HbA1c, SBP and BMI are provided for all treatment arms in Figure 21, Figure 22 and Figure 23, respectively.

Table 10: Summary of EAG preferred base case results for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	13.122	8.715	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	705	0.059	0.100	7,073	0.064
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	644	0.046	0.079	8,182	0.047
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	628	0.030	0.058	10,891	0.026
Semaglutide 0.5 mg	[redacted]	13.075	8.634	682	0.047	0.081	8,401	0.047
Semaglutide 1.0 mg	[redacted]	13.096	8.673	708	0.026	0.042	16,817	0.007
Oral semaglutide 7 mg	[redacted]	13.049	8.595	742	0.073	0.120	6,202	0.083
Oral semaglutide 14 mg	[redacted]	13.074	8.642	719	0.048	0.073	9,873	0.037
Liraglutide 1.2 mg	[redacted]	13.032	8.581	672	0.090	0.134	5,021	0.100
Liraglutide 1.8 mg	[redacted]	13.054	8.600	-409	0.068	0.115	Dominant	0.135

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year. * for tirzepatide versus comparator.

Table 11: Summary of EAG preferred base case results for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	13.155	8.768	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	1,723	0.092	0.153	11,272	0.067
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,662	0.079	0.132	12,599	0.049
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,646	0.063	0.111	14,851	0.029
Semaglutide 0.5 mg	[redacted]	13.075	8.634	1,700	0.080	0.134	12,651	0.049
Semaglutide 1.0 mg	[redacted]	13.096	8.673	1,726	0.059	0.095	18,115	0.009
Oral semaglutide 7 mg	[redacted]	13.049	8.595	1,760	0.106	0.173	10,183	0.085
Oral semaglutide 14 mg	[redacted]	13.074	8.642	1,737	0.081	0.126	13,786	0.039

Liraglutide 1.2 mg	[redacted]	13.032	8.581	1,690	0.123	0.187	9,038	0.102
Liraglutide 1.8 mg	[redacted]	13.054	8.600	609	0.101	0.168	3,625	0.138

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year. * for tirzepatide versus comparator.

Table 12: Summary of EAG preferred base case results for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	13.176	8.808	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	2,047	0.113	0.192	10,642	0.090
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,987	0.100	0.171	11,586	0.072
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,970	0.084	0.150	13,104	0.052
Semaglutide 0.5 mg	[redacted]	13.075	8.634	2,025	0.101	0.174	11,641	0.073
Semaglutide 1.0 mg	[redacted]	13.096	8.673	2,051	0.080	0.135	15,209	0.032
Oral semaglutide 7 mg	[redacted]	13.049	8.595	2,085	0.127	0.212	9,815	0.108
Oral semaglutide 14 mg	[redacted]	13.074	8.642	2,061	0.102	0.166	12,453	0.062
Liraglutide 1.2 mg	[redacted]	13.032	8.581	2,014	0.144	0.227	8,893	0.126
Liraglutide 1.8 mg	[redacted]	13.054	8.600	934	0.122	0.208	4,498	0.161

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Figure 18: Cost-effectiveness frontier for tirzepatide 5 mg versus comparators

The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: GBP: Great British Pounds; QALY: quality-adjusted life year; TZP: tirzepatide.

Figure 19: Cost-effectiveness frontier for tirzepatide 10 mg versus comparators

The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: GBP: Great British Pounds; QALY: quality-adjusted life year; TZP: tirzepatide.

Figure 20: Cost-effectiveness frontier for tirzepatide 15 mg versus comparators

The broken line indicates the cost-effectiveness frontier with the corresponding incremental cost-effectiveness ratio for tirzepatide versus the most-effective comparator. Comparators above the line can be considered less cost-effective.
Abbreviations: QALY: quality-adjusted life year; TZP: tirzepatide.

Figure 21: HbA1c progression for each treatment arm in the cost-effectiveness analysis based on UKPDS OM2 risk factor progression

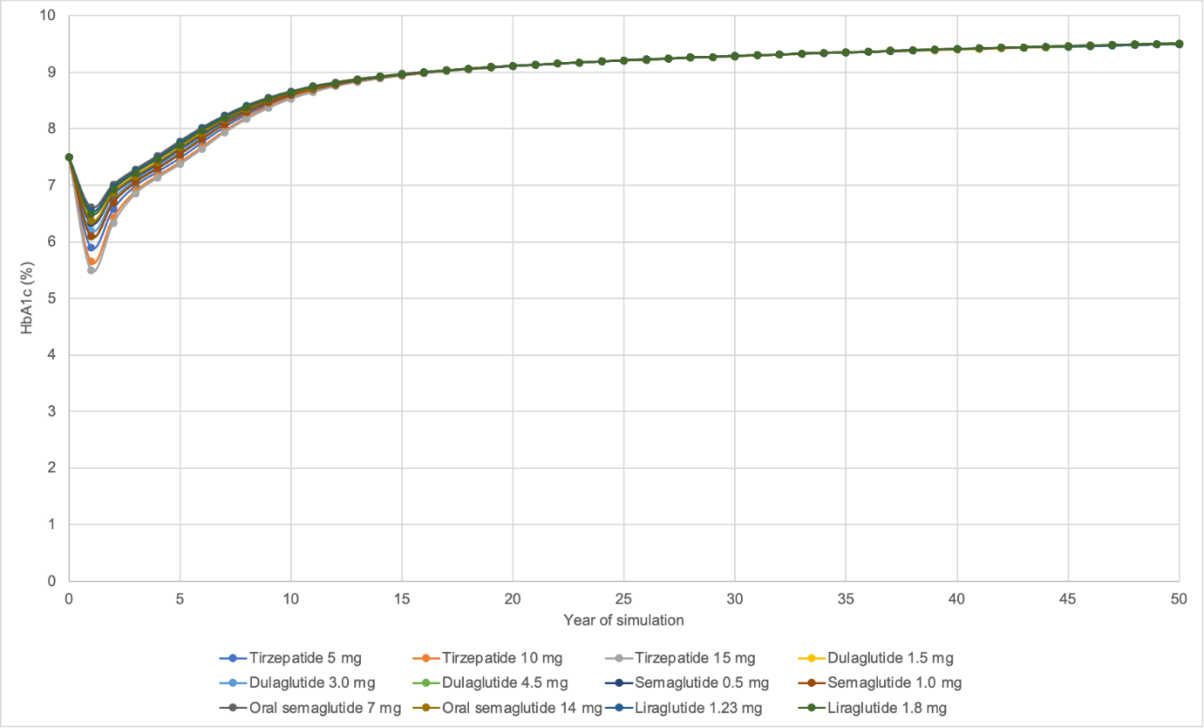


Figure 22: Systolic blood pressure progression for each treatment arm in the cost-effectiveness analysis based no progression during treatment and UKPDS OM2 risk factor progression after intensification to basal insulin

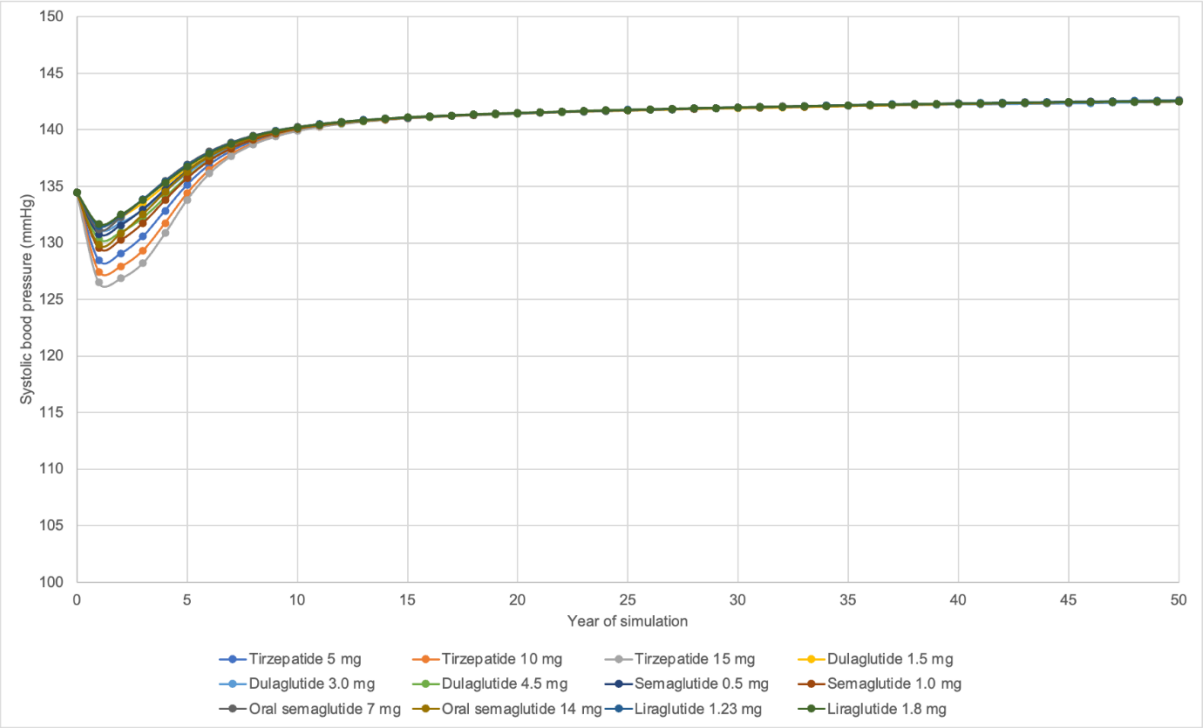
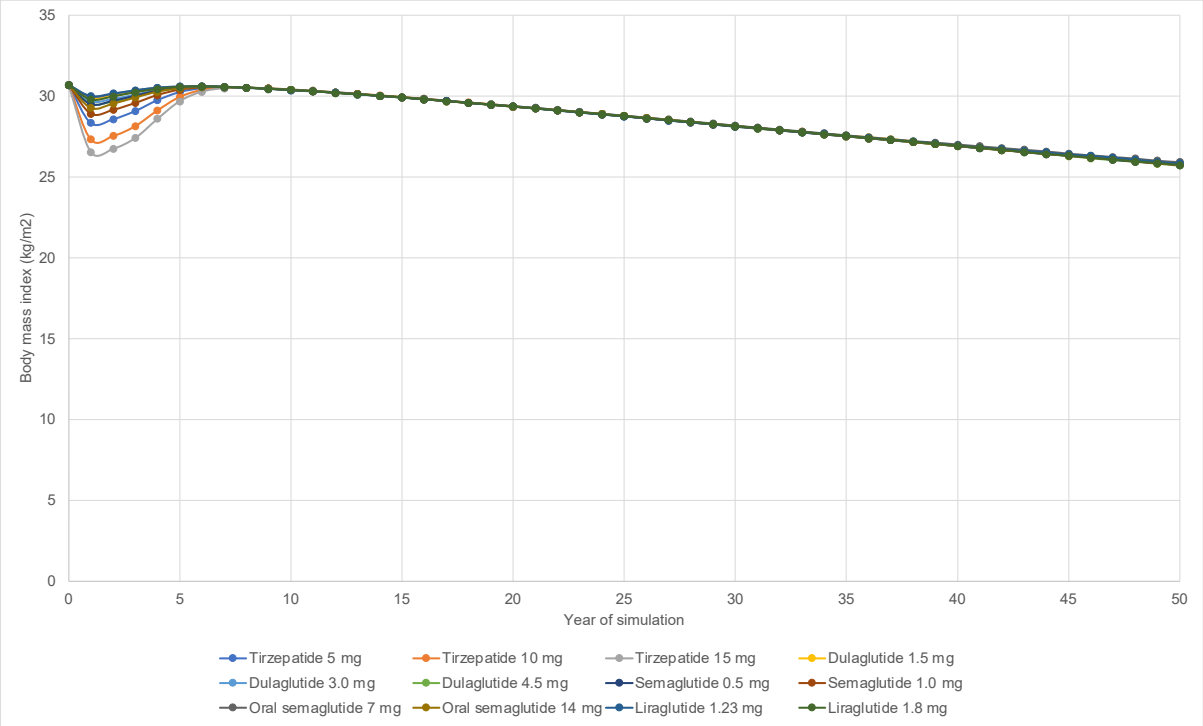


Figure 23: Body mass index progression for each treatment arm in the cost-effectiveness analysis based no progression during treatment and UKPDS OM2 risk factor progression after intensification to basal insulin



Probabilistic Sensitivity Analysis Results

For PSA, the model reported results based on a nonparametric bootstrapping approach, in which samples from 1% of the simulated population (in this case comprising 3,000 patients) were drawn 1,000 times from the full patient data set at the end of the simulation. Sampling was performed with replacement. The population mean and confidence intervals were then calculated by rerunning the cost and quality-of-life estimators on each sampled population and generating a set of descriptive statistics. PSA results for the comparisons of tirzepatide 5 mg with semaglutide 0.5 mg and tirzepatide 10 and 15 mg with semaglutide 1.0 mg (comparators selected based on cost-effectiveness frontiers, see Figure 18, Figure 19 and Figure 20 for details) are provided in Table 13.

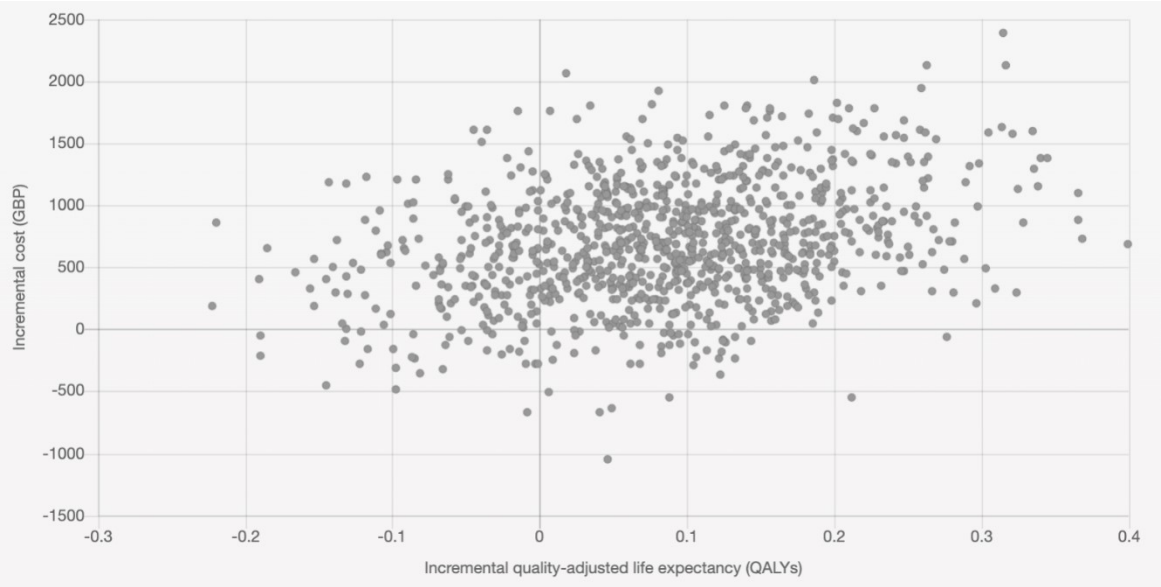
PSA indicated that there was a 70.6%probability that tirzepatide 5 mg would be cost-effective versus semaglutide 0.5 mg, assuming a willingness to pay threshold of £20,000 per QALY. Scatterplots and acceptability curves for the comparisons of tirzepatide 5 mg with semaglutide 0.5 mg are provided in Figure 24 and Figure 25. Similarly, PSA for tirzepatide 10 mg versus semaglutide 1.0 mg suggested that there was a 65.3% probability that tirzepatide 10 mg would be cost-effective, assuming a willingness to pay threshold of £20,000 per QALY gained. The incremental cost-effectiveness scatter plot and acceptability curve for the tirzepatide 10 mg versus semaglutide 1.0 mg are shown in Figure 26 and Figure 27, respectively. For tirzepatide 15 mg, PSA indicated that there was a 77.3% probability that tirzepatide would be cost-effective against semaglutide 1.0 mg, assuming a willingness to pay threshold of £20,000 per QALY gained. The incremental cost-effectiveness scatter plot and acceptability curve for the tirzepatide 15 mg versus semaglutide 1.0 mg are shown in Figure 28 and Figure 29, respectively.

Table 13: Summary of probabilistic sensitivity analysis results for tirzepatide versus comparators

	Direct costs (£)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental QALYs*	ICER* (£ per QALY gained)	Probability of tirzepatide being cost-effective**
Tirzepatide 5 mg	[redacted]	7.224 (7.151 – 7.296)				
Semaglutide 0.5 mg (versus tirzepatide 5 mg)	[redacted]	7.138 (7.069 – 7.207)	707 (216 – 1,180)	0.087 (-0.015 – 0.186)	8,149	70.6%
Tirzepatide 10 mg	[redacted]	7.286 (7.219 – 7.363)				
Semaglutide 1.0 mg (versus tirzepatide 10 mg)	[redacted]	7.174 (7.104 – 7.240)	1,585 (1,165 – 2,064)	0.112 (0.020 – 0.218)	14,137	65.3%
Tirzepatide 15 mg	[redacted]	7.331 (7.262 – 7.402)				
Semaglutide 1.0 mg (versus tirzepatide 15 mg)	[redacted]	7.174 (7.104 – 7.240)	1,801 (1,359 – 2,269)	0.157 (0.064 – 0.256)	11,506	77.3%

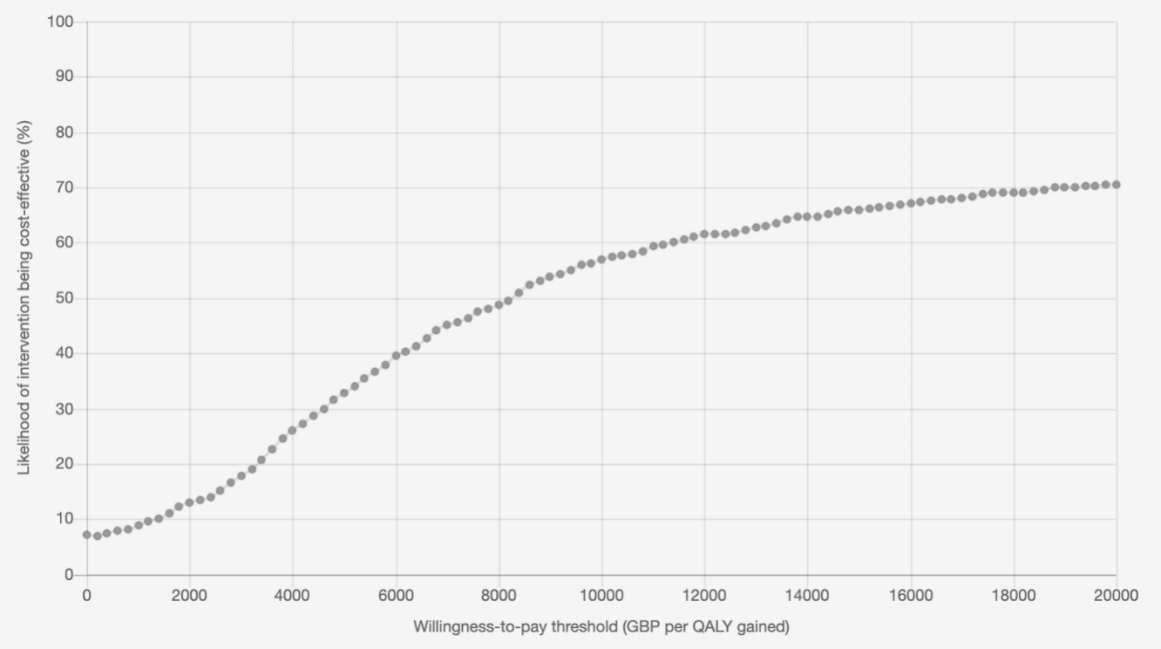
Values shown are means with 95% credible intervals in parentheses. * for tirzepatide versus comparator; **assuming a willingness to pay threshold of £20,000 per QALY again.
Abbreviations: QALY: quality-adjusted life year.

Figure 24: Cost-effectiveness scatterplot from probabilistic sensitivity analysis of tirzepatide 5 mg versus semaglutide 0.5 mg



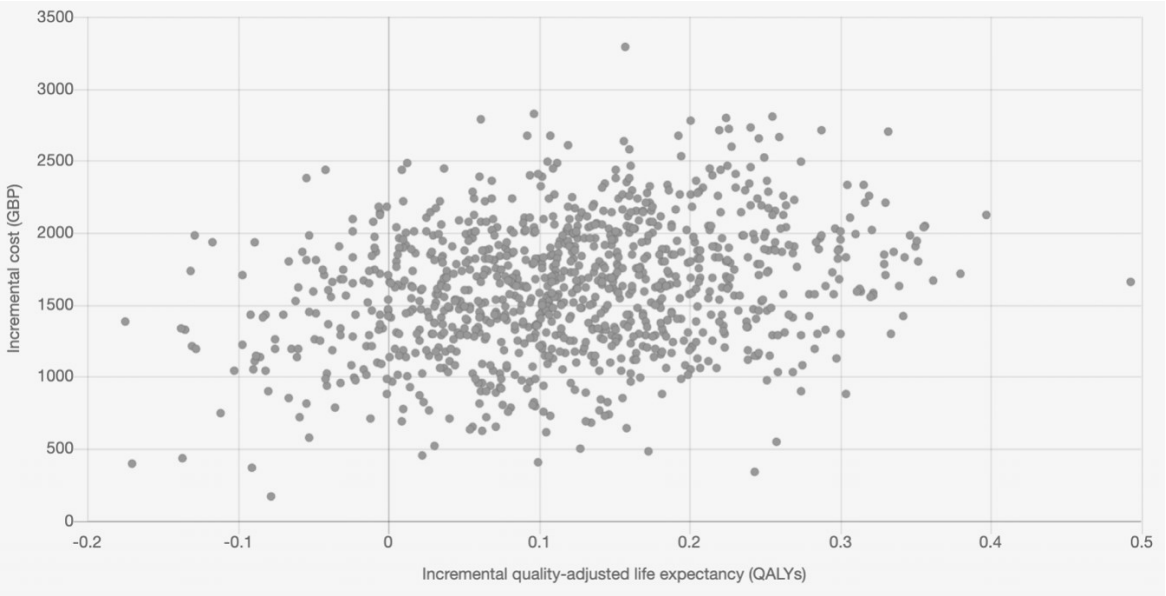
Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 25: Cost-effectiveness acceptability curve based on the probabilistic sensitivity analysis of tirzepatide 5 mg versus semaglutide 0.5 mg



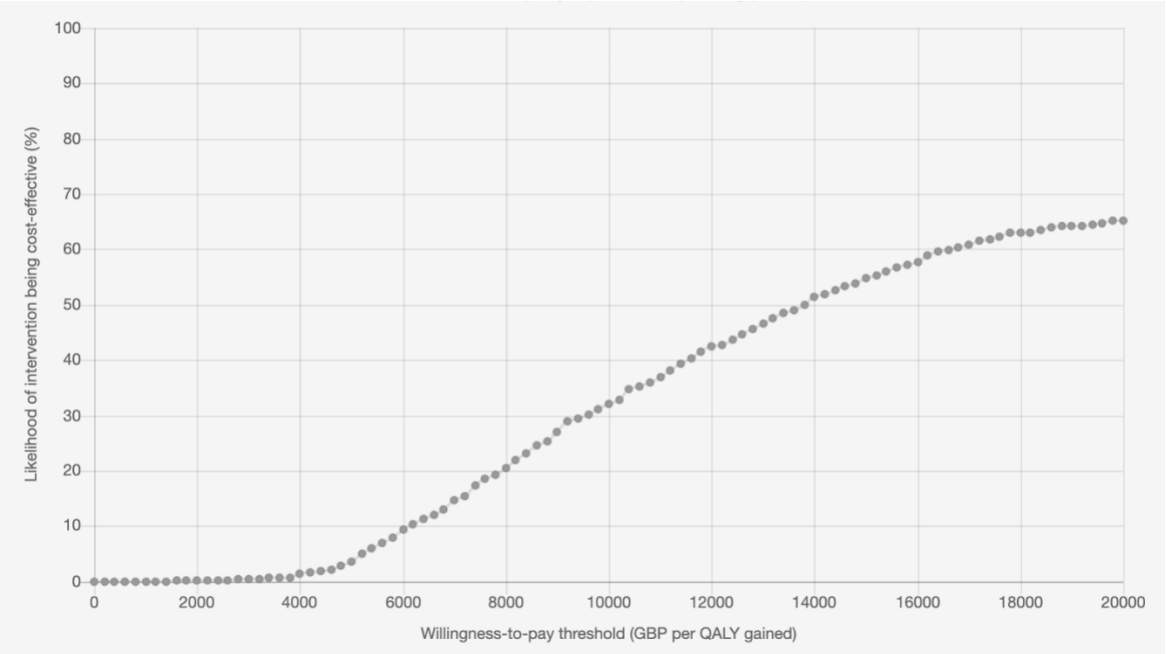
Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 26: Cost-effectiveness scatterplot from probabilistic sensitivity analysis of tirzepatide 10 mg versus semaglutide 1.0 mg



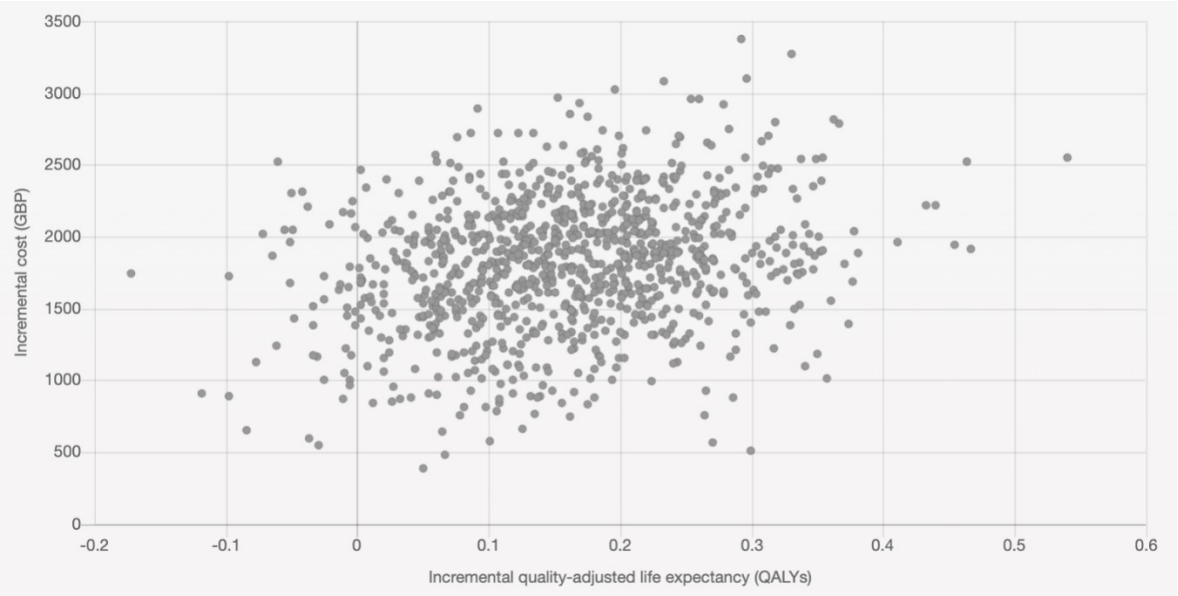
Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 27: Cost-effectiveness acceptability curve based on the probabilistic sensitivity analysis of tirzepatide 10 mg versus semaglutide 1.0 mg



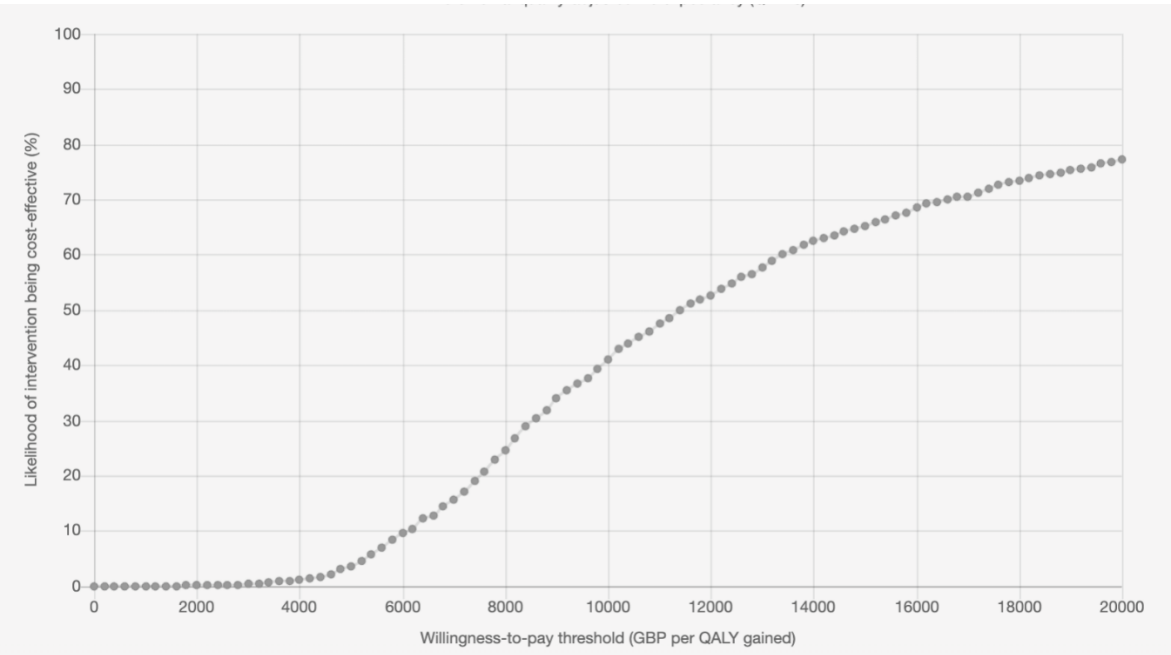
Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 28: Cost-effectiveness scatterplot from probabilistic sensitivity analysis of tirzepatide 15 mg versus semaglutide 1.0 mg



Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Figure 29: Cost-effectiveness acceptability curve based on the probabilistic sensitivity analysis of tirzepatide 15 mg versus semaglutide 1.0 mg



Abbreviations: GBP: Great British Pounds; QALYs: quality-adjusted life years.

Scenario Analysis Results

An overview of scenario analyses based on the preferred base case simulations is provided in Table 14.

Table 14: Overview of scenario analyses

EAG Suggestion (Section 6.1.2)	Action taken or response
12. Using the CORE Diabetes model (consistent with NG28). See Section 4.2.2 EAG comments a and b, as well as key issue 6.	No simulations were run using the CORE Diabetes Model. The rationale is outlined in the responses to *key issue 6* and *key issue 7*. In addition, time constraints for the review process meant a complete re-analysis in a new model environment was impossible.
13. Selecting single predictive models based on the best match of the derivation cohort to the decision problem (as requested in clarification question B4c). See Section 4.2.2 EAG comment c, as well as key issue 7.	No simulations were run in response to this comment in line with the rationale outlined in the responses to *key issue 6* and *key issue 7*.
14. Assuming the population characteristics from the SURPASS-2 trial (instead of based on THIN second intensification cohort). See Section 4.2.3 EAG comment a.	This scenario was run and the results are summarized below.
15. Assuming waning of the relative treatment effect while on the initial treatment. See Section 4.2.6 EAG comment b, as well as key issue 8.	No simulations were run in response to this comment in line with the rationale provided in response to *key issue 8*.
16. Assuming no difference in HDL and LDL between tirzepatide and dulaglutide. See Section 4.2.6 EAG comment f.	This scenario was run and the results are summarized below.
17. Only assume a utility decrement for higher BMI values, as was done in NG28 HE report. See Section 4.2.8 EAG comment d. This adjustment will likely increase the estimated ICER. Assuming no body weight change utility would increase the estimated ICER by roughly £600 for the comparison tirzepatide 10 mg versus semaglutide 1.0 mg, (CS Table 106).	This assumption is in line with preferred base case simulations (see above for details). A scenario analysis was run integrating the Boye *et al.* (2021) weight loss utilities during the first year on therapy to investigate the impact on cost-effectiveness (summarized below).
18. Exploring the impact of including costs associated with nausea in a scenario analysis. See Section 4.2.9 EAG comment c. This adjustment will likely increase the estimated ICER. The magnitude of the impact is unclear as including costs associated with nausea was not explored by the company.	This scenario was run and the results are summarized below.
19. Exploring the impact of including T2D health state costs. See Section 4.2.9 EAG comment d.	This scenario was run and the results are summarized below.

Scenario analysis 1 – SURPASS-2 population

Long-term projections with the PRIME T2D Model using SURPASS-2 cohort characteristics showed, as in the base case, that all three doses of tirzepatide were associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 15, Table 16 and Table 17). Tirzepatide 5 mg was associated with greater lifetime direct costs than most comparators, with incremental costs ranging between £433 and £537 and incremental cost-effectiveness ratios (ICERs) ranging between £3,936 and £11,287 per QALY gained. The exception was the comparison with liraglutide 1.8 mg, where tirzepatide 5 mg was projected to be cost saving and therefore dominant. Tirzepatide 10 mg was also associated with higher direct costs than the comparators, with ICERs for tirzepatide 10 mg

ranging between £4,279 and £14,236 per QALY gained. A similar pattern of results was projected for tirzepatide 15 mg, with ICERs ranging between £4,336 and £11,031 per QALY gained versus comparators. Incremental results between tirzepatide and comparators as well as estimates of net health benefit (NHB) are provided for the present analysis in Table 15, Table 16 and Table 17.

Table 15; Summary of SURPASS-2 cohort scenario analysis for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	14.169	9.400	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	14.118	9.306	457	0.051	0.094	4,873	0.071
Dulaglutide 3.0 mg	[redacted]	14.129	9.329	433	0.040	0.071	6,072	0.050
Dulaglutide 4.5 mg	[redacted]	14.147	9.351	433	0.022	0.050	8,737	0.028
Semaglutide 0.5 mg	[redacted]	14.127	9.325	537	0.042	0.075	7,153	0.048
Semaglutide 1.0 mg	[redacted]	14.141	9.356	498	0.028	0.044	11,278	0.019
Oral semaglutide 7 mg	[redacted]	14.098	9.280	473	0.071	0.120	3,936	0.097
Oral semaglutide 14 mg	[redacted]	14.139	9.333	461	0.030	0.067	6,862	0.044
Liraglutide 1.2 mg	[redacted]	14.103	9.284	461	0.066	0.116	3,970	0.093
Liraglutide 1.8 mg	[redacted]	14.111	9.295	-138	0.058	0.106	Dominant	0.113

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 16: Summary of SURPASS-2 cohort scenario analysis for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	14.195	9.446	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	14.118	9.306	1,245	0.077	0.140	8,896	0.078
Dulaglutide 3.0 mg	[redacted]	14.129	9.329	1,221	0.066	0.117	10,394	0.056
Dulaglutide 4.5 mg	[redacted]	14.147	9.351	1,221	0.048	0.096	12,752	0.035
Semaglutide 0.5 mg	[redacted]	14.127	9.325	1,325	0.068	0.121	10,929	0.055
Semaglutide 1.0 mg	[redacted]	14.141	9.356	1,286	0.054	0.090	14,236	0.026
Oral semaglutide 7 mg	[redacted]	14.098	9.280	1,261	0.097	0.166	7,579	0.103

Oral semaglutide 14 mg	[redacted]	14.139	9.333	1,249	0.056	0.113	11,021	0.051
Liraglutide 1.2 mg	[redacted]	14.103	9.284	1,249	0.092	0.162	7,696	0.100
Liraglutide 1.8 mg	[redacted]	14.111	9.295	649	0.084	0.152	4,279	0.119

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 17: Summary of SURPASS-2 cohort scenario analysis for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	14.226	9.491	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	14.118	9.306	1,446	0.108	0.184	7,842	0.112
Dulaglutide 3.0 mg	[redacted]	14.129	9.329	1,422	0.097	0.162	8,781	0.091
Dulaglutide 4.5 mg	[redacted]	14.147	9.351	1,423	0.079	0.140	10,142	0.069
Semaglutide 0.5 mg	[redacted]	14.127	9.325	1,526	0.099	0.166	9,209	0.089
Semaglutide 1.0 mg	[redacted]	14.141	9.356	1,488	0.085	0.135	11,031	0.060
Oral semaglutide 7 mg	[redacted]	14.098	9.280	1,463	0.128	0.211	6,935	0.138
Oral semaglutide 14 mg	[redacted]	14.139	9.333	1,450	0.087	0.158	9,189	0.085
Liraglutide 1.2 mg	[redacted]	14.103	9.284	1,451	0.123	0.207	7,014	0.134
Liraglutide 1.8 mg	[redacted]	14.111	9.295	851	0.115	0.196	4,336	0.154

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Scenario analysis 2 – No differences in HDL of LDL between tirzepatide and dulaglutide

Long-term projections with the PRIME T2D Model where changes from baseline in HDL and LDL for tirzepatide were matched the values for dulaglutide treatment showed, as in the base case, that all three doses of tirzepatide were associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 18, Table 19 and Table 20). Tirzepatide 5 mg was associated with greater lifetime direct costs than most comparators, with incremental cost-effectiveness ratios (ICERs) ranging between £5,315 and £16,449 per QALY gained. The exception was the comparison with liraglutide 1.8 mg, where tirzepatide 5 mg was projected to be cost saving and therefore dominant. Tirzepatide 10 mg was also associated with higher direct costs than the comparators, with ICERs for tirzepatide 10 mg ranging between £3,896 and £19,724 per QALY gained. A similar pattern of results was projected for tirzepatide 15 mg, with ICERs ranging between £4,588 and £15,796 per QALY gained versus comparators. Incremental results between tirzepatide and comparators as well as estimates of net health benefit (NHB) are provided for the scenario analysis in Table 18, Table 19 and Table 20.

Table 18: Summary of the scenario analysis of HDL and LDL changes for tirzepatide matched to dulaglutide for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	13.135	8.720	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	770	0.072	0.105	7,366	0.066
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	710	0.059	0.084	8,483	0.048
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	693	0.043	0.063	11,081	0.028
Semaglutide 0.5 mg	[redacted]	13.075	8.634	748	0.060	0.086	8,681	0.049
Semaglutide 1.0 mg	[redacted]	13.096	8.673	774	0.039	0.047	16,449	0.008
Oral semaglutide 7 mg	[redacted]	13.049	8.595	808	0.086	0.125	6,483	0.084
Oral semaglutide 14 mg	[redacted]	13.074	8.642	784	0.061	0.078	10,090	0.039
Liraglutide 1.2 mg	[redacted]	13.032	8.581	737	0.103	0.139	5,315	0.102
Liraglutide 1.8 mg	[redacted]	13.054	8.600	-343	0.081	0.120	Dominant	0.137

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 19: Summary of the scenario analysis of HDL and LDL changes for tirzepatide matched to dulaglutide for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	13.147	8.761	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	1,742	0.084	0.146	11,928	0.059
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,681	0.071	0.125	13,437	0.041
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,665	0.055	0.104	16,006	0.021
Semaglutide 0.5 mg	[redacted]	13.075	8.634	1,719	0.072	0.128	13,475	0.042
Semaglutide 1.0 mg	[redacted]	13.096	8.673	1,745	0.051	0.088	19,724	0.001

Oral semaglutide 7 mg	[redacted]	13.049	8.595	1,779	0.098	0.166	10,715	0.077
Oral semaglutide 14 mg	[redacted]	13.074	8.642	1,756	0.073	0.119	14,733	0.031
Liraglutide 1.2 mg	[redacted]	13.032	8.581	1,709	0.115	0.180	9,485	0.095
Liraglutide 1.8 mg	[redacted]	13.054	8.600	628	0.093	0.161	3,896	0.130

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 20: Summary of the scenario analysis of HDL and LDL changes for tirzepatide matched to dulaglutide for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	13.171	8.802	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	2,041	0.108	0.187	10,917	0.085
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,981	0.095	0.166	11,926	0.067
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,964	0.079	0.145	13,549	0.047
Semaglutide 0.5 mg	[redacted]	13.075	8.634	2,019	0.096	0.169	11,977	0.068
Semaglutide 1.0 mg	[redacted]	13.096	8.673	2,045	0.075	0.129	15,796	0.027
Oral semaglutide 7 mg	[redacted]	13.049	8.595	2,079	0.122	0.207	10,041	0.103
Oral semaglutide 14 mg	[redacted]	13.074	8.642	2,055	0.097	0.160	12,835	0.057
Liraglutide 1.2 mg	[redacted]	13.032	8.581	2,008	0.139	0.221	9,082	0.121
Liraglutide 1.8 mg	[redacted]	13.054	8.600	928	0.117	0.202	4,588	0.156

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Scenario analysis 3 – Weight loss utilities included in year 1

Long-term projections with the PRIME T2D Model where quality of life utilities associated with weight loss (Boye *et al.* 2022) tirzepatide and all comparators showed, as in the base case, that all three doses of tirzepatide were associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 21, Table 22 and Table 23.).[32] Tirzepatide 5 mg was associated with greater lifetime direct costs than most comparators, with incremental cost-effectiveness ratios (ICERs) ranging between £4,435 and £14,823 per QALY gained. The exception was the comparison with liraglutide 1.8 mg, where tirzepatide 5 mg was projected to be cost saving and therefore dominant. Tirzepatide 10 mg was also associated with higher direct costs than the comparators, with ICERs for tirzepatide 10 mg ranging between £3,243 and £16,337 per QALY gained. Similar results were projected for tirzepatide 15 mg, with ICERs ranging between £4,075 and £13,957 per QALY gained versus comparators. Incremental results between tirzepatide and comparators as well as estimates of net health benefit (NHB) for each comparison are provided for the scenario analysis in Table 21, Table 22 and Table 23.

Table 21: Summary of the scenario analysis with weight loss utilities included in year 1 for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	13.122	8.767	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.651	705	0.059	0.116	6,065	0.081
Dulaglutide 3.0 mg	[redacted]	13.076	8.677	644	0.046	0.090	7,150	0.058
Dulaglutide 4.5 mg	[redacted]	13.092	8.700	628	0.030	0.067	9,355	0.036
Semaglutide 0.5 mg	[redacted]	13.075	8.673	682	0.047	0.094	7,256	0.060
Semaglutide 1.0 mg	[redacted]	13.096	8.719	708	0.026	0.048	14,823	0.012
Oral semaglutide 7 mg	[redacted]	13.049	8.629	742	0.073	0.138	5,378	0.101
Oral semaglutide 14 mg	[redacted]	13.074	8.683	719	0.048	0.084	8,560	0.048
Liraglutide 1.2 mg	[redacted]	13.032	8.616	672	0.090	0.151	4,435	0.118
Liraglutide 1.8 mg	[redacted]	13.054	8.637	-409	0.068	0.130	Dominant	0.150

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 22: Summary of the scenario analysis with weight loss utilities included in year 1 for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	13.155	8.825	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.651	1,723	0.092	0.174	9,895	0.088
Dulaglutide 3.0 mg	[redacted]	13.076	8.677	1,662	0.079	0.148	11,231	0.065
Dulaglutide 4.5 mg	[redacted]	13.092	8.700	1,646	0.063	0.125	13,167	0.043
Semaglutide 0.5 mg	[redacted]	13.075	8.673	1,700	0.080	0.152	11,192	0.067
Semaglutide 1.0 mg	[redacted]	13.096	8.719	1,726	0.059	0.106	16,337	0.019
Oral semaglutide 7 mg	[redacted]	13.049	8.629	1,760	0.106	0.196	8,984	0.108

Oral semaglutide 14 mg	[redacted]	13.074	8.683	1,737	0.081	0.142	12,243	0.055
Liraglutide 1.2 mg	[redacted]	13.032	8.616	1,690	0.123	0.209	8,072	0.125
Liraglutide 1.8 mg	[redacted]	13.054	8.637	609	0.101	0.188	3,243	0.157

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 23: Summary of the scenario analysis with weight loss utilities included in year 1 for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	13.176	8.866	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.651	2,047	0.113	0.215	9,506	0.113
Dulaglutide 3.0 mg	[redacted]	13.076	8.677	1,987	0.100	0.189	10,497	0.090
Dulaglutide 4.5 mg	[redacted]	13.092	8.700	1,970	0.084	0.166	11,851	0.068
Semaglutide 0.5 mg	[redacted]	13.075	8.673	2,025	0.101	0.193	10,481	0.092
Semaglutide 1.0 mg	[redacted]	13.096	8.719	2,051	0.080	0.147	13,957	0.044
Oral semaglutide 7 mg	[redacted]	13.049	8.629	2,085	0.127	0.237	8,789	0.133
Oral semaglutide 14 mg	[redacted]	13.074	8.683	2,061	0.102	0.183	11,256	0.080
Liraglutide 1.2 mg	[redacted]	13.032	8.616	2,014	0.144	0.251	8,037	0.150
Liraglutide 1.8 mg	[redacted]	13.054	8.637	934	0.122	0.229	4,075	0.182

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Scenario analysis 4 – Costs associated with nausea included

For this scenario, each simulated patient experiencing nausea was assumed to receive 20 minutes with a community-based band 6 nurse and a course of 28 prochlorperazine 5 mg tablets (£18.33 plus £1.24 = 19.57) based on estimates from the PSSRU Unit Costs Database of Health and Social Care Professionals 2020/21 (https://www.pssru.ac.uk/project-pages/unit-costs/) and the NHS Electronic Drug Tariff (https://www.drugtariff.nhsbsa.nhs.uk/#/00837338-DC/DD00837127/Part%20VIIIA%20products%20P). Long-term projections with this cost included in the analysis produced results very similar to the base case, with all three doses of tirzepatide were associated with improvements in life expectancy and quality-adjusted life expectancy versus all comparators evaluated (Table 24, Table 25 and Table 26).[32] Tirzepatide 5 mg was associated with greater lifetime direct costs than most comparators, with incremental cost-effectiveness ratios (ICERs) ranging between £5,070 and £16,866 per QALY gained. The exception was the comparison with liraglutide 1.8 mg, where tirzepatide 5 mg was projected to be cost saving and therefore dominant. Tirzepatide 10 mg was associated with ICERs ranging between £3,687 and £18,205 per QALY gained. Similar results were projected for tirzepatide 15 mg, with ICERs ranging between £4,562 and £15,294 per QALY gained versus comparators. Incremental results between tirzepatide and comparators as well as estimates of net health benefit (NHB) for each comparison are provided for the scenario analysis in Table 24, Table 25 and Table 26.

Table 24: Summary of the scenario analysis with nausea costs included for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	13.122	8.715	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	707	0.059	0.100	7,092	0.064
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	645	0.046	0.079	8,189	0.047
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	628	0.030	0.058	10,889	0.026
Semaglutide 0.5 mg	[redacted]	13.075	8.634	685	0.047	0.081	8,438	0.047
Semaglutide 1.0 mg	[redacted]	13.096	8.673	710	0.026	0.042	16,866	0.007
Oral semaglutide 7 mg	[redacted]	13.049	8.595	748	0.073	0.120	6,246	0.082
Oral semaglutide 14 mg	[redacted]	13.074	8.642	721	0.048	0.073	9,898	0.037
Liraglutide 1.2 mg	[redacted]	13.032	8.581	678	0.090	0.134	5,070	0.100
Liraglutide 1.8 mg	[redacted]	13.054	8.600	-405	0.068	0.115	Dominant	0.135

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 25; Summary of the scenario analysis with nausea costs included for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	13.155	8.768	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	1,731	0.092	0.153	11,328	0.066
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,669	0.079	0.132	12,654	0.048
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,652	0.063	0.111	14,909	0.028
Semaglutide 0.5 mg	[redacted]	13.075	8.634	1,710	0.080	0.134	12,722	0.049
Semaglutide 1.0 mg	[redacted]	13.096	8.673	1,735	0.059	0.095	18,205	0.009
Oral semaglutide 7 mg	[redacted]	13.049	8.595	1,772	0.106	0.173	10,250	0.084

Oral semaglutide 14 mg	[redacted]	13.074	8.642	1,745	0.081	0.126	13,852	0.039
Liraglutide 1.2 mg	[redacted]	13.032	8.581	1,703	0.123	0.187	9,108	0.102
Liraglutide 1.8 mg	[redacted]	13.054	8.600	620	0.101	0.168	3,687	0.137

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 26: Summary of the scenario analysis with nausea costs included for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	13.176	8.808	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	2,059	0.113	0.192	10,701	0.089
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,997	0.100	0.171	11,644	0.072
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,979	0.084	0.150	13,165	0.051
Semaglutide 0.5 mg	[redacted]	13.075	8.634	2,037	0.101	0.174	11,712	0.072
Semaglutide 1.0 mg	[redacted]	13.096	8.673	2,062	0.080	0.135	15,294	0.032
Oral semaglutide 7 mg	[redacted]	13.049	8.595	2,099	0.127	0.212	9,883	0.107
Oral semaglutide 14 mg	[redacted]	13.074	8.642	2,072	0.102	0.166	12,520	0.062
Liraglutide 1.2 mg	[redacted]	13.032	8.581	2,030	0.144	0.227	8,963	0.125
Liraglutide 1.8 mg	[redacted]	13.054	8.600	947	0.122	0.208	4,562	0.160

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Scenario analysis 5 – Health state costs associated with T2D included

For the scenario analysis including health state costs associated with type 2 diabetes, an annual cost of £1,114.63 was assumed for each patient in the simulation.[36] The simulations provided much higher estimates of overall costs but produced incremental values similar to the base case analysis as shown in Table 27, Table 28 and Table 29. Tirzepatide 5 mg was associated with greater lifetime direct costs than most comparators, with incremental cost-effectiveness ratios (ICERs) ranging between £5,742 and £17,482 per QALY gained. The exception was the comparison with liraglutide 1.8 mg, where tirzepatide 5 mg was projected to be cost saving and therefore dominant. Tirzepatide 10 mg was associated with ICERs ranging between £4,274 and £18,792 per QALY gained. Similar results were projected for tirzepatide 15 mg, with ICERs ranging between £5,129 and £15,851 per QALY gained versus comparators. Incremental results between tirzepatide and comparators as well as estimates of net health benefit (NHB) for each comparison are provided for the scenario analysis in Table 27, Table 28 and Table 29.

Table 27: Summary of the scenario analysis with type 2 diabetes state costs included for tirzepatide 5 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 5 mg	[redacted]	13.122	8.715	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	768	0.059	0.100	7,705	0.061
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	694	0.046	0.079	8,814	0.044
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	660	0.030	0.058	11,457	0.025
Semaglutide 0.5 mg	[redacted]	13.075	8.634	733	0.047	0.081	9,019	0.045
Semaglutide 1.0 mg	[redacted]	13.096	8.673	736	0.026	0.042	17,482	0.005
Oral semaglutide 7 mg	[redacted]	13.049	8.595	821	0.073	0.120	6,859	0.079
Oral semaglutide 14 mg	[redacted]	13.074	8.642	771	0.048	0.073	10,585	0.034
Liraglutide 1.2 mg	[redacted]	13.032	8.581	768	0.090	0.134	5,742	0.095
Liraglutide 1.8 mg	[redacted]	13.054	8.600	-336	0.068	0.115	Dominant	0.132

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 28: Summary of the scenario analysis with type 2 diabetes state costs included for tirzepatide 10 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 10 mg	[redacted]	13.155	8.768	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	1,822	0.092	0.153	11,923	0.062
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	1,749	0.079	0.132	13,253	0.045
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	1,715	0.063	0.111	15,475	0.025
Semaglutide 0.5 mg	[redacted]	13.075	8.634	1,787	0.080	0.134	13,295	0.045
Semaglutide 1.0 mg	[redacted]	13.096	8.673	1,791	0.059	0.095	18,792	0.006
Oral semaglutide 7 mg	[redacted]	13.049	8.595	1,875	0.106	0.173	10,848	0.079

Oral semaglutide 14 mg	[redacted]	13.074	8.642	1,825	0.081	0.126	14,486	0.035
Liraglutide 1.2 mg	[redacted]	13.032	8.581	1,823	0.123	0.187	9,749	0.096
Liraglutide 1.8 mg	[redacted]	13.054	8.600	718	0.101	0.168	4,274	0.132

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

Table 29: Summary of the scenario analysis with type 2 diabetes state costs included for tirzepatide 15 mg versus comparators

	Direct costs (£)	Life expectancy (years)	Quality-adjusted life expectancy (QALYs)	Incremental costs (£)	Incremental life years*	Incremental QALYs*	ICER* (£ per QALY gained)	NHB (QALYs)
Tirzepatide 15 mg	[redacted]	13.176	8.808	--	--	--	--	--
Dulaglutide 1.5 mg	[redacted]	13.063	8.615	2,169	0.113	0.192	11,274	0.084
Dulaglutide 3.0 mg	[redacted]	13.076	8.636	2,095	0.100	0.171	12,219	0.067
Dulaglutide 4.5 mg	[redacted]	13.092	8.657	2,061	0.084	0.150	13,710	0.047
Semaglutide 0.5 mg	[redacted]	13.075	8.634	2,134	0.101	0.174	12,266	0.067
Semaglutide 1.0 mg	[redacted]	13.096	8.673	2,137	0.080	0.135	15,851	0.028
Oral semaglutide 7 mg	[redacted]	13.049	8.595	2,222	0.127	0.212	10,461	0.101
Oral semaglutide 14 mg	[redacted]	13.074	8.642	2,172	0.102	0.166	13,120	0.057
Liraglutide 1.2 mg	[redacted]	13.032	8.581	2,169	0.144	0.227	9,577	0.118
Liraglutide 1.8 mg	[redacted]	13.054	8.600	1,065	0.122	0.208	5,129	0.154

Abbreviations: NHB: net health benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; * for tirzepatide versus comparator.

References

1. National Institute of Health and Care Excellence (NICE). Guideline: type 2 diabetes management. Available at: https://www.nice.org.uk/guidance/ng28/ [Last accessed 18 February 2022].
2. European Medicines Agency (EMA). Clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus - Scientific guideline. Available at: https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus-scientific. [Accessed 11 May 2023].
3. Ozempic Summary of Product Characteristics (SmPC). Available at https://www.ema.europa.eu/en/documents/product-information/ozempic-epar-product-information_en.pdf. [Accessed 13 February 2023].
4. All Wales Medicines Strategy Group. AWMSG Secretariat Assessment Report. Semaglutide (Ozempic®)[1842]. Available at: https://awttc.nhs.wales/files/appraisals-asar-far/appraisal-report-semaglutide-ozempic-1842/. [Accessed 5 May 23]. 2018.
5. Scottish Medicines Consortium (SMC). Medicines Advice. Dulaglutide (Trulicity®) SMC1110/15. Available at: https://www.scottishmedicines.org.uk/medicines-advice/dulaglutide-trulicity-fullsubmission-111015/. Accessed 15 May 2023. 2016.
6. Scottish Medicines Consortium (SMC). Medicines Advice. Semaglutide (Ozempic®) SMC2092. Available at: https://www.scottishmedicines.org.uk/medicines-advice/semaglutide-ozempic-fullsubmission-smc2092/. [Accessed 15 May 2023]. 2019.
7. von Arx L-B, Rachman J, Webb J, et al. Therapeutic inertia related to the injectable glucagon-like peptide-1 receptor agonists dulaglutide and semaglutide in patients with type 2 diabetes in UK primary care. Diabetes, Obesity and Metabolism 2023;25:1331-1340.
8. clinicaltrials.gov. A Study of Tirzepatide (LY3298176) Compared With Dulaglutide on Major Cardiovascular Events in Participants With Type 2 Diabetes (SURPASS-CVOT). Available at: https://clinicaltrials.gov/ct2/show/NCT04255433/ [Accessed 25 April 2022].
9. Ryan R; Cochrane Consumers and Communication Review Group. Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage. Available at: https://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/heterogeneity_subgroup_analyses_revising_december_1st_2016.pdf. [Last accessed: 04May23]. 2016.
10. Eli Lilly Data on File. Auto-Correlation Issues.
11. Evans M, Berry S, Nazeri A, et al. The challenges and pitfalls of incorporating evidence from cardiovascular outcomes trials in health economic modelling of type 2 diabetes. Diabetes Obes Metab 2023;25:639-648.
12. Pollock RF, Norrbacka K, Boye KS, et al. The PRIME Type 2 Diabetes Model: a novel, patient-level model for estimating long-term clinical and cost outcomes in patients with type 2 diabetes mellitus. J Med Econ 2022;25:393-402.
13. Si L, Willis MS, Asseburg C, et al. Evaluating the ability of economic models of diabetes to simulate new cardiovascular outcomes trials: a report on the ninth mount hood diabetes challenge. Value in Health 2020;23:1163-1170.
14. Keng MJ, Leal J, Mafham M, et al. Performance of the UK Prospective Diabetes Study Outcomes Model 2 in a Contemporary UK Type 2 Diabetes Trial Cohort. Value Health 2022;25:435-442.
15. Saunders R, Boye K, Bae J, Norrbacka K, Timlin L, Pollock R, Valentine W. How can we best predict cardiovascular disease risk in type 1 diabetes? A comparison of risk equations. Abstract and oral presentation A-14-691-EASD at the European Association for the Study of Diabetes 50th Annual Meeting, Vienna, 2014.
16. Mount Hood Diabetes Challenge Network. Diabetes simulation modeling database. Available at: https://www.mthooddiabeteschallenge.com/registry. [Accessed 12 May 2023].

17. Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). Diabetologia 2004;47:1747-59.
18. Fishman GS. Discrete-event simulation: modeling, programming, and analysis: Springer, 2001.
19. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. Value in Health 2012;15:812-820.
20. Gerstein HC, Colhoun HM, Dagenais GR, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. Lancet 2019;394:121-130.
21. Marso SP, Bain SC, Consoli A, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med 2016;375:1834-1844.
22. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2016;375:311-22.
23. Courtney H, Nayar R, Rajeswaran C, et al. Long-term management of type 2 diabetes with glucagon-like peptide-1 receptor agonists. Diabetes Metab Syndr Obes 2017;10:79-87.
24. Palmer AJ, Roze S, Valentine WJ, et al. The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. Curr Med Res Opin 2004;20 Suppl 1:S5-26.
25. Corro Ramos I, Hoogendoorn M, Rutten-van Molken MPMH. How to address uncertainty in health economic discrete-event simulation models: an illustration for chronic obstructive pulmonary disease. Med Decis Making 2020;40:619-632.
26. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. Value Health 2012;15:835-42.
27. National Institute for Health and Care Excellence. Type 1 and 2 diabetes in adults: diagnosis and management. Economic modelling for periodontal treatment in adults with type 1 and type 2 diabetes. NICE guideline NG17, NG28. Economic model report. Available at: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-on-periodontal-treatment-in-adults-with-type-1-and-type-2-diabetes-pdf-11131191037. [Accessed 12 May 2023]. London: NICE, 2022:33.
28. National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. Economic modelling for continuous glucose monitoring in adults with type 2 diabetes. NICE guideline NG28. Economic model report. Available at: https://www.nice.org.uk/guidance/ng28/evidence/economic-model-report-pdf-11013295213 [Accessed 12 May 2023], 2022.
29. Salpeter SR, Greyber E, Pasternak GA, et al. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus: systematic review and meta-analysis. Arch Intern Med 2003;163:2594-602.
30. National Institute for Health and Care Excellence. Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: Response to request for clarification from the ERG. London: NICE, 2023:151.
31. Boye KS, Matza LS, Stewart KD, et al. Patient preferences and health state utilities associated with dulaglutide and semaglutide injection devices among patients with type 2 diabetes in Italy. J Med Econ 2019;22:806-813.
32. Boye KS, Matza LS, Stewart KD, et al. Health state utilities associated with weight loss in type 2 diabetes and obesity. J Med Econ 2022;25:14-25.
33. Ara R, Brazier J. Estimating health state utility values for comorbid health conditions using SF-6D data. Value Health 2011;14:740-5.

34. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. Diabetes Care 2004;27:2262-2265.
35. National Health Service (NHS). 2019/20 National Cost Collection Data Publication. Available at: https://www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/ [Accessed 11 July 2022].
36. Alva ML, Gray A, Mihaylova B, et al. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). Diabet Med 2015;32:459-66.

Appendix A

Table 30: Pairwise meta-analysis of MD: BMI heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.5523
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	78.1%	0.0837	0.0104
Placebo and exenatide 10 mcg BID	AWARD-1 Derosa 2012b/2013c/d	0%	0	0.4575
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5 AWARD-2	61.3%	0.023	0.0513
Placebo and liraglutide 1.8 mg	LIRA−ADD2SGLT2i PIONEER 4	80.8%	0.267	0.0226
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.4047
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6769
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6769

Abbreviations: BID: twice daily; BMI: body mass index; QW: once weekly.

Table 31: Pairwise meta-analysis of MD: body weight heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-2 LEAD-4	19.6%	0.0349	0.2882
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	0%	0	0.6863
Placebo and exenatide 10 mcg BID	Apovian 2010 AWARD-1 DeFronzo 2005 Derosa 2012b/2013c/d Kendall 2005 Liutkus 2010	67.4%	0.5409	0.0091
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	27.1%	0.0625	0.2537
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	73.4%	0.5134	0.0235
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5 SUSTAIN-7	30.6%	0.0894	0.2174
Exenatide 10 mcg BID and glargine	Bunck 2009/2010/2011 Gurkan 2014 GWAA	0%	0	0.9355
Placebo and exenatide 5 mcg BID	DeFronzo 2005 Kendall 2005	8.3%	0.015	0.2963
Exenatide 10 mcg BID and exenatide 5 mcg BID	DeFronzo 2005	66%	0.475	0.0865

	Kendall 2005			
Exenatide 10 mcg BID and glimepiride	Derosa 2011b EUREXA	27.2%	0.4422	0.2412
Liraglutide 1.8 mg and glargine	EAGLE LEAD-5	89.1%	1.1404	0.0025
Placebo and lixisenatide 20 mcg	GetGoal-F1 GetGoal-M GetGoal-P GetGoal-S	0%	0	0.7164
Placebo and liraglutide 1.2 mg	LEAD-2 LEAD-4	0%	0	0.3692
Placebo and liraglutide 1.8 mg	LEAD-2 LEAD-4 LEAD-5 LIRA-ADD2SGLT2i PIONEER 4	57.4%	0.3182	0.0519
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.4556
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.5475
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.8822
Semaglutide 1.0 mg QW and semaglutide 0.5 mg QW	SUSTAIN 2 SUSTAIN 4 SUSTAIN 7	0%	0	0.9952

Abbreviations: BID: twice daily; QW: once weekly.

Table 32: Pairwise meta-analysis of MD: eGFR heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.4984
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6269
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.8458

Abbreviations: BID: twice daily; eGFR: estimated glomerular filtration rate; QW: once weekly.

Table 33: Pairwise meta-analysis of MD: HbA1c heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-2 LEAD-4	37.8%	0.0104	0.2003
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	0%	0	1
Placebo and exenatide 10 mcg BID	Apovian 2010 AWARD-1 DeFronzo 2005 Derosa 2012b/2013c/d Kendall 2005 Liutkus 2010	78.8%	0.0514	0.0003
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	74.3%	0.0247	0.0204
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	83.1%	0.0392	0.0027
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5 SUSTAIN-7	0%	0	0.8683
Exenatide 10 mcg BID and glargine	Bunck 2009/2010/2011 Gurkan 2014 GWAA	0%	0	0.7464
Placebo and exenatide 5 mcg BID	DeFronzo 2005 Kendall 2005	62%	0.0243	0.1048
Exenatide 10 mcg BID and exenatide 5 mcg BID	DeFronzo 2005	3.4%	0.0006	0.309

	Kendall 2005			
Exenatide 10 mcg BID and glimepiride	Derosa 2011b EUREXA	0%	0	0.4381
Liraglutide 1.8 mg and glargine	EAGLE LEAD-5	86.1%	0.0654	0.0074
Placebo and lixisenatide 20 mcg	GetGoal-F1 GetGoal-M GetGoal-P GetGoal-S	36.8%	0.0061	0.1913
Placebo and liraglutide 1.2 mg	LEAD-2 LEAD-4	0%	0	0.7642
Placebo and liraglutide 1.8 mg	LEAD-2 LEAD-4 LEAD-5 LIRA-ADD2SGLT2i PIONEER 4	45.4%	0.014	0.1199
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6833
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.3587
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	42%	0.0035	0.1893
Semaglutide 1.0 mg QW and semaglutide 0.5 mg QW	SUSTAIN 2 SUSTAIN 4 SUSTAIN 7	60.8%	0.0071	0.0779

Abbreviations: BID: twice daily; HbA1c: glycated haemoglobin; QW: once weekly.

Table 34: Pairwise meta-analysis of MD: HDL heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-4	0%	0	0.7595
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	0%	0	0.954
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	73.5%	0.0012	0.023
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	43.7%	0.0004	0.1426
Placebo and exenatide 10 mcg BID	AWARD-1 Derosa 2012b/2013c/d Liutkus 2010	0%	0	0.6197
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5	30.2%	0.0001	0.2311
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6171
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.3173
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6171

Abbreviations: BID: twice daily; HDL: high density lipoprotein; QW: once weekly.

Table 35: Pairwise meta-analysis of MD: LDL heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-4	0%	0	0.4992
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	0%	0	0.927
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.9947
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.7926
Placebo and exenatide 10 mcg BID	AWARD-1 Derosa 2012b/2013c/d Liutkus 2010	0%	0.0004	0.4043
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5	0%	0	0.4811
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	72.3%	0.0061	0.0576
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	87.2%	0.0157	0.0052
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.377

Abbreviations: BID: twice daily; LDL: low density lipoprotein; QW: once weekly.

Table 36: Pairwise meta-analysis of RR: nausea heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-2 LEAD-4	0%	0	0.8746
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	68.1%	0.1958	0.0765
Placebo and exenatide 10 mcg BID	Apovian 2010 AWARD-1 DeFronzo 2005 Kendall 2005 Liutkus 2010	21.5%	0	0.2776
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.4721
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.9432
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5	31.9%	0.0202	0.2208
Placebo and exenatide 5 mcg BID	DeFronzo 2005 Kendall 2005	0%	0	0.481
Exenatide 5 mcg BID and exenatide 10 mcg BID	DeFronzo 2005 Kendall 2005	0%	0	0.9928
Liraglutide 1.8 mg and glargine	EAGLE LEAD-5	0%	0	0.9407
Placebo and lixisenatide 20 mcg	GetGoal-F1	39.5%	0.0416	0.175

	GetGoal-M GetGoal-P GetGoal-S			
Placebo and liraglutide 1.8 mg	LEAD-5 LIRA-ADD2SGLT2i	0%	0	0.8825
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	84.8%	0.1425	0.0103
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	80.6%	0.1013	0.0233
Tirzepatide 10 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6776

Abbreviations: BID: twice daily; QW: once weekly.

Table 37: Pairwise meta-analysis of MD: SBP heterogeneity results

Comparison	Studies	Heterogeneity Results		
		I^2	t^2	p
Liraglutide 1.2 mg and liraglutide 1.8 mg	1860-LIRA-DPP-4 LEAD-2 LEAD-4	0%	0	0.8134
Liraglutide 1.8 mg and sitagliptin 100 mg	1860-LIRA-DPP-4 LIRA-SWITCH	28.2%	0.6155	0.238
Placebo and exenatide 10 mcg BID	Apovian 2010 AWARD-1 DeFronzo 2005 Derosa 2012b/2013c/d Liutkus 2010	69.3%	6.4329	0.0206
Placebo and dulaglutide 0.75 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.5978
Placebo and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-5	0%	0	0.9697
Dulaglutide 0.75 mg and dulaglutide 1.5 mg	AWARD-1 AWARD-10 AWARD-2 AWARD-5	0%	0	0.7524
Liraglutide 1.8 mg and glargine	EAGLE LEAD-5	12.7%	0.1507	0.2845
Placebo and liraglutide 1.2 mg	LEAD-2 LEAD-4	76.8%	8.13	0.0377
Placebo and liraglutide 1.8 mg	LEAD-2 LEAD-4 LEAD-5 LIRA-ADD2SGLT2i	43.4%	1.6308	0.1321

	PIONEER 4			
Tirzepatide 5 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.4746
Tirzepatide 5 mg QW and tirzepatide 15 mg QW	SURPASS-2 SURPASS-3	0%	0	0.8108
Tirzepatide 15 mg QW and tirzepatide 10 mg QW	SURPASS-2 SURPASS-3	0%	0	0.6336

Abbreviations: BID: twice daily; SBP: systolic blood pressure; QW: once weekly.

1 SENSITIVTY ANALYSIS TABLES REQUESTED BY THE EAG (ON 24 MAY, 2023

One-way and multi-way sensitivity analysis results from simulations based on the EAG preferred base case and corresponding to Tables 103, 104, 105 and 106 in the CS are provided on the following pages. In cases where the previous sensitivity analysis in Table 106 could not be performed (due to changes in base case assumptions), this is noted in the table and the converse analysis was run with a view to identifying the contribution this change would make to base case outcomes.

This sensitivity analysis shows what the important contributors of cost-effectiveness are and remain aligned with the original base case, specifically:

- The HbA1c benefit is important (including how improved HbA1c delays treatment intensification versus semaglutide)
- The BMI benefit is important (including it influencing quality of life)
- It takes time for the risk factor benefits to reduce complication risk (and therefore to TZP to become cost-effective)

Table103 (EAG preferred BC version): Summary of one-way sensitivity analysis results with tirzepatide 5 mg versus semaglutide 0.5 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 5 mg	Semaglutide 0.5 mg	Incremental value	Tirzepatide 5 mg	Semaglutide 0.5 mg	Incremental value	
EAG preferred base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	*8,401*
Sensitivity analysis							
5-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,169
10-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	11,665
15-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,700
20-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,912
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	6,655
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,624
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	7,153

Abbreviations: QALY: quality-adjusted life years.

Table 104 (EAG preferred BC version): Summary of one-way sensitivity analysis results with tirzepatide 10 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
EAG preferred base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18,115
Sensitivity analysis							
5-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	33,518
10-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	24,853
15-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	20,693
20-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	19,331
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14,602
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	20,391
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14,236

Abbreviations: QALY: quality-adjusted life years.

Table 105 (EAG preferred BC version): Summary of one-way sensitivity analysis results with tirzepatide 15 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 15 mg	Semaglutide 1.0 mg	Incremental value	
EAG preferred base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	*15,209*
Sensitivity analysis							
5-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	28,093
10-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	20,600
15-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,250
20-year time horizon	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	15,695
0% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,744
6% discount rate on costs and clinical benefits	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	16,312
SURPASS-2 cohort	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	11,031

Abbreviations: QALY: quality-adjusted life years.

Table 106 (EAG preferred BC version): Summary of additional one-way and multi-way sensitivity analysis results for tirzepatide 10 mg versus semaglutide 1.0 mg

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
EAG preferred base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	*18,115*
Clinical drivers							
No HbA1c difference	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	39,085
No SBP difference	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	19,474
No serum lipids difference	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18,433
No BMI difference	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	30,878
Only HbA1c difference between treatments	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	35,059
Only BMI difference between treatments	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	53,280
Only HbA1c and BMI differences between treatments	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	22,924
Duration of therapy							
Intensification to insulin after 3 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,512
Intensification to insulin after 5 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	23,939
Second intensification to basal-bolus therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	15,845
Intensification at HbA1c 8.5% threshold	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	27,251
Intensification at HbA1c 9.5% threshold	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	33,008

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Quality of life utilities							
No weight change utility	No utility associated with weight changes was included in the EAG preferred base case simulations						
Weight change utility included	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	16,337
No weight/BMI utilities	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	27,997
No device utility	No administration device-related utility was included in the EAG preferred base case simulations						
Device utility included	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	16,893
No nausea utilities	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,577
No hypoglycaemia utilities	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	21,224
QALY age-adjustment based on Ara and Brazier	Age-adjustment for utilities was included in the EAG preferred base case simulations						
No age-adjustment on utilities	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	16,938
Multiplicative approach to combining utilities (with age-adjustment)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	24,911
Other base case assumptions							
Cohort ethnic groups changed from Black to Afro-Caribbean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,206
Sulfonylurea added to background therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18,416
Change in BMI values taken directly from NMA	Change in BMI values were taken directly from the NMA for the EAG preferred base case simulations						

	Direct costs (£)			Quality-adjusted life expectancy (QALYs)			ICER (£ per QALY gained)
	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	Tirzepatide 10 mg	Semaglutide 1.0 mg	Incremental value	
Change in BMI values estimated from body weight changes	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18,846
UKPDS OM2 risk factor progression for all risk factors	UKPDS OM2 risk factor progression was included in the EAG preferred base case simulations						
Complication costs taken from alternative sources (lit. review)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,685
UKPDS OM2 renal failure estimation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,939
UKPDS OM2 eGFR progression and renal failure estimation	As UKPDS OM2 eGFR progression is used in the EAG preferred base case, this scenario is identical to the *UKPDS OM2 renal failure estimation* simulation presented in the row above						
UKPDS OM2 mortality risk estimation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	18,157
Cause-subtracted life tables for mortality risk estimation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14,278

Abbreviations: BMI: body mass index; eGFR: estimated glomerular filtration rate; NMA: network meta-analysis; QALY: quality-adjusted life years; UKPDS OM2: United Kingdom Prospective Diabetes Study Outcomes Model 2.

in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]

EAG critique of Response to EAG Report May 2023

Produced by Kleijnen Systematic Reviews (KSR) Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University Medical Center (UMC+)

Authors Nigel Armstrong, Health Economics Manager, KSR Ltd, United Kingdom (UK)
Bram Ramaekers, Health Economist, Maastricht UMC+
Evangelos Danopoulos, Systematic Reviewer, KSR Ltd, UK
Sabine Grimm, Health Economist, Maastricht UMC+
Andrea Fernandez Coves, Health Economist, Maastricht UMC+
Mirre Scholte, Health Economist, Maastricht UMC+
Xiaoyu Tian, Health Economist, KSR Ltd, UK
Jiongyu Chen, Health Economist, KSR Ltd, UK
Lisa Stirk, Information Specialist, KSR Ltd, UK
Rachel Croft, Information Specialist, KSR Ltd, UK
Manuela Joore, Health Economist, Maastricht UMC+, the Netherlands
Robert Wolff, Managing Director, KSR Ltd, UK

Correspondence to Nigel Armstrong, Kleijnen Systematic Reviews Ltd
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, YO19 6FD
United Kingdom

Date completed 17/03/2023

Source of funding: This report was commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme as project number NIHR135771.

Declared competing interests of the authors: None.

Acknowledgements

We gratefully acknowledge the expert clinical advice input from Dr Mohammed Kamrudeen, Honorary Senior Lecturer and Consultant Endocrinologist, Hull University Teaching Hospitals NHS Trust.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Confidential comparator prices are highlighted in green throughout the report.

Any de-personalised data are highlighted in pink throughout the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Armstrong N, Ramaekers B, Danopoulos E, Grimm S, Fernandez Coves A, Scholte M, Tian X, Chen J, Stirk L, Croft R, Joore MA, Wolff R. Tirzepatide for the treatment of patients with type 2 diabetes [ID3938]: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2023.

Contributions of authors

Nigel Armstrong acted as project lead on this assessment, critiqued the clinical effectiveness evidence and economic evaluation and contributed to the writing of the report. Bram Ramaekers acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Sabine Grimm, Andrea Fernandez Coves, Mirre Scholte and Manuela Joore acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Evangelos Danopoulos, Xiaoyu Tian and Jiongyu Chen acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Lisa Stirk and Rachel Croft critiqued the search methods in the submission and contributed to the writing of the report. Thomas Debray acted as JAVA expert and biostatistician, critiqued the statistical evidence, and facilitated the critique of the economic model. Robert Wolff critiqued the

company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report, and supervised the project.

Abbreviations

ADA	American Diabetes Association
AE	adverse event
AIC	academic in confidence
APPADL	Ability to Perform Physical Activities of Daily Living
AWMSG	All Wales Medicines Strategy Group
BG	blood glucose
BID	twice a day
BMI	body mass index
BP	blood pressure
BPM	beats per minute
BRAVO	Building, Relating, Assessing, and Validating Outcomes
CADTH	Canadian Agency for Drugs and Technologies in Health
CEA	cost effectiveness analysis
CFB	change from baseline
CI	confidence interval
CIC	commercial in confidence
CKD	chronic kidney disease
COPD	chronic obstructive pulmonary disease
CRD	Centre for Reviews and Dissemination
CrI	credible interval
CS	company submission
CSR	Clinical Study Report
CV	cardiovascular
CVD	cardiovascular disease
CVOT	Cardiovascular Outcomes Trial
DBP	diastolic blood pressure
DIC	deviance information criterion
DPP-4	dipeptidyl-peptidase 4
DPP-4i	dipeptidyl-peptidase 4 inhibitors
DSA	deterministic sensitivity analyses
EASD	European Association for the Study of Diabetes
ECG	electrocardiogram
eGFR	estimated Glomerular Filtration Rate
EPI	epidemiology
EQ-5D	European Quality of Life-5 Dimensions
EAG	Evidence Assessment Group
ESRD	end stage renal disease
EU	Europe
EUR	Erasmus University Rotterdam
FAS	full analysis set
FE	fixing errors
FPG	fasting plasma glucose
FSG	fasting serum glucose
FV	fixing violations
GI	gastrointestinal
GLP-1	glucagon-like peptide-1
Hb	haemoglobin
HbA1c	glycated haemoglobin
HDL	high density lipoprotein
HDL-C	high density lipoprotein-cholesterol
HE	Health Economic
HQO	Health Quality Ontario
HR	hazard ratio

HR	heart rate
HRQoL	health-related quality of life
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
ICER	Institute for Clinical and Economic Review
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IWQOL-LITE-CT	Impact of Weight on Quality of Life-Lite Clinical Trials Version
JAGS	Just Another Gibbs Sampler
KDIGO	Kidney Disease Improving Global Outcomes
KSR	Kleijnen Systematic Reviews Ltd
LDL	low density lipoprotein
LDL-C	low density lipoprotein-cholesterol
LIRA	liraglutide
LYs	life years
MACE	major adverse cardiovascular events
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical Subject Headings
MHRA	Medicines and Healthcare Products Regulatory Agency
mITT	modified Intent-to-Treat
MJ	matters of judgement
NASH	non-alcoholic steatohepatitis
N/A	not applicable
NCPE	National Centre for Pharmacoeconomics
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	network meta-analysis
NPH	neutral protamine Hagedorn
NR	not reported
OAD	oral antidiabetic drug
ORs	odds ratios
OUS	outside the USA
PAS	Patient Access Scheme
PBAC	Pharmaceutical Benefits Advisory Committee
PBS	Pharmaceutical Benefits System
PICO	population, intervention, comparator and outcome
PPG	postprandial glucose
PRESS	Peer Review of Electronic Search Strategies
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life year
QD	once a day
QoL	quality of life
QW	once weekly
RA	receptor agonist
RCT	randomised controlled trial
RR	relative risk; risk ratio
SAE	serious adverse event
SBP	systolic blood pressure
SBU	Swedish Agency for Technology Assessment and Assessment of Social Services
ScHARR	School of Health and Related Research

SD	standard deviation
SEMA	semaglutide
SGLT-2i	sodium-glucose co-transporter-2 inhibitor
SIGN	Scottish Intercollegiate Guidelines Network
SLR	systematic literature review
SMD	standardised mean difference
SMQ	standardised MedDRA query
SmPC	summary of product characteristics
SoC	standard of care
SOC	system organ class
SPSL	severe pressure sensation loss
STA	Single Technology Appraisal
SU	sulfonylurea
T1D	type 1 diabetes
T2D	type 2 diabetes
TEAE	treatment emergent adverse events
TZD	thiazolidinediones
TZP	tirzepatide
UK	United Kingdom
UKPDS	United Kingdom Prospective Diabetes Study
UMC+	University Medical Center+
US	United States
USA	United States of America
WBC	white blood cell count
WHO	World Health Organization

Table of Contents

A1.1 The Decision Problem

Key Issue 1: Mismatch between scope and decision problem in terms of line of therapy and comparators might lead to a lack of evidence for the scope of interest in decision making

The company have confirmed that the population in the decision problem should be: *"…the same as the NG28 GLP-1 RA-eligible population* [triple therapy not effective, not tolerated or contraindicated]*, representing a narrower population than that specified in the marketing authorisation wording for tirzepatide and the NICE final scope for this evaluation."* (p. 3) This might not be a Key Issue unless the committee would like to make a decision on the wider population of the NICE scope.

Key Issue 2: Mismatch between decision problem and evidence in terms of line of therapy/OAD therapy experience might lead to an overestimate of the effectiveness and cost effectiveness of tirzepatide

The company again cite the SURPASS-4 subgroup analysis of baseline OADs and the NMA meta-regression to support the argument that treatment effect is independent of baseline therapy. However, as was noted in the EAG report, there was a significant effect on HbA1c in the subgroup analysis and there was also limited ability to test the hypothesis of baseline OAD independence given that so few patients received three OADs (about [redacted]) in SURPASS-4 and none in the NMA. The EAG also notes that, although tirzepatide would be expected to replace one of the three OADs, thus leaving only two concomitant background OADs, there is still a mismatch in line of therapy according to the company's decision problem given that experience with triple therapy would be expected, except for those patients where it is contraindicated. Indeed, the company confirms that it is *"… expected that clinicians in UK clinical practice would use tirzepatide in patients with type 2 diabetes (T2D) that is inadequately controlled with metformin and two OADs, as a more efficacious option whenever GLP-1 RAs would otherwise be considered. More specifically, if triple therapy with metformin and two OADs is not effective, not tolerated or contraindicated, clinicians would consider switching one of these drugs to a GLP-1 RA or tirzepatide*" (p. 3). They then also refer to the NG28 clinical guideline diagram which clearly illustrates that GLP-1 RAs are to be used in a later line of therapy than those in the SURPASS trials and the NMA. It is also the case that the trials included in the NMA mostly included patients with even less experience and fewer background OADs: out of 53 studies only one was only of patients with more than one OAD, 27 were of patients with a mixture of one or two OADs and 25 were of patients with metformin as the only OAD (see Table 3.23, EAG report). This therefore remains a Key Issue.

A.1.2 The Clinical Effectiveness Evidence

Key Issue 3: Mismatch between the administration of tirzepatide in clinical practice by titration and the tirzepatide trial evidence, the NMA and the CEA, according to maintenance dose strata, is likely to lead to biased estimates of effectiveness and cost effectiveness in an unknown direction

The company have now accepted the mismatch between tirzepatide dosing in the trials and as it would be in clinical practice: *"The company acknowledges that, given the SURPASS trial programme was based on pre-specified maintenance doses but in clinical practice, tirzepatide will be administered via titration between maintenance doses, there is a mismatch in terms of administration here."* (p.8) They argue that comparisons within the maintenance dose strata are a way to mitigate this problem, which is supported by few patients de-escalating. However, the mismatch applies even if there is no de-escalation given that the comparison between treatments that are titrated depends not only on the relative effectiveness between two maintenance doses, but the proportion of patients who escalate, which might also vary. The EAG would again suggest that, in the absence of titrated treatment evidence, the company's comparison in the tirzepatide middle dose (10mg) stratum might be the closest approximation. Nevertheless, this still remains a Key Issue.

Key Issue 4: Lack of comparative evidence on micro and macrovascular complications

The company continue to maintain that evidence from the SURPASS-CVOT trial is still not available. This therefore remains a Key Issue.

Key Issue 5: NMA of high risk of bias due to lack of feasibility assessment/assessment of trial comparability and insufficient sensitivity analyses

The validity of an NMA is based on the assumption that all the studies included in the network are similar in all the factors that may affect the relative effects (i.e. disease and patient characteristics that are potentially effect modifiers). The decision to execute an NMA must predominantly be based on a clinical judgment of the differences between studies. The EAG requested at clarification stage, that the company would address all potential treatment effect modifiers, including at least: concomitant therapy, HbA1c, comorbidities e.g., CVD/CV high risk, obesity, non-alcoholic fatty liver disease, sex, age, weight, BMI, duration of diabetes, race and ethnicity in a feasibility assessment.

The company continue to maintain that their feasibility assessment and sensitivity analyses were adequate. They state that "*Studies were identified in a systematic way to align with the PICO statement set out initially.*" (p. 9), *initially* i.e., for the SLR that was not executed for the decision problem in this CS. They provide further details of the studies that were excluded from the network because of comorbidities, although this was not the main concern of the EAG. They also provided a list of studies excluded because of extreme values of baseline characteristics, such as BMI and diabetes duration. However, there continues to be no analysis of the degree of variation between trials that were included, which appeared to be large: for example, the range of mean baseline HbA1c values varied from 7.4% to 10.3% and, out of the 136 arms presented, 11 arms reported values below 8% and 20% above 8.5% (see EAG report, Section 3.3). Also, baseline diabetes duration varied from 0.6 - 10.1 years and fourteen arms reported duration less than 6 years while 55 reported more than 8 years. The company also argues that the meta-regression showed that the treatment effect was not influenced by baseline characteristics. However, the meta-regression results were limited to only one factor i.e. number of OADs.

The company also mention the sensitivity analysis where trials with unknown proportion of metformin and patients on three OADs. However, this does not address the heterogeneity of background OADs in the main analysis mentioned in the EAG report.

The company also state the EAG have misunderstood what is meant by substantial heterogeneity, believing it to relate to the whole network as opposed to between direct (within trial) comparisons, but regardless of where in the network it lay, the point remains that the company regarded it as substantial. They have now decided to present the statistical heterogeneity results for pairwise meta-analysis (of trials making the same direct comparison) regarding seven characteristics (BMI, body weight, eGFR, HbA1c, HDL, nausea and SBP) in Tables 30-37 in Appendix A. The company states that "*As shown in Appendix A, out of all the treatment comparisons that were assessed for heterogeneity only a small portion had substantial heterogeneity.*" (p. 16). Nevertheless, out of the 91 I^2 results presented across the analyses, 21 were >60% (substantial/considerable heterogeneity), as shown for the first time in Appendix A. The company concluded that "*Due to this small portion of studies/treatment comparisons yielding increased heterogeneity, sensitivity analyses for the removal of these studies were not performed*." (p. 9). However, they presented three sensitivity analyses only for HbA1c, body weight and BMI. In these sensitivity analyses, it appears that, instead of removing only sufficient trials to reduce the I^2 to below the threshold, as identified in the respective heterogeneity assessment, they seem to have removed the entire comparison across the three sensitivity analyses, which might perhaps explain the convergence issue with *"sigma"* (presumably the between studies variance). It is not clear why this choice was made. However, the EAG note that this sensitivity analysis produces results that could be regarded as quite similar to the base case with no change in direction of effect, a small change in magnitude and some change in confidence interval overlap of the point of no difference, which is only for BMI, albeit with no estimates for some comparisons.

The company have also now provided the results of tests of model fit, residual deviance and deviance information criterion (DIC), which shows that the meta-regression model including baseline risk (placebo response) has a similar fit to the main analysis with the former providing marginally the better fit when including HbA1c or BMI, but not body weight.

Some further clarification was also provided by the company on matters that were related to the NMA:

- Heart rate and total cholesterol were not included as outcomes. The company acknowledged that these endpoints were not related to this CS and were mentioned in error.
- Three studies, Ji 2013, Kadowaki 2011 and Li 2014 were not included in any analysis because they included three OADs and were Asian. The ERG stated that it is not clear which these studies are as the full references are not provided. The company has yet to provide this information. The ERG also questioned whether the company's Asian studies, SURPASS-J-Mono and SURPASS-J-Combo, were included, but they did not answer this question either.
- The source of heterogeneity in insulin glargine trials was identified as total insulin dose.
- The company stated that they identified the trials that were included in the analysis (main or sensitivity) is in the Excel file provided by them in response to clarification question A24. The EAG had inquired which studies were part of **each** of the main NMAs and which the sensitivity analyses. Every outcome in the main analyses and every sensitivity analysis included different studies. It is still not clear which studies were included in each analysis.
- Studies of GLP-1 RAs that did not connect to any other treatment in the network were excluded. The EAG noted that a list of these treatments was not provided in the CS and has still not been offered.
- The time-course model with the best fit was the unequally spaced piecewise linear. The company stated that this model was chosen for the random models for change from baseline in HbA1c and for change from baseline in weight. The choice was made by comparing model fit statistics and examination of graphs. No data were presented by the company to this effect.

The resulting random effects models were compared to the fixed effects model via posterior mean residual deviance to the total number of data points, and the DICs between inconsistency model and the consistency model. The random effects model was chosen over the fixed effects but again no data were presented by the company.

- The studies that reported use of estimand or where estimand could be inferred, that being efficacy estimand (until rescue therapy) were listed. The company acknowledges that among the 45 studies included in the NMA, the estimand was clearly defined only in 10 studies, in 11 further studies the definition of the estimand was inferred while in the rest 24 studies the estimand was not defined and only the available data was included. Thus, there might be an impact from using two fundamentally different effect sizes in the same NMA (two endpoints), which has not been explored in the analysis.
- The meta-regression model was only run adjusting for OADs. No further data or explanation was offered.

In conclusion, the company have added one sensitivity analysis, where all trials making the same direct comparison with high heterogeneity between them are excluded, which does seem to show little difference to the main analysis. However, there still seems to be a large amount of clinical and statistical heterogeneity in the network, particularly in terms of baseline characteristics and type of OAD, as opposed to number of OADs, which means that the external and internal validity of the NMA is in question and high risk of bias in the NMA remains a Key Issue.

A.1.3 Cost-effectiveness issues

In general, the EAG wants to note that the onus to provide a comprehensive and transparent submission explaining the company's economic model and justification for the approach adopted is on the company. Merely providing access to the model (JSON) files does not release the company from the obligation to provide detailed explanation and justification related to the model development, technical implementation, analyses, and validation. Notably, the EAG received the model files early 2023 and was able to locally reproduce the company base-case (which is typically only the starting point of the EAG model assessment) after submitting the EAG report (mid-April 2023). From commencing work on this STA (January 2023), the EAG had access to the health economic model through a web interface. However, despite the fact that this web interface might be convenient for model users to run analyses, validating and scrutinizing a health economic model through such an interface is inherently challenging: e.g., not all input parameters can be adjusted, and it is difficult to examine the technical model implementation and associated assumptions.

Key issue 6: Model approach adopted by the company

All models have limitations (as highlighted by the company). However, as stated in the original EAG report, it is unclear to the EAG that the developed de novo model, specifically the current implementation as in the CS, is superior compared with existing diabetes models, e.g., has a better performance to predict complications, including CV events. For example, the company notes the poor performance of the UKPDS OM2 model to predict long-term cardiovascular outcomes for the ASCEND trial population, but does not provide evidence of a better performance of the PRIME T2D model for this population.

Key issue 7: Selection and use of risk models to estimate complications

The company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.

Key issue 8: Extrapolation of treatment effectiveness

The company did implement risk factor progression for the EAG preferred scenario and did provide additional justification for assuming no treatment waning by stating that: *"Long-term data from the CVOTs for dulaglutide, semaglutide and liraglutide show that body weight and SBP remain stable whilst on GLP-1 receptor agonist therapy".*

Key issue 9: Treatment discontinuation/intensification

The company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.

Key issue 10: Adverse events: not all incorporated for all treatments

The company provided the following response: "*rates of hypoglycaemia were not reported in the NMA due to many studies reporting zero events; therefore rates of hypoglycaemia were set to zero for tirzepatide and all comparators in the base case analysis. This assumption is likely to be a reasonable approximation for the interventions included in the present analysis based on the very low hypoglycaemia rates observed in the SURPASS trial programme and clinical studies of other T2D medications such as GLP-1 Ras."* (p.45) The EAG agrees that the impact of hypoglycaemia on the cost-effectiveness is likely to be limited due to the very low number of events. The EAG still prefers both nausea and vomiting to be included, instead of only nausea with a disutility corresponding to a vomiting health state.

Key issue 11: Age-adjustment for utility values: none for older age

The suggestion to include age-adjusted utility values in the EAG preferred base case was adopted by the company. The baseline utility value for T2D remains relatively high (0.815 as compared to 0.804 for the general population at the same age). A recent meta-analysis by Redenz et al. 2022 showed an average EQ-5D-3L utility of 0.772 (based on 19 studies) and an EQ-5D-3L utility reduction of 0.037 compared to the general population without T2D (2 studies).{Redenz, 2023 #497}

Key issue 12: Discrepancies related to utility and cost values

The discrepancies related to utility and cost values have been resolved.

Key issue 13: Potentially inappropriate PSA

According to the EAG's understanding of the probabilistic sensitivity analysis (PSA) in the PRIME T2D Model, first and second order uncertainty are mixed. Subsequently, non-parametric bootstrapping was used.

As stated in the EAG report, the company's approach deviates from standard PSA methods. To illustrate this point, NICE TSD 15 on patient level simulation does not mention bootstrapping, rather "*When evaluating the decision uncertainty in a patient-level simulation using PSA it is usually necessary to run two nested simulation loops.*" Regarding implications on results that the company's PSA approach might have: in short, the estimated mean results might be correct, but the distribution around the results is distorted and it probably underestimates uncertainty. Halpern et al. described in 2000:{Halpern, 2000 #498}

> *"Hunink et al. proposed a shortcut employing a subsampling or "bootstrap" technique. They ran the program once, varying both first- and second-order parameters. They recorded the results for each of the individual patients in this run, a total of 30,000 simulated patients. They then randomly selected 3,000 groups of 1,000 patients from the 30,000 and used the mean cost and effectiveness of each group as a point in the plot. It appears, though, that this subsampling shortcut may distort the underlying distribution and underestimate the variability in mean cost and effectiveness attributable to second order uncertainty."*
>
> *....*
>
> *"The problem arises because random selection at each stage of the two-stage procedure for each patient is equivalent to a single random selection of the patient outcomes from a third distribution, namely the marginal distribution derived from the joint distribution of both first- and second-order variables."*
>
> *....*
>
> *"First, the two distributions need not resemble each other in shape or mathematical expression. Second, whether the two resemble each other or not, the traditional simulation of a finite group of patients using both first- and second order uncertainty results in greater variability than a simulation that varies the parameter from patient to patient. (This point is derived in greater detail in the appendix.) Finally, the mean values of the two methods of simulation are the same. Therefore, the error identified here results in underestimating the value of information about the mean, but does not lead to incorrect conclusions for immediate decisions regarding individuals or groups."*
>
> *....*

"probabilistic sensitivity analysis, when implemented concurrently with first-order Monte Carlo simulation, as in Hunink et al., can lead to misleading results. The practice of independently selecting both the parameters and the outcomes for each patient at best underestimates the uncertainty due to ignorance regarding the parameters, and at worst leads to inaccurately shaped distributions of incremental cost and effectiveness."

....

"We have shown that the shortcut of simultaneously drawing from the parameter distribution and simulating individual outcomes leads to underestimation of the uncertainty attributable to the parameters. The correct approach is a two-stage simulation in which 1) parameters are drawn from their distributions, and 2) a Monte Carlo simulation is run, conditional upon these parameter values, and of sufficient size to make negligible the errors in the estimates of mean effectiveness and mean cost (conditional upon the parameter values). Thus, the second stage substitutes for a deterministic cohort analysis of the expected effectiveness and expected cost, and all the inferences about the effects of parameter uncertainty are captured by the first stage."

Later studies also confirmed that combining first and second order uncertainty (i.e. 'single loop' PSA in individual patient models) can be used to obtain the expected value (e.g. estimate of the ICER), but nested simulations are required if we are interested in the distribution of the expected outcome (reflecting parameter uncertainty).{Groot Koerkamp, 2011 #499;Vemer, 2014 #500}

In addition, the paper by Corro-Ramos describes some challenges and proposed solutions when incorporating uncertainty in patient level simulation (see EAG clarification question B2c for more details). When implementing two nested simulation loops in the PSA (as stated in NICE TSD 15 on patient level simulation), it would be helpful to clarify how these challenges are addressed by the company.

Key issue 14: No full deterministic one-way sensitivity analyses provided

The company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.

Key issue 15: Technical verification insufficient/model results not reproducible

Post submission of the EAG report, the EAG was able to run the model locally with the assistance of the company and reproduce the company's base-case results. This resolves part of the critique in the EAG report. However, with regards to the other critique points (lack of clarity how BMI related utility is implemented in the model, no complete overview of all model inputs, face, internal and external validity checks likely incomplete), the company did not provide new compelling evidence or arguments. Thus, most comments in the EAG report (with exception of the problems relating to reproducing the company's results locally) remain applicable.

Other Points Made by the EAG Requiring Clarification

The company provided some additional comments, listed by section of the EAG report, some of which are also connected to a Key issue:

- Section 4.2.2 Model structure: the table presented by the company that compares the PRIME T2D model with UKPDS OM2 and CORE Diabetes model, and updated schematic diagram of the PRIME T2D Model provided by the company provides clarification regarding which complications are incorporated in the PRIME T2D model.

- Section 4.2.3 Population: the company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.
- Section 4.2.4 Interventions and Comparators: the company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.
- Section 4.2.6 Treatment effectiveness and extrapolation: see Key issue 9.
- Section 4.2.7 Adverse events: see Key issue 10.
- Section 4.2.8 Health-related quality of life:
    - Regarding section 4.2.8 c) and d), the company has removed the utility values associated with weight loss in year 1 and administration of tirzepatide and dulaglutide in the EAG preferred base case simulations.
    - In response to section 4.2.8 e): with regards to the method used to combine disutility values, the company refers to other NICE guidelines and reports on diabetes that also use an additive approach to combining utilities. However, as also mentioned in the NICE DSU technical support document 12, the multiplicative method of combining utilities is preferred in the absence of a conclusive evidence base. The critique from the EAG therefore remains the same as in the original EAG report, except for the study that was referred to, which should have been Ara and Brazier 2010.{Ara, 2010 #495}
    - In response to section 4.2.8 f), the company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.
- Section 4.2.9 Treatment costs:
    - In response to section 4.2.9.2. a) and b): the company inflated all costs to 2022 values and corrected any inconsistencies between the values used in the model and the original sources. These corrections were included in the EAG preferred base case simulations.
    - In response to section 4.2.9.2. c) and d): two scenario analysis exploring 1) the costs of nausea and 2) annual T2D health state costs were included in the EAG preferred base case simulations as suggested.
    - In response to section 4.2.9.2. f): the company briefly responded to this point, explaining that costs were selected to best align with the NICE health economic evaluation for NG28.
- Section 5.1: the company did not provide new compelling evidence or arguments. Thus, the comments in the EAG report remain applicable.
- Section 5.2: see Key issue 14.
- Section 5.3: see Key issue 15.

A.1.4 EAG Preferred Base Case Simulations

According to the company's addendum Table 5, the company implemented EAG adjustments, 1, 5, 6, 7, 9 and 10. However, the other EAG adjustments were unfortunately not implemented nor were the analyses performed step-by-step showing the impact of the individual adjustments. Nevertheless, assuming that the adjustments suggested by the EAG were implemented correctly, this would be a step towards the EAG base-case.

Similarly, according to the company's addendum Table 6, the company performed EAG scenario analyses, 14, 16, 17-19. However, the other EAG scenario analyses were unfortunately not performed.