standard sources from the literature. The model has been very well validated against external data.

Weaknesses and Areas of uncertainty

The manufacturer's decision problem does not fully meet the scope in terms of interventions and comparators included. Principally, the submission does not cover people with T2DM inadequately controlled on monotherapy with a sulfonylurea. It is stated that relatively few patients would receive sulfonylurea monotherapy (most would receive metformin), and those that did would not be suitable for canagliflozin due to renal impairment. Expert clinical advice to the ERG suggests that not all people taking a sulfonylurea monotherapy would do so because metformin was contraindicated on this basis. Omission of this intervention could exclude a significant minority of people with diabetes who may be treated in the NHS. (Note that the only trial identified by the manufacturer evaluating a canagliflozin on a sulfonylurea background would not meet the scope in the view of the ERG as there was no active comparator.) Furthermore, there is no direct evidence comparing canagliflozin with other anti-diabetic drugs as add-on to insulin presented in the submission.

The NMA has been conducted to a reasonable standard, and is reported in adequate detail. However, caution is advised due to the low volume of direct evidence for some of the links in the networks. It should be noted that not all of the results of the NMA are used in the economic model (HbA_{1c} and BMI are the two NMA outcomes that are used most consistently in the model).

The manufacturer's quality assessment of trials included in the NMA noted some uncertainties in methodological quality, presumably due to poor reporting in the primary publications. There is potential for the results of the NMA to be subject to bias (though trials which were not double-blinded were excluded as a sensitivity analysis).

The economic model is complicated, consisting of around 20,000 lines of R code (in 29 separate files) with an Excel front- and back-end which contain additional lines of visual basic for applications (VBA) code. There are around 800 user-specifiable inputs. It has not been possible in the time available to fully check all inputs or model wiring.

The model uses patient-level simulation. This is common in models of T2DM due to the potentially complex nature of the disease pathway. However base case results are reported from simulations which use only 1000 patients per cohort and which may not therefore robustly capture the ICERs. Base case results also incorporate parameter uncertainty.

The manufacturer does not supply a fully incremental base case analysis and consequently the place of canagliflozin in the cost-effectiveness hierarchy is not clear.

The base case results reported in the MS generally show few incremental QALYs accruing to canagliflozin over a lifetime horizon of 40 years, at small incremental cost. Canagliflozin generally appears cost-effective in these results. However because of its small incremental benefit even slight variation to parameter assumptions can lead to a large change in the ICER.

Parameter distributions used in probabilistic sensitivity analysis (PSA) are not well documented. The extent to which PSA explores uncertainty in the decision problem is not clear.

Summary of additional work undertaken by the ERG

The ERG conducted additional analyses to:

- examine the variation in final ICER associated with re-running the base cases
- examine the variation in final ICER arising with no parameter uncertainty and 100,000 patients per cohort
- perform fully incremental analyses for dual therapy
- identify the probability that canagliflozin is cost effective in comparison with dapagliflozin
- examine the effect of modifications to the assumed clinical efficacy of glimepiride and exenatide

Fully incremental analysis conducted by the ERG for dual therapy in combination with metformin found that, with no parameter uncertainty and 100,000 patients per cohort, canagliflozin 100 mg is associated with an ICER of £84,800 compared to the next best option, DPP-4-i. The incremental differences for this comparison are £245 in total costs and 0.003 total QALYs over the lifetime horizon of the model. Canagliflozin 300 mg is associated with an ICER of £17,903

licenced indication and NICE's final scope – which states that the aim of this appraisal is to examine the effectiveness of canagliflozin within its licensed indication. The SmPC does not limit use of canagliflozin to only dual and triple therapy regimens including metformin; it states that canagliflozin can be used as an add-on therapy to other anti-hyperglycaemic agents, including insulin. Therefore the use of canagliflozin specified in the decision problem is narrower than defined in the licenced indication and may not reflect all the potential uses of canagliflozin in clinical practice. Furthermore, the ERG notes that the combination of canagliflozin, metformin and thiazolidinedione may be of limited relevance to clinical practice. Clinical expert advice to the ERG is that the use of thiazolidinediones is declining due to concerns about weight gain, heart failure and osteoporosis.

Comparators

For the assessment of dual therapy with a combination of canagliflozin and metformin, the comparators stated in the manufacturer's decision problem are all those specified in the final scope: sulfonylureas, thiazolidinediones (although the scope more specifically states 'pioglitazone'), DPP-4-inhibitors, GLP-1 receptor agonists and dapagliflozin, all in combination with metformin. These comparators are appropriate for the NHS and reflect the drug combinations recommended by NICE in CG 87,[7] TA 203,[10] TA 248,[9] and TA 288.[8]

For the assessment of triple therapy with a combination of canagliflozin, metformin and a sulfonylurea, the comparators stated in the manufacturer's decision problem are: DPP-4-inhibitors and GLP-1 analogues, both in combination with metformin and a sulfonylurea. The decision problem erroneously states that insulin is not a comparator for this triple therapy regimen, however insulin is included as a comparator in the manufacturer's clinical and cost-effectiveness analysis. These comparators are appropriate for the NHS and reflect the drug combinations recommended by NICE in CG 87,[7] TA 248,[9] and TA 203.[10] The comparator specified in the scope that have not been considered by the manufacturer is thiazolidinediones (although, again, the scope more specifically states 'pioglitazone') combined with metformin and a sulfonylurea. The manufacturer states that this has not been included as "insufficient clinical data were found to populate the NMA [network meta-analysis]" (MS p. 41). For the assessment of the triple therapy combination of canagliflozin, metformin and thiazolidinediones, the comparator is DPP-4-inhibitors combined with metformin and thiazolidinediones. Clinical expert advice to the ERG is that this combination is not generally used in clinical practice, and, as noted above, the use of thiazolidinediones is declining. The ERG therefore suggests that this comparator is of less relevance to clinical practice.

slightly higher in DIA3006 than in DIA3009 (MS p. 67), and the ERG agrees with this. The ERG notes that, as might be expected, proportionally more patients in the triple therapy study (DIA3015) than in the dual therapy studies had an HbA_{1c} of > 9% and had experienced diabetes-related complications. These patients also had a longer duration of disease.

The MS states that there is currently only one ongoing study of canagliflozin for T2DM: the CANVAS trial (DIA3008) of patients with a diagnosis of T2DM who have an elevated cardiovascular risk. This a safety study carried out in 4,330 patients which will report in 2017. The ERG notes that this trial is not relevant to the decision problem or scope, as it compares canagliflozin plus standard of care and placebo plus standard of care.

The MS appears to have identified all relevant head-to-head trials of canagliflozin. Other than this, the ERG identified one further ongoing RCT which is possibly relevant from searches of the UK Clinical Research Network (UKCRN) database (UKCRN study ID 7616). The study is a four-arm Phase 3 RCT comparing the efficacy, safety and tolerability of canagliflozin with placebo and sitagliptin in the treatment of patients with T2DM that is inadequately controlled on metformin and sulfonylurea. The ERG were unable to identify any further information about this study from internet searches and were unable to establish contact with the lead investigator, but note that it is funded by the manufacturer (Johnson & Johnson Pharmaceutical Research and Development).

Finally, although the ERG deemed the three placebo-controlled trials DIA3002, DIA3012, and DIA3008 (insulin sub-study) as not directly relevant to this report, it is noted that these trials are included in the NMA with the placebo arms used to connect treatments together indirectly. Furthermore, the economic model requires treatment effectiveness versus placebo to be specified. Whilst the ERG has not focused on these three trials in this report for the reasons stated above, given their inclusion in the NMA and the use of placebo comparison in the economic model, there is a brief discussion of their methodological quality in section 3.1.4.

3.1.4 Description and critique of the approach to validity assessment

The MS quality assessed all the included trials, and they used the NICE criteria (MS section 6.4). The canagliflozin clinical trials appear to be of good methodological quality and the ERG generally agrees with the manufacturer’s assessment of study quality (Table 3).

comparison in the economic model. Whilst the ERG has not undertaken an independent quality assessment of these three trials it is noted that they are similar in design to the active comparator trials, having been conducted as part of the canagliflozin clinical trial programme. It would therefore be reasonable to assume they are of similar good quality to the other trials in the programme.

3.1.5 Description and critique of manufacturer's outcome selection

The outcomes addressed in the manufacturer's decision problem include all of those specified in the scope, and are generally reported in the manufacturer's systematic review of the literature (though see below).

Generally the outcomes measured in the trials are reported in the MS, however see the specific exceptions noted in Table 3 (Manufacturer and ERG assessment of trial quality). Body Mass Index (BMI) is not reported in the MS but is included in the NMA.

MS Appendix 10.4.1 (Table 7) reports that a variety of exploratory HRQoL measures and Patient Reported Outcomes (PRO) were assessed using questionnaires, in the trials. These included established generic instruments such as the EQ-5D and SF-36, and disease-specific instruments: Current Health Satisfaction Questionnaire (CHES-Q); Impact of Weight on Quality of Life Lite (IWQOL-Lite); Multi-dimensional Diabetes Questionnaire (MDQ); Diabetes Utility Index (DUI); Diabetes Intention, Attitude, and Behavior Questionnaire (DIAB-Q).

The HRQoL and PRO outcomes are not reported in the main submission document (or in the trial journal publications), however MS Appendix 14 reports various sets of analyses of these instruments, some of which are pooled across trials, and some which are reported by trial (e.g. mainly trial DIA3009). Regression analyses are also reported to identify predictors of HRQoL. A description of the HRQoL and PRO instruments with supporting references is also provided. Some of the instruments are reported to have been validated, with some having been developed by the manufacturer.

direct comparisons are provided in the accompanying SLR report,[3] but do not appear to be in the main MS.

Outcomes of relevance to the economic model (i.e. change in HbA_{1c}, weight, BMI and SBP, and incidence of hypoglycaemia) are reported in the main submission (see MS p. 111) and further outcomes such as FPG are reported in the accompanying SLR report[3]. However, as noted later in this report (section 4.2.4) the economic model appears to use data from the NMA for HbA_{1c} and BMI, but for other outcomes input data is taken from the NMA or from specific clinical trials. Consequently, of all the outcomes included in the NMA the results for HbA_{1c} and BMI have most influence on the cost-effectiveness results in the submission.

Table 4 reports the ERG's critical appraisal of the manufacturer's NMA. In general the NMA is judged to be of reasonable quality according to the criteria, with a few caveats as noted in the table. The key caveats are:

- For the triple therapy networks, studies which did not measure outcomes at 26 +/- 10 weeks were excluded from the NMA after full text screening. This raises the possibility that some of those excluded could have been used to construct 52 week networks for triple therapy. This would have meant less reliance on 26 week effectiveness data in the economic model, and therefore the assumption of maintenance of treatment effects between 26 and 52 weeks (see point 4 in Table 4).
- From a brief examination of the manufacturer's tabulated critical appraisal judgments it appears that the methodological quality is reasonable, however, for a number of studies quality was judged to be unclear (see point 7 in Table 4).
- Some of the networks include few studies, particularly for triple therapy. The volume of direct evidence for some of the loops in the networks was low (approximately 2 trials for each loop, typically), with several treatments connected directly by only one trial (see point 8 in Table 4).
- Statistically significant heterogeneity was noted in a number of the direct comparisons across the networks, and was suggested to be caused by differences in variables such as HbA_{1c} and weight/BMI at baseline between studies (see point 9 in Table 4).

- It is not stated whether the inconsistency tests reported were adjusted to take account of multi-arm studies or multiple testing in networks with multiple loops. Very little information is given on the methods adopted for these tests, which do not appear to follow the methods outlined by the Decision Support Unit (DSU).

Table 4: ERG appraisal of Network Meta Analysis (NMA)

Appraisal criteria	Criteria met
A. CONCEPTUAL BASIS	
1. Is a justification given for conducting an MTC?	No. An explicit rationale is not given. However, given the number of comparators included in the scope it would be expected that canagliflozin would not yet have been directly compared to all of them, hence the justification for indirect comparison.
B. SYSTEMATIC PROCESSES	
2. Is a comprehensive and transparent search strategy reported?	Yes. The manufacturer conducted one overall literature search (MS section 6.1, and MS Appendix 10.2) to inform their clinical SLR of canagliflozin trials (MS sections 6.2 to 6.5) and their NMA (MS section 6.7). The ERG considers the search to be generally comprehensive (see ERG report section 3.11).
3. Are inclusion / exclusion criteria adequately reported?	Yes. MS Table 9 specifies the inclusion criteria for the clinical systematic literature review, and additional criteria for the NMA are specified in MS section 6.7.2.1. The additional criteria were proposed to ensure comparable pooling across trials (e.g. in terms of background therapy, trial length, patient co-morbidities). The ERG considers that the additional criteria are generally appropriate. The criteria exclude treatments not recommended in the UK by NICE or the Scottish Medicines Consortium.
4. Is the number of included /excluded studies from the MTC reported, with reasons for exclusions?	Yes, Figure 1 of the accompanying SLR[3] shows a PRISMA diagram, reporting the number of studies included (n=64 in the base case, plus a further in 21 in sensitivity analysis) and excluded (n=123) from the NMA and the reasons for exclusion. The manufacturer supplied a list of references excluded from the NMA at the request of the ERG. The ERG notes from the PRISMA flow chart that a total of 51 references were discarded from the NMA for not meeting the inclusion criteria. Specifically, studies with a time assessment not within the ranges of 26 weeks (+/- 10 weeks) or 52 weeks (+/- 10 weeks - only for the metformin background) were excluded. It is possible that some of these discarded studies could have been used to construct 52 week networks for triple therapy (i.e. MET and SU, and MET and TZD) and insulin therapy background, though the ERG has not conducted a thorough check of the list of excluded references provided due to time constraints.
5. Is a visual representation of the data networks provided?	Yes. MS Figures 34 to 38 illustrate the evidence networks for the five separate networks. The accompanying SLR report[3] provides separate network diagrams for each outcome included in each network.
6. Are the data from included studies extracted and tabulated?	Yes. Data were extracted by one reviewer. All data included in the meta-analysis were checked by a second analyst. Tables 48 to 104 in MS Appendix 10.7 provide data on study and patient characteristics, and data inputs into the NMA. The ERG consider that the level of detail provided is generally adequate.
7. Is the quality of the included studies assessed?	Yes. The manufacturer used the NICE recommended critical appraisal criteria as also used for the canagliflozin trials in MS section 6.4. The results of the appraisal are in MS Appendix 10.7.8 (Tables 105 to Table 108). No commentary or overall summary of study methodological quality is provided. It should be noted that critical appraisal appears from the MS to have been conducted by one reviewer and only a 20% sample checked by a second (After completion of this report the manufacturer confirmed this was an error and that the critical appraisal of all trials was checked by a second reviewer). The ERG has not performed an independent critical appraisal of the studies included in the NMA due to the large number of studies.

evidence?	to cause inconsistency were excluded in sensitivity analysis (see MS appendix 10.7.12 and 10.7.13) While DSU Technical Support Document 4[28] does indeed indicate that comparisons of direct and indirect evidence (the latter derived using the method described by Bucher et al[29]) may be used for assessing consistency they raise concerns over their use in networks containing multi-arm trials (such as DIA3006 and DIA3009) and the multiple testing required in networks with multiple loops. Neither the MS nor the accompanying SLR report[3] state whether the inconsistency tests reported were adjusted to take account of multi-arm studies or multiple testing in networks with multiple loops. Both reports provide very little information on the methods adopted for these tests, which do not appear to follow the methods outlined in the DSU document.[28]

3.2 Summary statement of manufacturer's approach

The quality of the manufacturer's systematic review is reasonable with some exceptions (see below), based on the CRD criteria (see Table 5). The MS provides a high level of detail on the procedures followed and the evidence included, though much of this detail can only be found in the appendices and also in a separate SLR report.[3]

There is a slight discrepancy in the description of the screening process for titles and abstracts between the MS and the separate SLR report[3] The MS states that titles and abstracts were screened by a first reviewer and a second reviewer quality checked the assessment. However, the SLR report[3] mentions that titles and abstracts 'were reviewed by two independent reviewers' (page 17), rather than them being checked (which would imply that the person checking would have seen the first reviewer's screening decision).

The MS does not explicitly state the process for screening full publications for inclusion in the systematic review. However, the SLR report[3] reports that screening decisions were quality checked by a second reviewer, with discrepancies resolved through discussion between the two reviewers (page 17).

The CRD recommend that as a minimum, one researcher should extract the data with a second researcher independently checking the data extraction forms for accuracy and detail. The MS reports that all data included in the meta-

had a probability of reducing weight of 16% compared to exenatide, but had a probability of at least 80% compared to the DPP-4-i sitagliptin, pioglitazone, and the SUs glimepiride and glibenclamide.

Table 27: NMA results for dual therapy, MET background at 52 weeks (BMI)

	Cana 100mg				Cana 300 mg				SUCRA
	Median	sd	CrI	Prob	Median	sd	CrI	Prob	
Exenatide 10µg BID	1.24	1.47	[-2.16 ; 3.95]	14%	1.10	1.47	[-2.30 ; 3.82]	16%	91%
Cana 300mg QD	0.14	0.96	[-1.86 ; 2.14]	38%					72%
Cana 100mg QD					-0.14	0.96	[-2.14 ; 1.86]	62%	67%
Vildagliptin 100mg QD	-0.19	2.00	[-4.63 ; 3.78]	57%	-0.32	2.00	[-4.75 ; 3.61]	62%	60%
Placebo	-0.49	1.47	[-3.82 ; 2.36]	73%	-0.63	1.47	[-3.94 ; 2.22]	77%	50%
Sitagliptin 100mg QD	-0.63	1.17	[-3.17 ; 1.75]	82%	-0.77	1.17	[-3.28 ; 1.60]	85%	45%
Pioglitazone 30mg QD	-1.53	2.28	[-6.01 ; 3.03]	78%	-1.66	2.28	[-6.13 ; 2.90]	80%	33%
Glimepiride	-1.72	1.17	[-4.08 ; 0.81]	95%	-1.85	1.17	[-4.22 ; 0.69]	95%	23%
Glibenclamide	-2.54	1.53	[-5.87 ; 0.52]	96%	-2.67	1.53	[-6.00 ; 0.40]	97%	9%

Triple therapy

In the triple therapy trial DIA3015, at 52 weeks, canagliflozin (300mg) combined with metformin and a sulfonylurea resulted in a statistically significant greater improvement in BMI than the DPP-4-i sitagliptin combined with metformin and a sulfonylurea (Table 28). Results for change in BMI from baseline were only reported in the NMAs (see Table 30 below).

Table 28: RCT results for triple therapy, MET and SU background at 52 weeks (weight [kg])

	Canagliflozin 300mg	Active comparator
DIA3015	(n = 377)	DPP-4-i (sitagliptin) (n = 378)
Value at Week 52 LOCF - Mean (SD)	85.27 (22.537)	89.65 (23.023)
Change from baseline - LS Mean (SE)	-2.5 (0.2)	0.3 (0.2)
Diff. of LS Means (SE) (minus sitagliptin) P-Value 95% CI	-2.8 (0.3) <0.001 (-3.3;-2.2)	

mITT analysis

Table 29 reports the weight change results of the random effects NMA for the metformin and sulfonylurea background. The network comprised eight trials, with no direct comparisons.

Critical appraisal of manufacturer's submitted economic evaluation

The ERG has considered the methods applied in the economic evaluation in the context of the critical appraisal questions listed in Table 51 below, drawn from common checklists for economic evaluation methods (e.g. Drummond and colleagues[30]). The critical appraisal checklist indicates that overall the manufacturer follows recommended methodological guidelines.

Table 51: Critical appraisal checklist of economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well-defined question?	Yes	In line with NICE scope, although some comparators were not included in the different therapy combinations
Is there a clear description of alternatives?	Yes	Comparisons made with SU, TZD, DPP-4-i, GLP-1-a, dapagliflozin and insulin according to their respective marketing authorisations
Has the correct patient group / population of interest been clearly stated?	Yes	For base case population: • people with T2DM requiring either dual therapy, triple therapy or as an add-on therapy with insulin • as per marketing authorisation Patients with an HbA_{1c} exceeding 9% (75mmol/mol) at baseline were considered as subgroups. (Discussed in sections 4.2.2)
Is the correct comparator used?	Partial	Scoped comparators for dual therapy with SU were not included; one scoped comparator (TZD) for triple therapy with metformin and SU was not included; two scoped comparators (GLP-1-a and insulin) for triple therapy with metformin and TZD were not included. (Discussed in section 2.3 and 4.2.3)
Is the study type reasonable?	Yes	
Is the perspective of the analysis clearly stated?	Yes	Noted in the decision problem as NHS and Personal Social Services (PSS)
Is the perspective employed appropriate?	Yes	Discussed in sections 4.2.6/4.2.7 for costs and 4.2.5 for outcomes.
Is effectiveness of the intervention established?	Yes	Discussed in section 4.2.4
Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	Yes	Discussed in section 4.2.1 A 40 year time horizon is used
Are the costs and consequences consistent with the perspective employed?	Yes	Discussed in sections 4.2.6/4.2.7 for costs and 4.2.5 for outcomes
Is differential timing considered?	Yes	3.5% for costs and health benefits, tested in sensitivity analysis at 0% and 6%
Is incremental analysis performed?	Yes	Described in section 4.1. Comparators discussed in section 4.2.3
Is sensitivity analysis undertaken and presented clearly?	Yes	Described in section 4.1. Discussed in section 4.2.9

In summary, the modelling approach adopted in the submission appears reasonable although no justification was provided for the final choice of model from other models which also met the inclusion criteria of the systematic review. The model structure is comprehensive and draws on well-documented risk equations including UKPDS 68 to model disease progression. However no details of the programming architecture of the 30 component program and data files, i.e. how the separate programs relate to each other, and the processing flow, are given in the Technical Report or User Guide. This has hindered model checking by the ERG. The model takes an appreciable amount of time to run with base case assumptions (15 hours plus). It is unclear if the base case use of 1000 patients per cohort is sufficiently large to robustly estimate the ICERs, or how the base case results might change when parameter uncertainty is switched off.

4.2.2 Patient Group

The patient group included in the economic evaluation is adults with T2DM. The principal sources used to describe patients' initial characteristics comes from RCTs included in the SLR (data reported in MS page 67, MS Appendix 4 and MS Appendix 9), see also Section 3.1.3 above. The populations correspond to the licensed indication for canagliflozin for add-on therapy. There are some differences in the populations in terms of their duration of diabetes and baseline HbA_{1c} reflecting the differences between the treatment sequencing of the trials (dual therapy, triple therapy, add-on insulin) as outlined below. The populations generally reflect the target population in clinical practice. Inadequate control of T2DM is defined in current guidelines[7] as HbA_{1c} >7.5% and the mean HbA_{1c} of participants in the included RCTs was above this threshold at baseline (although in the dual therapy group there may have been some participants below this as the mean HbA_{1c} was 7.8 and SD was 0.843).

For dual therapy with metformin background two RCTs (DIA3009/DIA3006) were 'pooled' (method not described); the populations of these trials were adults with T2DM. The inclusion criteria of these trials differed slightly in terms of the baseline range of HbA_{1c} eligible (which depended on the type and doses of prior treatment) but the mean age of participants overall was 56 years, 50% were male, 74% white and the mean baseline HbA_{1c} was 7.9%. The duration of diabetes was 6.7 years. For triple therapy with metformin and SU background two RCTs were pooled (DIA3002/DIA3015). Participants were all adults with T2DM, inclusion criteria differed slightly but participants mean age was 57 years, 54% were male, and the mean baseline HbA_{1c} was approximately 8.1%. The proportions of white participants was 69% overall (Appendix 9, Table 183) although the rates differed between the two trials, with around 82% in

- SU do not have renal impairment (see section 2.3 and 4.4 of this report for a discussion of this).
- In triple therapy comparisons in combination with a DPP-4-1 or GLP-1-a. [redacted]

A summary of the economic modelling comparisons conducted is shown in Table 55.

Table 55: Economic modelling comparisons conducted (reproduced from MS Table 32 p175).

Comparator	Cana dose	Dual therapy	Triple therapy		Add-on to insulin
		MET	MET+SU	MET+TZD	
SU	100mg	✓	*n/a*		
	300mg	✓	*n/a*		
TZD	100mg	✓		*n/a*	
	300mg	✓		*n/a*	
DPP-4-i	100mg	✓	✓	✓	✓
	300mg	✓	✓	✓	✓
GLP-1-a	100mg	✓	✓		✓
	300mg	✓	✓		✓
Dapagliflozin	100mg	✓			✓
	300mg	✓			✓
Insulin	100mg		✓		
	300mg		✓		

Legend: DPP-4-i: Dipeptidyl peptidase-4 inhibitor; GLP-1-a: Glucagon-like peptide-1 agonist ; n/a: not applicable; SU: sulfonylurea; TZD: Thiazolidinediones

The comparators are routinely used in UK NHS, although clinical advice to the ERG suggests that pioglitazone (a TZD) is used infrequently in clinical practice now. Two of the specific comparators used in the economic model (the SU glimepiride and DPP-4-I exenatide BID) are not the most widely prescribed in their class in the UK, and are not wholly equivalent clinically to other drugs in their class.

4.2.4 Clinical Effectiveness

The primary outcome from the efficacy trials used by the model is the first year change from baseline in HbA_{1c} (%) (MS p67). Secondary outcomes used by the model are first year BMI

State	Utility	SE	Reference	Justification
(per Episode)			disutility per symptomatic no help event; not chronic	87) also used utility estimates from Currie et al. 2006.MS ref 60; 128. Curie et al (2006) relates to a study that collected data from UK patients using both the hypoglycaemia fear survey and EQ-5D. The utility values were converted to annual values as follows: -0.047 x (3/12) year = -0.0118 per year.
Severe Hypoglycaemic Episode (per Episode)	-0.0118		Currie et. al 2006. annualized disutility per severe event; not chronic	
Obesity				
Excessive weight (disutility weight for each 1 kg/m^2 above a threshold of 25 kg/m^2. No disutility is applied to lower values)	-0.0061	0.001	CODE 2	Values used to inform utility values in the NICE CG-87
Adverse Events				
Lower urinary tract infections (male) (per Episode)	-0.0012		Janssen UK Study (regression analysis corrected for state duration)	No utility values related to urinary tract infections were identified in the systematic literature review. Therefore, a TTO utility study was conducted.(MS ref 164; 165)
Lower urinary tract infections (female) (per Episode)	-0.0012			
Upper urinary tract infections (male) (per Episode)	-0.0073			
Upper urinary tract infections (female) (per Episode)	-0.0073			
Balanoposthitis (male) (per Episode)	-0.0046			No utility values related to genital mycotic infections were identified in the systematic literature review. Therefore, a TTO utility study was conducted. (MS ref 164; 165)
Candidal vulvovaginitis (female) (per Episode)	-0.0046			

No HRQoL data for the canagliflozin clinical trials were reported in the SLR although MS Appendix 14 reports limited data on EQ-5D from three RCTs. These were for dual therapy (DIA3009) where the baseline utility was 0.798 (0.2049) and the week 52 utility was 0.814 (0.2140); and for triple therapy (DIA3002 + DIA3015 pooled) where the baseline utility was 0.788 (0.1929) and the week 52 utility was 0.797 (0.1957). These data are for all patients, rather than for just patients receiving canagliflozin.

The health benefits were measured and valued consistent with NICE's reference case and the utilities incorporated into the model appear to be appropriate. While the EQ-5D was used, no details of the value set adopted was reported in the MS, although as noted above the MS states

The ERG observes that whilst the scenario for no weight/BMI rebound following treatment discontinuation (MS scenarios 12 and 13) generally renders canagliflozin more cost-effective versus comparators (lower ICER than base case - for example MS Tables 78, 79, 80 and 81), this does not happen in the scenarios for canagliflozin 100mg and canagliflozin 300mg versus dapagliflozin in dual therapy (MS Tables 84 and 85 p261). In these scenarios the ICER is £24,697 compared to £8,674 in the base case for canagliflozin 100mg (MS Table 84), and £33,368 compared to £27,419. These results reflect data given in MS Table 185 which indicate that dapagliflozin is better at controlling BMI than both canagliflozin 100mg and canagliflozin 300mg (MS Table 185, Appendix p299).

Results of the scenario analysis for canagliflozin 300mg versus dapagliflozin in dual therapy show that canagliflozin 300mg is not cost-effective at a WTP threshold of £30,000/QALY in the scenarios for HbA_{1c} switch threshold; patient baseline characteristics; BMI rebound (as noted above); and disutility from weight gain (MS Table 85 p261).

The MS does not discuss the results of most of the scenario analyses specifically but concludes that the incremental gain offered by canagliflozin 100mg and 300mg is consistent across a broad set of conditions as the ICERs in the vast majority of cases remain cost-effective (MS p272). The ERG agrees with this assessment in general, with the exceptions noted above. Dapagliflozin is more cost-effective than canagliflozin 300mg in several scenarios. Parameter uncertainty was not switched off in scenario analysis.

Probabilistic sensitivity analysis

As stated earlier, a model PSA using 1000 cohorts of 1000 patients with a lifetime horizon of 40 years takes over 15 hours to run on a high specification desktop computer (Intel i7 16GM RAM). The MS states that PSA is a default feature of the ECHO-T2DM model and was run on all simulations conducted (MS p216). All base case simulations thus incorporate PSA and their results are given in Table 48, Table 49 and Table 50. Cost-effectiveness planes and acceptability curves are presented in MS Figures 72 to 87 (MS p245-254), for selected base case comparisons. Table 65 summarises the PSA results for the comparisons which are reported in the MS.

The submission does not identify the locations where the various PSA distributions are specified in the R code and due to time constraints the ERG did not check their coding.

Correlation is addressed for initial age and biomarker values. For the UKPDS 68 mortality and macro-vascular risk equations second order (parameter) uncertainty is captured using a set of 1,000 boot-strap replications of the regression coefficients (Technical Report p67). Correlation for the UKPDS 82 equations has not yet been applied as these have not yet been made available by the study authors (manufacturer's clarification letter to NICE and ERG).

In summary, although all relevant variables appear to have been included in PSA, the distributional assumptions for these variables are not well described in the MS and are not transparent from the input data spreadsheet or R code. Consequently it is not clear to what extent uncertainty in the decision problem has been explored. The standard errors used for the HbA1c % treatment effect parameters in the first year are too small as the drift estimate has been added to this effect in the first year but the standard error of the drift estimate is not reflected in the HbA1c% treatment effect standard error. Cost-effectiveness acceptability data are not presented for all base-case comparisons.

PSA using 1000 iterations is included in the model base-case results (Section 4.1). The ERG reports base-case results with no second-order uncertainty, and cost-effectiveness acceptability data for key comparisons not reported in the MS, in Section 4.3.

4.2.10 Comment on validity of results with reference to methodology used

The ERG considers that the economic model captures all important aspects of the disease pathway and extrapolates intermediate outcomes to final outcomes in a robust and consistent manner, drawing upon standard sources from the literature. The model has been very well validated against external data.

The model is run with 1000 patients per cohort in the base case. It is unclear if this is sufficient to robustly estimate the ICERs for all treatment comparisons. The base case ICERs incorporate parameter uncertainty. Base case results are presented as pairwise treatment comparisons. Fully incremental analyses are not supplied. Parameter distributions used in PSA are not well documented and the parameter ranges used for some variables in one-way sensitivity analysis do not reflect much uncertainty. PSA results are not presented for all base case treatment comparisons.