NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Contents:

The following documents are made available to consultees and commentators:

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

1. **Company submission** from Janssen

2. **Clarification questions and company responses**

3. **Patient group, professional group and NHS organisation submissions** from:
   a. Myeloma UK
   b. UK Myeloma Forum

4. **Evidence Review Group report** prepared by Southampton Health Technology Assessments Centre

5. **Evidence Review Group report – factual accuracy check**

6. **Evidence Review Group report – addendum**

7. **Technical engagement response from company**
   a. Technical engagement response
   b. Technical engagement response appendix

8. **Technical engagement responses from experts:**
   a. Dr Karthik Ramasamy – clinical expert, nominated by Janssen
   b. Dr Neil Rabin - clinical expert, nominated by UK Myeloma Forum
   c. Mrs Rosie Dill – patient expert, nominated by Myeloma UK

9. **Technical engagement responses from consultees and commentators:**
   a. Myeloma UK

10. **Evidence Review Group critique of company response to technical engagement** prepared by Southampton Health Technology Assessments Centre

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

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Document B

Company evidence submission

September 2020

File name	Version	Contains confidential information	Date
ID1510_Company Document B_REDACTED	1.0	Yes	02/09/2020

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 guide to the methods of technology appraisal and the NICE guide to the processes of technology appraisal.

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 [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.

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 [redacted] 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

Contents 3
List of tables 5
List of figures 7
List of abbreviations 9
B.1 Decision problem, description of the technology and clinical care pathway 14
B.1.1 Decision problem 14
B.1.2 Description of the technology being appraised 16
B.1.3 Health condition and position of the technology in the treatment pathway 17
B.1.3.1 Disease overview 17
B.1.3.2 Epidemiology 18
B.1.3.3 Effect of MM on patients and carers 19
B.1.3.4 Description of the clinical care pathway 20
B.1.4 Equality considerations 23
B.2 Clinical effectiveness 24
B.2.1 Identification and selection of relevant studies 24
B.2.2 List of relevant clinical effectiveness evidence 24
B.2.3 Summary of methodology of the relevant clinical effectiveness evidence 25
B.2.3.1 Study design 25
B.2.3.2 Baseline characteristics of trial participants 34
B.2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence 38
B.2.4.1 Study population and patient disposition 38
B.2.4.2 Statistical analyses 39
B.2.5 Quality assessment of the relevant clinical effectiveness evidence 44
B.2.6 Clinical effectiveness results of the relevant trials 45
B.2.6.1 Response analyses 46
B.2.6.2 Survival analyses 51
B.2.6.3 Landmark analyses for survival by response 61
B.2.6.4 Health-related quality of life assessment 66
B.2.7 Subgroup analysis 70
B.2.7.1 Response analyses 70
B.2.7.2 Survival analyses 71
B.2.8 Meta-analysis 71
B.2.9 Indirect and mixed treatment comparisons 72
B.2.9.1 Overview of the indirect treatment comparisons conducted 72
B.2.9.2 Identification of relevant studies 73
B.2.9.3 Compatibility of included studies and matching 75
B.2.9.4 Results of the matching-adjusted indirect comparisons 82
B.2.9.5 MAIC conclusion and external validation 84
B.2.10 Adverse reactions 85
B.2.11 Ongoing studies 94
B.2.12 Innovation 97
B.2.13 Interpretation of clinical effectiveness and safety evidence 98
B.3 Cost effectiveness 102
B.3.1 Published cost-effectiveness studies 102
B.3.2 Economic analysis 104
B.3.2.1 Patient population 104
B.3.2.2 Model structure 105
B.3.2.3 Intervention technology and comparators 111
B.3.3 Clinical parameters and variables 113
B.3.3.1 Induction and consolidation: response and survival 113
B.3.3.2 Survival inputs for post-consolidation 114
B.3.3.3 Adverse events 131

B.3.4 Measurement and valuation of health effects 132
B.3.4.1 Health-related quality-of-life data from clinical trials 132
B.3.4.2 Mapping 134
B.3.4.3 Health-related quality-of-life studies 135
B.3.4.4 Adverse reactions 135
B.3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis 136
B.3.5 Cost and healthcare resource use identification, measurement and valuation 138
B.3.5.1 Intervention and comparators' costs and resource use 138
B.3.5.2 Health-state unit costs and resource use 157
B.3.5.3 Adverse reaction unit costs and resource use 157
B.3.5.4 Miscellaneous unit costs and resource use 158
B.3.6 Summary of base-case analysis inputs and assumptions 158
B.3.6.1 Summary of base-case analysis inputs 158
B.3.6.2 Assumptions 162
B.3.7 Base-case results 169
B.3.7.1 Base-case incremental cost-effectiveness analysis results 169
B.3.8 Sensitivity analyses 170
B.3.8.1 Probabilistic sensitivity analysis 170
B.3.8.2 Deterministic sensitivity analysis 171
B.3.8.3 Scenario analysis 174
B.3.8.4 Summary of sensitivity analyses results 175
B.3.9 Subgroup analysis 175
B.3.10 Validation 176
B.3.10.1 Validation of cost-effectiveness analysis 176
B.3.11 Interpretation and conclusions of economic evidence 177
References 180
Appendices 188

List of tables

Table 1: The decision problem 15
Table 2: Technology being appraised 16
Table 3: Clinical effectiveness evidence(43) 24
Table 4: Key Study Characteristics for CASSIOPEIA(44, 46) 27
Table 5: IMWG criteria for MRD(64) 34
Table 6: Baseline patient demographics and disease characteristics in CASSIOPEIA (ITT population)(43, 44) 35
Table 7: Summary of IMWG Revised ISS Staging in MM (ITT population)(44) 38
Table 8: Summary of data sets analysed(44) 39
Table 9: Statistical methods for the Primary Analysis for Part 1 of CASSIOPEIA(44, 46) 40
Table 10: Summary of CASSIOPEIA data cuts(44, 45) 42
Table 11: Quality assessment of the CASSIOPEIA trial 44
Table 12: Summary of key clinical efficacy results(43-45, 65) 45
Table 13: Summary of response rates based on computerised algorithm (ITT population)(44, 45) 47
Table 14: Time to response among patients treated with BTd compared with DBTd (Response-evaluable analysis set, median follow-up = 18.8 months)(44) 48
Table 15: Duration of response among patients treated with BTd compared with DBTd (Response-evaluable analysis set, median follow-up = 18.8 months)(44) 49
Table 16: Disposition of best response of patients with post-consolidation MRD MFC Negative at 10^{-5} but were not in post-consolidation CR/sCR (ITT population, median follow-up = 29.2 months)(45) 51
Table 17: PFS results with and without IPW adjustments (ITT population, median follow-up = 18.8 months)(44) 53
Table 18: Summary of sensitivity analysis for PFS(44) 53
Table 19: Comparison of updated PFS (PHA vs PA1), regardless of 2nd randomisation (ITT population)(44, 45) 54
Table 20: PFS IPW adjusted results (ITT population, median follow-up = 29.2 months)(65) 55
Table 21: Comparison of CASTOR data-cuts: PFS(69-74) 56
Table 22: Time to disease progression among patients treated with BTd compared with DBTd (ITT population, assessed by computerised algorithm)(43, 45) 56
Table 23: OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months)(44) 59
Table 24: OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(65) 60
Table 25: Cox proportional hazard models results (landmark analysis) 64
Table 26: EORTC QLQ C30 and EQ-5D-5L compliance rates at study time points (ITT population)(44) 66
Table 27: Baseline values for the EORTC QLQ-C30 (CASSIOPEIA, ITT population)(44) 67
Table 28: EQ-5D-5L utility and VAS change from baseline (ITT population)(80) 70
Table 29: Excluded studies and rationale for exclusion from the MAIC 74
Table 30: Summary of the trials used in the MAICs 74
Table 31: Quality assessment scores 75
Table 32: MAIC Baseline characteristics(81) 75
Table 33: Baseline characteristics before and after matching DBTd/BTd to Bd(81) 78
Table 34: Baseline characteristics before and after matching DBTd/BTd to BCd(81) 80
Table 35: Results of the naive comparison and MAIC (DBTd/BTd versus Bd)(81) 83
Table 36: Results of the naive comparison and MAIC (DBTd/BTd versus BCd)(81) 84
Table 37: Naïve comparison of survival outcomes for BTd, BCd and Bd from PHE datasets(2). 85
Table 38: Treatment exposure (CASSIOPEIA, safety analysis set)(44) 86
Table 39: Summary of TEAEs[a] during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44) (43, 81) 87
Table 40: TEAEs[a] by MedDRA system organ class and preferred term during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44) 88

Table 41: Most common (≥2%) serious TEAEs[a] by MedDRA system organ class and preferred term during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44) 89
Table 42: Treatment-emergent IRRs during induction/ASCT/consolidation phase by system organ class and preferred term and maximum toxicity grade (CASSIOPEIA, safety population)(43, 44) 90
Table 43: Clinical trials for the evaluation of daratumumab in NDTE MM patients(105-108) 95
Table 44: Summary list of published cost-effectiveness studies 103
Table 45: Patient baseline characteristics in the cost-utility analysis(44) 105
Table 46: Features of the economic analysis 108
Table 47: Proportion of patients receiving induction, ASCT and consolidation therapy(43, 44) 113
Table 48: MRD negativity at post-consolidation assessment(80, 81) 114
Table 49: Goodness-of-fit statistics for BTd MRD+ PFS (landmark analysis) survival models.. 119
Table 50: Comparison of predicted survival rates for BTd MRD+ PFS (landmark analysis) survival models 119
Table 51: Hazard ratios for modelling PFS 121
Table 52: Goodness-of-fit statistics for BTd MRD+ OS (landmark analysis) survival models ... 122
Table 53: Comparison of predicted survival rates for BTd MRD+ OS (landmark analysis) survival models 122
Table 54: Hazard ratios for modelling overall survival 124
Table 55: Survival rates predicted by the model compared to those observed in CASSIOPEIA (May 2019 data cut; PHA) and the PHE dataset 129
Table 56: Comparison of patient characteristics from the Public Health England dataset (at diagnosis) and BTd arm of the CASSIOPEIA trial (at baseline) 130
Table 57: Incidence of adverse events included in the model (induction, ASCT and consolidation)(44) 131
Table 58: Mean EQ-5D utility during induction/ASCT/consolidation phase of CASSIOPEIA 132
Table 59: Duration for health-state utility values (induction and post-induction to post-consolidation response) based on CASSIOPEIA(44) 133
Table 60: Health-state utility values included in the model (base case) 133
Table 61: Health-state utility values included in the model (scenario analyses) 134
Table 62: Duration and utility decrements associated with adverse events included in the model 135
Table 63: Summary of utility values for cost-effectiveness analysis 137
Table 64: Summary of dosing regimens for first-line treatment included in the model 140
Table 65: Drug acquisition costs for first-line treatment (daratumumab and bortezomib) 141
Table 66: Drug acquisition costs for first-line treatment (thalidomide and dexamethasone) 142
Table 67: Summary of drug acquisition costs for first-line treatment – induction and consolidation 142
Table 68: Concomitant medication costs 143
Table 69: Stem-cell transplantation costs 144
Table 70: Distribution of patients to subsequent therapies (by line of therapy) 148
Table 71: Median TTP/PFS for subsequent lines of treatment 149
Table 72: Summary of dosing regimens for subsequent treatments 150
Table 73: Drug acquisition costs for subsequent treatments (ixazomib, lenalidomide, pomalidomide and panobinostat) 152
Table 74: Drug acquisition costs for subsequent treatments (cyclophosphamide) 153
Table 75: Summary of drug acquisition costs for subsequent treatments – drug costs per cycle (Ld, Bd, BCd, PBd, Pd, and ILd) 153
Table 76: Summary of drug acquisition costs for subsequent treatments – drug cost per cycle (DBd and daratumumab monotherapy) 155
Table 77: Administration costs 156
Table 78: Monitoring costs 157
Table 79: Adverse event costs 158
Table 80: Summary of variables applied in the economic model 159
Table 81: Assumptions used in the cost-utility analysis 163
Table 82: Deterministic base case results 169
Table 83: Average probabilistic cost-effectiveness results 170

Table 84: Summary of results from scenario analyses 174

List of figures

Figure 1: Disease and treatment progression of multiple myeloma(23) 18
Figure 2: Treatment effects most desire by patients(35) 20
Figure 3: Proposed positioning of daratumumab combination therapy for transplant eligible patients in the NHS England clinical pathway of care 22
Figure 4: Overview of the CASSIOPEIA Study Design(44) 26
Figure 5: Overview of CASSIOPEIA dosing schedule 27
Figure 6: Representation of depth of response 33
Figure 7: Participant flow in CASSIOPEIA(44) 39
Figure 8: Summary of post-consolidation* response rates based on computerised algorithm (ITT population, median follow-up = 18.8 months)(44) 46
Figure 9: Summary of response rates, based on time of assessment (ITT population)(43, 45) .. 47
Figure 10: MRD-negative rate (10^{-5}) post-consolidation[a] by MFC (ITT population, median follow-up = 18.8 months)(44) 50
Figure 11: Kaplan-Meier plot for PFS from 1st randomisation for induction/ASCT/consolidation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months;)(43, 44) 52
Figure 12: Kaplan-Meier plot for PFS from 1st randomisation for induction/ASCT/consolidation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months;)(65) 55
Figure 13: Kaplan-Meier plot of time to progression from 1st randomisation regardless of 2nd randomisation based on computerised algorithm (ITT population, median follow-up = 29.2 months)(45) 57
Figure 14: Kaplan-Meier plot for PFS2 from 1st randomisation regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(45) 58
Figure 15: Kaplan-Meier plot for OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months)(44) 60
Figure 16: Kaplan-Meier plot for OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(65) 61
Figure 17: Landmark analysis: PFS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months) 63
Figure 18: Landmark analysis: OS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months) 64
Figure 19: Exploratory analysis: pooled front-line and RRMM PFS analysis from daratumumab studies, stratified by MRD status and treatment group 65
Figure 20: EORTC QLQ-C30 GHS change from baseline among patients treated with either DBTd or BTd (mixed effects model for repeated measures)(80) 68
Figure 21: EORTC QLQ-C30 change from baseline in pain subscale scores (mixed effects model for repeated measures)(80) 68
Figure 22: EORTC QLQ-C30 change from baseline in emotional function subscale scores (mixed effects model for repeated measures)(80) 69
Figure 23: EORTC QLQ-C30 change from baseline in cognitive function subscale scores (mixed effects model for repeated measures)(80) 69
Figure 24: Kaplan-Meier plot of PFS among patients who did and did not achieve sCR (ITT population, median follow-up = 18.8 months)(43) 71
Figure 25: PFS Network of Evidence 72
Figure 26: The multiple mechanisms of action of daratumumab 97
Figure 27: Partitioned survival model structure 106
Figure 28: Response-based approach incorporating landmark analysis 107
Figure 29: Extrapolation of OS for DBTd (ITT population, median follow-up = 29.2 months) 114
Figure 30: Extrapolation of OS for BTd (ITT population, median follow-up = 29.2 months) 115

Figure 31: Approach to modelling survival by response and treatment arm in the base case analysis 117
Figure 32: Extrapolation of PFS for BTd MRD+ (landmark analysis) 120
Figure 33: Extrapolation of OS for BTd MRD+ (landmark analysis) 123
Figure 34: Comparison of OS and PFS for DBTd predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined) 127
Figure 35: Comparison of OS and PFS for BTd predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined) 128
Figure 36: Cost-effectiveness plane for DBTd versus BTd 170
Figure 37: Cost-effectiveness acceptability curve 171
Figure 38: ICER tornado plot from deterministic sensitivity analyses – top 10 parameters 173

List of abbreviations

Abbreviation	Definition
1PL	One prior line
2L/3L/4L	2nd/3rd/4th line
ADCC	Antibody-dependent cell-mediated cytotoxicity
ADCP	Antibody-dependent cellular phagocytosis
AE	Adverse event
AIC	Akaike information criterion
ANC	Absolute neutrophil count
ASCT	Autologous stem cell transplant
B	Bortezomib
BA	Base case analysis
BCd	Bortezomib, cyclophosphamide and dexamethasone
Bd	Bortezomib and dexamethasone
BIC	Bayesian information criteria
BLd	Bortezomib, lenalidomide and dexamethasone
BNF	British National Formulary
BSA	Body surface area
BTd	Bortezomib, dexamethasone and thalidomide
BW	Bi-weekly
Cd	Carfilzomib and dexamethasone
CDC	Complement-dependent cytotoxicity
CDF	Cancer Drugs Fund
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CR	Complete response
CSR	Clinical study report
CT	Computed tomography
CTd	Cyclophosphamide, thalidomide and dexamethasone
CUA	Cost-utility analysis
D	Daratumumab
d	Dexamethasone
DAD	Detailed Advice Document

DBd	Daratumumab, bortezomib and dexamethasone
DBMP	Daratumumab, bortezomib, melphalan and prednisone
DBTd	Daratumumab, bortezomib, thalidomide and dexamethasone
DCEP	Dexamethasone cyclophosphamide-etoposide-cisplatin
DLd	Daratumumab, lenalidomide and dexamethasone
DOR	Duration of response
DSA	Deterministic sensitivity analysis
DSU	Decision Support Unit
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EMA	European Medicines Agency
eMIT	Electronic market information tool
EORTC-CLQ-C30	European Organisation for Research and Treatment of Cancer-Quality of Life Questionnaire-C30
EQ-5D-5L	EuroQol-5D, 5 levels
ESMO	European Society of Medical Oncology
ESS	Effective sample size
FEV	Forced Expiratory Volume
FISH	Fluorescent in situ hybridization
FLC	Free light chains
GCP	Good Clinical Practice
GCSF	Granulocyte-colony stimulating factor
GHS	Global health status
GPM	General population mortality
HDT	High-dose therapy
HIV	Human immunodeficiency virus
HOVON	Dutch-Belgium Cooperative Trial Group for Hematology Oncology
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICD-O-3	International Classification of Disease of Oncology 3rd edition
ICER	Incremental cost-effectiveness ratio
ICH	The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

IDMC	Independent data monitoring committee
IFM	Intergroupe Francophone du Myelome
IgG1κ	Human immunoglobulin G1 kappa
ILd	Ixazomib, lenalidomide and dexamethasone
IMWG	International Myeloma Working Group
IPD	Individual patient data
IPW	Inverse Probability Weighting
IRR	Infusion-related reactions
ISS	International Staging System
ITT	Intention-to-treat
IV	Intravenous
IWRS	Interactive web response system
KM	Kaplan-Meier
LCD	Light-chain disease
Ld	Lenalidomide plus dexamethasone
LDH	Lactate dehydrogenase
LEN-2Y	Lenalidomide maintenance for 2 years
LEN-CR	Lenalidomide until complete response
LS	Least-squares
LY	Life year
LYG	Life years gained
mAb	Monoclonal antibody
MAIC	Matching adjusted indirect comparison
MFC	Multiparametric flow cytometry
MGUS	Monoclonal gammopathy of undetermined significance
MM	Multiple myeloma
MRD	Minimal residual disease
MRI	Magnetic resonance imaging
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NCRAS	National Cancer Registration and Analysis Service
NDTE	Newly diagnosed transplant-eligible
NE	Not evaluable/estimable
NGF	Next generation flow
NGS	Next-generation sequencing

NMA	Network meta-analysis
NR	Not reported
ONS	Office for National Statistics
ORR	Overall response rate
OS	Overall Survival
PA1	Primary Analysis for Part 1
PAd	Doxorubicin-dexamethasone
PAS	Patient Access Scheme
PBd	Panobinostat, bortezomib and dexamethasone
Pd	Pomalidomide and dexamethasone
PCR	Polymerase chain reaction
PD	Progressed disease
PET	Positron emission tomography
PF	Progression free
PFS	Progression-free survival
PFS2	Progression-free survival on subsequent line of therapy
PHA	Post-hoc Interim Analysis
PHE	Public Health England
PO	Per os (oral)
PR	Partial response
PSA	Probabilistic sensitivity analysis
Q2W	Every 2 weeks
QALY	Quality adjusted life year
QD	Once daily
QoL	Quality of life
QW	Every week
RBC	Red blood cell
RCT	Randomised control trial
R-ISS	Revised International Staging System
RRMM	Relapsed/refractory multiple myeloma
RTDS	Radiotherapy Dataset
SA1	Sensitivity Analysis 1
SA2	Sensitivity Analysis 2
SACT	Systemic Anti-Cancer Therapy

SC	Subcutaneous
sCR	Stringent complete response
SCT	Stem-cell transplantation
SD	Standard deviation
SD	Stable disease
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of Product Characteristics
SOC	Standard of care
STA	Single Technology Appraisal
SUV	Standardised uptake value
T	Thalidomide
Td	Thalidomide and dexamethasone
TEAE	Treatment-emergent adverse event
TFI	Treatment-free interval
TSD	Technical support document
TTP	Time to progression
ULN	Upper limit of normal
URTI	Upper respiratory tract infections
VAd	Vincristine, doxorubicin-dexamethasone
VAS	Visual analogue scale
VAT	Value added tax
VGPR	Very good partial response
WHO	World Health Organization

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1 *Decision problem*

The submission covers the technology's full marketing authorisation for this indication: daratumumab in combination with bortezomib, thalidomide and dexamethasone (DBTd) for the treatment of adult patients with newly diagnosed multiple myeloma (MM) who are eligible for autologous stem cell transplant (ASCT).

The decision problem addressed in this submission, compared to that defined in the final scope issued by NICE, is summarised in Table 1. The company submission differs from the final NICE scope and the NICE reference case with respect to the included comparators only.

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	People with previously untreated MM who are eligible for ASCT	Adult patients with newly diagnosed MM who are eligible for ASCT	This population is considered to be in line with the full marketing authorisation for this indication
Comparator(s)	• Bortezomib with dexamethasone (Bd) or with dexamethasone and thalidomide (BTd) • Bortezomib with cyclophosphamide and dexamethasone (BCd) (off-label) • Cyclophosphamide with thalidomide and dexamethasone (CTd) (off-label)	• Bd • BTd • BCd (off-label)	Janssen does not consider CTd a relevant comparator to DBTd in this indication following clinical expert feedback that CTd is rarely used as an induction therapy for NDTE MM patients in England.(1) Real-world evidence supports limited CTd usage, with steady decline in prescribing and less than 2% of NDTE MM patients in England treated with CTd since 2018.(2) Furthermore, CTd is not recommended by NICE, or recognised by international or European clinical practice guidelines.

Key: ASCT = autologous stem cell transplant; Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, dexamethasone and thalidomide; CTd = cyclophosphamide, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MM = multiple myeloma; NDTE = newly diagnosed transplant-eligible.

B.1.2 *Description of the technology being appraised*

A description of the technology being appraised, DBTd, is presented in Table 2.

Table 2: Technology being appraised

UK approved name and brand name	Daratumumab (Darzalex®)
Mechanism of action	Daratumumab is a first-in-class, fully human immunoglobulin G1 kappa (IgG1κ) monoclonal antibody (mAb) that binds to CD38, a cell surface glycoprotein found on the surface of many immune cells, including white blood cells.(3, 4) Preclinical data suggests that daratumumab binding to CD38 induces tumour cell death through multiple mechanisms, including direct on-tumour and indirect immunomodulatory actions.(5) These processes include immune-mediated mechanisms of action (i.e. complement-dependent cytotoxicity [CDC], antibody-dependent cell-mediated cytotoxicity [ADCC] and antibody-dependent cellular phagocytosis [ADCP]), as well as induction of myeloma cell apoptosis and various immunomodulatory mechanisms. The concept of clonal heterogeneity contributing to disease progression in MM led to the strategy of adopting combination therapies to eradicate both the dominant and minor clones. Combination treatment strategies are now recommended for routine clinical practice by the International Myeloma Working Group (IMWG). CD38 is a distinct target from those of other approved agents for MM and this together with its high efficacy and favourable safety profile make daratumumab an ideal candidate for combination therapy.
Marketing authorisation/CE mark status	Marketing authorisation was granted on 20th January 2020
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	The licenced indications for daratumumab are: • "in combination with lenalidomide and dexamethasone (DLd) or with bortezomib, melphalan and prednisone (DBMP) for the treatment of adult patients with newly diagnosed multiple myeloma who are ineligible for autologous stem cell transplant" • "in combination with bortezomib, thalidomide and dexamethasone for the treatment of adult patients with newly diagnosed multiple myeloma who are eligible for autologous stem cell transplant" • "in combination with lenalidomide and dexamethasone (DRd), or bortezomib and dexamethasone (DBd), for the treatment of adult patients with multiple myeloma who have received at least one prior therapy" • "as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, whose prior therapy included a proteasome inhibitor and an immunomodulatory agent and who have demonstrated disease progression on the last therapy."(5)
Method of administration and dosage	Daratumumab 1,800 mg is available as a solution for injection.(5) Daratumumab is available as a fixed dose with each 15 mL vial of solution for injection containing 1,800 mg (120 mg daratumumab per mL). Daratumumab is administered once weekly for the first two cycles (weeks 1-8), followed by every two weeks for cycles 3-4 and cycles 5-6. Drug administration should be done by a healthcare professional, and the first dose should be administered in an environment where resuscitation facilities are available.(5)

	Daratumumab 20 mg/ml is also available as a solution for infusion.(6) Daratumumab administered via infusion is available in two single dose vials 100 mg/5 ml (20 mg/ml) and 400 mg/20 ml (20 mg/ml). The recommended dose of daratumumab is 16 mg/kg body weight administered as an intravenous (IV) infusion according to the same dosing schedule described above (as solution for injection) and requires dilution and administration by a healthcare professional.(6)
Additional tests or investigations	A one-off blood sample to type and screen patients' serum is required prior to starting daratumumab.(5)
List price and average cost of a course of treatment	List Price 1,800 mg (fixed-dose vial) = £4,320.00 (excl. VAT). This is equivalent to the cost of a 1,200 mg IV infusion (i.e. cost parity assuming an average daratumumab patient weight of 75 kg). List Price 100 mg (IV infusion) = £360.00 (excl. VAT) List Price 400 mg (IV infusion) =£1,440.00 (excl. VAT)
Patient access scheme (if applicable)	[redacted]
Key: ADCC = antibody-dependent cell-mediated cytotoxicity; ADCP = antibody-dependent cellular phagocytosis; CDC = complement-dependent cytotoxicity; DBd = daratumumab, bortezomib and dexamethasone; DBMP = daratumumab, bortezomib, melphalan and prednisone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; DLd = daratumumab, lenalidomide and dexamethasone; IgG1κ = human immunoglobulin G1 kappa; IMWG = International Myeloma Working Group; IV = intravenous; mAb = monoclonal antibody; PAS = Patient Access Scheme; VAT = value added tax.	

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

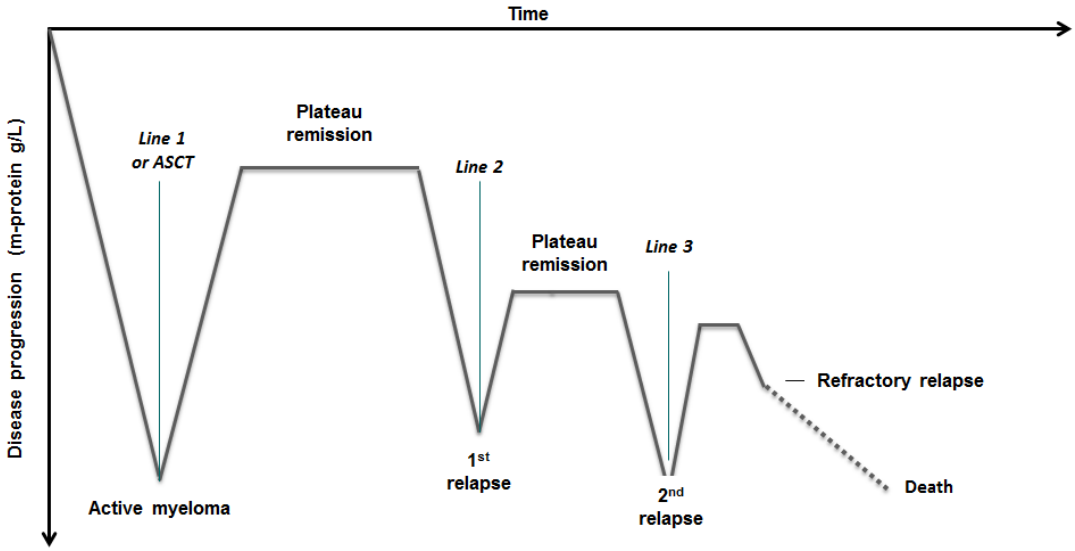
B.1.3.1 Disease overview

MM is a rare and incurable blood cancer with orphan disease designation in both the USA and Europe.(5, 7, 8) It is characterised by the excessive proliferation of malignant plasma cells within the bone marrow and the overproduction of M-protein.(9-11) Over time, these components accumulate in the bones, blood and multiple organs throughout the body, leading to serious complications which require immediate medical treatment, including elevated calcium levels (hypercalcemia), renal impairment, anaemia and bone disease.(9, 12) Additional presenting features include fatigue and unremitting bone pain, recurrent or persistent infection and hyperviscosity (i.e. increased blood viscosity).(9, 12, 13)

MM is genetically complex and develops from the continued accumulation of genetic abnormalities over time.(14) The genetic heterogeneity of MM means it is a difficult disease to treat and that clinical outcomes, including overall survival (OS), vary depending on a number of prognostic factors including: International Staging System (ISS) stage and whether the patient is considered high-risk.(15, 16) MM follows a relapsing-remitting course where all newly diagnosed patients eventually become refractory to therapy over time.(17-20) With each relapse, it becomes more difficult to induce deep and durable responses to treatment and attrition rates increase.(21, 22) Consequently, the prognosis of patients with relapsed/refractory disease is much poorer than

those with newly diagnosed MM, with the prognosis worsening with each successive relapse (Figure 1). It is therefore important to use the most effective treatments in the front-line setting, as patients may not survive or be fit enough to receive treatment at later lines.

Figure 1: Disease and treatment progression of multiple myeloma(23)



Key: ASCT = autologous stem cell transplant.

B.1.3.2 Epidemiology

In 2017, there were 5,034 new cases of MM in England, accounting for 2% of all new cancer cases.(24) Over the last decade, MM incidence rates have increased by approximately 15% in the UK and are projected to rise a further 11% between 2014 and 2035; this increase is largely a reflection of the changing prevalence of risk factors and improvements in diagnosis.(24) For patients with newly diagnosed MM, high-dose therapy (HDT) followed by an ASCT represents standard of care for those patients who are fit enough to receive these interventions.(25, 26) HDT-ASCT is an intensive treatment option and involves giving high doses of chemotherapy (typically melphalan) to kill the myeloma cells and then infusing stem cells back into the patient, allowing the bone marrow to recover.

The majority of patients in the UK with MM are diagnosed at a later stage in life (74% are diagnosed aged ≥65 years), and so may not be fit enough to receive HDT and ASCT.(24) Age alone does not however determine eligibility for ASCT, and according to NICE NG 35 guidelines, frailty and performance status measures that include comorbidities should be considered when assessing suitability.(25) Approximately one third of patients with newly diagnosed MM are expected to be eligible for ASCT, based on clinical expert feedback following a recent advisory board meeting involving three UK clinicians.(1)

Considerable progress in the treatment of MM has improved patient survival, however, MM remains incurable and all surviving patients will eventually relapse. In England, the 5- and 10-year survival rates for adults with newly diagnosed MM are approximately 52.3% and 29.1% respectively (2013-2017).(24) For patients who are eligible for ASCT, outcomes have improved significantly following the introduction of novel agent-based combinations as induction therapy prior to HDT and ASCT (e.g. bortezomib with dexamethasone), and response rates have been shown to improve with the addition of a third-agent.(26) However, since the recommendation by NICE in 2014 of bortezomib in combination with dexamethasone (Bd), or with thalidomide and dexamethasone (BTd), no other treatments have received European Medicines Agency (EMA)

licence approval as induction therapy for adult patients with previously untreated MM who are eligible for HDT and ASCT.(27) With all patients eventually relapsing with currently available therapies, and given the poorer prognosis associated with relapsed/refractory disease, there still remains a high level of unmet need for effective, well tolerated new treatment options in the front-line setting.

B.1.3.3 Effect of MM on patients and carers

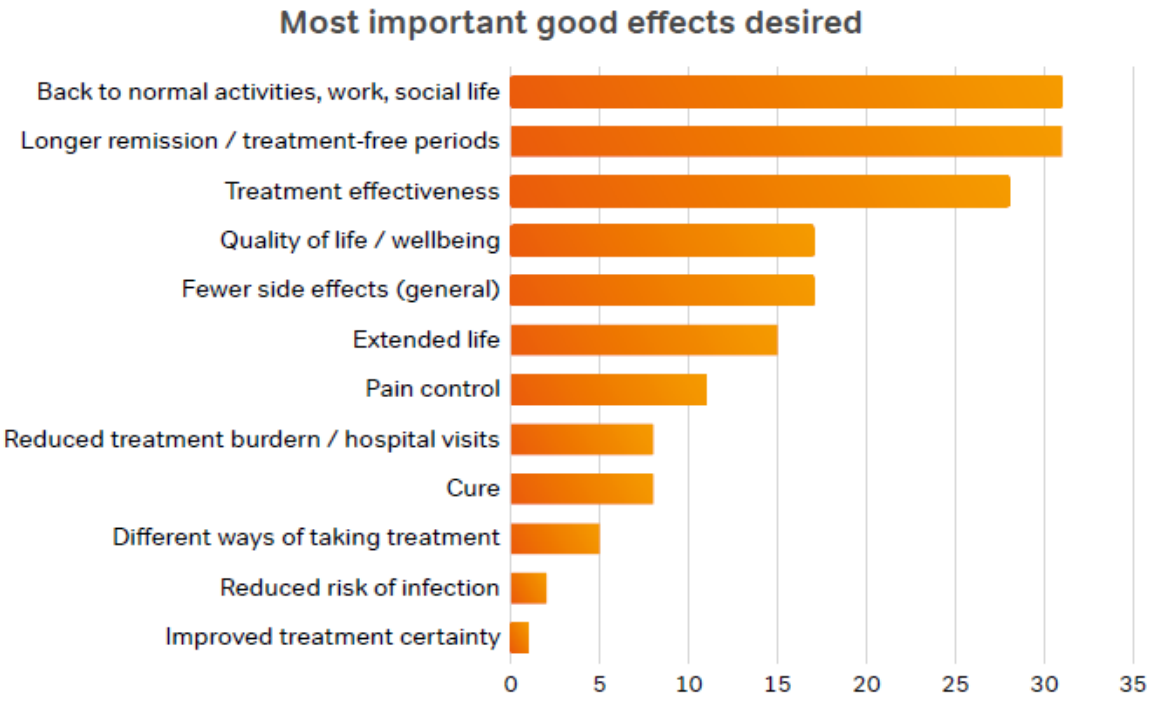
There is evidence that patients with myeloma report worse symptoms and health-related quality of life (HRQoL) than those with other haematological cancers, including lymphoma or leukaemia.(28) The clinical burden of MM is influenced by both progressive disease symptoms and treatment-associated complications such as weakness, fatigue, bone pain, weight loss, confusion, excessive thirst and constipation.(29)

Patients with MM live in fear of relapse.(30) Uncertainty about the future causes ongoing anxiety and often affects patients' relationships with family and friends who may act as informal caregivers.(30, 31) This leads to decreased independence and increased social isolation.(30) Treatments that achieve a lasting remission, offer maximum life expectancy and freedom from the emotional burden of the disease (to "not always think of the disease") are therefore highly valued by patients.

Achieving prolonged remission following first-line treatment is critical for improving and maintaining the HRQoL of patients. Indeed, the symptomatic burden for patients with relapsed/refractory disease is greater than newly diagnosed MM due to the progressive nature of the disease and the cumulative adverse effects of subsequent treatment.(32) Observational data from a UK study, which included responses from 370 patients with MM, demonstrated that patient HRQoL is reduced following progression from their first treatment-free interval (TFI) to second-line treatment and subsequent lines of therapy.(33) This study also showed that a longer TFI was significantly associated with improved HRQoL.(33)

In a recent European study of patient perceptions regarding MM and its treatment in patients with newly diagnosed and relapsed/refractory MM (N=30), patient preferences on key efficacy and safety outcomes were elicited.(34) The results of qualitative interviews revealed increased life expectancy (87%), remission/response (80%) and reduced fatigue (80%) as the most important treatment preferences. Symptoms of fatigue and bone pain were most often discussed while, among patients with NDMM, cognitive impairment was the most frequently mentioned side-effect (94% of respondents). Duration of treatment was most often discussed in the context of treatment burden (mentioned by 83% of NDMM respondents), indicating that a sustained period of treatment-free remission would be highly valued by patients. This finding is consistent with results from a recent qualitative survey undertaken by NICE's Science Policy and Research programme in collaboration with Myeloma UK. In the survey of 97 UK MM patients, respondents were asked what the most important good effects (or characteristics) they would want from any treatment for myeloma with the joint top-ranked response being a longer remission / treatment-free period (Figure 2).

Figure 2: Treatment effects most desire by patients(35)



The symptom burden associated with MM was also highlighted in the responses from this survey, with fatigue and tiredness; other symptoms and side effects; mobility and daily activities; and pain and discomfort, being reported by patients as the aspects of MM that has the greatest impact on their lives.(35) The negative effects of treatment that patients would most want to avoid were also assessed as part of the survey, thus highlighting the need for treatments that themselves have minimal disruption on patient's health (i.e. avoidance of adverse events) and normal activities. Across both studies, it is clear that longer remission and treatment-free intervals are goals of therapy that are highly valued by patients with MM, in addition to increased life expectancy and reduced symptom burden.

Most of the clinical management of MM is provided in the outpatient setting; therefore the bulk of care is informal and provided by caregivers.(36) Caregivers may perform complicated technical procedures (e.g. dressing changes, intravenous line care and injections), assist the patient with daily living, attend appointments and take in complex information.(36) Therefore, the detrimental effects of MM on working life are not only experienced by patients, but also their caregivers.(37) Almost half (49%) of the partners of patients with MM report symptoms of anxiety and 14% report symptoms of depression.(37) The emotional impact experienced by caregivers of patients with MM further hinders their ability to work. The unmet need in supportive care is considerable and carers have specifically reported a need for help to manage the side effects and complications experienced by patients due to treatment for MM.(37)

B.1.3.4 Description of the clinical care pathway

MM is a treatable but incurable disease. Patients typically require multiple lines of treatment, usually involving drug combinations with proteasome inhibitors (PIs; i.e. bortezomib, carfilzomib or ixazomib) and/or immunomodulatory agents (IMiDs; i.e. thalidomide, lenalidomide or pomalidomide), with dexamethasone added to both treatment classes to further alleviate symptom burden, with or without SCT. Almost all surviving patients with MM eventually relapse

from, or become refractory to, existing treatment options.(38) Consequently, the aims of treatment are to induce remission, delay progression, prolong survival and maximise quality of life.(26)

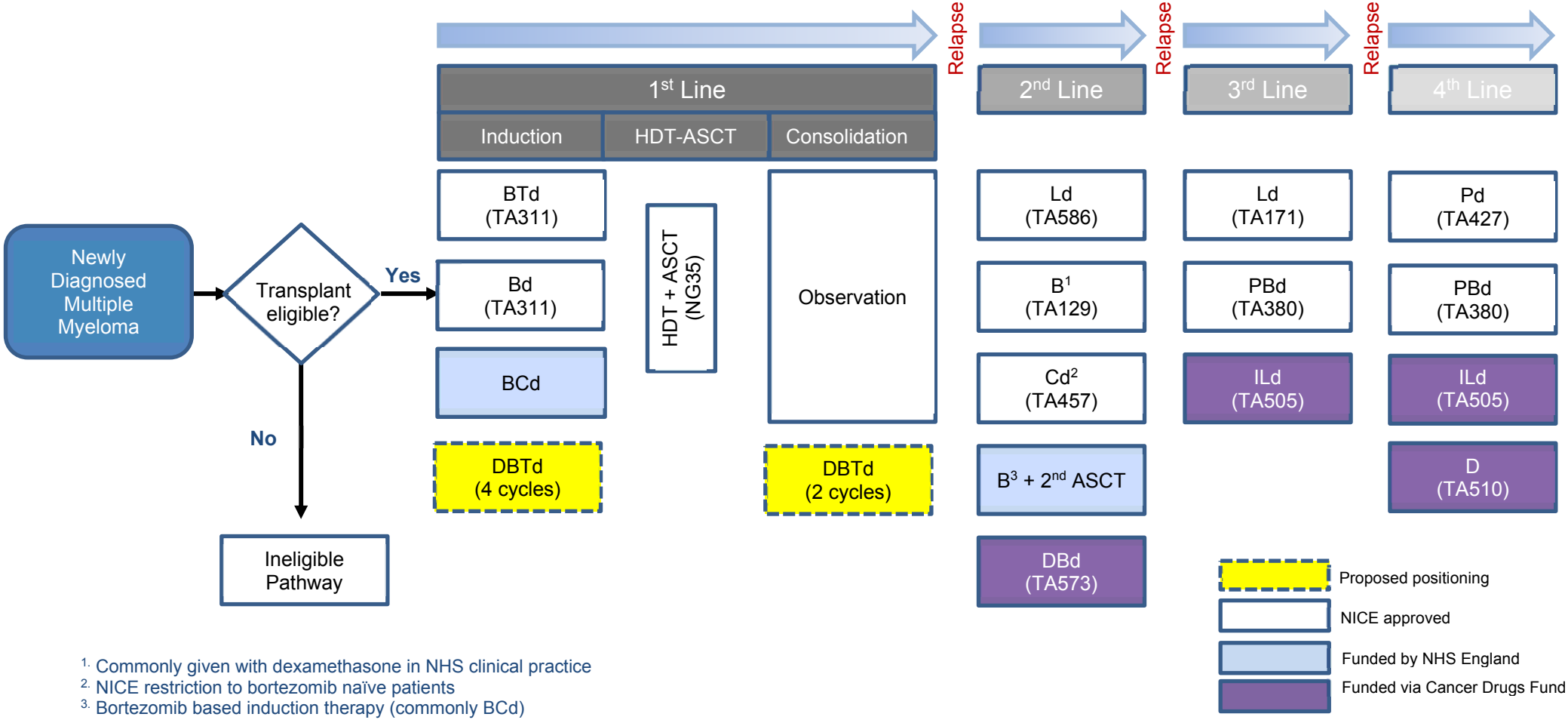
Treatment guidelines for the management of MM are available from the European Society of Medical Oncology (ESMO), European Myeloma Network (EMN), National Comprehensive Cancer Network (NCCN) and NICE (refer to NG35).(25, 26, 39, 40) Several regimens have been recommended by NICE for the treatment of MM, predominantly relating to the relapsed/refractory setting (refer to Figure 3 below).(25) For newly diagnosed patients who are fit enough to receive intensive treatment, HDT followed by ASCT represents the current standard of care.(25, 26) Approximately one third of patients with newly diagnosed MM are expected to be eligible for HDT-ASCT, based on clinical expert feedback following a recent advisory board meeting involving three UK clinicians.(1)

Prior to receiving HDT-ASCT, patients receive induction therapy to reduce the number of (malignant) plasma cells in the bone marrow and achieve some form of remission.(41) The only induction therapy recommended by NICE for patients with previously untreated MM who are eligible for ASCT is bortezomib; either in combination with dexamethasone, or with thalidomide and dexamethasone.(27) Clinical outcomes for patients receiving triplet therapy are superior to those receiving Bd alone, and a recent advisory board meeting involving three UK clinicians confirmed that treatment with BTd represents standard of care (SOC) induction therapy for newly diagnosed patients who are eligible for ASCT.(1, 26) For a minority of patients where thalidomide is not considered suitable (e.g. due to challenging thrombosis or baseline neuropathy/neurotoxicity), clinician feedback is that bortezomib in combination with cyclophosphamide and dexamethasone (BCd; off-label) may be administered instead with the doublet therapy, Bd, rarely used.(1)

Despite improvement in patient outcomes following the introduction of bortezomib-based induction therapy, MM remains incurable and all patients eventually relapse. One of the challenges of treatment to date has been to find options that effectively target and eliminate all clonal and subclonal mutations. Daratumumab binds to CD38, a protein that is overexpressed on the surface of MM cells. It works by targeting the tumour directly and indirectly, as well as uniquely modulating the immune system.(3, 4) It is this combination of direct and immunomodulatory effects that harnesses the body’s own immune system to fight the disease that explains the deep responses and step-change in efficacy observed with daratumumab for this indication.

The current clinical care pathway for MM patients in England is presented in Figure 3, including the proposed positioning of DBTd for front-line transplant-eligible patients. The EMA licence for DBTd includes 4 cycles of induction therapy, ASCT, followed by 2 cycles of consolidation therapy. Consolidation therapy is not part of routine clinical practice in the NHS in England, however it is included in the licence for DBTd and is therefore part of the evidence considered in the submission. Consolidation therapy is generally given for a short duration (2-4 cycles) after ASCT to further deepen responses, and aims to provide long-term disease control.(42) The current clinical care pathway is based on recommendations made by NICE as part of previous technology appraisals.

Figure 3: Proposed positioning of daratumumab combination therapy for transplant eligible patients in the NHS England clinical pathway of care



1. Commonly given with dexamethasone in NHS clinical practice
2. NICE restriction to bortezomib naïve patients
3. Bortezomib based induction therapy (commonly BCd)

Key: ASCT = autologous stem cell transplant; Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; Cd = carfilzomib and dexamethasone; CDF = Cancer Drugs Fund; D = daratumumab; DBd = daratumumab, bortezomib, and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HDT = high-dose therapy; ILd = ixazomib, lenalidomide and dexamethasone; ITT = intention-to-treat; Ld = lenalidomide and dexamethasone; NICE = National Institute for Health and Care Excellence; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone.

B.1.4 *Equality considerations*

No equality issues related to the use of daratumumab combination therapy (i.e. DBTd) for the treatment of newly diagnosed transplant-eligible (NDTE) MM patients have been identified.

B.2 Clinical effectiveness

B.2.1 *Identification and selection of relevant studies*

A systematic literature review (SLR) of the published literature was conducted to identify the relevant clinical efficacy and safety data for DBTd (and comparators) as a treatment for NDTE MM patients (refer to Appendix D where the full SLR methodology and results are presented). One randomised control trial was identified, MMY3006 (CASSIOPEIA), that included NDTE MM patients receiving DBTd, with results from the Primary Analysis for Part 1 reported in Moreau et al. (2019).(43) In addition to the published evidence sources, the following non-published evidence have also been included within this submission:

- the trial clinical study report (CSR)(44)
- results from a post-hoc interim analysis (PHA) performed to support EMA regulatory approval(45)
- results from a post-hoc landmark analysis of progression free survival (PFS) and OS to support economic model development (refer to Section B.2.6.3)

B.2.2 *List of relevant clinical effectiveness evidence*

CASSIOPEIA is an ongoing phase III randomised, open-label, active-controlled, European multicentre trial (see Table 3).(43) Evidence from the CASSIOPEIA trial was used as the primary source of data to support the use of DBTd in this indication in the marketing authorisation application to the EMA. Pre-specified analysis for Part 1 applied a clinical cut-off date of 19th June 2018, representing a median follow-up of 18.8 months. During the regulatory process, Janssen received a Request for Supplementary Information (RSI) from the EMA which resulted in an unplanned post-hoc interim analysis with a clinical cut-off of 1st May 2019, representing an additional 10.4 months of study follow-up (total median follow-up of 29.2 months).

Clinical inputs used in the cost-effectiveness model were derived from the CASSIOPEIA trial (refer to Section B.3.3)

Table 3: Clinical effectiveness evidence(43)

Study	CASSIOPEIA (NCT02541383)
Study design	Randomised, open-label, active-controlled, parallel-group, multicentre, Phase III trial. In Part 1, patients were randomised to receive four 28-day cycles of induction therapy with DBTd or BTd prior to HDT-ASCT, followed by two 28-day cycles of consolidation therapy with DBTd or BTd.
Population	Adults with previously untreated MM who are eligible for autologous stem cell transplantation.
Intervention(s)	**DBTd (N = 543):** • Daratumumab (16 mg/kg) was administered by IV infusion weekly for two 28-day cycles, then every 2 weeks for the remaining induction and consolidation cycles based on treatment assignment BTd in the DBTd arm was administered as described below for the comparator
Comparator(s)	**BTd (N = 542):** • Bortezomib was administered SC at a dose of 1.3 mg/m^2 twice a week for four 28-day induction cycles, and two consolidation cycles

	• Thalidomide was administered orally at 100 mg daily for four 28-day induction cycles and two 28-day consolidation cycles • Dexamethasone was administered orally or via IV infusion at 40 mg on days 1,2,8,9,15,16,22, and 23 of cycles 1 and 2. In cycles 3 and 4, dexamethasone was administered at 40 mg on days 1,2 and 20 mg on subsequent dosing days. Dexamethasone 20 mg was administered in cycles 5 and 6 cycles (consolidation cycles)		
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	CASSIOPEIA represents the primary source of efficacy and safety data for DBTd in this indication. Data reported from CASSIOPEIA are relevant to the decision problem and have been used in the model.		
Reported outcomes specified in the decision problem[a]	**Primary Endpoint:** • Proportion of patients achieving stringent complete response (sCR) after consolidation therapy **Secondary Endpoints:** • Post-consolidation complete response (CR) rate • **Post-consolidation minimal residual disease (MRD) negative rate** • Post-induction sCR rate • Post-induction MRD rate (exploratory efficacy analysis) • **PFS** • **OS** • Adverse effects (AEs) of treatment • HRQoL (patient-reported perception of global health)		
All other reported outcomes	**Other secondary endpoints:** • Post-induction overall response rate (ORR) and rate of very good partial response (VGPR) or better • Duration of CR and sCR • Time to response (CR and sCR) • TTP • Time to second objective disease progression (PFS2)		
Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; CR = complete response; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HDT = high-dose therapy; HRQoL = health-related quality of life; MM = Multiple myeloma; MRD = minimal residual disease; ORR = overall response rate; PFS = progression-free survival; PFS2 = second progression-free survival; SC = subcutaneous; sCR = stringent complete response; TTP = Time to progression; VGPR = very good partial response. [a] Bold text signifies those efficacy outcomes included in the cost-utility analysis. The ASCT rate was not a specified clinical outcome of the CASSIOPEIA trial but was reported (see Section B.2.4.1).			

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

B.2.3.1 Study design

CASSIOPEIA was designed as a 2-part clinical study comparing DBTd with BTd in newly diagnosed MM patients who are eligible for ASCT. The study consists of three phases as follows:(44)

- Screening Phase: extends up to 28 days prior to Cycle 1, Day 1
- Treatment Phase: conducted in two parts:

- Part 1: Induction/ASCT/Consolidation phase (1:1 Randomisation). The consolidation phase of treatment began approximately 30 days after ASCT with response evaluated at Day 100 post ASCT
- Part 2: Maintenance phase (1:1 Re-randomisation of patients achieving at least a partial response [PR] after consolidation). Patients who have not achieved a response enter the Follow-up Phase and are followed until disease progression or death, even if they receive subsequent treatment

- Follow-up Phase: extends from treatment discontinuation until death, loss to follow-up, withdrawal of consent, or study end, whichever occurs first

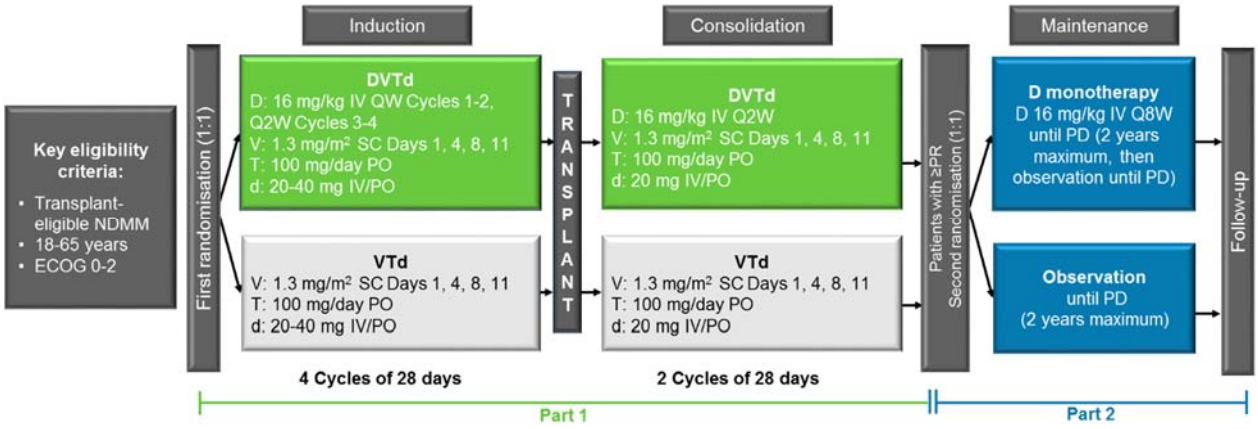
The licence for this indication covers Part 1 only (induction and consolidation phase), and data presented in this submission are from the pre-specified June 2018 data cut with additional supportive data presented from the May 2019 data cut, upon which EMA granted marketing authorisation for DBTd (see Section B.2.6). Whilst Part 2 of the study remains blinded, Janssen does not have access to individual patient-level data and is unable to perform any additional statistical analysis for Part 1 which may account for events that occur in Part 2 e.g. re-randomisation to maintenance therapy.

Patients in CASSIOPEIA were randomised 1:1 to receive either DBTd or BTd using a permuted block randomisation. The stratification factors included were as follows:(44)

- Site affiliation (Intergroupe Francophone du Myelome [IFM] or Dutch-Belgium Cooperative Trial Group for Hematology Oncology [HOVON])
- ISS staging (I, II, or III)
- Cytogenetic risks (standard risk or high risk as defined by presence of del17p or t(4;14), as centrally confirmed during screening)

An overview of the study design for CASSIOPEIA is shown in Figure 4 and the key study characteristics are presented in Table 4.

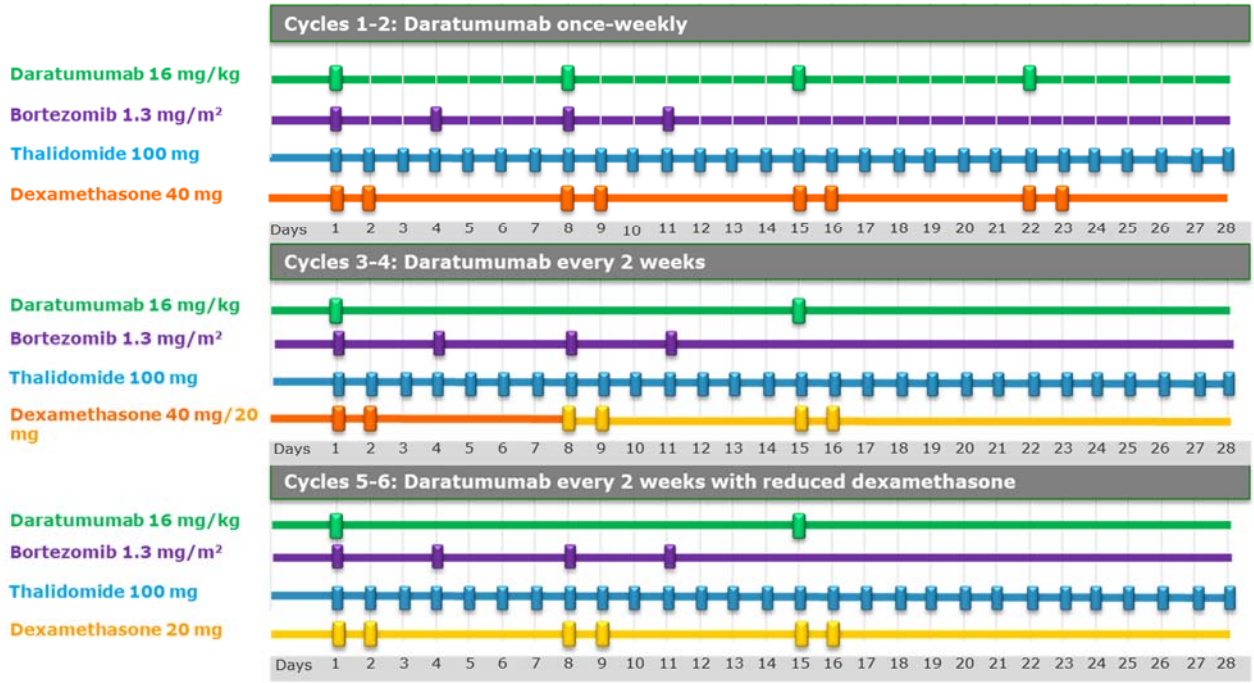
Figure 4: Overview of the CASSIOPEIA Study Design(44)



Key: D = daratumumab; d = dexamethasone; DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ECOG = Eastern Cooperative Oncology Group; IV = intravenous; NDMM = newly diagnosed multiple myeloma; QW = weekly; Q2W = every 2 weeks; SC = subcutaneous; PO = *per os* (oral); Q8W = every 8 weeks; PD = progressive disease; T = thalidomide; V = bortezomib (referred to as B throughout the submission); VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

A schematic representation of the dosing schedule is provided in Figure 5 with further details described in Table 4. The Treatment Phase for Part 1 consisted of up to a maximum of six 28-day (4-week) cycles, split between four induction cycles and two consolidation cycles. Patients were treated for the allowed maximal treatment period or until disease progression or unacceptable toxicity.

Figure 5: Overview of CASSIOPEIA dosing schedule



Note: Cycle duration was 4 weeks (28 days).

Table 4: Key Study Characteristics for CASSIOPEIA(44, 46)

CASSIOPEIA	
Location	Patients were treated across 111 European sites including: France (70), Belgium (13), and the Netherlands (28).
Trial design	Randomised, open-label, active-controlled, parallel-group, multicentre, phase III study to investigate the efficacy and safety of DBTd in patients with previously untreated MM eligible for ASCT. The 'Treatment Phase' was conducted in two parts with Part 1 covering the induction/ASCT/consolidation phase.
Method of allocation	In Part 1, patients were randomly assigned (1:1) to DBTd or BTd using a permuted block randomisation. Stratification factors included site affiliation (IFM) or (HOVON), ISS disease stage (I, II, III) and cytogenetic risk status (presence [high risk] or absence [standard risk] of del17p or t[4;14] cytogenetic abnormalities).
Key inclusion criteria	• Patients aged between 18 and 65 years. • Patients with documented MM satisfying the CRAB or biomarkers of malignancy criteria and measurable disease defined by: • Monoclonal plasma cells in the bone marrow ≥10% or presence of a biopsy proven plasmacytoma AND any one or more of the following myeloma defining events: o Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than upper limit of normal (ULN) or >2.75 mmol/L (>11 mg/dL) o Renal insufficiency: creatinine clearance <40 mL/min or serum creatinine >177 µmol/L (>2 mg/dL) o Anaemia: haemoglobin >2 g/dL below the lower limit of normal or haemoglobin <10 g/dL o Bone lesions: one or more osteolytic lesions on skeletal radiography, computed tomography (CT), or positron emission tomography CT (PET-CT) o Clonal bone marrow plasma cell percentage ≥60% o Involved: uninvolved serum free light chain ratio ≥100

	○ >1 focal lesion on (magnetic resonance imaging) MRI studies • Measurable disease as defined by any of the following: ○ IgG MM: Serum monoclonal paraprotein (M-protein) level ≥200 mg/24 hours; or ○ IgA, IgE, IgD, or IgM MM: serum M-protein level ≥0.5 g/dL or urine M-protein level ≥200 mg/24 hours; or ○ IgD MM: serum M-protein level <0.5 g/dL and Serum immunoglobulin free light chain ≥10 mg/dL and abnormal serum immunoglobulin kappa lambda free light chain ratio; or ○ Light chain MM without measurable disease in the serum or the urine: Serum immunoglobulin free light chain ≥10 mg/dL and abnormal serum immunoglobulin kappa lambda free light chain ratio • Newly diagnosed patients eligible for high dose therapy and ASCT. • Patients with an Eastern Cooperative Oncology Group (ECOG) performance status score of 0, 1, or 2. • Patients must have pre-treatment clinical laboratory values meeting the following criteria during the Screening Phase: ○ Haemoglobin ≥7.5 g/dL (≥5 mmol/L; prior red blood cell [RBC] transfusion or recombinant human erythropoietin use permitted) ○ Absolute neutrophil count (ANC) ≥1.0 x 10^9/L (GCSF use permitted) ○ AST ≤2.5 x ULN ○ Total bilirubin ≤1.5 x ULN (except in subjects with congenital bilirubinemia, such as Gilbert syndrome, direct bilirubin ≤1.5 x ULN); ○ Calculated creatinine clearance ≥40 mL/min/1.73 m^2 ○ Corrected serum calcium ≤14 mg/dL (<3.5 mmol/L); or free ionized calcium ≤6.5 mg/dL (≤1.6 mmol/L) ○ Platelet count ≥70 x 10^9/L for patients in whom <50% of bone marrow nucleated cells are plasma cells; otherwise platelet count >50x10^9/L (transfusions were not permitted to achieve this minimum platelet count) • Women who are partners of men and of childbearing potential must commit to either absolute and continuous abstinence confirmed to her physician on a monthly basis or practice one of the advised methods of birth control. Contraception must begin 4 weeks before start of therapy. • Woman of childbearing potential must have 2 negative serum or urine pregnancy tests at Screening, first within 10 to 14 days prior to dosing and the second within 24 hours prior to dosing.
Key exclusion criteria	• Patient has received daratumumab or other anti-CD38 therapy previously. • Patient has a diagnosis of primary amyloidosis, monoclonal gammopathy of undetermined significance, smouldering MM, or solitary plasmacytoma. • Patient has a diagnosis of Waldenstrom's macroglobulinemia, or other conditions in which IgM M-protein is present in the absence of a clonal plasma cell infiltration with lytic bone lesions. • Patient has prior or current systemic therapy of SCT for any plasma cell dyscrasia, with the exception of an emergency use of a short course (equivalent to dexamethasone 40 mg/day for a maximum 4 days) of corticosteroids before treatment. • Patient has peripheral neuropathy or neuropathic pain Grade 2 or higher, as defined by the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4. • Patient has had any prior or concurrent invasive malignancy (other than MM) within 10 years of study start except adequately treated basal cell or squamous cell carcinoma of the skin, carcinoma in situ of the cervix, localised prostate adenocarcinoma diagnosed ≥3 years and without evidence of biochemical failure, or other cancer for which the subject has

	undergone potentially curative therapy and has no evidence of that disease for ≥10 years. • Patient has had radiation therapy within 14 days of Cycle 1, Day 1. • Patient has had plasmapheresis within 28 days of Cycle 1, Day 1. • Patient is exhibiting clinical signs of meningeal involvement of MM. • Patient has known chronic obstructive pulmonary disease (COPD) with a Forced Expiratory Volume in 1 second (FEV1) <50% of predicted normal. Note that FEV1 testing is required for patients suspected of having COPD and patients must be excluded if FEV1 <50% of predicted normal. • Patient has known moderate or severe persistent asthma within the past 2 years, or currently has uncontrolled asthma of any classification. • Patient is known to be seropositive for history of human immunodeficiency virus (HIV) or known to have active hepatitis B or hepatitis C. • Patient has any concurrent medical or psychiatric condition or disease (e.g. active systemic infection, uncontrolled diabetes, acute diffuse infiltrative pulmonary disease) that is likely to interfere with the study procedures or results, or that in the opinion of the investigator, would constitute a hazard for participating in this study. • Patient has clinically significant cardiac disease, including: o Myocardial infarction within 1 year before randomisation, or an unstable or uncontrolled disease/condition related to or affecting cardiac function o Uncontrolled cardiac arrhythmia or clinically significant electrocardiogram (ECG) abnormalities o Screening 12-lead ECG showing a baseline QT interval as corrected by Fridericia's formula (QTcF) >470 msec • Patient has known allergies, hypersensitivity, or intolerance to boron or mannitol, corticosteroids, monoclonal antibodies or human proteins, or their excipients, or known sensitivity to mammalian-derived products. Or patient has known hypersensitivity to thalidomide. • Patient has plasma cell leukaemia or POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, and skin changes). • Patient is known or suspected of not being able to comply with the study protocol. • Patient is a woman who is pregnant, or breast-feeding, or planning to become pregnant while enrolled in this study or within 4 months after the last dose of any component of the treatment regimen. Or, subject is a man who plans to father a child while enrolled in this study or within 4 months after the last dose of any component of the treatment regimen. • Patient has had major surgery within 2 weeks before randomisation or will not have fully recovered from surgery, or has surgery planned during the time the patient is expected to participate in the study. Kyphoplasty or Vertebroplasty are not considered major surgery. • Patient has received an investigational drug or used an invasive investigational medical device within 4 weeks before randomisation or is currently enrolled in an interventional investigational study. • Patient has contraindications to the use of any components of the backbone treatment regimens, per local prescribing information. • Incidence of gastrointestinal disease that may significantly alter the absorption of oral drugs. • Patients unable or unwilling to undergo antithrombotic prophylactic treatment.
Study drugs (refer to Figure 5)	**In the DBTd arm:** • Daratumumab (16 mg/kg) was administered by IV infusion weekly on days 1,8,15 and 22 for two 28-day cycles, then every 2 weeks for the remaining induction and consolidation cycles based on treatment assignment. **In both the DBTd and BTd arms:**

	• Bortezomib was administered SC at a dose of 1.3 mg/m^2 twice a week (Days 1, 4, 8 and 11) for four 28-day induction cycles (Cycles 1 to 4), and two consolidation cycles (Cycles 5 and 6), with an option to change the schedule from twice weekly to once weekly, should toxicity be experienced. Cycles remained 28 days regardless of injection interval. On treatment days when both bortezomib and daratumumab were administered, bortezomib was administered after the end of the daratumumab infusion. • Thalidomide was administered orally at 100 mg daily for four 28-day induction cycles and two 28-day consolidation cycles. • Dexamethasone was administered at 40 mg on days 1,2,8,9,15,16,22,23 of cycles 1 and 2. In cycles 3 and 4, dexamethasone was administered at 40 mg on days 1,2 and 20 mg on subsequent dosing days (8,9,15,16). Dexamethasone 20 mg was administered on days 1,2,8,9,15,16 of cycles 5 and 6.
Permitted and disallowed concomitant medications	Throughout the study, investigators could prescribe any concomitant medications or treatments deemed necessary to provide adequate supportive care except for those listed as prohibitive therapies. Prohibitive therapies included: • Other antineoplastic therapy for treating MM, including medications that target CD38 • Continuation of study treatment during or after emergency orthopaedic surgery or radiotherapy • Investigational agents including agents with activity against or under investigation for MM, including systemic corticosteroids • Nonsteroidal anti-inflammatory agents
Primary outcome	Post-consolidation sCR rate: assessed by computer algorithm and defined as the percentage of patients achieving CR in addition to having a normal serum free light chain (FLC) ratio and an absence of clonal cells in bone marrow by immunohistochemistry, immunofluorescence or 2- to 4-color flow cytometry. Post-consolidation response was assessed at Day 100 post-ASCT.
Secondary outcomes	Major secondary endpoints for Part 1 included: • PFS: defined as the time from the initial randomisation to either confirmed PD per the International Myeloma Working Group criteria or death, whichever comes first • TTP: defined as the time from the initial randomisation to confirmed PD per the IMWG criteria, or death due to progressive disease, whichever occurs first • CR rate: defined as the proportion of patients who achieved CR or better by the end of consolidation assessed by computer algorithm in accordance with IMWG criteria • MRD-negative rate: defined as the proportion of patients who achieved MRD-negative status by the end of consolidation assessed by computer algorithm in accordance with IMWG criteria • Post-induction sCR rate: defined as the proportion of patients who achieved sCR prior to high-dose therapy/ASCT assessed by computer algorithm in accordance with IMWG criteria • PFS2: defined as the time from initial randomisation to time of subsequent progression on next-line of therapy after disease progression on study treatment • OS: measured from the date of initial randomisation to the date of patient's death. If the patient is alive or vital status is unknown, then the patient's data was censored at the date the patient was last known to be alive Other secondary endpoints included: • Post-induction ORR and rate of VGPR or better: defined as the proportion of patients who have achieved PR or better by the end of induction assessed by computer algorithm in accordance with IMWG criteria • Duration of CR and sCR: calculated from the date of the initial

	documentation of a CR or sCR to the date of the first documented evidence of relapse of CR or disease progression, assessed by computer algorithm in accordance with IMWG criteria, whichever occurs first • HRQoL (patient-reported perception of global health)
Pre-specified subgroups	• Sex (male, female) • Age (<50 years, ≥50 years) • Site (IFM, Hovon) • ISS staging (I, II, III) • Cytogenetic risk (high risk, standard risk) • Baseline renal function (CrCl) (>90 mL/min, ≤90 mL/min) • Baseline hepatic function (normal, impaired) • Type of MM (IgG, non-IgG) • ECOG performance score (0, ≥1)
Efficacy and safety evaluations	Efficacy evaluations included measurements of tumour burden/residual disease, myeloma proteins, bone marrow examinations, skeletal surveys, assessment of extramedullary plasmacytomas and serum calcium corrected for albumin. Disease evaluations were required to be performed as outlined in the Time and Events Schedules on the scheduled assessment day (±3 days) (refer to Appendix L). Disease evaluations scheduled for treatment days were collected before the study drug was administered. Disease evaluations were mainly performed by a central laboratory. Blood samples for calculating serum calcium corrected for albumin, and bone marrow examination for clinical staging were, for example, performed locally. Disease response was assessed based on IMWG consensus recommendations for MM treatment response criteria. For quantitative immunoglobulin at baseline, M-protein, and immunofixation measurements in serum and 24-hour urine, the investigator used results provided by the central laboratory. Patients believed to have attained a sCR had this confirmed centrally by a minimum of 4 colour flow cytometry, requiring a fresh bone marrow aspirate. All response categories (CR, sCR, VGPR, PR and PD) require 2 consecutive assessments made at any time before the institution of any new therapy; CR, sCR, VGPR, PR, and SD categories also required no known evidence of progressive or new bone lesions if radiographic studies were performed. Disease progression was based on assessments from IMWG Guidelines. For continuation of treatment, the IMWG response was determined on an ongoing basis by the investigator. For data analysis and reporting, however, the study team used a validated computer algorithm to provide consistent review of the data necessary to determine disease progression and response according to the IMWG criteria. Safety evaluations included AE monitoring, physical examinations, ECGs monitoring, clinical laboratory parameters (haematology and chemistry), vital sign measurements (pulse, blood pressure and temperature), and ECOG performance status.

Key: AE = Adverse event; ANC = absolute neutrophil count; ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; COPD = chronic obstructive pulmonary disease; CR = complete response; CT = computed tomography; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ECG = electrocardiogram; ECOG = Eastern Cooperative Oncology Group; FEV = Forced Expiratory Volume; FLC = free light chains; GCSF = granulocyte-colony stimulating factor; HIV = human immunodeficiency virus; HOVON = Dutch-Belgium Cooperative Trial Group for Hematology Oncology; HRQoL = health-related quality of life; IFM = Intergroupe Francophone du Myelome; IMWG = International Myeloma Working Group; ISS = International Staging System; IV = intravenous; MM = multiple myeloma; MRD = minimal residual disease; MRI = magnetic resonance imaging; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events; ORR = overall response rate; OS = overall survival; PD = progressive disease; PET = positron emission tomography; PFS = progression-free survival; PO = *per os* (oral); PR = partial response; RBC = red blood cell; SC = subcutaneously; sCR = stringent complete response; ULN = upper limit of normal; VGPR = very good partial response.

Evaluation of response and disease progression

Assessment of response and disease progression was performed by a central laboratory (as opposed to local laboratories), and a validated computerised algorithm was used in line with the IMWG criteria of response.(47, 48) This approach allows for a stricter, registration-quality rigour and objective evaluation of response as opposed to investigator assessments which is based on clinical judgement. However, this more stringent evaluation method can also result in higher reported rates of VGPR relative to CR. For example, if there was missing data related to CR response criteria or an inconclusive assessment leading to an inability to declare a CR, a VGPR was declared. As a sensitivity analysis, additional investigator assessments of response and disease progression per the IMWG response criteria were performed.[I]

Refer to Appendix L for further details regarding evaluation of response and disease progression in CASSIOPEIA, including the response criteria used for the primary efficacy assessment.

Rational for sCR as a primary endpoint

Survival outcomes in NDTE MM have improved considerably over time as new treatments have become available, with 4-year survival rates exceeding 80%.(49, 50) Therefore, it is increasingly difficult to demonstrate a significant improvement in OS in this patient population over the short duration of a clinical trial. The level of tumour burden reduction has been demonstrated to be a useful measure for predicting long-term survival outcomes in MM.(51-56)

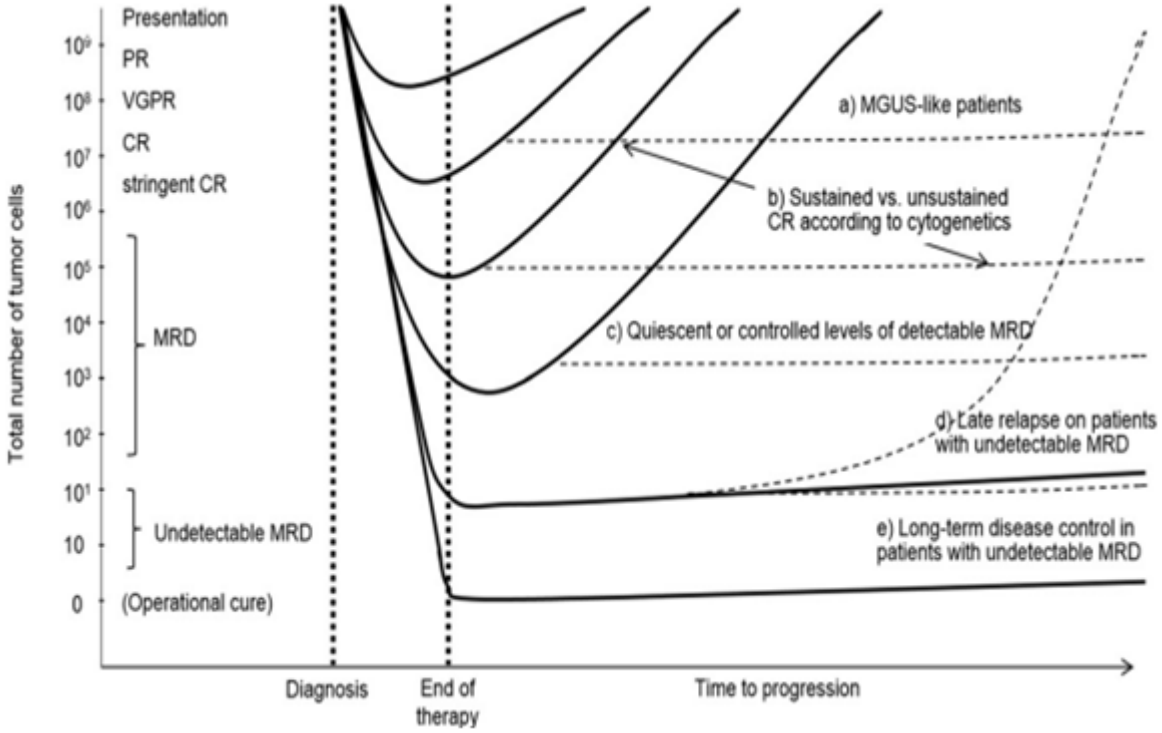
Since the introduction of effective triplet therapies, such as BTd, most NDTE MM patients are able to achieve VGPR or CR however all patients eventually relapse.(21, 38) In order to measure deeper levels of response, including the possibility of complete cancer cell eradication, more stringent definitions of response are required. In 2006, the IMWG introduced sCR as a new stringent measure of response in MM, reflecting a deeper level of response than previous definitions.(48) The achievement of sCR in patients with NDMM has been shown to strongly correlate with improved PFS and OS.(57, 58) Therefore, sCR, as defined by IMWG uniform response criteria(59), allows the detection of response beyond the CR level and predicts long-term survival, representing a useful and meaningful endpoint in clinical trials in NDTE patients. In CASSIOPEIA, the primary endpoint of sCR was assessed post-consolidation, allowing the efficacy of induction and consolidation treatment to be measured, without including the effect of maintenance treatment.

Minimal residual disease

As discussed in Section B.1.3.1, despite improvements in treatment, all patients eventually relapse. Relapse is due to some cancerous cells that resist treatment and undergo clonal expansion and evolution, resulting in tumour repopulation in a patient. This population of remaining cells that contribute to relapse is known as “minimal residual disease” (MRD). The state of “MRD negativity” is one where no remaining clonal or sub-clonal cancerous cells can be detected using currently available measurement techniques, and therefore relapse is less likely with long-term disease control achieved for some patients (Figure 6).

[I] Similar with the centralised assessment, results from CASSIOPEIA based on investigator assessments of response are not comparable to other MM studies as they were quality controlled for agreement with the response category determined by the centralised lab, as opposed to local labs.

Figure 6: Representation of depth of response



Key: CR = complete remission; MGUS = monoclonal gammopathy of undetermined significance; MRD = minimal residual disease; PR = partial response; VGPR = very good partial response.
Filled lines illustrate the paradigm for the positive correlation between increasing depth of response and increasing progression-free survival. Dotted lines indicate distinct biological subgroups that differ from the paradigm: (a) patients with a baseline MGUS-like signature and prolonged survival irrespectively of CR; (b) patients with unsustained CR (high-risk cytogenetics and persistent MRD); (c) MRD-positive patients who may also experience extended outcomes if small residual clones are quiescent (MGUS-like) or under control (e.g., by immune cells); (d) an MRD-negative result does not preclude the risk of relapse, and optimization of MRD monitoring together with follow-up MRD studies are likely crucial to predict relapses early on; (e) long-term disease control (i.e., functional cure) could potentially be achieved if therapy eradicates detectable MRD levels.
Source: Paiva et al. (2015).(60)

In addition to traditional assessment of response, IMWG guidelines now recommend consideration of MRD after each treatment stage in patients with a CR. MRD is a new, more sensitive measure of disease compared with established definitions of clinical response in MM, where residual tumour cells are identified in the bone marrow based on the IMWG criteria described in Table 5.(61-63) Within CASSIOPEIA, MRD post-consolidation was assessed as a key secondary endpoint for all patients in Part 1. MRD was primarily evaluated in CASSIOPEIA by EuroFlow-based multiparametric flow cytometry (MFC) and additionally with next-generation sequencing (NGS) of bone marrow aspirates.

Table 5: IMWG criteria for MRD(64)

Response subcategory	Response criteria
Sustained MRD-negative	MRD negativity in the bone marrow confirmed ≥1 year apart by NGF, NGS, or both and by imaging (see flow MRD-negative category)
Flow MRD-negative	Absence of phenotypically aberrant clonal plasma cells by NGF on bone marrow aspirates using EuroFlow (or validated equivalent method) with a minimum sensitivity of 1 in 10^5 nucleated cells or higher
Sequencing MRD-negative	Absence of clonal plasma cells by NGS on bone marrow aspirate Presence of a clone is defined as <2 identical sequencing reads from bone marrow aspirates using the LymphoSIGHT platform (or validated equivalent method) with a minimum sensitivity of 1 in 10^5 nucleated cells or higher
Imaging-positive MRD-negative	MRD negativity as defined by NGF or NGS, plus at least one of the following criteria: • Disappearance of every area of increased tracer uptake found at baseline or a preceding PET/CT • Decrease to less mediastinal blood pool SUV • Decrease to less than that of surrounding normal tissue

Key: CT = computed tomography; IMWG = International Myeloma Working Group; MRD = minimal residual disease; NGF = next generation flow; NGS = next generation sequencing; PET = positron emission tomography; SUV = standardised uptake value.
These criteria are based on those used by Zamagni and colleagues and expert panel (IMPetUs; Italian Myeloma criteria for PET Use). Baseline positive lesions were identified by presence of focal areas of increased uptake within bones, with or without any underlying lesion identified by CT and present on ≥2 consecutive slices. Alternatively, SUVmax=2.5 within osteolytic CT areas >1 cm in size, or SUVmax=1.5 within osteolytic CT areas ≤1 cm in size were considered positive. Imaging should be performed once MRD negativity is determined by multiparameter flow cytometry or NGS.

B.2.3.2 Baseline characteristics of trial participants

Baseline demographic and disease characteristics were well balanced between the treatment groups (Table 6). The median age of patients in the study was 58.0 years (range 22-65), with 84.1% of patients being 50 years of age or older.(43, 44) Baseline ECOG scores of 0 or 1 were reported for 90.0% of patients.(43) The majority of patients had serum measurable disease in IgG (59.4%) and IgA (16.5%).(44) One-hundred sixty-eight (15.5%) of patients had a high-risk cytogenetic abnormality.(43) ISS staging was 39.8%, 45.0% and 15.2% for stage I, II and III respectively, with a numerically higher proportion of patients classified as Stage II in the DBTd (47.0%) arm compared with the BTd arm (43.0%).(43, 44)

Table 6: Baseline patient demographics and disease characteristics in CASSIOPEIA (ITT population)(43, 44)

Characteristic	BTd (n=542)	DBTd (n=543)	Total (n=1,085)
Sex (Female), n (%)	223 (41.1%)	227 (41.8%)	450 (41.5%)
Age, years, n (%)			
<50	90 (16.6%)	83 (15.3%)	173 (15.9%)
≥50-65	452 (83.4%)	460 (84.7%)	912 (84.1%)
Mean (SD)	56.7 (7.03)	56.8 (6.93)	56.6 (6.98)
Median	58.0	59.0	58.0
Range	(26; 65)	(22; 65)	(22; 65)
Weight (kg), n (%)			
<50	13 (2.4%)	10 (1.8%)	23 (2.1%)
50-64	123 (22.7%)	131 (24.1%)	254 (23.4%)
65-85	268 (49.4%)	270 (49.7%)	538 (49.6%)
>85	138 (25.5%)	132 (24.3%)	270 (24.9%)
Mean (SD)	75.83 (15.605)	75.52 (15.632)	75.67 (15.612)
Median	75.00	74.00	74.50
Range	(44.0; 142.5)	(46.0; 135.0)	(44.0; 142.5)
Height (cm)			
Mean (SD)	170.2 (9.66)	169.9 (10.02)	170.0 (9.84)
Median	170.0	170.0	170.0
Range	(146; 201)	(143; 201)	(143; 201)
Body surface area (m^2)			
Mean (SD)	1.886 (0.2298)	1.880 (0.2258)	1.883 (0.2277)
Median	1.880	1.870	1.870
Range	(1.39; 2.71)	(1.40; 2.61)	(1.39;2.71)
Baseline ECOG score, n (%)			
0	257 (47.4%)	265 (48.8%)	522 (48.1%)
1	230 (42.4%)	225 (41.4%)	455 (41.9%)
2	55 (10.1%)	53 (9.8%)	108 (10.0%)
Type of myeloma by immunofixation, n (%)			
IgG	333 (61.4%)	351 (64.6%)	684 (63.0%)
IgA	104 (19.2%)	87 (16.0%)	191 (17.6%)
IgM	2 (0.4%)	1 (0.2%)	3 (0.3%)
IgD	13 (2.4%)	5 (0.9%)	18 (1.7%)
Light chain	66 (12.2%)	83 (15.3%)	149 (13.7%)

Kappa	46 (8.5%)	53 (9.8%)	99 (9.1%)
Lambda	20 (3.7%)	30 (5.5%)	50 (4.6%)
Biclonal	19 (3.5%)	12 (2.2%)	31 (2.9%)
Negative immunofixation	5 (0.9%)	4 (0.7%)	9 (0.8%)
Type of measurable disease[a], n (%)			
IgG	314	331	645
IgA	99	80	179
Other[b]	22	13	35
Urine only	67	70	137
Serum FLC only	40	48	88
Unknown	0	1	1
ISS staging, n (%)			
I	228 (42.1%)	204 (37.6%)	432 (39.8%)
II	233 (43.0%)	255 (47.0%)	488 (45.0%)
III	81 (14.9%)	84 (15.5%)	165 (15.2%)
Time since initial diagnosis to randomisation (months)			
Mean (SD)	1.37 (2.184)	1.33 (2.984)	1.35 (2.614)
Median	0.95	0.92	0.92
Range	(0.2; 31.0)	(0.2; 66.6)	(0.2; 66.6)[f]
Number of lytic bone lesions, n (%)			
None	86 (15.9%)	81 (15.0%)	167 (15.5%)
1-3	153 (28.3%)	176 (32.6%)	329 (30.5%)
4-6	110 (20.4%)	98 (18.1%)	208 (19.3%)
>7	191 (35.4%)	185 (34.3%)	376 (34.8%)
Presence of diffuse myeloma-related osteopenia, n (%)			
Yes	49 (9.1%)	53 (9.8%)	102 (9.4%)
No	491 (90.9%)	487 (90.2%)	978 (90.6%)
Presence of extramedullary plasmacytomas, n (%)			
Yes	2 (0.4%)	8 (1.5%)	10 (0.9%)
No	540 (99.6%)	535 (98.5%)	1,075 (99.1%)
Presence of evaluable bone marrow assessment, n (%)			
Yes	533 (98.3%)	533 (98.2%)	1,066 (98.2%)
No	9 (1.7%)	10 (1.8%)	19 (1.8%)
% Plasma cells, bone marrow biopsy/aspirate, n (%)			
<10	17 (3.2%)	20 (3.8%)	37 (3.5%)
10-30	249 (46.7%)	245 (46.0%)	494 (46.3%)

>30	267 (50.1%)	268 (50.3%)	535 (50.2%)
% Plasma cells, bone marrow biopsy, n (%)			
<10	2 (2.3%)	3 (3.1%)	5 (2.7%)
10-30	35 (40.7%)	32 (33.0%)	67 (36.6%)
>30	49 (57.0%)	62 (63.9%)	111 (60.7%)
% Plasma cells, bone marrow aspirate, n (%)			
<10	37 (7.1%)	43 (8.2%)	80 (7.7%)
10-30	242 (46.3%)	250 (47.9%)	492 (47.1%)
>30	244 (46.7%)	229 (43.9%)	473 (45.3%)
Bone marrow cellularity, n (%)			
Hypercellular	155 (29.0%)	136 (25.6%)	291 (27.3%)
Normocellular	223 (41.8%)	244 (46.0%)	467 (43.8%)
Moderately cellular	116 (21.7%)	107 (20.2%)	223 (20.9%)
Severely acellular	23 (4.3%)	27 (5.1%)	50 (4.7%)
Indeterminate	17 (3.2%)	17 (3.2%)	34 (3.2%)
Cytogenetics profile			
T(4; 14)			
N[c]	503	501	1,004
Normal	450 (89.5%)	450 (89.8%)	900 (89.6%)
Abnormal	53 (10.5%)	51 (10.2%)	104 (10.4%)
Del17p			
N[d]	503	501	1,004
Normal	464 (92.2%)	459 (91.6%)	923 (91.9%)
Abnormal	39 (7.8%)	42 (8.4%)	81 (8.1%)
Risk result			
N[e]	540	542	1,082
High risk	86 (15.9%)	82 (15.1%)	168 (15.5%)
Standard risk	454 (84.1%)	460 (84.9%)	914 (84.5%)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; FLC = free light chains; ECOG = Eastern Cooperative Oncology Group; ISS = International Staging System; SD = standard deviation.
Note: Percentages are calculated with the number of patients in each group with available data as denominator.
[a] Includes patients without measurable disease in serum and urine.
[b] Includes IgD, IgM, IgE and biclonal.
[c] Subjects with t(4; 14) measured (normal or abnormal).
[d] Patients with Del17p measured (normal or abnormal).
[e] Includes patients with risk results available.
[f] Incorrect "time to initial diagnosis" data were entered into the database for 4 patients. These data errors did not affect the median reported in this analysis.

The stratification factors for randomisation described in Section B.2.3.1 were well-balanced between the two treatment groups. After initiation of the study, a revised ISS (R-ISS) was published. In addition to albumin and β-2-microglobulin, the R-ISS uses additional information consisting of lactate dehydrogenase (LDH) and cytogenetic risk. Based on the dataset, a post-hoc calculation was performed to build the revised ISS. In the post-hoc revised ISS, more patients were classified as Stage II in the DBTd (71.6%) arm compared with the BTd arm (63.7%) indicating a poorer prognosis for patients treated with daratumumab.(44, 65)

Table 7: Summary of IMWG Revised ISS Staging in MM (ITT population)(44)

	BTd (n, %)	DBTd (n, %)	Total (n, %)
Analysis set: ITT	542	543	1,085
IMWG Revised ISS Staging[a]			
N	540	535	1,075
I	146 (27.0%)	103 (19.3%)	249 (23.2%)
II	344 (63.7%)	383 (71.6%)	727 (67.6%)
III	50 (9.3%)	49 (9.2%)	99 (9.2%)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; IMWG = International Myeloma Working Group; ISS = International Staging System; ITT = intention-to-treat.
Note: Percentages are calculated with the number of patients in each group with available data as denominator.
[a] Determination is based on three factors: International Staging System; presence of chromosomal abnormalities of t(4; 14), or del17p by FISH testing and serum LDH at Pre-induction Baseline.

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

B.2.4.1 Study population and patient disposition

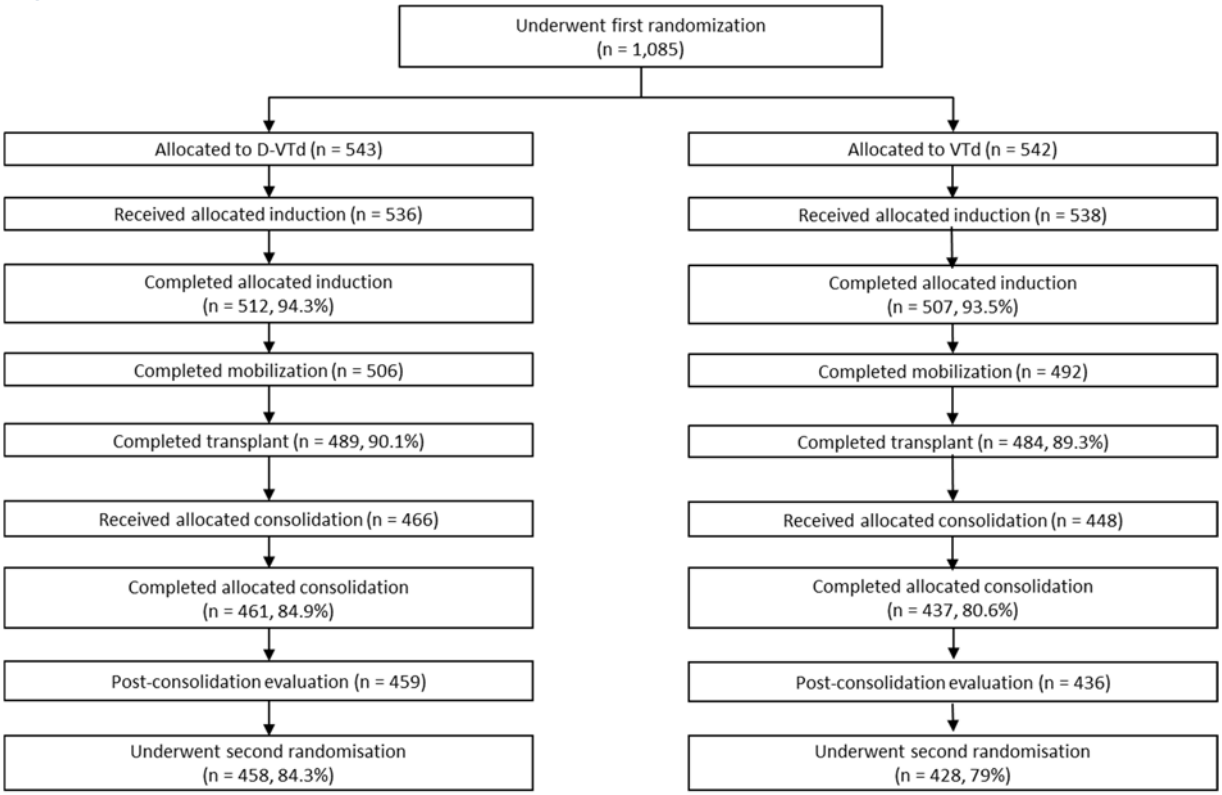
A total of 1,085 patients (DBTd: 543; BTd: 542) were randomised between 22 September 2015 and 1 August 2017 at 111 European sites across France (70), Belgium (13) and the Netherlands (28).(43, 44) The patient flow is shown in Figure 7.

As of the clinical cut-off date for the Primary Analysis for Part 1 (PA1, June 2018), 536 patients in the DBTd group and 538 patients in BTd groups were treated (98.7% and 99.3% of the total number of patients randomised in each group, respectively).(43) Among these, 461 patients (84.9%) in the DBTd group and 437 patients (80.6%) in the BTd group had completed all four cycles of induction treatment and both cycles of consolidation treatment.(43) In the DBTd group, 489 patients (90.1%) had undergone ASCT, compared with 484 patients (89.3%) in the BTd group.(43)

Among patients who were randomised, disease progression (1.3% in the DBTd group, 0.7% in BTd group) and unacceptable and/or severe adverse events (AEs) (1.7% in DBTd; 0.4% in BTd group) were the most common reasons for not proceeding to the transplant stage after receiving induction treatment and stem cell mobilisation.(44)

Refer to Section B.2.10 for discussion on safety outcomes from the CASSIOPEIA study.

Figure 7: Participant flow in CASSIOPEIA(44)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

The study populations used for the analysis of outcomes from the CASSIOPEIA trial are presented in Table 8.

Table 8: Summary of data sets analysed(44)

Study population	Description	DBTd, N	BTd, N
Intention-to-treat (ITT) analysis set	Included all randomised patients	543	542
Safety analysis set	Included all randomised patients who received at least one dose of study drug and contributed any safety data after the start of study treatment	536	538
Response-evaluable analysis set	Included patients who have a confirmed diagnosis of MM and measurable disease at baseline or screening visit. In addition, patients must have received at least one component of study treatment and have adequate post-baseline disease assessments to allow for the assessment of disease	536	535

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention-to-treat; MM = multiple myeloma.

B.2.4.2 Statistical analyses

Details of the statistical methods for the Primary Analysis for Part 1 of CASSIOPEIA are presented in Table 9.

Table 9: Statistical methods for the Primary Analysis for Part 1 of CASSIOPEIA(44, 46)

CASSIOPEIA	
Hypothesis Objective	The primary efficacy analysis was performed by testing the null hypothesis that there was no difference in the post-consolidation sCR rate between DBTd and BTd in patients with newly diagnosed MM who are eligible for ASCT.
Statistical analysis	The primary efficacy endpoint for Part 1 was the post-consolidation sCR rate (i.e. 100 days post-ASCT). Comparisons between the DBTd and BTd arms with respect to the post-consolidation sCR rate were made using the Cochran-Mantel-Haenszel chi-square test. A Mantel-Haenszel odds ratio, along with corresponding 2-sided 95% confidence intervals and the p-value from the Cochran-Mantel-Haenszel test were calculated. Analysis of primary and secondary efficacy variables were based on the intention-to-treat (ITT) population, which included all patients randomised in the first randomisation. All safety analysis was based on the safety analysis set. The safety population for Part 1 was defined as all patients randomised in the first randomisation who received at least 1 dose of study drug. A separate Type 1 error rate (alpha) was assigned at the level of 0.05 for each part of the study reflecting the 2 distinct hypotheses of interest for Part 1 (induction/consolidation) and Part 2 (maintenance). No interim analysis was planned for the primary endpoint for Part 1. The alpha level for each endpoint was 0.05 (2-sided). The alpha spending was performed to strongly control the overall Type 1 error rate. For PFS, an alpha level of 0.0001 was assigned for the pre-specified Primary Analysis for Part 1, and 0.0499 for the final Part 1 PFS analysis (planned at the same time as the interim PFS analysis for Part 2). For OS, only descriptive analysis was performed for the pre-specified Primary Analysis for Part 1, and an alpha level of 0.0005 and 0.0495 was assigned to the interim and final OS from first randomisation analyses, respectively. The interim OS from first randomisation analyses will occur at the same time as the Part 2 interim analysis, and the final analysis for OS from first randomisation will occur at the same time as the Part 2 final analysis. For key secondary endpoints, a pre-specified hierarchical testing procedure was followed, the order of which was: post-consolidation rate of MRD negativity; post-consolidation rate of CR or better; PFS from first randomisation; and OS from first randomisation. PFS was also analysed using the Inverse Probability Weighting (IPW) method in order to adjust for the potential confounding impact of the second randomisation on PFS from first randomisation, as described in Section B.2.6.2.
Sample size, power calculation	The sample size for CASSIOPEIA took into consideration the statistical power required for the primary comparisons in both stages of the study. Taking into account the required sample size for Part 2, and assuming 75% of patients in the induction/ASCT/consolidation stage would be eligible to be randomised for maintenance, 1,080 patients (540/arm) were randomised in the first randomisation. This sample size provided at least 85% power to detect an improvement in sCR rate from 25% to 35% at a 2-sided alpha of 0.05.
Data management, patient withdrawals	Patients were withdrawn from the study for any of the following reasons: • Lost to follow-up • Withdrawal of consent for study participation • Death • Sponsor terminates the study • Screening failure Reasons for withdrawal were documented on the eCRF and source document. If a patient was lost to follow-up, the measures taken to contact the patient and

	determine the reason for discontinuation/withdrawal also had to be documented. Patients who did not achieve a response entered the Follow-up Phase and were followed until disease progression or death, even if they received subsequent treatment. Patients who withdrew consent from the study before disease progression were censored at the last disease assessment before withdrawal of consent. Patients lost to follow-up were censored at the least disease assessment before the patient was lost to follow-up. Patients who had not progressed and were still alive at the cut-off date for the analysis were censored at the last disease assessment. Patients without any post-baseline disease assessment were censored at randomisation.
Key**:** AE = Adverse event; ANC = absolute neutrophil count; ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; COPD = chronic obstructive pulmonary disease; CR = complete response; CT = computed tomography; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ECG = electrocardiogram; ECOG = Eastern Cooperative Oncology Group; eCRF = Electronic case report form; FEV = Forced Expiratory Volume; FLC = free light chains; GCSF = granulocyte-colony stimulating factor; HIV = human immunodeficiency virus; HOVON = Dutch-Belgium Cooperative Trial Group for Hematology Oncology; IFM = Intergroupe Francophone du Myelome; IMWG = International Myeloma Working Group; IPW = Inverse Probability Weighting; ISS = International Staging System; ITT = intention-to-treat; IV = intravenous; MM = multiple myeloma; MRD = minimal residual disease; MRI = magnetic resonance imaging; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events; ORR = overall response rate; OS = overall survival; PD = progressive disease; PET = positron emission tomography; PFS = progression-free survival; PO = *per os* (oral); PR = partial response; RBC = red blood cell; SC = subcutaneous; sCR = stringent complete response; ULN = upper limit of normal; VGPR = very good partial response.	

Summary of data cuts presented in the submission

During the regulatory process, Janssen received a Request for Supplementary Information (RSI) from EMA which resulted in an unplanned post-hoc interim analysis (median follow-up = 29.2 months). Due to the relatively short follow-up from the Primary Analysis for Part 1, this section includes results from both data cuts with the post-hoc interim analysis providing an additional 10.4 months of study follow-up. Table 10 presents a summary of the two data cuts upon which the evidence for the clinical efficacy of DBTd versus BTd is based.

Table 10: Summary of CASSIOPEIA data cuts(44, 45)

Data cut description	Median follow-up	Population included	Outcomes assessed	Rational for inclusion
Primary Analysis for Part 1 (PA1)	18.8 months (Clinical cut-off 19 June 2018)	ITT	**Primary Endpoint:** • Post-consolidation sCR **Major Secondary Endpoints:** • PFS (IPW adjusted / multi-variate analysis / unadjusted) • TTP • Post-consolidation CR rate • Post-consolidation MRD-negative rate (10^{-4} and 10^{-5})[a] • Post-induction sCR rate • PFS2 • OS **Other Secondary Endpoints:** • Post-induction ORR and rate of VGPR or better • Duration of CR and sCR • Patient reported perception of global health **Pre-specified subgroup analysis:** • sCR • PFS • MRD (10^{-5} was analysed for the entire ITT population using MFC, with 10^{-6} performed as an exploratory analysis for a selection of patients according to NGS)	Pre-specified. In the submission, PA1 provides a baseline for the longitudinal improvement observed consistently in trial outcomes at subsequent post-hoc analyses.
Post-hoc Interim Analysis	29.2 months (Clinical cut-off 1 May 2019)	ITT	• Post-consolidation MRD-negative rate in patients with post-consolidation sCR/CR or better/VGPR or better/Overall response (10^{-4} and 10^{-5}) • Best response for patient with Post-consolidation MRD negativity (10^{-5}) without CR/sCR • PFS (IPW adjusted / unadjusted)[b] • PFS2 • PFS from Post-induction MRD assessment • PFS by Post-induction MRD status (10^{-5})	Along with PA1, this post-hoc interim analysis provided the data necessary to support marketing authorisation by the EMA. PHA provides an additional 10.4 months of study follow-up, and gives further evidence supporting deepening response rates over time and

			• PFS by Post-induction MRD status (10^{-5}) and treatment group • OS[b] • TTP • Best response over time	sustained survival benefit.

Key: EMA = European Medicines Agency; ITT = intention-to-treat; MRD = minimal residual disease; MFC = multiparametric flow cytometry; NGS = Next generation sequencing; PA1 = Primary Analysis for Part 1; PHA = Post-hoc Interim Analysis.
[a] MRD response rates have been implemented in the economic model (MFC 10^{-5}, refer to Section B.3)
[b] Landmark analysis of OS and PFS by post-consolidation MRD status has been implemented in the economic model (refer to Section B.3.3.2)

B.2.5 *Quality assessment of the relevant clinical effectiveness evidence*

The CASSIOPEIA trial and other relevant comparator trials were assessed for quality using the Cochrane Risk-of-Bias assessment tool.(66) The results of these quality assessments are presented in Appendix D. The overall risk of bias in the CASSIOPEIA trial was considered to be low.

A summary of the quality of the CASSIOPEIA trial is also presented in Table 11, using the criteria adapted from Systematic reviews: CRD's guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination).

Table 11: Quality assessment of the CASSIOPEIA trial

	CASSIOPEIA	
	Response	**Risk of bias**
Was randomisation carried out appropriately?	Yes, randomisation was carried out as per the pre-specified randomisation method; patients were randomised using a central IWRS	Low
Was the concealment of treatment allocation adequate?	CASSIOPEIA was open label	Low, as patients were randomised using a central IWRS
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes, demographic and baseline characteristics were well balanced between the two treatment groups with no categories having a difference of ≥10%	Low
Were the care providers, participants and outcome assessors blind to treatment allocation?	CASSIOPEIA was open label and only Janssen was blinded to the results	Low, as an IDMC reviewed the data
Were there any unexpected imbalances in drop-outs between groups?	No, of the 1,085 patients randomised (543 in the DBTd group and 542 in the BTd group), 1,074 received study treatment: 536 patients received DBTd and 538 patients received BTd	Low
Is there any evidence to suggest that the authors measured more outcomes than they reported?	None	Low
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, the ITT population was used for analysis of the primary endpoint and other time-to-event efficacy endpoints, which included all randomised patients	Low

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; IDMC = independent data monitoring committee; ITT = intention-to-treat; IWRS = Interactive web response system; RCT = randomised controlled trial.
Note: Adapted from Systematic reviews: CRD's guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination).

B.2.6 *Clinical effectiveness results of the relevant trials*

CASSIOPEIA is the first phase III trial of daratumumab in patients with newly diagnosed, transplant-eligible, MM and demonstrated that the addition of daratumumab to BTd resulted in a statistically significant and clinically meaningful improvement in depth of response post consolidation (sCR, CR or better, VGPR or better and MRD-negative rate).(43) These responses translated into a statistically significant and clinically meaningful improvement in PFS (see Section B.1.3.3).(43) As a fixed treatment duration, DBTd offers a sustained treatment-free interval which, as reported in Section B.1.3.3, combined with a longer remission is the most important positive effect of treatment and as such is highly valued by patients. Whilst the OS data remains immature, there is already a clear trend supporting the clinical benefit of DBTd compared with BTd.

Efficacy analysis, including the assessment of MRD negativity, were performed for the ITT population that included all patients that underwent the first randomisation. A summary of the key clinical efficacy results from PA1 and the Post-hoc Interim Analysis is presented in Table 12, with full discussion of each endpoint provided in the remainder of this section.

Table 12: Summary of key clinical efficacy results(43-45, 65)

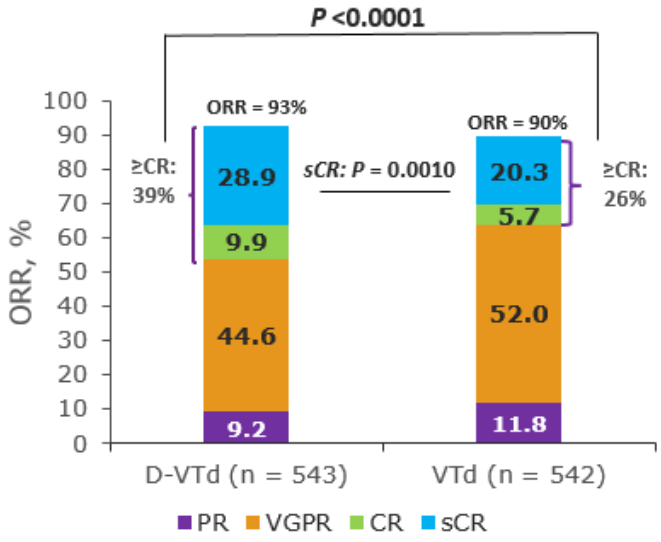
	PA1 (median follow-up = 18.8 months)		PHA (median follow-up = 29.2 months)	
	BTd	**DBTd**	**BTd**	**DBTd**
Response				
Post-consolidation sCR rate	110 (20.3%)	157 (28.9%)	n/a	n/a
Post-consolidation sCR Odds ratio (95% CI)	n/a	1.60 (1.21, 2.12) p=0.0010	n/a	n/a
MRD-negative status (10^{-5})[a]				
Post-consolidation MRD-negative rate regardless of response	236 (43.5%)	346 (63.7%)	n/a	n/a
Post-consolidation MRD-negative rate Odds ratio (95% CI)	n/a	2.27 (1.78, 2.90) p<0.0001	n/a	n/a
Survival outcomes				
PFS HR (95% CI); P-value	n/a	0.47 (0.33-0.67); p<0.0001	n/a	0.495 (0.378-0.647); p<0.0001
OS HR (95% CI); P-value	n/a	0.43 (0.23-0.80); p=0.0065	n/a	0.52 (0.33-0.85); p=0.0070
Health-related quality of life				
EORTC-CLQ-C30 GHS subscale LS mean change from baseline to 100 days post-ASCT (95% CI)	8.7 (6.5-11)	9.7 (7.4-11.9)	n/a	n/a
P-value	p=0.4523		n/a	

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EORTC-CLQ-C30 = European Organisation for Research and Treatment of Cancer-Quality of Life Questionnaire-C30; GHS = global health status; HR = hazard ratio; LS = least-squares; MRD = minimal residual disease; n/a = not applicable; OS = overall survival; PA1 = Primary Analysis for Part 1; PHA = Post-hoc Interim Analysis; PFS = progression-free survival; sCR = stringent complete response.
[a] Post-consolidation MRD-negative rate was measured by a standardised Euroflow based multiparametric assay.

B.2.6.1 Response analyses

Treatment with DBTd was associated with a statistically significant and clinically meaningful improvement in the rate of post-consolidation sCR (primary endpoint) compared with BTd alone (28.9% vs 20.3%; OR: 1.60; 95% CI: 1.21, 2.12; p=0.0010; Figure 8).(43)

Figure 8: Summary of post-consolidation* response rates based on computerised algorithm (ITT population, median follow-up = 18.8 months)(44)

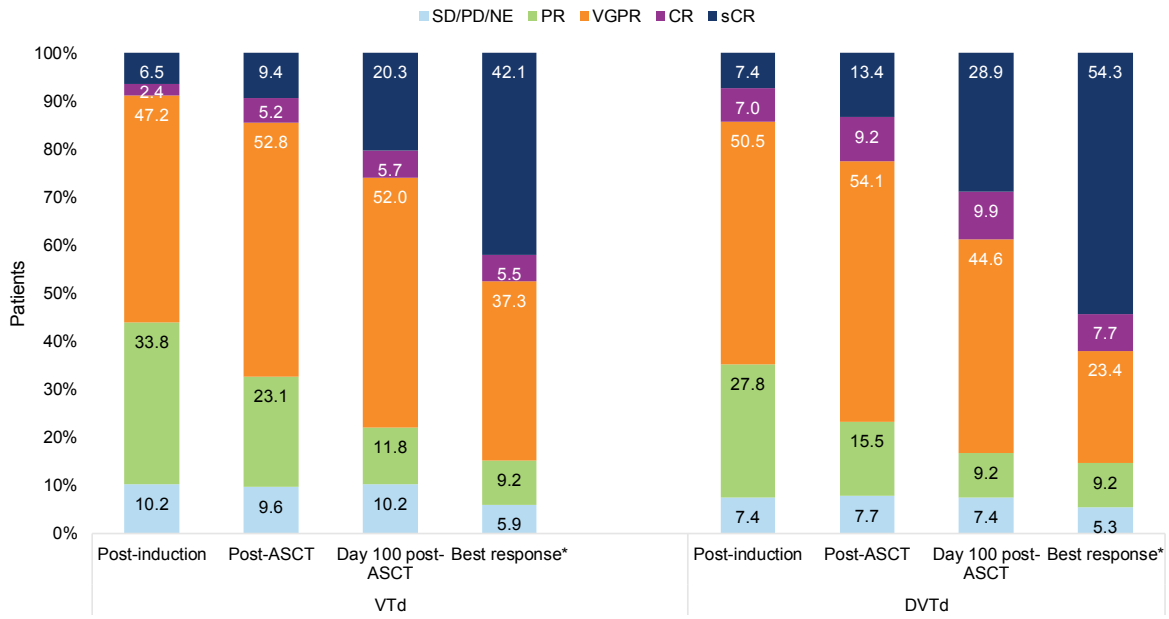


Key: CR = complete response; DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; ORR = overall response rate; P = P-value; PR = partial response; sCR = stringent complete response; VGPR = very good partial response; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).
*Post-consolidation response (100 days post ASCT) was measured based on a strict computerised algorithm, in line with IMWG criteria of response.(47, 48)

Post-consolidation CR or better (≥CR) and VGPR or better (≥VGPR) also showed significantly deeper levels of response for DBTd compared to BTd. Similar results were achieved when response was evaluated by investigator assessment with the proportion of patients achieving sCR and ≥CR in the DBTd group significantly higher than those in BTd group.(44)

For both DBTd and BTd, deeper responses were observed over time with each study phase in CASSIOPEIA.(43) For overall best response (median follow-up = 29.2 months), statistically significantly greater sCR and ≥CR rates were achieved for DBTd compared to BTd (sCR: 54.3% vs 42.1%, respectively; OR=1.64; 95% CI: 1.29, 2.09; p<0.0001; ≥CR: 62.1% vs 47.6% respectively; OR: 1.80 with 95% CI: 1.41, 2.30; p<0.0001).(45) An overview of response rates over time in CASSIOPEIA is provided in Figure 9 and Table 13.

Figure 9: Summary of response rates, based on time of assessment (ITT population)(43, 45)



Key: ASCT = autologous stem cell transplant; CR = complete response; Day 100 post-ASCT = post-consolidation; DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; NE = not evaluable; PD = progressive disease; PR = partial response; sCR = stringent complete response; SD = stable disease; VGPR = very good partial response; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).
*All data shown are from the Primary Analysis for Part 1 (June 2018 data cut), with the exception of best response for overall study which was analysed at the Post-hoc Interim Analysis (PHA, May 2019 data cut) and regardless of second randomisation.
Note: assessment of response in CASSIOPEIA was based on a strict computerised algorithm, in line with IMWG criteria of response.(47, 48)

Table 13: Summary of response rates based on computerised algorithm (ITT population)(44, 45)

Response category	DBTd (n=543)	BTd (n=542)	OR (95% CI)	P-value
Post-induction, n (%)				
sCR	40 (7.4%)	35 (6.5%)	1.16 (0.72, 1.86)	0.5344
≥CR (sCR + CR)	78 (14.4%)	48 (8.9%)	1.73 (1.18, 2.53)	0.0048
≥VGPR (sCR + CR + VGPR)	352 (64.8%)	304 (56.1%)	1.44 (1.13, 1.84)	0.0033
Overall response (sCR+CR+VGPR+PR)	503 (92.6%)	487 (89.9%)	1.41 (0.92, 2.17)	0.1057
Post-transplant, n (%)				
sCR	73 (13.4%)	51 (9.4%)	1.5 (1.02, 2.19)	0.0356
≥CR (sCR+CR)	123 (22.7%)	79 (14.6%)	1.72 (1.26, 2.35)	0.0006
≥VGPR (sCR+CR+VGPR)	417 (76.8%)	365 (67.3%)	1.6 (1.23, 2.09)	0.0005
Overall response (sCR+CR+VGPR+PR)	501 (92.3%)	490 (90.4%)	1.26 (0.82, 1.93)	0.2806
Post-consolidation, n (%)				
sCR	157 (28.9%)	110 (20.3%)	1.60 (1.21, 2.12)	0.0010
≥CR (sCR+CR)	211 (38.9%)	141 (26.0%)	1.82 (1.40, 2.36)	<0.0001

≥VGPR (sCR+CR+VGPR)	453 (83.4%)	423 (78.0%)	1.41 (1.04, 1.92)	0.0239
Overall response (sCR+CR+VGPR+PR)	503 (92.6%)	487 (89.9%)	1.41 (0.92, 2.16)	0.1085
Best response over time (PHA, May 2019 data cut), n (%)				
sCR	295 (54.3%)	228 (42.1%)	1.64 (1.29, 2.09)	<0.0001
≥CR (sCR+CR)	337 (62.1%)	258 (47.6%)	1.80 (1.41, 2.30)	<0.0001
≥VGPR (sCR+CR+VGPR)	464 (85.5%)	460 (84.9%)	1.05 (0.75, 1.47)	0.7781

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; CR = complete response; ITT = intention-to-treat; PHA = Post-hoc Interim Analysis; PR = partial response; sCR = stringent complete response; VGPR = very good partial response.

Time to response (other secondary endpoint)

At median follow-up of 18.8 months, the median time to response (PR or better) in CASSIOPEIA was approximately one month (Table 14). Similar median time to response was observed for both treatment groups for ≥VGPR (DBTd: 2.14 months, BTd: 2.83 months) and for ≥CR (DBTd: 7.23 months, BTd: 7.38 months).(44, 65)

Table 14: Time to response among patients treated with BTd compared with DBTd (Response-evaluable analysis set, median follow-up = 18.8 months)(44)

	BTd	DBTd
Responders (≥PR)	510	513
Time to first response[a,b] (months)		
N	510	513
Median (range)	1.05 (0.8; 10.1)	1.02 (0.7; 10.0)
Time to VGPR or better[a] (months)		
N	434	454
Median (range)	2.83 (0.9; 10.3)	2.14 (0.9; 10.6)
Time to CR or better[a] (months)		
N	144	211
Median (range)	7.38 (1.9; 11.4)	7.23 (1.9; 10.6)
Time to sCR[a] (months)		
N	113	157
Median (range)	7.98 (3.6; 10.8)	7.98 (3.5; 11.2)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; CR = complete response; ITT = intention-to-treat; PR = partial response; sCR = stringent complete response; VGPR = very good partial response.
Note: Response-evaluable set includes subjects who have a confirmed diagnosis of multiple myeloma and measurable disease at baseline or screening. In addition, subjects must have received at least one component of study treatment and have adequate post-baseline disease assessments.
[a] Response (PR, VGPR, CR, sCR) needs to be achieved by Day 100 post-ASCT (or imputed date if missing) + 30 days.
[b] Response PR of better.

Duration of response (other secondary endpoint)

At median follow-up of 18.8 months, the median duration of response (DOR) was not reached for either treatment group (Table 15).

Table 15: Duration of response among patients treated with BTd compared with DBTd (Response-evaluable analysis set, median follow-up = 18.8 months)(44)

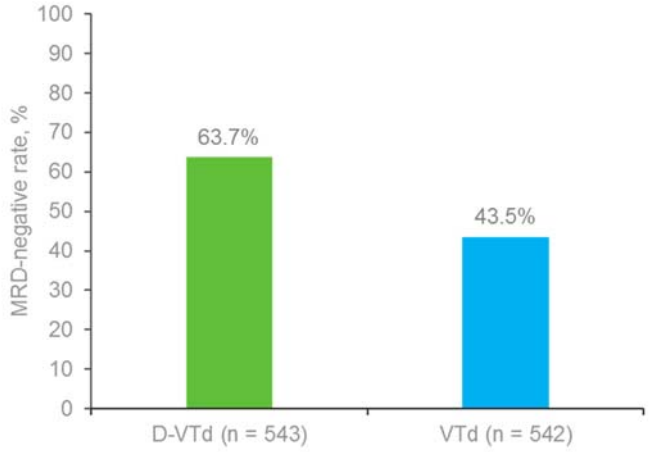
	BTd	DBTd
Responders (≥PR)	510	513
Duration of response[a,b] (months)		
Number of events (%)	67 (13.1%)	29 (5.7%)
Median (95% CI)[c]	NE (29.8, NE)	NE (NE, NE)
Duration of CR or better[a] (months)		
Number of events (%)	7 (4.9%)	4 (1.9%)
Median (95% CI)[c]	NE (NE, NE)	NE (NE, NE)
Duration of sCR[a] (months)		
Number of events (%)	5 (4.4%)	2 (1.3%)
Median (95% CI)[c]	NE (NE, NE)	NE (NE, NE)

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; CR = complete response; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; NE = Not estimable PR = partial response; sCR = stringent complete response; VGPR = very good partial response.
Note: Number of events refers to number of responders (or complete responders, or stringent complete responders) who developed disease progression or died due to disease progression.
[a] Response (PR, VGPR, CR, sCR) needs to be achieved by Day 100 post-ASCT (or inputted date if missing) + 30 days.
[b] First response PR of better.
[c] Based on Kaplan-Meier product limit estimates.

Post-consolidation MRD (key secondary endpoint)

As noted in Section B.2.3.1 above, MRD was assessed in all patients in the ITT population, regardless of response. This contrasts with previous MM trials where MRD was assessed in patients who achieved a pre-specified level of response (e.g. patients with suspected CR).(43) A statistically significant higher rate of post-consolidation MRD negativity, evaluated using MFC, was observed with DBTd compared with BTd alone at a threshold of 1 tumour cell per 10^{-5} white cells (63.7% vs 43.5%; OR: 2.27; 95% CI: 1.78, 2.90; p<0.0001) (Figure 10).(43, 44)

Figure 10: MRD-negative rate (10^{-5}) post-consolidation[a] by MFC (ITT population, median follow-up = 18.8 months)(44)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; MRD = minimal residual disease; MFC = multiparametric flow cytometry; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).
[a] Post-consolidation (100 days post ASCT) was measured by a standardised Euroflow based multiparametric assay.

An exploratory evaluation of MRD using next generation sequencing (NGS) indicated a similar trend, with significantly deeper responses for patients treated with DBTd compared with BTd (NGS 10^{-5}: 56.6% vs 36.8%; OR: 2.26; 95% CI: 1.68, 3.05; p<0.0001)[II].(44) At the higher sensitivity threshold of 10^{-6}, DBTd almost doubled the rate of MRD negativity compared to BTd (NGS 10^{-6}: 39.1% vs 22.8%; OR: 2.18; 95% CI: 1.58, 3.01; p<0.0001) indicating a significantly deeper level of response achieved with DBTd. Such strikingly deep levels of response can be expected to translate to long-term disease control and the hope of functional cure for these patients.(44) Indeed, patients achieving MRD negativity at the highest sensitivity (10^{-6}) have been reported to have the longest PFS compared to MRD negativity at sensitivities 10^{-5} and 10^{-4}.(67) The value to patients of long-term disease control and a sustained period of treatment-free remission is strongly linked to established patient preferences (see Section B.1.3.3), and the hope that is attached to a potential functional cure is not something that is captured in the cost per quality adjusted life year (QALY) calculations in Section B.3.

It is notable that the rates of MRD negativity observed in CASSIOPEIA were higher than ≥CR rates in both treatment arms.(43) This is due to a timing difference of clearing malignant plasma cells from the bone marrow (required for MRD-negative assessment) and clearing any detectable trace of paraprotein from blood serum and urine (required for CR). Given the half-life of paraproteins, the paraprotein can still be detectable despite the fact that the malignant clone in the bone marrow is eliminated. In other words, the elimination of plasma cells in the bone marrow typically occurs earlier than the elimination of paraprotein in the blood. A sign of this is the fact that, with additional follow-up, the proportion of patients who are MRD-negative at post consolidation but not in CR or better reduces as more patients reach CR or better over time (Table 16). Of the 289 patients who were MRD-negative and not in CR or better at PA1, 184 (64%) had reached CR or better at PHA (median follow-up = 29.2 months).(45)

[II] An evaluation of the agreement between MFC and NGS methods in all patients, regardless of response, showed good agreement for MRD determination (at sensitivity threshold of 10^{-5}). Observed agreement, calculated as the (total number of patients positive by both NGS and MFC + total number of patients negative by both NGS and MFC) / total number of patients with NGS and MFC results, was 83.5%.

Table 16: Disposition of best response of patients with post-consolidation MRD MFC Negative at 10^{-5} but were not in post-consolidation CR/sCR (ITT population, median follow-up = 29.2 months)(45)

	Induction/ASCT/consolidation	
	BTd	**DBTd**
Patients who had post-consolidation MRD-negative at 10^{-5} but who were not in post-consolidation CR/sCR	127	162
Best response		
VGPR or worse	51 (40.2)	54 (33.3)
CR or sCR	76 (59.8)	108 (66.7)

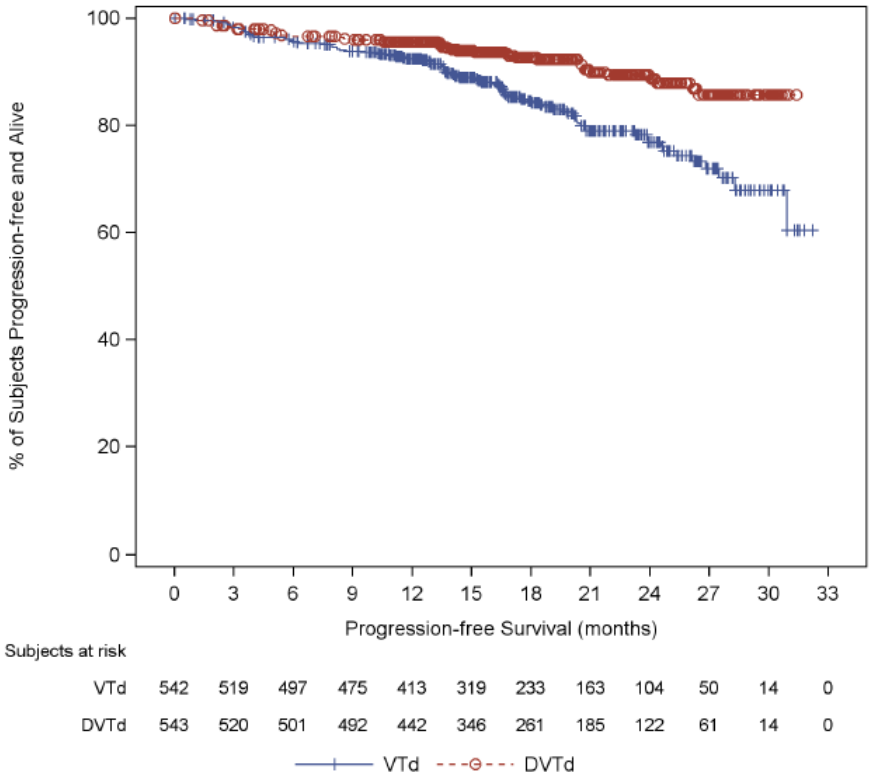
Key: DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention-to-treat; MRD = minimal residual disease; MFC = multiparametric flow cytometry; BTd = bortezomib, thalidomide and dexamethasone.

B.2.6.2 Survival analyses

Survival analysis: Progression-free survival (PA1) (key secondary endpoint)

After a median follow-up of 18.8 months, a total of 45 (8.3%) PFS events had occurred in the DBTd arm compared to 91 (16.8%) events in the BTd arm.(43) Treatment with DBTd was associated with a statistically significant, and clinically meaningful improvement in the risk of disease progression or death compared with BTd (HR: 0.47; 95% CI: 0.33, 0.67; p<0.0001).(43) DBTd resulted in a 53% reduction in the risk of disease progression or death compared with BTd, with 2-year PFS rates of 89.4% and 76.9% respectively.(44) Figure 11 presents the Kaplan-Meier (KM) plot for PFS from PA1 of CASSIOPEIA.

Figure 11: Kaplan-Meier plot for PFS from 1st randomisation for induction/ASCT/consolidation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months;)(43, 44)



Key: ASCT = autologous stem cell transplant; DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; PFS: progression-free survival; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

PFS adjusted for maintenance(43, 44)

To mitigate potential bias to the PFS outcomes for Part 1 caused by study maintenance, a per-protocol pre-specified statistical analysis was performed using the inverse probability weighting (IPW) method to adjust for the second randomisation (Table 17). The IPW method provides an unbiased PFS estimate and maintains the Type 1 error rate by stratifying two groups based on their maintenance treatment (i.e. DBTd versus BTd for patients who received daratumumab maintenance, and DBTd versus BTd for patients who received observation maintenance).(68) All patients including those who were not re-randomised were included in this PFS analysis. This analysis was performed and reviewed by a sequestered group independent of the study team to protect the integrity of the Part 2 analysis.

Consistent results in favour of DBTd versus BTd were seen when PFS was analysed after adjustment for the second randomisation, demonstrating that the observed treatment effect is attributable to the 1st part of the study (HR: 0.47; 95% CI: 0.33, 0.67; p<0.0001).(44) The similarity of adjusted and unadjusted analyses results was expected given the high proportion of patients re-randomised in both treatment groups and the relatively short duration of maintenance therapy to date. Refer to Appendix L for details regarding the IPW methodology.

Table 17: PFS results with and without IPW adjustments (ITT population, median follow-up = 18.8 months)(44)

IPW Analysis; DBTd versus BTd	
HR (95% CI)	0.47 (0.33, 0.67)
P-value	<0.0001
Analysis without adjustment for second randomisation; DBTd versus BTd	
HR (95% CI)	0.47 (0.33, 0.67)
P-value	<0.0001

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; IPW = inverse probability weighting; ITT = intention-to-treat; PFS = progression-free survival.

PFS sensitivity analysis

A number of sensitivity analyses for PFS were performed and results are presented in Table 18. Results from the PFS piecewise analysis by study phase demonstrate a consistent benefit of DBTd versus BTd across the different treatment phases. To evaluate the impact of transplant, multivariate analysis was performed which shows that by including transplant as a time varying covariate in the model, the treatment effect of daratumumab was maintained (HR: 0.48; 95% CI: 0.34, 0.69; p<0.0001).(44) This result is expected given the similarity in the proportion of patients in the DBTd and BTd arms receiving transplant, 90.1% and 89.1% respectively. An additional multivariate analysis was performed which included the interaction of treatment and transplant, where the PFS benefit of DBTd compared with BTd was similarly preserved (HR: 0.38; 95% CI: 0.18, 0.80; p=0.011).(44) These analyses indicate that the observed PFS benefit of daratumumab is over and above that obtained from induction and consolidation with BTd, and over and above the impact of ASCT on long-term outcomes.

Table 18: Summary of sensitivity analysis for PFS(44)

	DBTd (n=543)	**BTd (n=542)**	**HR (95% CI)[a]**	**P-value[b]**
PFS investigator assessment				
Number of events	48 (8.8%)	93 (17.2%)	0.49 (0.35, 0.69)	<0.0001
PFS censored for subsequent therapy				
Number of events	43 (7.9%)	85 (15.7%)	0.48 (0.33, 0.69)	<0.0001
PFS piecewise				
Induction phase[c]				
Number of events	16 (2.9%)	19 (3.5%)	0.84 (0.43, 1.63)	-
Consolidation phase[d]				
Number of events	5 (0.9%)	13 (2.4%)	0.38 (0.13, 1.06)	-
Maintenance phase[e]				
Number of events	24 (4.4%)	59 (10.9%)	0.37 (0.23, 0.60)	-
PFS Multivariate (including transplant in the model)[f]				
HR (95% CI)	-	-	0.48 (0.34, 0.69)	<0.0001
PFS Multivariate (testing for interaction of treatment and transplant)				
HR (95% CI)	-	-	0.38 (0.18, 0.80)	0.0110

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; PFS = progression-free survival.
Note: PFS events included confirmed progressive disease (per IMWG criteria) or death
[a] HR and 95% CI from a Cox regression analysis with treatment as the sole explanatory variable.
[b] P-value is based on the log-rank test.
[c] Events occurred after 5.3 months (median time to the start of ASCT) were censored.
[d] Excluded the subjects who had the event or censored before 5.3 months. Events occurred after 9.5 months (median time to 2nd randomisation) were censored
[e] Excluded the subjects who had the event or censored before 9.5 months.
[f] HR, 95% CI, and p-value are from a Cox regression analysis. Both treatment and time-varying transplantation are explanatory variables in the model. A hazard ratio < 1 of treatment group indicates an advantage for DBTd.

PFS updated results (PHA)

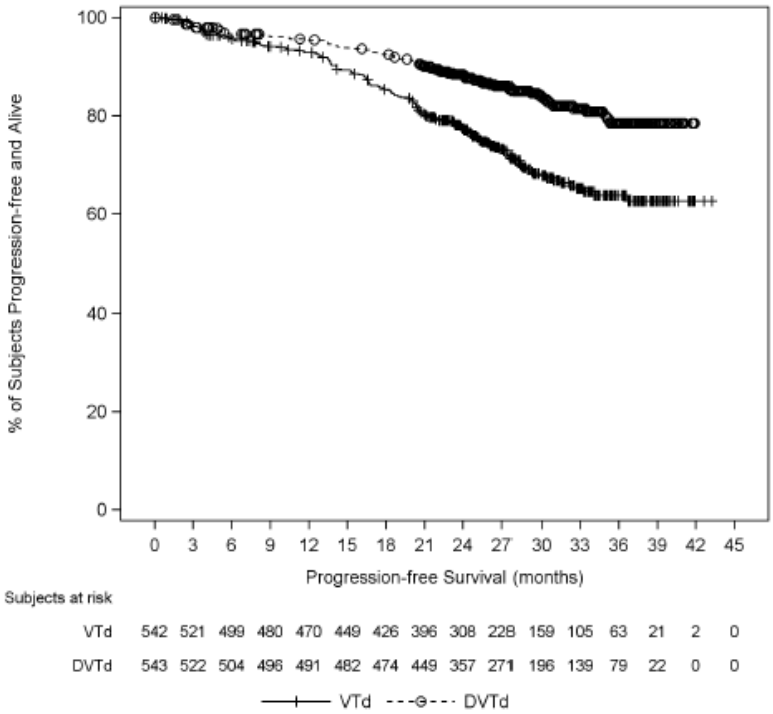
At the time of clinical cut-off for the PHA, a total of 83 (15.3%) PFS events had occurred in the DBTd group, and 151 (27.9%) events in the BTd group.(45) A comparison of PFS results from the PA1 and PHA without adjustment for the second randomisation is presented in Table 19 with the associated Kaplan-Meier plot shown in Figure 12. After a median follow-up of 29.2 months, median PFS was not reached in either treatment group.(45)

Table 19: Comparison of updated PFS (PHA vs PA1), regardless of 2nd randomisation (ITT population)(44, 45)

	PA1 (median follow-up = 18.8 months)		PHA (median follow-up = 29.2 months)	
	BTd	DBTd	BTd	DBTd
n/N (%)	91/542 (16.8%)	45/543 (8.3%)	151/542 (27.9%)	83/543 (15.3%)
Median (95% CI)	NE (941.00, NE)	NE (NE, NE)	NE (NE, NE)	NE (NE, NE)
HR (95% CI)	0.47 (0.33, 0.67)		0.495 (0.38, 0.65)	
P-value	<0.0001		<0.0001	
6-month PFS rate % (95% CI)	95.8 (93.7, 97.2)	96.6 (94.6, 97.8)	95.8 (93.8, 97.2)	96.6 (94.7, 97.8)
12-month PFS rate % (95% CI)	92.4 (89.8, 94.4)	95.6 (93.5, 97.1)	92.9 (90.3, 94.8)	95.4 (93.3, 96.9)
18-month PFS rate % (95% CI)	84.6 (80.7, 87.7)	92.7 (89.8, 94.7)	85.3 (82.0, 88.1)	92.5 (89.9, 94.5)
24-month PFS rate % (95% CI)	76.9 (71.5, 81.3)	89.4 (85.6, 92.3)	77.4 (73.4, 80.8)	88.4 (85.3, 90.9)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; ITT = intention-to-treat; PA1 = Primary Analysis for Part 1; PHA = Post-hoc Interim Analysis; PFS = progression-free survival.

Figure 12: Kaplan-Meier plot for PFS from 1st randomisation for induction/ASCT/consolidation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months;)(65)



Key: ASCT = autologous stem cell transplant; DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; PFS = progression-free survival; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

The additional 10.4 months follow-up demonstrates a consistent benefit for DBTd over BTd in terms of PFS with a 51% reduction in the risk of disease progression or death (HR: 0.495; 95% CI 0.38, 0.65; p<0.0001), while the 2-year PFS rates remain stable at 88.4% and 77.4% respectively.(45) At the request of the Committee for Medicinal Products for Human Use (CHMP), the updated PFS results were adjusted based on the IPW method with consistent results indicating minimal impact of the second randomisation on the PFS outcomes for Part 1 with longer study follow-up (Table 20). Whilst Part 2 of the study remains blinded, Janssen does not have access to, and is unable to perform, any additional statistical analysis to adjust survival outcomes attributable to Part 1 for the second randomisation.

Table 20: PFS IPW adjusted results (ITT population, median follow-up = 29.2 months)(65)

	Induction/ASCT/Consolidation	
	HR (95% CI)[a]	P-value[b]
Overall	0.50 (0.34, 0.75)	0.0005

Key: ASCT = autologous stem cell transplant; CI = confidence interval; HR = hazard ratio; IPW = inverse probability weighting; ITT = intention-to-treat; PFS = progression-free survival.
[a] Hazard ratio and 95% CI from a Cox regression analysis with treatment as the sole explanatory variable.
[b] The p-value is based on the log-rank test.
Patients randomised to daratumumab maintenance at the second randomisation were censored at the date of the second randomisation.

The maintenance of treatment benefit with longer follow-up is in line with other phase III daratumumab trials including CASTOR for relapsed/refractory MM (RRMM) where the treatment benefit for PFS stabilised approaching two years of study follow-up (Table 21).

Table 21: Comparison of CASTOR data-cuts: PFS(69-74)

DBd vs Bd	Median follow-up=7.4 months (IA1)	Median follow-up=19.4 months	Median follow-up=26.9 months (IA2)	Median follow-up=40.0 months (ASH18)	Median follow-up=50.2 months (ASH19)
ITT HR (95% CI)	0.39 (0.28-0.53)	0.31 (0.24-0.39)	0.32 (0.25-0.40)	0.31 (0.25-0.40)	0.31 (0.24-0.39)
1PL HR (95% CI)	0.31 (0.18-0.52)	0.19 (0.12-0.29)	0.23 (0.16-0.33)	0.22 (0.15-0.32)	0.21 (0.15-0.31)

Key: Bd = bortezomib and dexamethasone; CI: confidence interval; DBd = daratumumab, bortezomib, and dexamethasone; HR: hazard ratio; IA = Interim Analysis; ITT = intention-to-treat; 1PL = one prior line.

Time to disease progression (other secondary endpoint)

At median follow-up of 18.8 months, a total of 118 patients had progressive disease or died due to progressive disease, including 42 patients (8%) in the DBTd group, and 76 patients (14%) in the BTd group.(43) TTP was significantly improved with DBTd and was associated with a 48% reduction in the risk of disease progression compared with BTd (HR=0.52; 95% CI: 0.36, 0.76; p=0.0006).(43) TTP was measured from first randomisation and assessed by computerised algorithm.

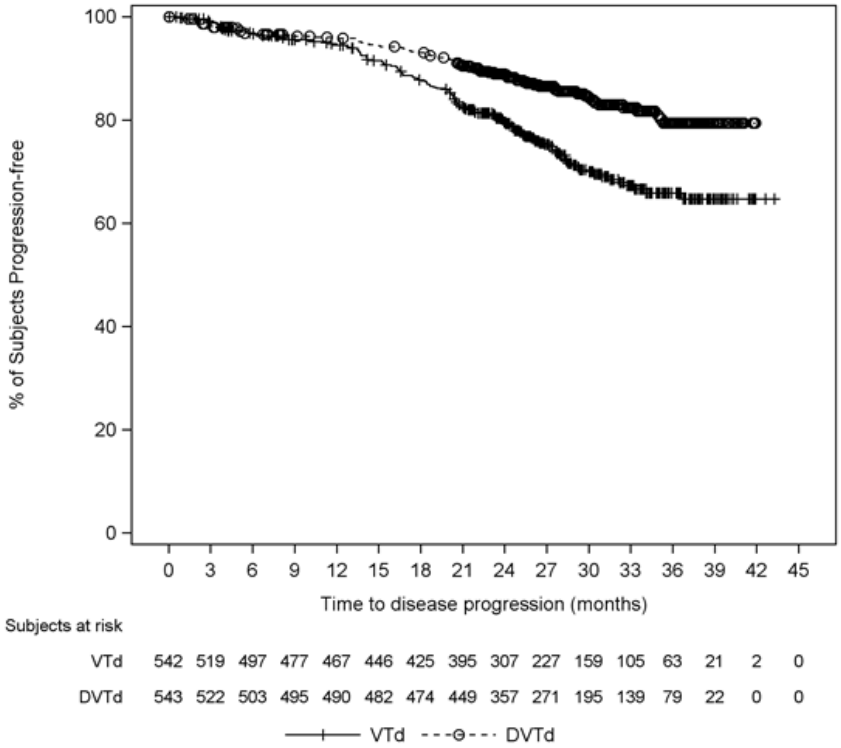
At the time of clinical cut-off for PHA (median follow-up = 29.2 months), a further 97 progression events had been recorded bringing the total to 215, including 79 patients (14.5%) in the DBTd group and 136 patients (25.1%) in the BTd group.(45) The significant improvement in TTP with DBTd versus BTd was maintained with the longer follow-up (HR: 0.52; 95% CI: 0.39, 0.68; p<0.0001).(75) Whilst median TTP was not reached for either treatment group, there was a clear trend for a longer progression-free interval with DBTd compared to BTd (Figure 13).

Table 22: Time to disease progression among patients treated with BTd compared with DBTd (ITT population, assessed by computerised algorithm)(43, 45)

	Median follow-up = 18.8 months		Median follow-up = 29.2 months	
	BTd (n=542)	DBTd (n=543)	BTd (n=542)	DBTd (n=543)
Time to disease progression (days)				
Number of events [a]	76 (14.0%)	42 (7.7%)	136 (25.1%)	79 (14.5%)
Median (95% CI) [b]	NE (941.0, NE)	NE (NE, NE)	NE (NE, NE)	NE (NE, NE)
P-value [c]	0.0006		<0.0001	
HR (95% CI) [d]	0.52 (0.36, 0.76)		0.52 (0.39, 0.68)	

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; NE = not estimable; ITT = intention-to-treat.
[a] Including all patients randomised in Induction/ASCT/Consolidation regardless of second randomisation.
[b] Based on Kaplan-Meier product limit estimates.
[c] P-value is based on the log-rank test.
[d] HR and 95% CI from Cox regression analysis with treatment as the sole explanatory variable.

Figure 13: Kaplan-Meier plot of time to progression from 1st randomisation regardless of 2nd randomisation based on computerised algorithm (ITT population, median follow-up = 29.2 months)(45)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

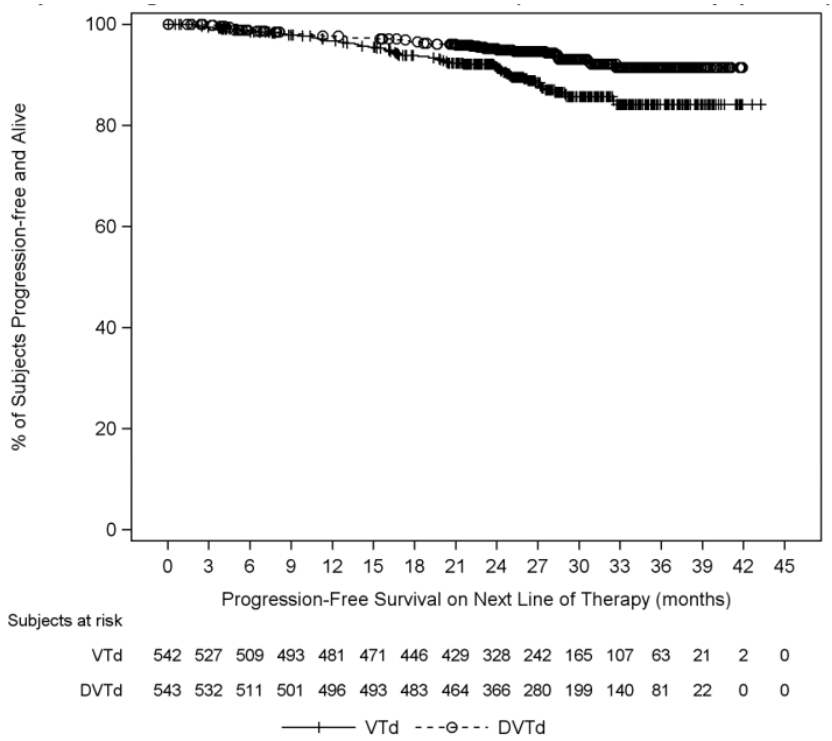
Analysis of TTP using disease progression assessed by the investigator was highly consistent with the results using disease progression by computerised algorithm. Refer to Appendix L for further details.

Progression-free survival on the subsequent line of therapy (other secondary endpoint)

Progression-free survival on the subsequent line of therapy (PFS2) represents the time interval between the date of randomisation to the date of progressive disease on the next line of subsequent treatment or death from any cause. At a median follow-up of 18.8 months, PFS2 data remained immature with only 18 (3.3%) events in the DBTd group and 37 (6.8%) events for BTd. Despite the low number of events, there was a strong trend for improved outcome for patients treated with DBTd (HR 0.46; CI 0.26, 0.82; p=0.0062).(44) These results demonstrate that the PFS benefit of DBTd is maintained beyond the next line of therapy received.

At a median follow-up of 29.2 months, PHA demonstrated consistent results with a significant improvement in PFS2 for patients treated with DBTd (HR: 0.51; 95% CI: 0.33, 0.78; p=0.0015).(75) At the time of the second analysis for Part 1 of the study, 33 (6.1%) PFS2 events had occurred in the DBTd group compared with 60 (11.1%) events in the BTd group.(45)

Figure 14: Kaplan-Meier plot for PFS2 from 1st randomisation regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(45)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; PFS = progression-free survival; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

Survival analysis: Overall survival (PA1) (key secondary endpoint)

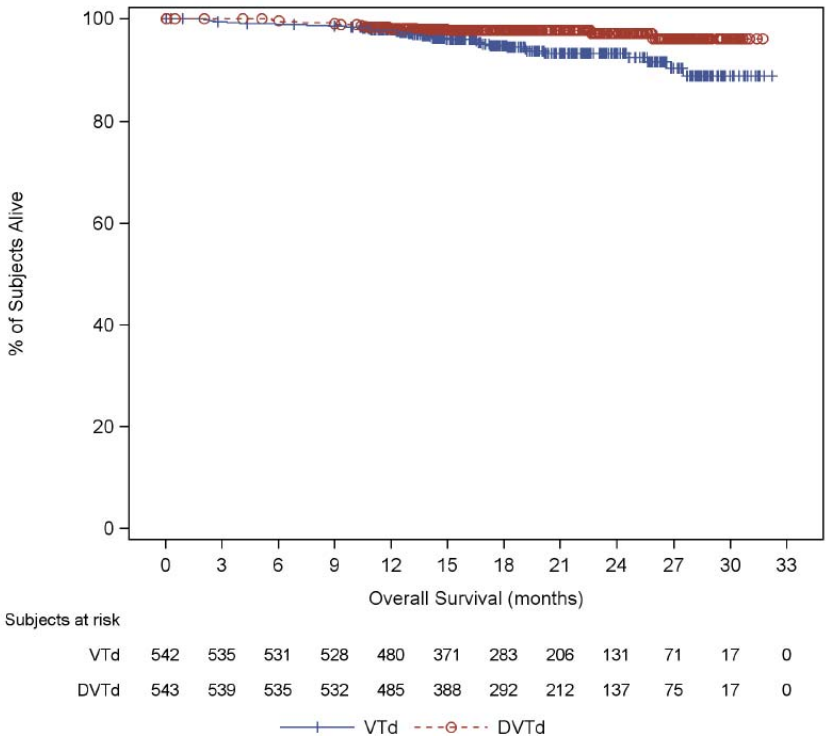
At the Primary Analysis for Part 1, a total of 46 death events had occurred, including 14 patients in the DBTd group and 32 patients in the BTd group (Table 23). Despite the immaturity of the survival data, a strong trend for improved OS was observed with DBTd compared with BTd with a 57% reduction in the risk of death (HR: 0.43; 95% CI: 0.23, 0.80; nominal p=0.0065, not adjusted for second randomisation).(43) Refer to Figure 15 for the KM plot from CASSIOPEIA after a median follow-up of 18.8 months.

Table 23: OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months)(44)

	BTd	DBTd
Analysis set: intention-to-treat	542	543
Overall survival (days)		
Number of events (%)[a]	32 (5.9%)	14 (2.6%)
Median (95% CI)[b]	NE (NE, NE)	NE (NE, NE)
P-value[c]	0.0065	
HR (95% CI)[d]	0.43 (0.23, 0.80)	
6-month Survival rate % (95% CI)[b]	98.9 (97.5, 99.5)	99.6 (98.5, 99.9)
12-month Survival rate % (95% CI)[b]	97.8 (96.1, 98.7)	98.1 (96.5, 99.0)
18-month Survival rate % (95% CI)[b]	94.7(92.2, 96.5)	97.6 (95.9, 98.7)
24-month Survival rate % (95% CI)[b]	93.2 (90.1, 95.3)	97.1 (94.7, 98.4)

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; NE = not estimable; ITT = intention-to-treat; OS = overall survival.
[a] Including all patients randomised in Part 1 regardless of second randomisation.
[b] Based on Kaplan-Meier product limit estimates.
[c] P-value is based on the log-rank test.
[d] HR and 95% CI from a Cox regression analysis with treatment as the sole explanatory variable.

Figure 15: Kaplan-Meier plot for OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 18.8 months)(44)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; OS = overall survival; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

OS updated results (PHA)

An updated (post-hoc) analysis of OS was performed with a median follow-up of 29.2 months, representing an additional 10.4 months of follow-up (Table 24). At the time of clinical cut-off for PHA, there were an additional 28 reported deaths resulting in a total of 74 cumulative deaths in the overall study (26 in the DBTd group and 48 in the BTd group).(65) Although OS data from CASSIOPEIA remains immature with median OS not reached on either arm, the treatment benefit in favour of DBTd was maintained with longer study follow-up, further supporting the overall clinical benefit of the daratumumab combination (HR: 0.52; 95% CI: 0.33, 0.85; nominal p=0.0070, not adjusted for second randomisation).(65) Refer to Figure 16 for the corresponding KM plot for OS.

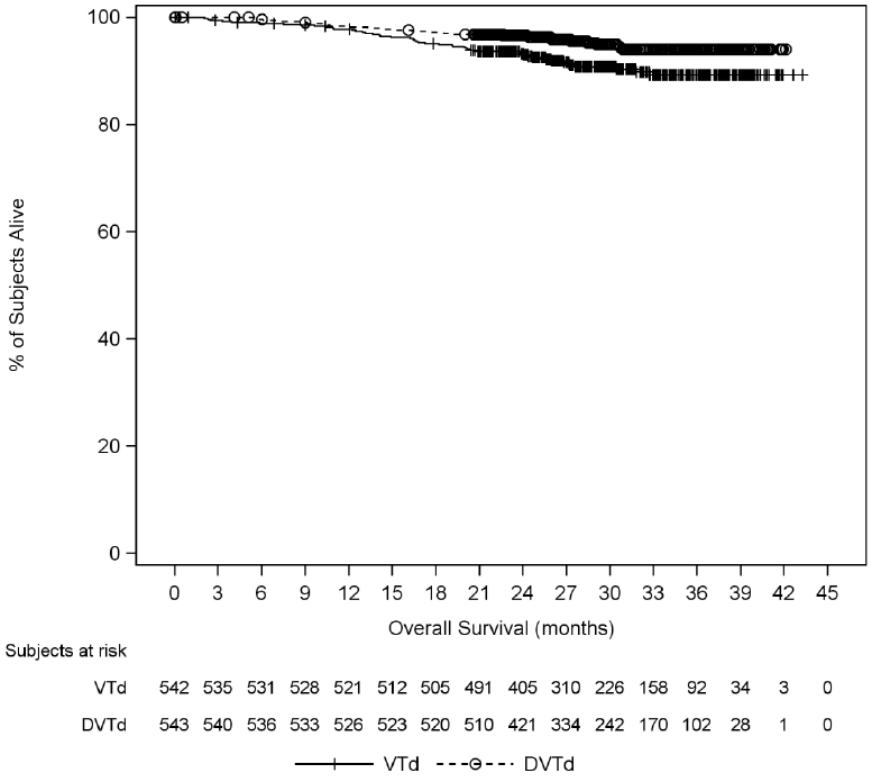
Table 24: OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(65)

	BTd	DBTd
Analysis set: intention-to-treat	542	543
Overall survival (days)		
Number of events (%) [a]	48 (8.9%)	26 (4.8%)
Median (95% CI) [b]	NE (NE, NE)	NE (NE, NE)
P-value [c]	0.0070	
HR (95% CI) [d]	0.52 (0.33, 0.85)	
6-month Survival rate % (95% CI) [b]	98.9 (97.5, 99.5)	99.6 (98.5, 99.9)

12-month Survival rate % (95% CI) [b]	97.8 (96.1, 98.7)	98.1 (96.6, 99.0)
18-month Survival rate % (95% CI) [b]	95.1 (92.9, 96.7)	97.2 (95.4, 98.3)
24-month Survival rate % (95% CI) [b]	93.2 (90.6, 95.0)	96.6 (94.7, 97.9)

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; NE = not estimable; ITT = intention-to-treat; OS = overall survival.
[a] Including all patients randomised in Part 1 regardless of second randomisation.
[b] Based on Kaplan-Meier product limit estimates.
[c] P-value is based on the log-rank test.
[d] HR and 95% CI from a Cox regression analysis with treatment as the sole explanatory variable.

Figure 16: Kaplan-Meier plot for OS from 1st randomisation, regardless of 2nd randomisation (ITT population, median follow-up = 29.2 months)(65)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; OS = overall survival; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

B.2.6.3 Landmark analyses for survival by response

Achieving deep and durable responses by eliminating as many clonal types as possible is one of the primary aims of treatment in the front-line setting and is associated with improved long-term outcomes for both survival and disease progression.(76) MRD is the most sensitive measure of response currently available and has been recommended in IMWG response assessment criteria.(64)

To explore the impact of MRD negativity on survival outcomes in the CASSIOPEIA trial, exploratory analyses were conducted to compare PFS and OS for patients who achieved MRD negativity at the time of the post-consolidation assessment versus those with an MRD-positive response. In order to mitigate the effect of immortal time bias (i.e. patients needed to live long enough to experience the

event), a landmark analysis was performed using individual patient data (IPD) from the CASSIOPEIA trial (PHA; median follow-up = 29.2 months) in which survival was assessed from the time of the post-consolidation response assessment, with patients who experienced the event of interest (i.e. death or progression) before this timepoint being excluded from the analysis. The 'landmark' point used was the time of the post-consolidation response assessment, which differed between each individual patient with respect to the time from randomisation. Treatment arm was also included in the analyses to explore whether the impact of MRD negativity on survival outcomes was the same for patients in the DBTd and BTd arms of the CASSIOPEIA trial. Cox proportional hazard models were calculated using the R package 'survival' to determine the effect of treatment in each of the MRD groups for PFS and OS.

Landmark analyses have been used in previous economic analysis in MM to overcome the time to response or immortal time bias that can occur due to the delayed clinical responses.(77, 78) Further, a review of endpoints and statistical considerations for immunomodulatory agents in MM highlighted the importance of using landmark survival analysis to benchmark long-term survival outcomes.(79)

Kaplan-Meier plots for PFS and OS from the time of the post-consolidation response assessment, by treatment arm and MRD status, are presented in Figure 17 and Figure 18, respectively. As these analyses were conducted for the purpose of informing the cost-effectiveness model, which utilised 28-day model cycles, time on these plots is expressed in terms of the number of 28-day cycles.

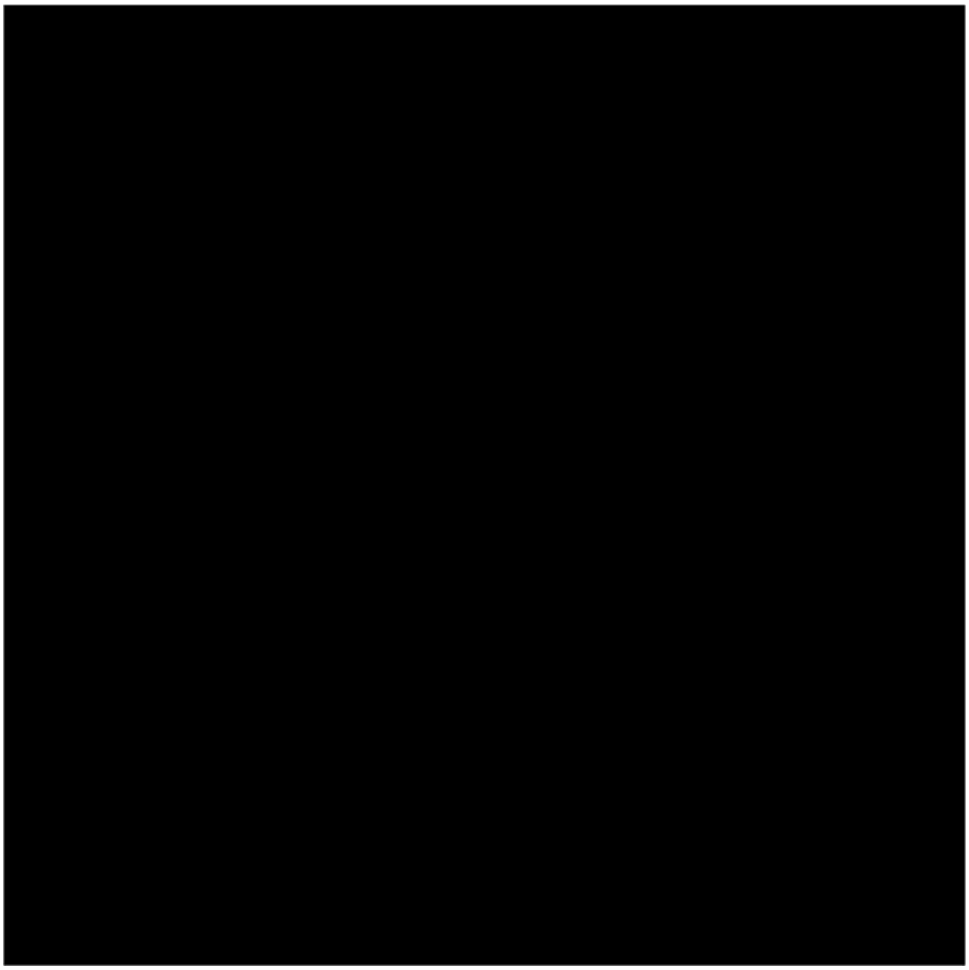
As shown in the Kaplan-Meier plots below, patients achieving post-consolidation MRD negativity demonstrated improved survival (PFS and OS) compared to those with an MRD-positive response. Furthermore, survival outcomes for patients treated with DBTd who achieve MRD negativity resemble the general population (when matched for age and gender), suggesting possible long-term disease control and providing hope of a functional cure for some patients (Figure 18).

The results of the Cox proportional hazard models show treatment with DBTd is associated with improvements in outcomes for both MRD-negative and MRD-positive patients when compared to the BTd arm (Table 25), indicating a daratumumab treatment effect regardless of MRD status. This reflects deeper responses for DBTd treated patients, as demonstrated by a higher proportion of patients achieving MRD negativity measured by NGS at a greater sensitivity threshold of 10^{-6} (see Section B.2.6.1), as well as the deeper conventional response according to IMWG achieved with daratumumab in MRD-positive patients. Indeed, evidence of a daratumumab treatment effect regardless of MRD status is [redacted] An exploratory, pooled analysis of daratumumab studies demonstrated [redacted] 19 Specifically, in the front-line transplant ineligible setting, a pooled analysis of ALCYONE and MAIA demonstrated [redacted]. Similarly, in the RRMM setting, a pooled analysis of CASTOR and POLLUX demonstrated a [redacted].

The deeper responses, and observed treatment effect attributable to daratumumab, reflect its unique mechanism of action, and specifically the combination of direct and immunomodulatory effects that harnesses the body's own immune system to target and eliminate malignant plasma cells. The results

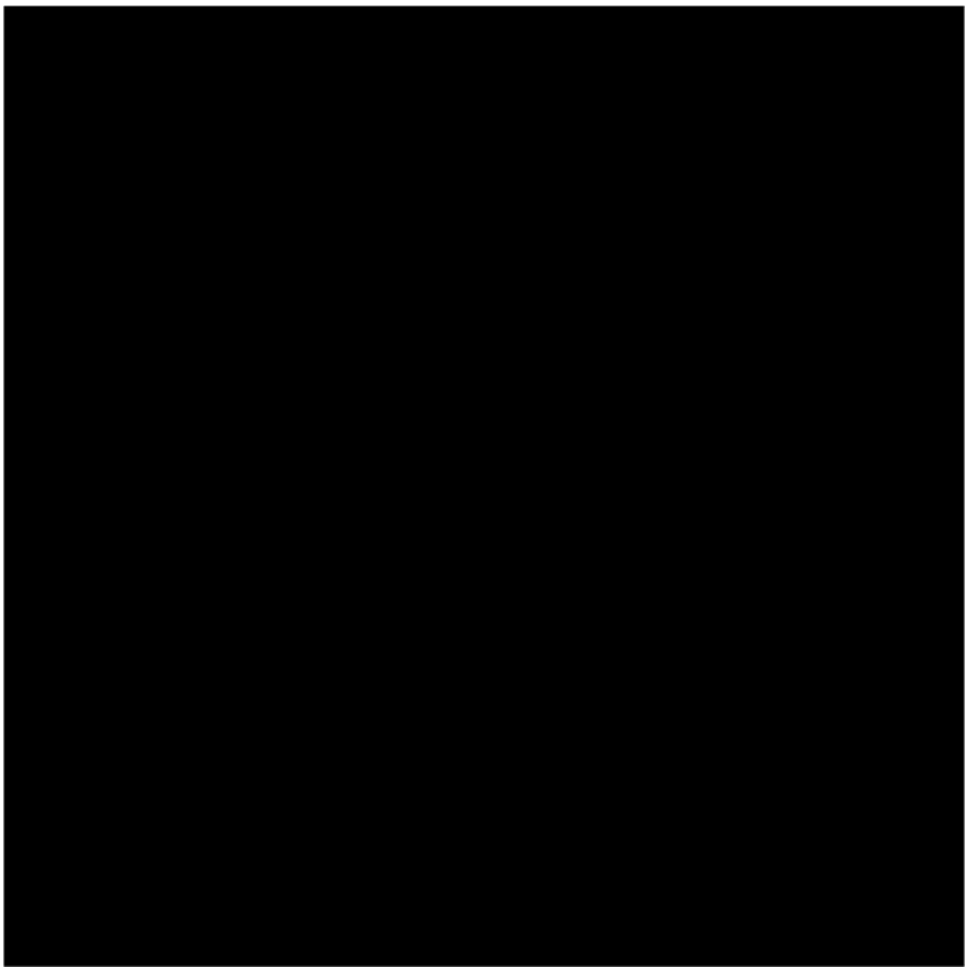
from these landmark analyses have been used to inform the survival inputs used in the cost-effectiveness analysis (see Section B.3.3.2).

Figure 17: Landmark analysis: PFS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months)



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.

Figure 18: Landmark analysis: OS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months)



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; GPM = general population mortality; MRD = minimal residual disease; OS = overall survival.

Table 25: Cox proportional hazard models results (landmark analysis)

DBTd versus BTd	HR for OS (95% CI)	HR for PFS (95% CI)
MRD-negative	[redacted]	[redacted]
MRD-positive	[redacted]	[redacted]

Key: CI = confidence interval; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MRD = minimal residual disease; OS = overall survival; PFS = progression-free survival.

Figure 19: Exploratory analysis: pooled front-line and RRMM PFS analysis from daratumumab studies, stratified by MRD status and treatment group

Key: Dara = daratumumab; HR = hazard ratio; MRD = minimal residual disease; NDMM = newly diagnosed multiple myeloma; PFS = progression-free survival; RRMM = relapsed/refractory multiple myeloma; TIE = transplant ineligible.

B.2.6.4 Health-related quality of life assessment

In Part 1 of CASSIOPEIA, pre-specified assessment of functional status and well-being were assessed using the EORTC QLQ-C30 and the EuroQol-5D, 5 levels (EQ-5D-5L) tools at:(44, 80)

- Screening (Baseline)
- Post-induction (Cycle 4 Day 28)
- Post-consolidation (Day 100 post ASCT)

Patients treated with both DBTd and BTd experienced meaningful and sustained improvements in HRQoL.(44) A statistically significant reduction in pain was seen with DBTd compared with BTd, while treatment with DBTd also resulted in significantly greater improvements in emotional functioning and a smaller decline in cognitive functioning on the EORTC QLQ-C30 subscales. As noted in Section B.1.3.3, bone pain was one of the symptoms most frequently reported in a recent European study of MM patient perceptions whilst cognitive impairment was the most frequently reported side-effect for NDMM. Improvements in pain and cognitive functioning for patients treated with DBTd are therefore closely aligned to MM patient preferences. Similarly, improvements in emotional functioning on the EORTC QLQ-C30 subscale may be indicative of a psychological impact of achieving sustained remission and a prolonged treatment-free interval. This benefit, and the value of hope for the future associated with no detectable disease and long-term disease control, is not intrinsically captured in the QALY framework.

The overall health state of patients, as measured by EQ-5D-5L, was improved in both treatment groups over the course of treatment.(44, 80, 81) Importantly, QoL assessment showed no adverse QoL impact of a quadruplet therapy over the standard BTd triplet. This means that patients treated with the DBTd quadruplet therapy combination benefit from improved PFS and OS with no significant detriment to overall HRQoL versus the existing SOC triplet therapy (BTd).

At the baseline and throughout the Part 1 of the study, both DBTd and BTd groups demonstrated high compliance rates for EORTC QLQ-C30 and EQ-5D-5L assessments (Table 26).

Table 26: EORTC QLQ C30 and EQ-5D-5L compliance rates at study time points (ITT population)(44)

	EORTC QLQ-C30		EQ-5D-5L	
	DBTd	BTD	DBTd	BTD
Baseline	94%	94%	93%	93%
Cycle 4 Day 28	84%	80%	82%	79%
Post-consolidation	90%	88%	89%	87%

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer-Quality of Life Questionnaire-C30; EQ-5D-5L = EuroQol-5D, 5 levels; GHS = global health status; ITT = intention-to-treat; SD = standard deviation.

EORTC QLQ-C30

The EORTC QLQ-C30 is a validated instrument that is widely used to measure QoL in patients with cancer.(82) This self-administered questionnaire captures symptoms that are relevant to MM patients and its results provide information about the possible side effects of treatment. It has five functional scales (physical, role, emotional, cognitive and social functioning), one Global Health Status (GHS) scale, three symptom scales (fatigue, nausea and vomiting, and pain) as well as single symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea).

Baseline values for all subscales of the EORTC QLQ-C30 were comparable for patients treated with DBTd and BTd (Table 27).

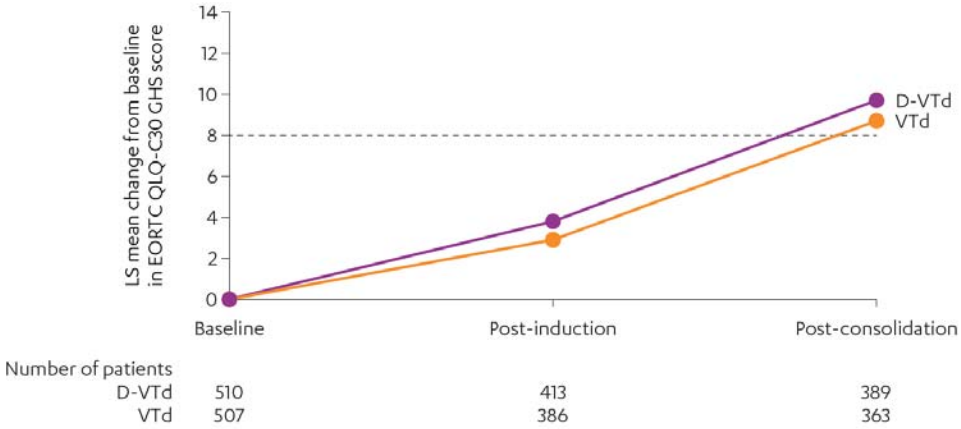
Table 27: Baseline values for the EORTC QLQ-C30 (CASSIOPEIA, ITT population)(44)

Subscale score, mean (SD)	DBTd	BTd
GHS	57.6 (24.2)	58.4 (24.5)
Symptom scales		
Fatigue	41.1 (28.4)	42.6 (29.6)
Nausea and vomiting	6.0 (15.2)	7.16 (17.0)
Pain score	47.4 (34.8)	46.4 (34.2)
Functional scales		
Cognitive functioning	84.8 (21.2)	85.6 (19.5)
Emotional functioning	67.9 (23.5)	65.7 (23.8)
Physical functioning	71.2 (27.5)	70.5 (28.4)
Role functioning	54.3 (37.8)	55.4 (36.9)
Social functioning	69.8 (34.4)	71.4 (32.7)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer-Quality of Life Questionnaire-C30; GHS = global health status; ITT = intention-to-treat; SD = standard deviation.
Note: Higher scores indicate better GHS, better functioning and more symptoms. Lower scores indicate worsening symptoms. The highest possible score is 100 at baseline.

At post-consolidation, both DBTd and BTd treatment groups had demonstrated improvements in overall HRQoL with regards to global health status (GHS), symptom, and function EORTC QLQ-C30 subscales.(44, 80) For GHS, there was an improvement in least-squares (LS) mean change from baseline for both DBTd and BTd through to Day 100 post ASCT, with change for both groups exceeding the minimally important difference (MID) of 8 points (LS mean change from baseline; DBTd = 9.7 [95% CI:7.4, 11.9], BTd = 8.7 [95% CI: 6.5,11]; p=0.4523).(44, 83) The difference between the DBTd and BTd groups was not statistically significant.(44, 80)

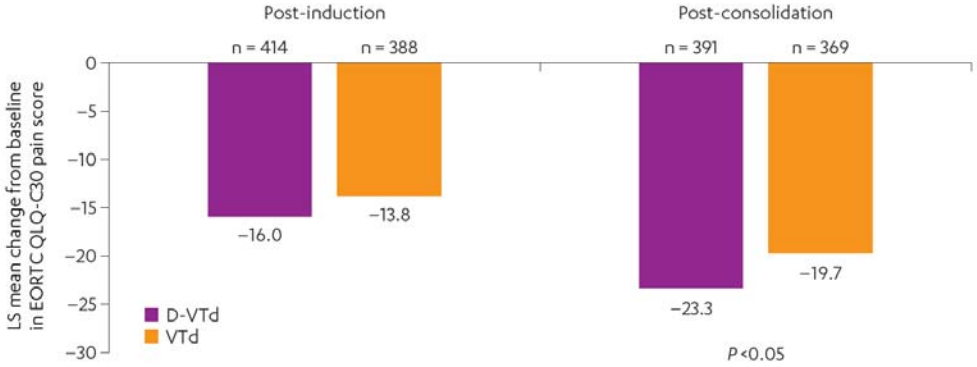
Figure 20: EORTC QLQ-C30 GHS change from baseline among patients treated with either DBTd or BTd (mixed effects model for repeated measures)(80)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Core Cancer Quality of Life Questionnaire-C30; GHS = Global Health Status; LS = least-squares; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).
Least square means are derived based on the mixed effects model with repeated measures, in which the dependent variable is change from baseline in score and independent variables are baseline, visit, treatment, visit by treatment interaction and randomisation stratification factors - ISS staging (I, II, III), region (Europe vs Other) and age (<75 years vs ≥75 years) as fixed effects and individual subject as random effect.

For patients in the DBTd group, a statistically significant reduction in pain symptoms compared with the BTd group was reported post-consolidation (LS mean change from baseline -23.3 and -19.7, respectively; p=0.0416) (Figure 21).

Figure 21: EORTC QLQ-C30 change from baseline in pain subscale scores (mixed effects model for repeated measures)(80)

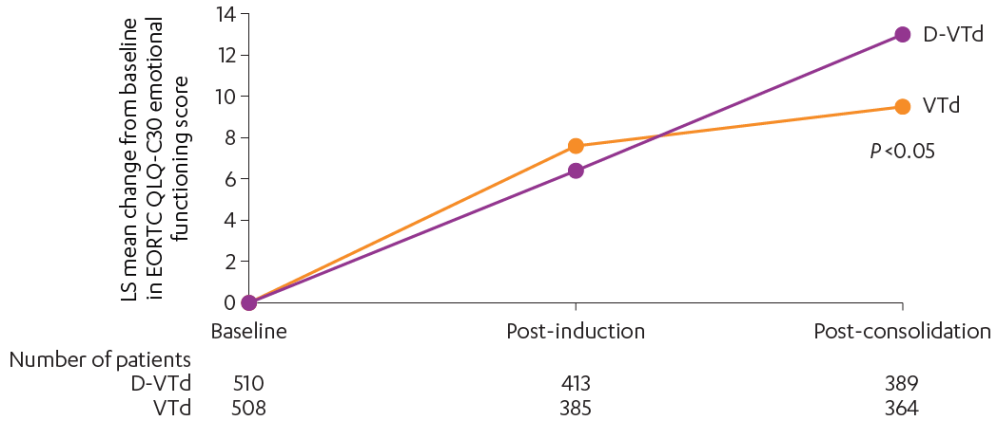


Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Core Cancer Quality of Life Questionnaire-C30; LS = least-squares; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

The reduction in pain symptoms score was clinically meaningful for both DBTd and BTd (exceeding a 15.7 point threshold for clinical significance),(84) with a particularly pronounced LS mean change from baseline over 20 points in the DBTd group, suggesting a large reduction in pain post-consolidation.(80) The proportion of patients using analgesics in the DBTd and BTd groups were similar (91.2% vs 92.1% respectively), indicating pain reduction was not confounded by use of concomitant pain management.(44)

For the EORTC QLQ-C30 functional scales, a statistically significant improvement in emotional functioning was reported in the DBTd group compared with that in the BTd group post-consolidation (LS mean change from baseline 13.0 vs 9.5 respectively; p=0.0131)(Figure 22).(44, 80)

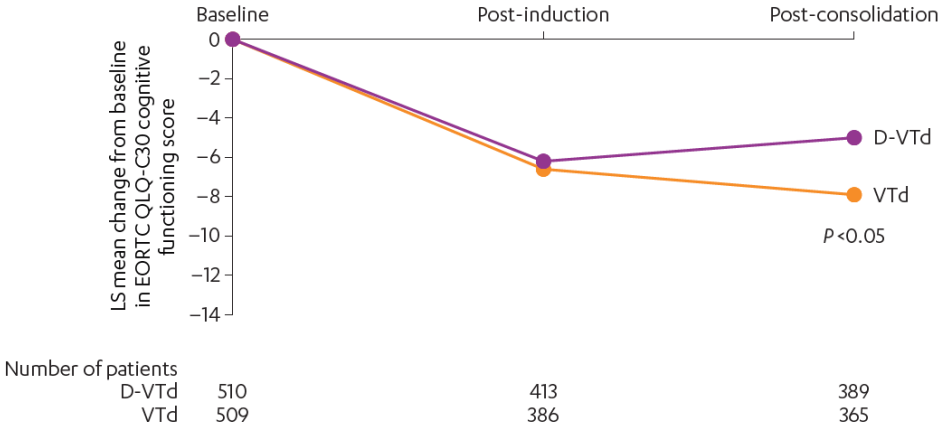
Figure 22: EORTC QLQ-C30 change from baseline in emotional function subscale scores (mixed effects model for repeated measures)(80)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Core Cancer Quality of Life Questionnaire-C30; LS = least-squares; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

Use of DBTd was associated with significantly less decline in cognitive function compared with BTd at Day 100 post ASCT (LS mean change from baseline -5.0 vs -7.9, respectively; p=0.0358) (Figure 23).(44, 80)

Figure 23: EORTC QLQ-C30 change from baseline in cognitive function subscale scores (mixed effects model for repeated measures)(80)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Core Cancer Quality of Life Questionnaire-C30; LS = least-squares; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

While a decline in cognitive function was observed in both DBTd and BTd groups, the mean change from baseline was not clinically meaningful based on the pre-specified threshold of 10 points or the 0.5 standard deviation threshold calculated using distribution-based criteria in the clinical trial population.(80)

Least square mean changes from baseline were not statistically significantly different between treatment groups for the other function (physical, role and social) and symptom scales (fatigue and nausea and vomiting). For further details refer to Appendix L.

EQ-5D-5L

Both EQ-5D-5L utility and VAS scores were comparable at baseline for patients treated with DBTd and BTd (refer to Appendix L for EQ-5D-5L utility). Over the course of treatment there was an improvement in EQ-5D-5L utility and VAS, measured by the LS mean change from baseline at Cycle 4 Day 28 and post-consolidation (Day 100 post-ASCT). Improvements were similar between the DBTd and BTd groups (Table 28).(80)

The EQ-5D provides a single measure across multiple domains of health and therefore does not highlight the benefits of treatment on specific aspects of health which may be most meaningful for patients. For example, although no statistically significant differences in EQ-5D-5L were observed between treatment arms, statistically significant and clinically meaningful reductions in pain and improvements in emotional functioning were observed for DBTd compared with BTd, as assessed by EORTC QLQ-C30.

Table 28: EQ-5D-5L utility and VAS change from baseline (ITT population)(80)

	DBTd LS Means of Change from Baseline (95% CI)	BTd LS Means of Change from Baseline (95% CI)	Difference Mean (95% CI)	P-values
Utility score				
Cycle 4 Day 28	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)	0.0 (-0.02, 0.02)	0.9695
Post-consolidation	0.17 (0.14, 0.19)	0.16 (0.13, 0.18)	0.01 (-0.01, 0.04)	0.2946
VAS				
Cycle 4 Day 28	2.7 (0.5, 4.8)	2.2 (0.1, 4.4)	0.4 (-1.8, 2.7)	0.7014
Post-consolidation	8.6 (6.5, 10.8)	7.7 (5.5, 9.9)	0.9 (-1.4, 3.2)	0.4408

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EQ-5D-5L = EuroQol-5D, 5 levels; ITT = intention-to-treat; LS = least-squares; VAS = visual analogue scale.

B.2.7 *Subgroup analysis*

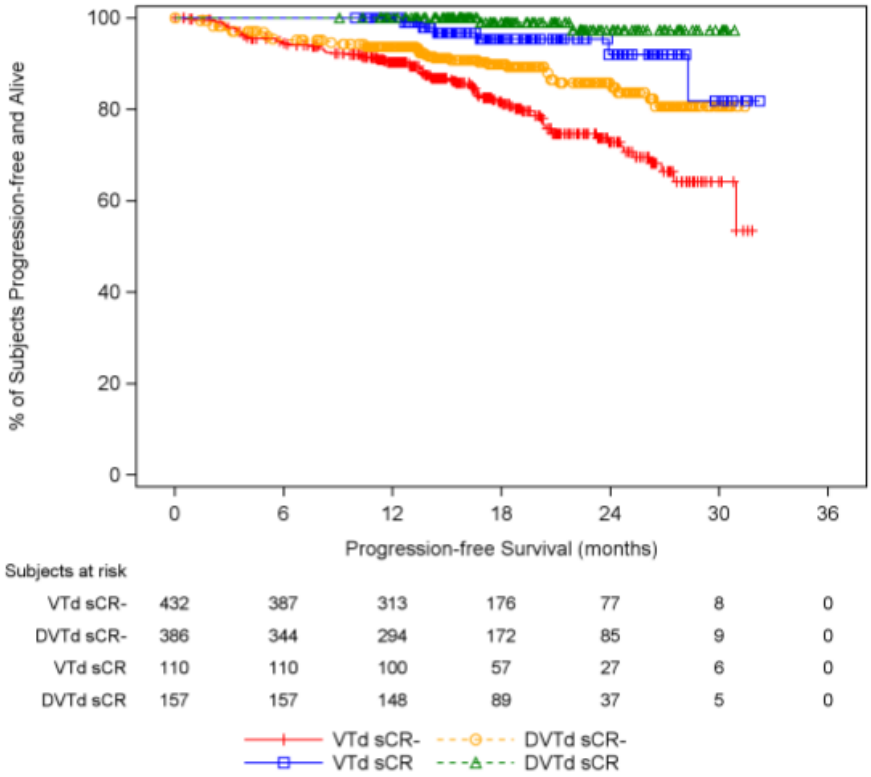
B.2.7.1 Response analyses

Pre-specified subgroup analysis of post-consolidation sCR demonstrated that the treatment effect was consistent across all patient subgroups analysed, with the exception of patients with a high-risk cytogenetic profile or ISS disease stage III (albeit an odds ratio of 1.07 for ISS disease stage III still favoured DBTd).(43) However, the CASSIOPEIA trial was not powered to assess differences between treatment arms in the each of the subgroups. Both subgroups included small numbers of patients (ISS Stage III: DBTd group n=84, BTd group n=81; high risk: DBTd group n=82, BTd group n=86) and the CIs for the ORs in these subgroups were wide (ISS Stage III: 0.54, 2.12; high risk: 0.42, 1.66).(43, 44)

After a median follow-up of 29.2 months, updated analysis of sCR demonstrated a benefit for all patient subgroups treated with DBTd compared to BTd which is in line with recently published data from CASTOR and POLLUX for RRMM that showed that longer observation is needed to show a benefit for the difficult to treat population with high-risk features.(45, 85, 86)

Importantly, post-hoc subgroup analysis based on response demonstrated that achievement of sCR was associated with prolonged PFS, supporting its clinical importance as a relevant prognostic factor (Figure 24).

Figure 24: Kaplan-Meier plot of PFS among patients who did and did not achieve sCR (ITT population, median follow-up = 18.8 months)(43)



Key: DVTd = daratumumab, bortezomib, thalidomide and dexamethasone (referred to as DBTd throughout the submission); ITT = intention-to-treat; MRD = minimal residual disease; MFC = multiparametric flow cytometry; VTd = bortezomib, thalidomide and dexamethasone (referred to as BTd throughout the submission).

Refer to Appendix E for further details on subgroup analysis for sCR.

Subgroup analysis of MRD

MRD-negative rates were consistently in favour of DBTd across all patient subgroups (including ISS disease stage III and patients with high-risk cytogenetic profile at trial entry) in prespecified subgroup analyses.(43, 44, 87)

Refer to Appendix E for further details on subgroup analysis for MRD.

B.2.7.2 Survival analyses

Prespecified subgroup analyses of PFS indicated similar PFS benefits with DBTd compared with BTd across patient subgroups, including patients with a high-risk cytogenetic profile or ISS disease stage III.(43) The Post-hoc Interim Analysis demonstrated that the PFS benefit for DBTd across all patient subgroups was maintained with longer study follow-up.(45)

Refer to Appendix E for further details on subgroup analysis for PFS.

B.2.8 *Meta-analysis*

As only one relevant trial evaluating DBTd was identified as part of the SLR, no meta-analysis is required.

B.2.9 *Indirect and mixed treatment comparisons*

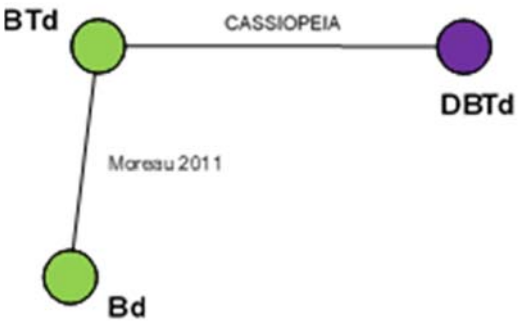
B.2.9.1 Overview of the indirect treatment comparisons conducted

As discussed in Section B.1.3, BTd represents SOC as induction treatment for newly diagnosed MM patients eligible for ASCT in England. For completeness, in this section, we present indirect evidence versus BCd and Bd which we understand are regimens used for a minority of patients where thalidomide is not considered suitable (e.g. due to challenging thrombosis or baseline neuropathy/neurotoxicity). Clinical expert opinion following the recent advisory board meeting involving three UK clinicians is that BTd is broadly comparable with BCd and superior to Bd.(1) This is supported by a naïve comparison of survival outcomes from a real-world evidence study utilising the Public Health England (PHE) datasets which also indicated inferior survival outcomes for Bd compared to BTd, as described in Section B.2.9.5 below.

There are no clinical studies directly comparing the efficacy of DBTd with either BCd or Bd. The feasibility of conducting a network meta-analysis (NMA) was explored to generate indirect evidence that could inform the comparison. To identify studies of daratumumab and potential comparator therapies for NDTE MM, a SLR of randomised clinical trial evidence was conducted. The original SLR was performed for the period 2015–2018 with an SLR update covering the period 2018–2020. In total, the SLRs identified 14 studies (including CASSIOPEIA) evaluating therapies for NDTE MM relevant to England (refer to Appendix D for the full list of studies and publications), of which five provided further supportive clinical evidence for BTd (IFM 2007-02, GIMEMA-MMY-3006, Ludwig et al. [2012], PETHEMA, and IFM 2013-04, in addition to CASSIOPEIA).(88-92) Of these, one study directly compared BTd to BCd (IFM 2013-04) and one study compared BTd to Bd (IFM 2007-02).(88, 92)

Major differences in the study design, including the timing of response assessment, response criteria used, single versus double transplant and maintenance (Y/N) meant that a NMA based on response outcomes was not possible (refer to Appendix D). Similarly, due to the incomplete and heavily restricted network, generating indirect evidence for survival outcomes (PFS/OS) via a standard NMA approach was not feasible. For the comparison of DBTd against BCd, no network was identified for either PFS or OS as data for these outcomes were not reported in the IFM2013-04 trial publication.(92) For Bd, a network using BTd as a common comparator for Bd and DBTd was possible for PFS only, as depicted in Figure 25, but not for OS (outcome not reported).(88)

Figure 25: PFS Network of Evidence



Key: Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

Unanchored Matching-Adjusted Indirect Comparison

In the absence of a viable network of studies with sufficient comparability to inform an NMA, unanchored matching adjusted indirect comparisons (MAIC) were explored. Given the use of a

response-based modelling approach for this submission (refer to Section B.3), the feasibility of conducting MAIC of response outcomes to inform the model was assessed. Specifically, the studies included in the SLR that included BCd or Bd were reviewed for whether they reported post-consolidation response, with a primary focus on MRD negativity, which is the response outcome used in the economic model. However, only data from the CASSIOPEIA trial reported data on post-consolidation MRD relevant to inform the economic analyses (refer to Appendix D). One study was identified in the SLRs that reported post-consolidation sCR/CR following BCd induction, however this was a small (n=64), single-arm, Japanese phase II study with no UK patients.(93) Furthermore, in this study, two cycles of BTd consolidation were administered and therefore the post-consolidation results are not reflective of BCd induction/consolidation therapy.

MAIC for survival outcomes (PFS and OS) were also explored in order to provide an estimate of the relative efficacy of DBTd and BTd versus BCd and Bd.(81) The 14 studies identified through the SLR were first assessed for feasibility to inform an unanchored MAIC for these outcomes (refer to Table 29). The following phase III studies were included in the comparisons, in addition to CASSIOPEIA:

- IFM 2005-01 (Bd vs VAd): patients received Bd induction (four 21-day cycles) with or without dexamethasone cyclophosphamide-etoposide-cisplatin (DCEP) consolidation therapy, or vincristine, doxorubicin-dexamethasone (VAd) induction (four 28-day cycles) with or without DCEP consolidation (two 28-day cycles); all patients with a partial response or better post-ASCT were to receive 2 months (or cycles) consolidation with lenalidomide followed by lenalidomide maintenance or placebo until relapse.(75)
- GMMG-MM5 (BCd vs PAd): patients received BCd induction (three 21-day cycles) or doxorubicin-dexamethasone (PAd; three 28-day cycles) followed by ASCT and then lenalidomide consolidation (two 28-day cycles) and maintenance for 2 years (LEN-2Y) or until CR (LEN-CR).(94, 95)

As both PFS and OS are influenced by differences in the maintenance therapies used, a comparison of the induction therapies alone was not possible. Instead, a comparison of the trials' treatment overall schemas, adjusted for population differences was explored.

The MAIC used IPD from CASSIOPEIA and published aggregate information on baseline characteristics and outcomes including KM curves from the IFM 2005-01 and GMMG-MM5 trials. The MAIC analysis followed the method described by Signorovitch et al. (2012)(96) and guidelines from the NICE Decision Support Unit (DSU).(97) This process involved the following three key steps:

1. Deriving balancing weights and applying them to estimate the average baseline characteristics that match the published aggregate characteristics of the comparator populations
2. Comparing adjusted outcomes for CASSIOPEIA versus comparators (IFM 2005-01 or GMMG-MM5)
3. Quantifying the relative treatment effect of CASSIOPEIA versus comparators (IFM 2005-01 or GMMG-MM5) across balanced study populations

Further details of the MAIC methodology are described in Appendix D.

B.2.9.2 Identification of relevant studies

Refer to Appendix D for full details of the original clinical SLR and SLR update.

Combined, the original SLR and SLR update identified 14 studies that investigated treatments for NDTE MM. Each study was reviewed for its suitability for inclusion in a MAIC, with consideration being given to the data reported (e.g. KM data for OS and PFS) and the comparability of baseline characteristics. Following this review, it was determined that three of the 14 studies identified were suitable for inclusion: CASSIOPEIA, IFM 2005-01 and GMMG-MM5.

The remaining studies and the reason for exclusion are presented in Table 29.

Table 29: Excluded studies and rationale for exclusion from the MAIC

Author, year	Trial name or clinical data source	Treatments	Reason for exclusion
Moreau, 2011(88)	IFM 2007-02	BTd vs. Bd	OS not reported
Kumar, 2012(98)	EVOLUTION	BLd vs. BCd	Neither PFS nor OS reported
Cavo, 2010(89)	GIMEMA-MMY-3006	BTd vs. Td	CASSIOPEIA used as primary evidence source for BTd efficacy
Ludwig, 2012(90)	NR	BTd vs. CBTd	
Rosinol, 2012(91)	PETHEMA	BTd vs. Td	
Moreau, 2016(92)	IFM 2013-04	BTd vs. BCd	Neither PFS nor OS reported
Gregersen, 2017(99)	CLAIM	Clarithromycin + BCd vs. Placebo + BCd	Neither PFS nor OS reported; BCd as placebo + BCd
El-Ghammaz, 2016(100)	NR	VAD vs. Bd	Survival outcomes only reported separately for patients with or without 17p deletion and/or t(4;14)
Kumar, 2019(101)	NR	BLd vs. BCd	Neither PFS nor OS reported
Sunami, 2019(93)	JSCT MM12	BCd	Single arm, phase II
Tanaka, 2019(102)	NR	BCd	Single arm, phase II

Key: Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BLd = bortezomib, lenalidomide and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; OS = overall survival; PFS = progression-free survival; VAD = vincristine, doxorubicin, and dexamethasone.

A summary of the three trials used to provide evidence for relevant treatments available in England and included in the MAICs is provided in Table 30.

Table 30: Summary of the trials used in the MAICs

	BTd	DBTd	BCd	Bd	Data source
CASSIOPEIA	Yes	Yes	-	-	MMY3006(43)
IFM 2005-01	-	-	-	Yes	Harousseau, 2010(75)
GMMG-MM5	-	-	Yes	-	Mai, 2015(95)

Key: Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

Quality assessment

The quality assessment process followed the Cochrane Risk-of-Bias assessment tool.(103) Key issues surrounding biases were taken into account when interpreting the results of the RCTs. Results of the quality assessment for studies included in the MAICs are presented in Table 31, with full results presented in Appendix D.

Table 31: Quality assessment scores

Trial name or clinical data source	Overall risk of bias
CASSIOPEIA(81)	Low
IFM 2005-01(75)	Low
GMMG-MM5(95)	Low

B.2.9.3 Compatibility of included studies and matching

In addition to quality assessment, compatibility assessment was performed to determine the feasibility of conducting MAICs with the available data. Compatibility assessment included a comparative review of the trial design, population profiles and outcome measures of the relevant studies: CASSIOPEIA, IFM 2005-01, and GMMG-MM5.

Baseline characteristics

Table 32 presents a comparison of the baseline patient characteristics for CASSIOPEIA, IFM 2005-01, and GMMG-MM5. Overall, the studies appear similar in most baseline characteristics. Notable differences include ISS stage and creatinine with a lower proportion of ISS stage III in CASSIOPEIA compared to both IFM 2005-01 and GMMG-MM5. Similarly, CASSIOPEIA had a slightly lower mean (78.9 DBTd, 76.2 BTd) and median (75 DBTd, 73 BTd) µmol/L creatinine compared to IFM 2005-01 (106.4 mean and 87 median) and GMMG-MM5 (median of 1 mg/dL, which is approximately equivalent to 88 µmol/L). Differences in baseline characteristics including ISS staging and creatinine were able to be adjusted for in the MAIC.

Table 32: MAIC Baseline characteristics(81)

	DBTd (CASSIOPEIA) (N=543)	BTd (CASSIOPEIA) (N=542)	Bd (IFM 2005-01) (N=240)	BCd (GMMG-MM5) (N=251)
Male, n (%)	319 (58.9)	316 (58.2)	139 (57.9)	153 (60.9)
Age				
≥65, n (%)	43 (7.9)	38 (7.0)	NR	NR
<65, n (%)	499 (92.1)	505 (93.0)	NR	NR
Mean (SD)	56.5 (7.0)	56.8 (6.93)	55.4 (NR)	NR
Median (min - max)	58 (26–65)	59 (22–65)	57.2 (NR)	58.7 (33–70)
ISS stage, n (%)				
I	228 (42.1)	204 (37.6)	102 (42.5)	94 (37.5)
II	233 (43.0)	255 (47.0)	81 (33.8)	82 (32.7)
III	81 (14.9)	84 (15.5)	52 (21.7)	75 (29.9)
Not determined/missing	0	0	5 (2.1)	NR
ECOG, n (%)				
0	257 (47.4)	265 (48.8)	NR	114 (45.4)
1	230 (42.4)	225 (41.4)	NR	116 (46.2)
2	55 (10.1)	53 (9.8)	NR	17 (6.8)

3	0	0	0	4 (1.6)
B2-microglobulin (mg/L)				
Missing, n	0	0	NR	NR
Median (min - max)	3.25 (1.2–21.2)	3.2 (1.2–18.4)	NR	NR
>3 mg/L, n (%)	300 (55.4)	296 (54.5)	137 (57.1)	NR
Cytogenetics, n (%)				
del(13) by FISH	NR	NR	101 (42.1)	NR
t(4;14) and/or del(17p)	86 (17.1)	82 (16.4)	40 (16.7)	NR
Adverse cytogenetics del17p, n (%)				
Performed	503 (100)	501 (100)	NR	222 (100)
Positive (% performed)	39 (7.8)	42 (8.4)	NR	23 (10.4)
Missing	39 (7.2)	42 (7.7)	NR	29 (11.6)
t(4;14), n (%)				
Performed	503 (100)	501 (100)	NR	219 (100)
Positive (% performed)	53 (10.5)	51 (10.2)	NR	22 (10.1)
Missing	39 (7.2)	42 (7.7)	NR	32 (12.8)
Hemoglobin (g/dL)				
Mean	11.5	11.2	10.9	NR
Median (min - max)	11.5 (5.9–17)	11.1 (7.0–16.1)	10.9	10.7 (6.0–16.3)
Anaemia				
Hb <10 g/dL or 2 g/dL <normal[a], n (%)	191 (35.2)	223 (41.1)	NR	138 (55)
Creatinine				
Mean (µmol/L)	78.9	76.2	106.4	NR
Median (µmol/L)	75	73	87	88
Median (Min - Max) (mg/dL)	0.8 (0.1–2.7)	0.8 (0.1–2.4)	NR	1 (0.4–11.3)
Renal insufficiency				
Creatinine >177 µmol/L, n (%)	2 (0.4)	1 (0.2)	NR	39 (15.5)
Calcium (mmol/L)				
Mean	2.4	2.4	2.4	NR
Median (min-max)	2.4 (1.8–3.7)	2.4 (0.2–3.6)	2.4	2.4 (1.7–5.4)
Missing, n (%)	22 (4.1)	9 (1.7)	NR	0
Calcium elevation				

Calcium >2.65 mmol/L, n (%)	38 (7.0)	55 (10.1)	NR	31 (12.3)
Missing	22 (4.1)	9 (1.7)	NR	0
LDH (serum), n (%)				
≤ULN	344 (63.5)[b]	302 (55.6)	NR	207 (82.5)
>ULN	189 (34.9)[b]	226 (41.6)	NR	44 (17.5)
Unknown	9 (1.7)[b]	15 (2.8)	NR	0 (0)
Platelets (per nL)				
Median (min–max)	250 (22–584)	241 (49–999)	NR	240 (22–533)
Bone disease				
Lytic lesions or myeloma-related osteopenia / osteoporosis, n (%)	462 (85.2)	465 (85.6)	NR	223 (88.8)
Missing	2 (0.4)	3 (0.6)	NR	NR
Heavy-chain isotype /Type of myeloma by Immunofixation, n (%)				
IgG	333 (61.4)	351 (64.6)	NR	148 (59.0)
IgA	104 (19.2)	87 (16.0)	NR	51 (20.3)
LCD	66 (12.2)	83 (15.3)	NR	47 (18.7)
Other	39 (7.2)	22 (4.1)	NR	5 (2.0)
Light-chain isotype, n (%)				
Kappa	NR	NR	NR	160 (63.8)
Lambda	NR	NR	NR	91 (36.2)
Gain 1q21 (>2 copies), n (%)				
Performed	NR	NR	NR	213 (100)
Positive (% performed)	NR	NR	NR	79 (37.1)
Missing	NR	NR	NR	38 (15.1)

[a] Calcium elevation, renal impairment, anaemia and bone disease are defined according to CRAB criteria for symptomatic MM.
[b] ULN for LDH was not reported for GMMG-MM5 trial; CASSIOPEIA was defined using patient-dependent cut-offs of 213 U/L or 225 U/L.
Key: BA = base case analysis; Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; ESS = effective sample size; ISS = International Staging System; FISH = fluorescent in situ hybridization; SA = sensitivity analysis; NR = Not reported; LDH = lactate dehydrogenase; ECOG = Eastern Cooperative Oncology Group; LCD = light-chain disease; ULN = upper limit of normal.

Overall, results from the compatibility assessment indicated that both IFM 2005-01 and GMMG-MM5 included a potentially higher risk population compared to CASSIOPEIA. The trials, however, had similar designs, comparable eligibility criteria and sufficient overlap in most baseline characteristics to conduct an MAIC analysis and avoid the need to investigate an alternative simulated treatment comparison (STC) approach.

Refer to Appendix D for full details of the compatibility assessment including a visual comparison of the trial designs.

MAIC matching: DBTd/BTd versus Bd

Table 33 presents the baseline characteristics before and after matching DBTd/BTd (CASSIOPEIA) to Bd (IFM 2005-01). After weighting, all baseline characteristics included in the base-case for matching were balanced for DBTd versus Bd.

The variables for adjustment in the base case MAIC analysis included all baseline characteristics commonly reported by the two studies except creatinine: age (mean and median), sex (proportion of males), ISS stage (proportion of patients with each ISS stage category), beta-2 macroglobulin (proportion of patients with >3 mg/L), cytogenetic profile (proportions of patients with t(4;14) and/or del(17p) cytogenetic profile), haemoglobin (mean and median), and calcium (mean and median). Creatinine was excluded due to a lack of similarity (or overlap) of the reported mean and median values between IFM-2005-01 and CASSIOPEIA, resulting in substantial ESS reduction (90% for DBTd and 77% for BTd). Adjusting for the full list of available baselines characteristics, including creatinine (mean and median), was however conducted as a sensitivity analysis.

Table 33: Baseline characteristics before and after matching DBTd/BTd to Bd(81)

Characteristic	Before Matching		After Matching		Target
	DBTd (n=543)	BTd (n=542)	DBTd (ESS_{BA}=415/ ESS_{SA}=57)	BTd (ESS_{BA}=393/ ESS_{SA}=122)	Bd (n=240)
Age (years)					
Mean	56.8	56.5	55.4	55.4	55.4
Median (Min–Max)	59.0 (22 - 65)	58.0 (26 - 65)	58.0 (22 - 65)	57.5 (26 - 65)	57.2 (NR–≤65)
Gender, %					
Male	58.2	58.9	57.9	57.9	57.9
ISS Stage, %					
I	37.6	42.1	43.4	43.4	43.4 (102/235) [a]
II	47	43	34.5	34.5	34.5 (81/235) [a]
III	15.5	14.9	22.1	22.1	22.1 (52/235) [a]
Not determined	0	0	0	0	0 [a]
B2-microglobulin, %					
>3 mg/L	54.5	55.4	57.1	57.1	57.1
Abnormal Cytogenetics, %					
t (4;14) and/or del(17p)	16.4	17.1	16.7	16.7	16.7
Missing	7.7	7.2	0	0	0
del (13) by FISH	NR	NR	NR	NR	42.1
Haemoglobin (g/dL)					
Mean	11.2	11.5	10.9	10.9	10.9

Median (Min–Max)	11.1 (7-16)	11.5 (5.9-17)	10.8 (NR-NR)	10.9 (NR-NR)	10.9 (NR-NR)
Creatinine$_{BA}$ (µmol/L)					
Mean	76.2	78.9	76.8	80.2	106.4
Median (Min–Max)	73.0 (5-213)	75.0 (6.5-235.1)	74.0 (NR-NR)	76.0 (NR-NR)	87.0 (NR-NR)
Creatinine$_{SA}$ (µmol/L)					
Mean	76.2	78.95	106.4	106.4	106.4
Median (Min–Max)	73.0 (NR-NR)	75.0 (NR-NR)	87.0 (NR-NR)	87.0 (NR-NR)	87.0 (NR-NR)
Calcium (mmol/L)					
Mean	2.4	2.4	2.4	2.4	2.4
Median (Min–Max)	2.4 (0.2-3.6)	2.4 (1.8-3.7)	2.4 (NR-NR)	2.4 (NR-NR)	2.4 (NR-NR)
Missing	1.7	4.1	0	0	0

Key: BA = base case analysis; Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ESS = effective sample size; ISS = International Staging System; FISH = fluorescent in situ hybridization; SA = sensitivity analysis; NR = Not reported.
[a] % were calculated out of patients with non-missing values (240-5=235) for Bd.

In the base-case weighting model, the effective sample size (ESS) was reduced from the original sample size by 128 (24%) for DBTd and 149 (27%) for BTd. In the base-case weighting model for DBTd, the rescaled weights ranged from 0 to 3.92, and median of 0.96. Importantly, the distribution did not reveal the presence of patients with very large weights suggesting outliers (refer to Appendix D). A similar weight distribution was observed for BTd with a range of 0 to 4.21 and median of 0.95 (refer to Appendix D). Some patients received a weight of 0 because they were excluded from the matching either due to exclusion criteria or due to missing values in the covariates being matched.

In the sensitivity analysis, which also matched on creatinine, ESS was reduced from the original sample size by 90% for DBTd and 77% for BTd. The distribution of the rescaled weights for DBTd was heavily skewed to the right (median of 0.47) and ranged from 0 to 60.17, revealing the presence of patients with very large weights (refer to Appendix D). Similarly, the distribution of rescaled weights for the BTd arm was heavily skewed to the right (median of 0.66) and ranged between 0 to 27.73, revealing the presence of patients with very large weights (refer to Appendix D). Therefore, the results from the sensitivity analyses should be interpreted with caution.

MAIC matching: DBTd/BTd versus BCd

Table 34 presents the baseline characteristics before and after matching DBTd/BTd (CASSIOPEIA) to BCd (GMMG-MM5). After weighting, all commonly reported baseline characteristics between CASSIOPEIA and the GMMG-MM5 study except anaemia and renal insufficiency were balanced for DBTd versus BCd.

The variables for adjustment in the base case MAIC analysis included all baseline characteristics commonly reported by the two studies except anaemia and renal insufficiency: age (median), sex (proportion of males), ISS stage (proportion of patients with each ISS stage category), ECOG/WHO status (proportion of patients with 0, 1, 2 but not 3), cytogenetic profile (proportions of patients with t(4;14) and/or del(17p) abnormality), creatinine (median), bone disease (proportion of patients with bone disease), calcium (median and proportion with calcium elevation (calcium >2.65 mmol/L), haemoglobin (median and range), platelets (median and range), LDH (proportion with >ULN), and heavy-chain isotype (proportion of patients with IgG, IgA, LCD, Other).

Anaemia was excluded due to lack of overlap (or similarity) in the reported values between GMMG-MM5 and CASSIOPEIA, resulting in substantial ESS reduction (51% for DBTd and 50% for BTd) after matching. Based on clinical feedback, it was determined that anaemia was not a critical aspect of prognosis compared to other factors and could be excluded from the base case analysis; mean haemoglobin concentration and platelet count were adjusted instead. Of note, as there was only one patient in each arm in CASSIOPEIA with renal insufficiency, this baseline characteristic could not be adjusted for. Differences in LDH between the two studies also posed a concern about potential substantial ESS reductions. LDH was based on local lab in CASSIOPEIA, whereas in GMMG-MM5, it was not reported. There is no uniform ULN for LDH. However, based on clinical feedback, it was determined that LDH was an important prognostic factor and should be included in the matching model.

Table 34: Baseline characteristics before and after matching DBTd/BTd to BCd(81)

Characteristic	Before Matching		After Matching		Target
	DBTd (N=543)	BTd (N=542)	DBTd (ESS_{BA}=206/ ESS_{SA1}=196)/ ESS_{SA2}=272)	BTd (ESS_{BA}=211/ ESS_{SA1}=207/ ESS_{SA2}=272)	BCd (N=251)[a]
Age (years)					
Median (Min–Max)	59.0 (22 - 65)	58.0 (26 - 65)	58.0 (35 - 65)	58.0 (34 - 65)	58.7 (33 - 70)
Gender, %					
Male	58.2	58.9	61.0	61.0	61.0
ECOG/WHO Performance Status, %					
0	48.8	47.4	45.4	45.4	45.4 (114/251)
1	41.4	42.4	46.2	46.2	46.2 (116/251)
2-3	9.8	10.1	8.4	8.4	8.4 (21/251)
Heavy-chain isotype /Type of myeloma by Immunofixation, %					
IgG	64.6	61.4	59	59	59
IgA	16	19.2	20.3	20.3	20.3
LCD	15.3	12.2	18.7	18.7	18.7
Other	4.1	7.2	2	2	2
Calcium elevation (calcium >2.65 mmol/L), %					
Yes	10.1	7	12.3	12.3	12.3
Missing	1.7	4.1	0	0	0
Renal insufficiency (creatinine >177 µmol/L)$_{BA}$, %					
Yes	0.2	0.4	0.8	0.4	15.5
Renal insufficiency (creatinine >177 µmol/L)$_{SA1}$, %					
Yes	0.2	0.4	0.8	0.4	15.5
Renal insufficiency (creatinine >177 µmol/L)$_{SA2}$, %					
Yes	0.2	0.4	0.5	0.4	15.5
Anaemia (Hb <10 g/dL or 2 g/dL <normal)$_{BA}$, %					
Yes	41.1	35.2	46.9	50.5	55

Anaemia (Hb <10 g/dL or 2 g/dL <normal)$_{SA1}$, %					
Yes	41.1	35.2	55	55	55
Anaemia (Hb <10 g/dL or 2 g/dL <normal)$_{SA2}$, %					
Yes	41.1	35.2	46.6	48.7	55
Bone disease (lytic lesions[†]), %					
Yes	85.6	85.2	88.8	88.8	88.8
Missing	0.6	0.4	0	0	0
ISS Stage, %					
I	37.6	42.1	37.4	37.4	37.4
II	47	43	32.7	32.7	32.7
III	15.5	14.9	29.9	29.9	29.9
LDH (serum)$_{BA}$, %					
>ULN	41.6	34.9	17.5	17.5	17.5
Unknown	2.8	1.7	0	0	0
LDH (serum)$_{SA1}$, %					
>ULN	41.6	34.9	17.5	17.5	17.5
Unknown	2.8	1.7	0	0	0
LDH (serum)$_{SA2}$, %					
>ULN	41.6	34.9	42.1	41.1	17.5
Unknown	2.8	1.7	2.0	0.9	0
Adverse cytogenetics - del17p, %					
Performed	92.3	92.8	100	100	100 (222/251)
Positive (% performed)	8.4	7.8	10.4	10.4	10.4 (23/222)
Missing	7.7	7.2	0	0	11.6 (29/251)
Adverse cytogenetics – t (4;14), %					
Performed	92.3	92.8	100	100	100 (219/251)
Positive (% performed)	8.4	10.5	10	10	10.0 (23/219)
Missing	7.7	7.2	0	0	12.7 (32/251)
Calcium (serum, mmol/L)					
Median (Min–Max)	2.4 (0.2 - 3.6)	2.4 (1.8 - 3.7)	2.4 (1.8 - 3.4)	2.4 (1.8 - 3.7)	2.4 (1.7 - 5.4)
% above 2.4	42.5	39.7	50	50	50
Missing	1.7	4.1	0	0	0
Creatinine (serum, mg/dL)					
Median (Min–Max)	0.8 (0.1 - 2.4)	0.8 (0.1 - 2.7)	1.0 (0.4 - 2.4)	1.0 (0.4 - 2.7)	1.0 (0.4 – 11.4)

[†] Or myeloma-related osteopenia/osteoporosis (Mai et al. 2015). Derived from CASSIOPEIA trial data as 'Baseline Presence of Diffuse Myeloma-related Osteopenia' or 'Baseline Number of Lytic Bone Lesions >1'.

% above 1.0	23	27.5	50	50	50
Hemoglobin (g/dL)					
Median (Min–Max)	11.1 (7.0 - 16.1)	11.5 (5.9 - 17.0)	10.6 (7.1 - 16.1)	10.7 (7.9 - 16.0)	10.7 (6.0 - 16.3)
% above 10.7	58.9	65.7	50	50	50
Platelets (per nL)					
Median (Min–Max)	241.0 (49.0 - 999.0)	250 (22 - 584)	238.0 (49.0 - 525.0)	239 (70 - 519)	240 (22 – 533)
% above 240	50.6	55.9	50	50	50

Key: BA = Base-case analysis (all baseline characteristics reported except anaemia and renal insufficiency); BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; ECOG = Eastern Cooperative Oncology Group; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; LCD = light-chain disease; SA1 = Sensitivity Analysis 1 (all baseline characteristics reported in both studies except renal insufficiency); SA2 = Sensitivity Analysis 2 (all baseline characteristics reported except anaemia, LDH and renal insufficiency); WHO = World Health Organization.
[a] Source: Baseline characteristics for the BCd arm were extracted from Mai et al. 2015.

In the base-case weighting model, ESS was reduced from the original sample size by 337 (62%) for DBTd and 331 (61%) for BTd. In the base-case, the rescaled weights were mostly small with some skewness to the right (median of 0.58) without presence of very large outliers (range 0.00-9.92) for DBTd (refer Appendix D) and mostly small with some skewness to the right (median of 0.64) without presence of very large outliers (range 0.00-10.01) for BTd (refer Appendix D). In sensitivity analysis 1, which matched on all characteristics included in the base-case plus anaemia, the ESS was reduced from the original sample size by 64% for DBTd and 62% for BTd. In sensitivity analysis 2, which matched on all characteristics included in the base-case except LDH, the ESS was reduced from the original sample size by 50% for DBTd and 50% for BTd.

B.2.9.4 Results of the matching-adjusted indirect comparisons

MAIC PFS/OS results for DBTd/BTd versus Bd (CASSIOPEIA; median follow-up = 29.2 months)

Both a naïve comparison (DBTd from CASSIOPEIA versus Bd from IFM 2005-01) and the MAIC (base case) results [redacted], demonstrating [redacted] for patients treated with the daratumumab combination (Table 35). [redacted] were also seen in the naïve comparison (BTd from CASSIOPEIA and Bd from IFM 2005-01) and the MAIC (base case) for BTd versus Bd, although [redacted].

The MAIC sensitivity analysis adjusting for creatinine showed similar results to the base case analysis however results should be interpreted with caution due to a reduced ESS, less table point estimates, wider CIs and outliers in weight distributions. Refer to Appendix D for KM survival analysis (PFS and OS) for both the base case and sensitivity analysis.

Table 35: Results of the naive comparison and MAIC (DBTd/BTd versus Bd)(81)

	Naïve comparison		MAIC (Base case)		MAIC (Sensitivity analysis)	
	PFS	OS	PFS	OS	PFS	OS
DBTd vs Bd						
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BTd vs Bd						
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MAIC = matching-adjusted indirect comparison; OS = overall survival; PFS = progression-free survival.

The MAIC analysis may be limited by unreported or unobserved confounding factors that could not be adjusted for (e.g. del[13] was not collected in CASSIOPEIA), as well as differences in study design. These results should also be considered in the context of overall treatment schema: post-induction treatments varied between CASSIOPEIA and IFM 2005-01, and 21% of patients in IFM 2005-01 received a second transplant. That Bd is associated with worse survival outcomes compared to BTd, as shown in the results of the MAIC, is however supported by feedback from UK clinical experts and data from the PHE linked datasets (refer to Section B.2.9.5).

MAIC PFS/OS results for DBTd/BTd versus BCd (CASSIOPEIA; median follow-up = 29.2 months)

Both a naïve comparison (DBTd from CASSIOPEIA and BCd from GMMG-MM5) and the MAIC results (base case) [redacted] DBTd versus BCd, demonstrating [redacted] PFS and OS for patients treated with the daratumumab combination (Table 36). By contrast, there was [redacted] in PFS or OS for BTd versus BCd in the naïve comparison (BTd from CASSIOPEIA and BCd from GMMG-MM5) or the MAIC (base case). The MAIC sensitivity analyses showed similar results to the base case analysis. Whilst a non-statistically significant difference in OS between DBTd and BCd was seen in sensitivity analysis 2, this is likely due to not adjusting for LDH (elevated levels of which are associated with poorer outcomes). Refer to Appendix D for Kaplan-Meier survival analysis (PFS and OS) for both the base case and sensitivity analysis.

Table 36: Results of the naive comparison and MAIC (DBTd/BTd versus BCd)(81)

	Naïve comparison		MAIC (Base case)		MAIC (Sensitivity analysis 1)		MAIC (Sensitivity analysis 2)	
	PFS	OS	PFS	OS	PFS	OS	PFS	OS
DBTd vs BCd								
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BTd vs BCd								
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: BCd-LEN-2Y = bortezomib, cyclophosphamide and dexamethasone with 2-year lenalidomide maintenance; BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MAIC = matching-adjusted indirect comparison; OS = overall survival; PFS = progression-free survival

B.2.9.5 MAIC conclusion and external validation

MAIC conclusion

The MAIC approach may be limited by unreported or unobserved confounding factors that could not be adjusted for, as well as differences in study design and overall survival follow-up lengths. That is, the MAIC approach is susceptible to potential bias but remains a useful tool to understand the relative efficacy of DBTd and BTd versus other regimens used in clinical practice given a lack of direct or indirect evidence necessary to form a network. The MAIC results [redacted] of DBTd as induction and consolidation therapy with ASCT for MM patients who are transplant eligible and showed [redacted] results for BTd versus BCd, with BTd also having [redacted] versus Bd.

External validation

To better understand survival outcomes based on existing standard of care treatment options in England, and to complement the MAIC, Janssen commissioned a real-world evidence study. This study utilised the PHE datasets to investigate PFS/OS for NDTE MM patients (refer to Section B.3.3.2 for details).[III] In total, [redacted] patients who were newly diagnosed with MM in England between 1st January 2015 and 31st December 2018 (inclusive) and who had received ASCT were included in the analysis. Follow-up for this cohort was analysed to 31 December 2019 with linkage to the Systemic Anti-Cancer Therapy Chemotherapy (SACT) dataset for [redacted] patients. Naïve comparison of outcomes from the PHE datasets indicates that PFS/OS are [redacted], and [redacted] (Table 37), which is consistent with the findings from the MAICs described above.

[III] Response data is not routinely available within the cancer registry. As such, response was not included as part of the analysis of the PHE dataset.

Table 37: Naïve comparison of survival outcomes for BTd, BCd and Bd from PHE datasets(2)

Survival rates	% alive (95% CI)			% alive and progression free (95% CI)[a]		
Time (months)	12	24	36	12	24	36
All first-line treatments (ASCT-positive) (n=[redacted])	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BTd as first-line treatment (ASCT-positive) (n=[redacted])	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BCd as first-line treatment (ASCT-positive) (n=[redacted])	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Bd as first-line treatment (ASCT-positive) (n=[redacted])	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key: ASCT = autologous stem cell transplant, CI = confidence interval, Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; PHE = Public Health England.

Although naïve comparison is not without bias, as any confounding factors are not adjusted for, it provides a useful validation of the MAIC analysis. The fact that MAIC, naïve comparison (within a comprehensive UK based real world evidence source) and clinical expert opinion all agree that [redacted] (with both regimens having [redacted] versus Bd), brings confidence that this is indeed the case.(1)

As such, and given the similarity in costs of these generic regimens, a pragmatic approach may be taken to economic modelling. That is, if DBTd is considered cost-effective versus BTd, the daratumumab quadruplet combination is also likely to be cost-effective versus BCd and Bd. This pragmatic approach avoids the significant challenges of incorporating MAIC based comparative effectiveness of DBTd versus BCd and Bd into the response-based economic model.

B.2.10 *Adverse reactions*

Safety was analysed as a secondary outcome in CASSIOPEIA. No additional studies are available to provide evidence of safety and tolerability of DBTd. Results from CASSIOPEIA indicate that the safety profile of DBTd is consistent with the known safety profile of BTd and that of daratumumab as a monotherapy.(43)

Treatment exposure

The median treatment duration in CASSIOPEIA during Part 1 of the study was 8.9 months and 8.7 months for the DBTd and BTd groups respectively (Table 38).(43, 81) For both the DBTd and the BTd groups, the median number of treatment cycles was 6. Median dose intensities were similar for bortezomib, thalidomide and dexamethasone between treatment groups.(43, 44)

Table 38: Treatment exposure (CASSIOPEIA, safety analysis set)(44)

	DBTd (n=536)	BTd (n=538)
Median duration of treatment (months)	8.9	8.7
Number of treatment cycles, total, median (range)	6 (1; 6)	6 (1; 6)
Treatment cycles at induction stage, median (range)	4 (1; 4)	4 (1; 4)
Treatment cycles at consolidation stage, median (range)	2 (1; 2)	2 (1; 2)
Daratumumab relative dose intensity, induction/consolidation (%)		
Mean (SD)	98.38 (6.306)	-
Median	99.72	-
Q1, Q3	(97.76; 100.78)	-
Range	(7.3; 113.1)	-
Bortezomib relative dose intensity, induction/consolidation (%)		
Mean (SD)	91.5 (12.057)	91.31 (11.211)
Median	96.77	96.30
Q1, Q3	(87.02; 99.45)	(84.73; 99.17)
Range	(24.5; 105.7)	(49.2; 106.7)
Thalidomide relative dose intensity, induction/consolidation (%)		
Mean (SD)	86.6 (19.30)	86.1 (18.36)
Median	96.4	95.4
Q1, Q3	(79.2; 100.0)	(78.0; 100.0)
Range	(2; 150)	(0; 104)
Dexamethasone relative dose intensity, induction/consolidation (%)		
Mean (SD)	96.8 (10.14)	96.2 (11.84)
Median	100.0	100.0
Q1, Q3	(96.7; 100.0)	(96.7; 100.0)
Range	(13; 120)	(0; 125)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; SD = standard deviation.

TEAE overall

At median follow-up of 18.8 months, almost all patients treated with DBTd or BTd had at least one treatment-emergent adverse event (TEAE) after the start of treatment (99.8% and 99.6%, respectively).(44) Slightly higher rates of grade 3 and 4 TEAEs were observed in the DBTd group compared to the BTd group (80.6% vs 75.8%), principally driven by haematological events including neutropenia and lymphopenia.(44) Serious TEAEs were comparable between groups (46.8% for DBTd and 47.4% for BTd).(43, 44) The percentage of patients who discontinued treatment because of at least one TEAE was marginally lower for DBTd compared to BTd (7.5% and 8.4%, respectively), while TEAEs leading to death occurred in 1 patient (0.2%) in the DBTd group and 9 patients (1.7%) in the

BTd group.(43) These results show that the addition of daratumumab to standard of care BTd is not linked to decreased tolerability or safety concerns. A summary of TEAEs at 18.8 months of follow-up is provided in Table 39.

Further information regarding deaths in the CASSIOPEIA trial is presented in Appendix F.

Table 39: Summary of TEAEs[a] during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44) (43, 81)

	DBTd (n=536)	BTd (n=538)
Any TEAE, n (%)	535 (99.8%)	536 (99.6%)
Grade 3/4 TEAE, n (A%)	432 (80.6%)	408 (75.8%)
Serious TEAE, n (%)	251 (46.8%)	255 (47.4%)
TEAE leading to discontinuation, n (%)	40 (7.5%)	45 (8.4%)
TEAEs leading to death, n (%)	1 (0.2%)	9 (1.7%)

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; TEAE = treatment-emergent adverse event.
Note: Adverse events emerging during ASCT phase related to the planned procedures were not reported.
[a] TEAEs during induction, ASCT, or consolidation Treatment Phase; incidence reflects the number of patients experiencing at least one TEAE associated with at least one of the study treatments.

TEAE by preferred term

Overall, the safety profile was similar between treatment groups, including the incidence of TEAEs occurring in ≥10% of patients in either treatment group. However, a higher frequency (≥5% difference) was reported in the DBTd group for nausea (DBTd: 30.2%; BTd 24.2%), neutropenia (DBTd: 29.3%; BTd 16.5%), thrombocytopenia (DBTd: 20.3%; BTd: 13.6%), lymphopenia (DBTd: 18.5%; BTd: 12.5%), and cough (DBTd: 17.2%; BTd: 10.4%). Other most common TEAEs (≥20% in either group) were balanced between the two treatment groups, including peripheral sensory neuropathy, paraesthesia, constipation, asthenia, peripheral oedema, and pyrexia.(43, 44)

Frequently reported Grade 3 or 4 TEAEs (occurring in ≥10% of patients in either treatment group) were neutropenia, lymphopenia, stomatitis and thrombocytopenia (Table 40).(43, 44) The incidence of Grade 3 or Grade 4 TEAEs was increased for patients receiving daratumumab, driven by the haematological events of neutropenia and lymphopenia, which occurred more frequently in the DBTd group compared with the BTd group (neutropenia: 27.6% vs 14.7%; lymphopenia: 17.0% vs 9.7%). The increased rate of neutropenia in patients receiving daratumumab was not associated with any increased risk of neutropenic fever, as patients in the both treatment groups reported comparable levels of febrile neutropenia.(44)

Table 40: TEAEs[a] by MedDRA system organ class and preferred term during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44)

	DBTd (n=536)		BTd (n=538)	
	All grades (≥10%)	Grade3/4 (≥5%)	All grades (≥10%)	Grade3/4 (≥5%)
Blood and lymphatic system disorders	**303 (56.5%)**	**249 (46.5%)**	**253 (47.0%)**	**196 (36.4%)**
Neutropenia	157 (29.3%)	148 (27.6%)	89 (16.5%)	79 (14.7%)
Thrombocytopenia	109 (20.3%)	59 (11.0%)	73 (13.6%)	40 (7.4%)
Lymphopenia	99 (18.5%)	91 (17.0%)	67 (12.5%)	52 (9.7%)
Anaemia	73 (13.6%)	n/a	81 (15.1%)	n/a
Febrile neutropenia	n/a	36 (6.7%)	n/a	28 (5.2%)
Infections and infestations	**351 (65.5%)**	**n/a**	**306 (56.9%)**	**n/a**
Bronchitis	102 (19.0%)	n/a	66 (12.3%)	n/a
General disorders and administration site conditions	**414 (77.2%)**	**n/a**	**398 (74.0%)**	**n/a**
Asthenia	171 (31.9%)	n/a	155 (28.8%)	n/a
Oedema peripheral	162 (30.2%)	n/a	148 (27.5%)	n/a
Pyrexia	140 (26.1%)	n/a	114 (21.2%)	n/a
Fatigue	70 (13.1%)	n/a	86 (16.0%)	n/a
Gastrointestinal disorders	**431 (80.4%)**	**124 (23.1%)**	**416 (77.3%)**	**131 (24.3%)**
Constipation	272 (50.7%)	n/a	262 (48.7%)	n/a
Nausea	162 (30.2%)	n/a	130 (24.2%)	n/a
Diarrhoea	103 (19.2%)	n/a	89 (16.5%)	n/a
Vomiting	87 (16.2%)	n/a	52 (9.7%)	n/a
Stomatitis	86 (16.0%)	68 (12.7%)	104 (19.3%)	88 (16.4%)
Musculoskeletal and connective tissue disorders	**245 (45.7%)**	**n/a**	**252 (46.8%)**	**n/a**
Bone pain	70 (13.1%)	n/a	82 (15.2%)	n/a
Back pain	59 (11.0%)	n/a	55 (10.2%)	n/a
Nervous system disorders	**437 (81.5%)**	**73 (13.6%)**	**456 (84.8%)**	**73 (13.6%)**
Peripheral sensory neuropathy	314 (58.6%)	47 (8.8%)	340 (63.2%)	46 (8.6%)
Paraesthesia	118 (22.0%)	n/a	108 (20.1%)	n/a
Tremor	71 (13.2%)	n/a	58 (10.8%)	n/a

Psychiatric disorders	**141 (26.3%)**	**n/a**	**153 (28.4%)**	**n/a**
Insomnia	61 (11.4%)	n/a	78 (14.5%)	n/a
Anxiety	58 (10.8%)	n/a	46 (8.6%)	n/a
Respiratory, thoracic and mediastinal disorders	**259 (48.3%)**	**n/a**	**185 (34.4%)**	**n/a**
Cough	90 (16.8%)	n/a	49 (9.1%)	n/a
Dyspnoea	77 (14.4%)	n/a	66 (12.3%)	n/a
Skin and subcutaneous tissue disorders	**255 (47.6%)**	**n/a**	**222 (41.3%)**	**n/a**
Rash	86 (16.0%)	n/a	67 (12.5%)	n/a
Erythema	61 (11.4%)	n/a	47 (8.7%)	n/a

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; n/a = not applicable; TEAE = treatment-emergent adverse event.
Note: AEs emerging during ASCT phase related to the planned procedures were not reported.
[a] TEAEs during induction, ASCT, or consolidation Treatment Phase; incidence reflects the number of patients experiencing at least one TEAE associated with at least one of the study treatments.

Serious TEAEs

Serious TEAEs occurred at similar rates in the DBTd group and the BTd group with overall incidence of 46.8% and 47.4% respectively (Table 41). The most commonly reported serious TEAEs (≥2%) in the CASSIOPEIA safety population included neutropenia (DBTd 3.9%, BTd 1.5%), pneumonia (DBTd 3.5%, BTd 1.7%), pyrexia (DBTd 2.8%, BTd 4.3%) and pulmonary embolism (DBTd 1.5%, BTd 3.7%).(43, 44)

Table 41: Most common (≥2%) serious TEAEs[a] by MedDRA system organ class and preferred term during the induction/ASCT/consolidation period (CASSIOPEIA, safety population)(43, 44)

	Proportion of patients, n (%)	
	DBTd (n=536)	**BTd (n=538)**
Total number of patients with serious TEAEs	251 (46.8%)	255 (47.4%)
Infections and infestations	80 (14.9%)	67 (12.5%)
Pneumonia	19 (3.5%)	9 (1.7%)
Sepsis	7 (1.3%)	11 (2.0%)
Blood and lymphatic system disorders	57 (10.6%)	44 (8.2%)
Neutropenia	21 (3.9%)	8 (1.5%)
Febrile neutropenia	12 (2.2%)	15 (2.8%)
Thrombocytopenia	12 (2.2%)	4 (0.7%)
Febrile bone marrow aplasia	7 (1.3%)	11 (2.0%)

Respiratory, thoracic and mediastinal disorders	38 (7.1%)	38 (7.1%)
Lung disorder	11 (2.1%)	6 (1.1%)
Pulmonary embolism	8 (1.5%)	20 (3.7%)
General disorders and administration site conditions	33 (6.2%)	37 (6.9%)
Pyrexia	15 (2.8%)	23 (4.3%)
Nervous system disorders	33 (6.2%)	44 (8.2%)
Peripheral sensory neuropathy	11 (2.1%)	15 (2.8%)

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; TEAE = treatment-emergent adverse event.
Note: Adverse events emerging during ASCT phase related to the planned procedures were not reported.
[a] TEAEs during induction, ASCT, or consolidation Treatment Phase; incidence reflects the number of patients experiencing at least one TEAE associated with at least one of the study treatments.

Infusion-related reactions

At median follow-up of 18.8 months, infusion-related reactions (IRRs) of any grade associated with daratumumab were observed in 35.4% of the patients, with 26.9% experiencing IRR at first infusion, 1.9% with the second infusion, and 11.7% cumulative with subsequent infusions (the latter mainly occurring at the first infusion after ASCT (10.7%)).(43, 44) (43, 81) The IRRs were mostly limited to Grade 1 or 2 events (Table 42). The preferred terms and severity of IRRs were consistent with those previously reported following daratumumab mono- and combination therapy. The most common IRRs were general disorders and administration site conditions which included chills (5.6%) and pyrexia (3.7%). Overall, IRRs were manageable with a low frequency of Grade 3 or 4 events (3.5%) low rates of discontinuation (0.6%) and no fatal events.

As referred to in Section B.1.2, a licence extension for a subcutaneous (SC) formulation of daratumumab was received in June 2020. Results from the non-inferiority phase III study COLUMBA demonstrated that the rate of IRRs was significantly reduced with SC versus IV (12.7% vs 34.5%; odds ratio, 0.28; 95% CI, 0.18-0.44; P <0.0001).(104) It is therefore anticipated that IRRs associated with administering DBTd will be substantially reduced following the availability of daratumumab as a SC injection.

Table 42: Treatment-emergent IRRs during induction/ASCT/consolidation phase by system organ class and preferred term and maximum toxicity grade (CASSIOPEIA, safety population)(43, 44)

	All Grades n (%)	Grade 3 n (%)	Grade 4 n (%)	Grade 5 n (%)
Total number of patients with IRRs	190 (35.4%)	17 (3.2%)	2 (0.4%)	0
Total number of patients with IRRs in more than 1 infusion	25 (4.7%)	0	0	0
MedDRA system organ class				
General disorders and administration site conditions	60 (11.2%)	2 (0.4%)	0	0
Chills	30 (5.6%)	0	0	0

Pyrexia	20 (3.7%)	1 (0.2%)	0	0
Chest discomfort	5 (0.9%)	0	0	0
Feeling hot	3 (0.6%)	0	0	0
Malaise	3 (0.6%)	0	0	0
Chest pain	2 (0.4%)	1 (0.2%)	0	0
Hyperthermia	2 (0.4%)	0	0	0
Infusion site extravasation	2 (0.4%)	0	0	0
Injection site swelling	1 (0.2%)	0	0	0
Non-cardiac chest pain	1 (0.2%)	0	0	0
Peripheral swelling	1 (0.2%)	0	0	0
Respiratory, thoracic and mediastinal disorders	58 (10.8%)	5 (0.9%)	2 (0.4%)	0
Dyspnoea	20 (3.7%)	2 (0.4%)	0	0
Cough	16 (3.0%)	0	0	0
Throat irritation	7 (1.3%)	0	0	0
Rhinorrhoea	6 (1.1%)	0	0	0
Bronchospasm	5 (0.9%)	1 (0.2%)	1 (0.2%)	0
Nasal pruritus	2 (0.4%)	0	0	0
Respiratory disorder	2 (0.4%)	0	0	0
Acute respiratory distress syndrome	1 (0.2%)	0	1 (0.2%)	0
Asthmatic crisis	1 (0.2%)	0	0	0
Choking sensation	1 (0.2%)	0	0	0
Laryngeal discomfort	1 (0.2%)	1 (0.2%)	0	0
Laryngeal oedema	1 (0.2%)	0	0	0
Laryngospasm	1 (0.2%)	0	0	0
Nasal congestion	1 (0.2%)	1 (0.2%)	0	0
Pharyngeal hypoesthesia	1 (0.2%)	0	0	0
Pulmonary embolism	1 (0.2%)	1 (0.2%)	0	0
Respiratory distress	1 (0.2%)	0	0	0

Respiratory symptom	1 (0.2%)	0	0	0
Rhinitis allergic	1 (0.2%)	0	0	0
Sneezing	1 (0.2%)	0	0	0
Gastrointestinal disorders	43 (8.0%)	0	0	0
Vomiting	25 (4.7%)	0	0	0
Nausea	18 (3.4%)	0	0	0
Abdominal pain	3 (0.6%)	0	0	0
Diarrhoea	3 (0.6%)	0	0	0
Odynophagia	1 (0.2%)	0	0	0
Palatal oedema	1 (0.2%)	0	0	0
Skin and subcutaneous tissue disorders	43 (8.0%)	4 (0.7%)	0	0
Rash	15 (2.8%)	2 (0.4%)	0	0
Urticaria	9 (1.7%)	0	0	0
Erythema	8 (1.5%)	0	0	0
Pruritus	7 (1.3%)	1 (0.2%)	0	0
Hyperhidrosis	3 (0.6%)	0	0	0
Angioedema	1 (0.2%)	1 (0.2%)	0	0
Dermatitis acneiform	1 (0.2%)	0	0	0
Dermatitis allergic	1 (0.2%)	0	0	0
Rash generalised	1 (0.2%)	0	0	0
Rash macular	1 (0.2%)	0	0	0
Rash macular-papular	1 (0.2%)	0	0	0
Rash vesicular	1 (0.2%)	0	0	0
Vascular disorders	30 (5.6%)	8 (1.5%)	0	0
Hypertension	19 (3.5%)	8 (1.5%)	0	0
Hypotension	6 (1.1%)	0	0	0
Hot flush	3 (0.6%)	0	0	0
Flushing	1 (0.2%)	0	0	0
Vasoconstriction	1 (0.2%)	0	0	0
Nervous system disorders	18 (3.4%)	2 (0.4%)	0	0
Tremor	6 (1.1%)	1 (0.2%)	0	0
Headache	3 (0.6%)	0	0	0

Paraesthesia	3 (0.6%)	0	0	0
Aphonia	1 (0.2%)	0	0	0
Burning sensation	1 (0.2%)	0	0	0
Dizziness	1 (0.2%)	0	0	0
Head discomfort	1 (0.2%)	0	0	0
Migraine	1 (0.2%)	1 (0.2%)	0	0
Somnolence	1 (0.2%)	0	0	0
Cardiac disorders	7 (1.3%)	0	0	0
Tachycardia	5 (0.9%)	0	0	0
Atrial fibrillation	1 (0.2%)	0	0	0
Sinus tachycardia	1 (0.2%)	0	0	0
Eye disorders	6 (1.1%)	1 (0.2%)	0	0
Vision blurred	4 (0.7%)	0	0	0
Asthenopia	1 (0.2%)	0	0	0
Eye swelling	1 (0.2%)	1 (0.2%)	0	0
Visual impairment	1 (0.2%)	0	0	0
Immune system disorders	5 (0.9%)	0	0	0
Hypersensitivity	5 (0.9%)	0	0	0
Musculoskeletal and connective tissue disorders	4 (0.7%)	0	0	0
Back pain	1 (0.2%)	0	0	0
Bone pain	1 (0.2%)	0	0	0
Muscle spasms	1 (0.2%)	0	0	0
Pain in jaw	1 (0.2%)	0	0	0
Psychiatric disorders	3 (0.6%)	0	0	0
Anxiety	3 (0.6%)	0	0	0
Infections and infestations	2 (0.4%)	0	0	0
Rhinitis	1 (0.2%)	0	0	0
Sinusitis	1 (0.2%)	0	0	0
Investigations	2 (0.4%)	0	0	0
Oxygen saturation decreased	2 (0.4%)	0	0	0
Ear and labyrinth disorders	1 (0.2%)	0	0	0
Vertigo	1 (0.2%)	0	0	0

Injury, poisoning and procedural complications	1 (0.2%)	0	0	0
Infusion related reaction	1 (0.2%)	0	0	0

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; IRR = infusion-related reactions; TEAE = treatment-emergent adverse event.
Note: AEs are reported using MedDRA version 20.0. During the transplant period, according to protocol, only limited AE were collected. Percentages are calculated with the number of patients in each group as denominator.

B.2.11 *Ongoing studies*

A summary of ongoing studies that should provide additional clinical evidence for daratumumab in front-line transplant-eligible MM is shown in Table 43.

Table 43: Clinical trials for the evaluation of daratumumab in NDTE MM patients(105-108)

Study	Target indication/ population	Primary objective	Phase	N	Efficacy hypothesis	Trial start date	Estimated trial completion date	Interim data before completion?
MMY3006 (CASSIOPEIA) – Part 2	Daratumumab in combination with bortezomib, thalidomide and dexamethasone for the treatment of adult patients with newly diagnosed MM who are eligible for high-dose chemotherapy with ASCT	To compare the efficacy of daratumumab as single agent in maintenance compared to observation only in terms of PFS in patients with newly diagnosed MM when used after ASCT and consolidation therapy	III	886	The study is designed to achieve a power of 80% to detect a 25% reduction in the risk of progression or death (i.e. a PFS increase from 45 months to 60 months corresponding to a 0.75 HR daratumumab maintenance versus observation with a log-rank test [two-sided alpha is 0.05]).	September 22, 2015	August, 2024	November 2020[a]
MMY3014 (PERSEUS)	Daratumumab in combination with bortezomib, lenalidomide and dexamethasone for the treatment of adult patients with newly diagnosed MM who are eligible for high-dose chemotherapy with ASCT.	To compare the efficacy of daratumumab when combined with bortezomib, lenalidomide and dexamethasone (DBLd) to that of bortezomib, lenalidomide and dexamethasone (BLd), in terms of PFS in patients with newly diagnosed MM.	III	690	The study is designed to achieve a power of 85% to detect a 31% reduction in the risk of progression or death (i.e. a PFS increase to 91 months from 63 months corresponding to a 0.69 HR DBLd versus BLd with a log-rank test [two-sided alpha is 0.05]). In addition, it will also achieve approximately 70% power to detect a 25% reduction in the risk of death (HR: 0.75) DBLd versus BLd with a log-rank test (two-sided alpha=0.05).	December 14, 2018	November, 2029	April/May 2021[a]

MMY2004 (GRIFFIN)	Daratumumab in combination with lenalidomide, bortezomib, and dexamethasone for the treatment of adult patients with newly diagnosed MM who are eligible for high-dose chemotherapy with ASCT.	To determine if the addition of daratumumab to lenalidomide-bortezomib-dexamethasone (DLBd) will increase the proportion of participants achieving sCR, by the time of completion of post ASCT consolidation treatment, compared with LBd alone.	II	224	The study is designed to achieve a power of 80% to detect an absolute 15% increase over 35% in post-consolidation sCR rate using a 1-sided likelihood ratio test at the 10% significance level.	August 29, 2016	January 25, 2022	2-year update expected December 2020
MMY2012 (LYRA)	Daratumumab in combination with cyclophosphamide, bortezomib, and dexamethasone (Dara-CyBorD) in previously untreated and relapsed patients with MM.	To evaluate CR+VGPR rate following 4 cycles of induction therapy of Dara-CyBorD, in previously untreated subjects, and in relapsed subjects with multiple myeloma.	II	101	The study is designed to achieve an 80% power to detect an absolute 20% increase over 60% with Dara-CyBorD in CR+VGPR rate using a 5% 1-sided significance level.	November 9, 2016	September 30, 2020	n/a

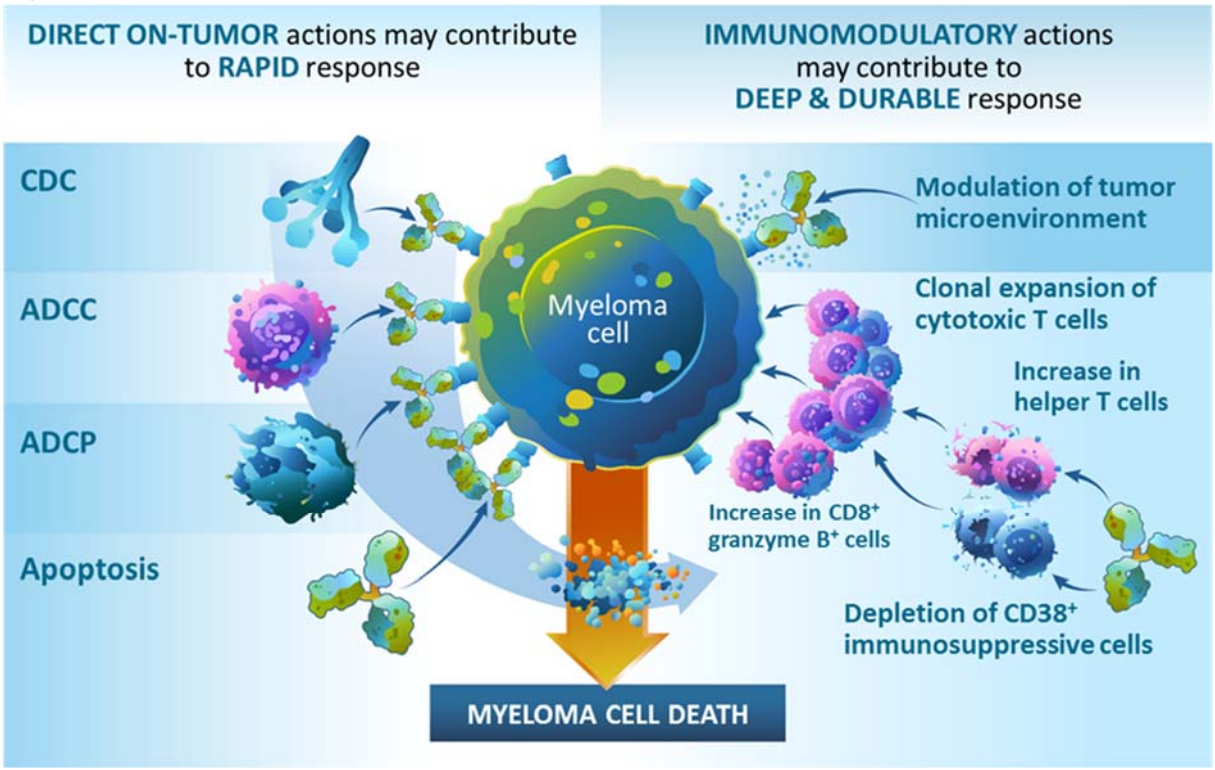
Key: ASCT = autologous stem cell transplant; DBLd = daratumumab, bortezomib, lenalidomide and dexamethasone; BLd = bortezomib, lenalidomide and dexamethasone; HR = hazard ratio; MM = multiple myeloma; n/a = not applicable; PFS = progression-free survival.
[a] These timelines represent a best estimate as the exact timing of interim analysis is event-driven.

B.2.12 *Innovation*

Currently, there is no cure for MM. The primary goal of therapy is therefore to induce remission and delay disease progression. With each relapse, it becomes more challenging to induce a deep and durable response to treatment, with high attrition rates between lines of therapy highlighting the need to treat patients with the most efficacious regimens first. Despite several new treatments having been approved in later lines during the past decade, there has been limited progress in the development of new effective regimens for the management of NDTE MM patients with no new licenced therapy approved since BTd in 2013.(44) All patients eventually relapse leading to poorer prognosis, highlighting the high level of unmet need that still exists. Treatments that can offer prolonged periods of remission and treatment-free intervals are highly valued by patients, as reported in patient preference surveys (refer to Section B.1.3.3).

Daratumumab is a first-in-class, fully human IgG1κ mAb that binds to CD38, a protein that is overexpressed on the surface of MM cells. It works by targeting the tumour directly and indirectly (refer to Figure 26), as well as uniquely modulating the immune system in a way that is not typically seen in monoclonal antibodies; put simply, it boosts patients' immune system.(3, 4) It is the combination of these direct and indirect immunomodulatory effects that explain the step-change in efficacy for this indication observed with daratumumab.

Figure 26: The multiple mechanisms of action of daratumumab



Key: ADCC = antibody-dependent cell-mediated cytotoxicity; ADCP = antibody-dependent cellular phagocytosis; CDC = complement-dependent cytotoxicity.

CD38 is a distinct and novel target from those of other approved agents for MM due to its universal expression in plasma and myeloma cells. This universal expression not only allows daratumumab to induce myeloma cell death through multifactorial mechanisms (see above), but also means daratumumab is effective, irrespective of clonal heterogeneity. Clonal heterogeneity is a consequence of the genetically complex nature of MM, which develops from the continued accumulation of genetic abnormalities over time. This results in sub clones of plasma cells with considerable genetic

heterogeneity that contribute to the progression of MM and the development of drug resistance.(13, 109-111) One of the challenges of treatment to date has been to find options that effectively target and eliminate all clonal and subclonal mutations – clones that remain following treatment will re-populate the disease via clonal expansion and evolution. The concept of clonal heterogeneity contributing to disease progression in MM led to the strategy of adopting combination therapies to eradicate both the dominant and minor clones. Combination treatment strategies are now recommended for routine clinical practice by the International Myeloma Working Group. As noted above, CD38 is a distinct and novel target from those of other approved agents for MM and this, together with its high efficacy and favourable safety profile, make daratumumab an ideal candidate for combination therapy.

In the CASSIOPEIA trial, DBTd, demonstrated a statistically significant improvement in efficacy versus existing SOC, BTd triplet therapy, in terms of response post-consolidation (sCR, CR or better, VGPR or better, and MRD-negative rate). These unparalleled responses have translated into a 51% reduction in the risk of progression or death for the DBTd group versus the BTd group after median follow-up of 29.2 months and already a trend for improved OS. Indeed there is evidence that the mortality rate for patients treated with DBTd who achieve MRD negativity resembles that of the general population, providing hope of long-term disease control and a functional cure for some patients (Section B.2.6).(43, 44) It is the combination of direct and indirect immunomodulatory effects which drive significantly deeper responses that explain the exceptional efficacy for this indication observed with daratumumab.

B.2.13 *Interpretation of clinical effectiveness and safety evidence*

Strengths and limitations of the clinical evidence base

CASSIOPEIA was a registrational quality phase III RCT that directly compared DBTd against the relevant active comparator BTd. Results from the PHE linked dataset analysis, along with market research data and UK clinical expert opinion indicate that BTd is SOC induction therapy for NDTE MM patients in England.(1, 2, 112)

CASSIOPEIA was a high-quality, active-controlled study conducted in line with ICH guidelines on Good Clinical Practice (GCP), and applicable regulatory and country-specific requirements. Steps taken to ensure the accuracy and reliability of the data included the selection of qualified investigators and appropriate study sites, review of protocol procedures with the investigator and study-site personnel before the study, periodic monitoring visits by sponsor representatives, and direct transmission of clinical laboratory data from a central laboratory into the sponsor's data base. The study had an open label design because of the difference in mode of administration for the trial drugs (daratumumab infusions are administered over a longer duration than bortezomib injections). However, the risk for bias was minimised since patients were randomised using a central interactive web response system (IWRS). In addition, outcomes were reviewed by an Independent Data Monitoring Committee (IDMC) which considered efficacy and safety outcomes to be robust, leading to regulatory approval by EMA. A summary of the quality of the CASSIOPEIA trial is presented in Section B.1.1.

Relevance of response outcomes and MRD status

The definitions of treatment response and disease progression developed by the IMWG in 2006, updated in 2014, are widely used in clinical practice.(48, 59) In recognition of the high levels of complete response being achieved with newer treatments, consensus criteria on the assessment of MRD negativity were published by IMWG in 2016.(64) MRD is a more sensitive measure of disease burden than the measures of clinical response defined by the IMWG revised uniform response criteria (including sCR, CR and VGPR).(61) It is linked to depth of response and long-term outcomes. Indeed, a recently expanded meta-analysis demonstrated that MRD-negative status is associated with prolonged PFS and OS in NDTE MM (refer to Section B.3.3.2).(45) In CASSIOPEIA, a statistically significant higher rate of patients achieving MRD negativity was observed for DBTd versus BTd,(43) with landmark

analysis of survival by MRD status indicating significantly better PFS and OS for MRD-negative patients relative to MRD-positive patients (refer to Section B.2.6.3).

Whilst the routine assessment of MRD negativity is not yet established in UK clinical practice, the positive link between MRD negativity and long-term survival outcomes means that MRD negativity is a highly relevant prognostic marker associated with substantial clinical benefit. Over and above depth of response, the landmark analysis from CASSIOPEIA also supports a daratumumab treatment effect, with improved survival in the DBTd arm versus BTd seen for both MRD-positive and MRD-negative patients. Evidence of a daratumumab treatment effect regardless of MRD response is [redacted] This treatment effect reflects deeper responses for DBTd treated patients, as indicated by a higher proportion of patients achieving MRD negativity measured by NGS at a greater sensitivity threshold of 10^{-6} (see Section B.2.6.1), as well as the deeper conventional response according to IMWG achieved with daratumumab in MRD-positive patients.

Generalisability of CASSIOPEIA to clinical practice in England

CASSIOPEIA was a multicentre, international trial that enrolled participants generally representative of NDTE MM patients in England. While all patients were recruited outside of the UK, all the sites were in countries with similar demographics to the UK (France, Belgium and the Netherlands). Expert clinical opinion indicated that while patients recruited in CASSIOPEIA were generally a little younger, and excluded patients with severe renal impairment, baseline demographic and disease characteristics were otherwise broadly similar to clinical practice in England.(1) This assessment is supported by a comparison of patient characteristics between the PHE linked dataset analysis and the BTd arm of CASSIOPEIA (see Section B.3.3.2). Whilst baseline demographic and clinical characteristics in CASSIOPEIA were well balanced between the two treatment arms, a revised international staging system for response (R-ISS) indicated a poorer prognosis for patients recruited into the DBTd arm of the study, with 71.6% classified as stage II compared to 63.7% for BTd (see Section B.2.3.2). In this regard, it is important to recognise the potential for the clinical efficacy results in terms of response rates and survival (PFS/OS) to be biased against daratumumab in favour of BTd.

The relevant comparator for this submission is BTd as induction therapy only, therefore the efficacy of BTd in CASSIOPEIA is not fully generalisable to UK clinical practice. When considering the response rates observed in CASSIOPEIA (post-induction, post ASCT and post consolidation) it is clear that the relative effectiveness of DBTd versus BTd in CASSIOPEIA will be an underestimate of the relative benefit of DBTd compared to current clinical practice. That is, induction, ASCT and consolidation with BTd delivers greater levels of sCR than BTd induction and ASCT only (20.3% post-consolidation sCR versus 9.4% post-transplant sCR).(44)

Maintenance treatment administered in Part 2 of CASSIOPEIA differs from clinical practice in England where maintenance therapy is not recommended by NICE or routinely commissioned by NHS England for NDTE patients. In CASSIOPEIA, 50% of patients on each arm who achieved at least a partial response were re-randomised to receive daratumumab monotherapy as maintenance treatment. To address potential confounding by the subsequent maintenance therapy, a prespecified IPW analysis of PFS was performed by a sequestered group independent of the study team to provide an unbiased estimate of treatment benefit for DBTd compared to BTd irrespective of subsequent maintenance treatment. As noted in Section B.2.6.2 the results of this analysis demonstrate that the relative treatment effect for PFS was maintained. Thus, whilst absolute survival outcomes for both DBTd and BTd may be better than expected in routine clinical practice in England, the relative treatment benefit for DBTd versus BTd is generalisable to outcomes expected in the real world. Indeed, as noted above, is likely to be a conservative estimate given BTd administered in CASSIOPEIA included 2 cycles of consolidation.

Finally, it is noted that the thalidomide daily dose of 100 mg/day used in CASSIOPEIA is different to the dosing schedule in the BTd label which recommends a gradual increase in thalidomide dose from 50 mg/day to 200 mg/day.(113) Clinical expert feedback has, however, confirmed the modified thalidomide dose used in CASSIOPEIA is consistent with BTd administration in clinical practice in England.(1, 114, 115) An MAIC analyses of the modified BTd dosing has also demonstrated similar or better efficacy (and a similar safety profile) compared with the BTd label (OS HR=0.640; 95% CI, 0.363–1.129, p=0.121; PFS HR=0.672; 95% CI 0.467–0.966, p=0.031 for BTd dosing in CASSIOPEIA versus BTd label), while a flat daily dose of 100 mg/day also helped mitigate the risk of heterogeneity in daily dosing administered in the CASSIOPEIA trial.(116)

Principal findings of the clinical evidence base

In the CASSIOPEIA trial, DBTd resulted in an unprecedented clinical benefit that was both statistically significant and clinically meaningful when compared with BTd alone (refer to Section B.2.6).(43, 44) The addition of daratumumab to BTd resulted in significantly deeper post-consolidation responses (100 days post-ASCT) as measured by the rate of sCR, CR or better and MRD negativity compared with BTd alone.(43, 44) Importantly, these benefits were observed consistently across all prespecified subgroups of the ITT population in the latest analysis of the CASSIOPEIA trial (PHA) (refer to Section B.2.7 and Appendix E), making them more likely to be reproducible in clinical practice. The relationship between MRD negativity and improved PFS and OS is established in front-line myeloma (including transplant-eligible patients), having been demonstrated through SLR and meta-analysis of studies reporting MRD status and survival outcomes (refer to Section B.3.3.2 and Appendix M for further details).(61, 63)

In CASSIOPEIA, the significant differences in post-consolidation MRD-negative rates between DBTd and BTd is already being translated to better long-term outcomes, as demonstrated by a 51% reduction in the risk of disease progression or death for patients treated with DBTd after median follow-up of 29.2 months (refer to Section B.2.6.2). (43, 44) Whilst overall survival data remains immature, there is already a strong trend towards improved outcomes for DBTd treated patients (refer to Section B.2.6.2). A statistically significant survival benefit can reasonably be expected given the highly statistically significant improvement in MRD negativity observed, and the relationship between MRD negativity and OS. In addition, there is evidence to support a daratumumab treatment effect regardless of MRD status, as shown in the CASSIOPEIA landmark analysis and also demonstrated in [redacted].2.6.3 Indeed, survival outcomes for patients treated with DBTd who achieve MRD negativity resemble general population mortality, providing hope of a functional cure for some patients.

In the absence of a viable network of studies with sufficient comparability to inform an NMA, an unanchored MAIC was performed to compare PFS and OS for DBTd versus both BCd and Bd (refer to Section B.2.9). The feasibility of conducting an MAIC based on response was explored, however only the CASSIOPEIA trial reported data on post-consolidation MRD negativity that could be used to inform the economic model. The MAIC analysis demonstrated a [redacted] in terms of both PFS and OS for DBTd compared with BCd and Bd in patients with NDTE MM. [redacted]. The results of the MAICs regarding the comparative efficacy of BTd (i.e. [redacted]) are consistent with clinical expert opinion and also a naïve comparison of real world data from the PHE linked dataset analysis (refer to Section B.2.9.5).

In CASSIOPEIA, HRQoL was generally maintained for patients treated with DBTd compared to BTd, with clinically and statistically significant improvement in pain, and statistically significant improvements in emotional functioning and cognitive decline (refer to Section B.2.6.4).(43, 44) As noted earlier, improvements in pain and cognitive functioning are closely aligned to MM patient preferences.(34, 35)

Importantly, HRQoL assessment showed no negative HRQoL impact of the quadruplet DBTd therapy over the standard BTd triplet, suggesting that patients treated with DBTd may achieve improved clinical outcomes (i.e. PFS and OS) versus SOC triplet therapy, without significant detriments in HRQoL as a result of the addition of daratumumab.

DBTd was also well-tolerated in CASSIOPEIA, with clinically manageable side effects consistent with the known safety profiles of daratumumab monotherapy and the BTd regimen (refer to Section B.2.10).(43, 44) No new safety signals were identified.(43) IRRs associated with the use of daratumumab were mild and manageable and are anticipated to reduce significantly with the use of SC daratumumab.(43) Furthermore, SC daratumumab is expected to improve convenience for patients with administration time reduced from several hours to approximately 5 minutes.(104)

Over and above the significant clinical benefits of DBTd, the fixed treatment duration and a substantial increase in the treatment-free period post induction/consolidation therapy is highly valued by patients and carers alike to allow quality time with loved ones and to ‘not always think of the disease’. The efficacy and safety of daratumumab, and the impact of achieving sustained remission on patient HRQoL, has been included as part of the cost-effectiveness analysis presented in Section B.3. However, the positive effect that treatment with DBTd could have on informal carers in terms of reduced anxiety/depression and being able to return to work is not captured as part of these analyses. Similarly, the psychological impact of achieving a sustained period of treatment-free remission, in terms of the sense of hope that patients and carers may experience in place of the fear of relapse, is not intrinsically captured as part of the QALY framework.

B.3 Cost effectiveness

B.3.1 *Published cost-effectiveness studies*

In order to identify evidence relating to the HRQoL and utility (humanistic burden) and cost/resource use (economic burden) that may be of relevance to this submission, two SLRs were conducted (refer to Appendices G, H and I). As part of the economic SLR, published economic evaluations of interventions for transplant-eligible patients with newly diagnosed MM were also identified.

No published economic evaluations of DBTd in this indication were identified in either the original or updated SLR searches, however published documents for UK HTA of bortezomib as BTd or Bd were identified (SMC ID 927/13 and NICE TA311) and have been used to inform inputs and assumptions for the model.(27, 117) A summary of the cost-effectiveness analyses submitted by Janssen to NICE and the Scottish Medicines Consortium (SMC) as part of TA311 and ID 927/13, respectively, is presented in Table 44.

Table 44: Summary list of published cost-effectiveness studies

Study	Year	Summary of model	Patient population	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
NICE TA311(27)	2014	Markov state-transition cohort model with health states based on response to induction therapy (CR, PR or NR), followed by health states representing subsequent lines of therapy. A proportion of patients were assumed to receive SCT following induction therapy. OS data from the pivotal trials were immature and so OS in the model was dependent on the level of post-induction response achieved. The cost-effectiveness of BTd was assessed versus Td. Comparisons were also made for PAD and Bd versus VAD.	Adult patients with previously untreated multiple myeloma, eligible for high dose therapy and SCT	Company submission base case: BTd: 4.00 Td: 3.06	Company submission base case: BTd: £72,815 Td: £49,414	Company submission base case: BTd versus Td: £24,683 per QALY gained
SMC ID 927/13(117)	2014	See above. The same model structure was used for the submission to the SMC.	Adult patients with previously untreated multiple myeloma, eligible for high dose therapy and SCT	Not reported in DAD Incremental QALYs for BTd versus Td: 1.04	Not reported in DAD Incremental costs for BTd versus Td: £24k	As reported in DAD: BTd versus Td: £23,077 per QALY gained

Key: Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; CR = complete response; DAD = Detailed Advice Document; ICER = incremental cost-effectiveness ratio; NR = no response; OS = overall survival; PAD = bortezomib, doxorubicin and dexamethasone; PR = partial response; QALYs = quality-adjusted life years; SCT = stem-cell transplantation; SMC = Scottish Medicines Consortium; Td = thalidomide and dexamethasone; VAD = vincristine, doxorubicin, dexamethasone.

B.3.2 *Economic analysis*

A *de novo* cost-utility analysis (CUA) has been conducted for the purpose of this appraisal and is described below.

The aim of this economic analysis was to determine the cost-effectiveness of DBTd versus BTd as a treatment for adult patients with newly diagnosed MM who are eligible for ASCT. The analyses have been conducted from the perspective of the NHS in England taking into account direct costs and benefits.

The economic evaluation was structured as follows:

- Health outcomes are measured both in terms of life years gained (LYG) and QALYs gained
- Primary outcome measure for the economic evaluation is the ICER (cost per QALY gained) for the comparison of DBTd versus BTd
- Clinical effectiveness for DBTd and BTd is measured through OS and PFS, which have been modelled to be dependent on post-consolidation MRD status using a landmark analysis approach
- All relevant treatment-specific costs are considered including cost of the medicine, administration costs, and adverse event costs
- Costs associated with concomitant medicines and medical resource use are also included
- The time horizon used is equivalent to a lifetime time horizon (the maximum age that could be reached in the model is 100 years old)
- The discount rate is set to 3.5% for both costs and benefits, as specified in the NICE reference case

B.3.2.1 Patient population

The patient population for the economic evaluation was adult patients with newly diagnosed MM who are eligible for ASCT. This is consistent with the licensed indication for daratumumab that is of interest for this submission and the patient population included in the CASSIOPEIA trial.(5, 43)

The characteristics of patients entering the model were based on the baseline demographic and disease characteristics of the ITT population recruited in CASSIOPEIA (refer to Section B.2.3.2). As noted in Section B.2.13, clinical expert feedback was that patients recruited in CASSIOPEIA were generally a little younger than expected in clinical practice in England, and excluded patients with severe renal impairment, but were otherwise considered broadly similar.(1)

- Age and gender are included in the model to determine general population mortality inputs (refer to Section B.3.3.2)
- Body weight and body surface area (BSA) are included in the model in order to calculate the drug acquisition costs of treatments that are dosed based on weight (e.g. daratumumab, in the scenario analysis) or BSA (e.g. bortezomib) (refer to Section B.3.5.1)

Table 45: Patient baseline characteristics in the cost-utility analysis(44)

Characteristic	Value
Mean age of patients (years)	56.6
Mean weight of patients (kg)	75.67
Mean BSA of patients (m^2)	1.88
Male (%)	58.5%
Key: BSA = body surface area.	

B.3.2.2 Model structure

The CASSIOPEIA trial represents the main source of evidence for this submission. A standard partitioned survival model was initially explored which directly extrapolated PFS and OS based on the observed trial data (refer to Section B.3.3.2). However, due to the general good prognosis of NDTE patients, survival data is immature (as described in Section B.2.6). Consequently, there was considerable variation and uncertainty in the long-term survival extrapolations.

A response-based model was therefore developed in which estimates of long-term survival were based on the level of response achieved by patients following receipt of induction, ASCT and consolidation therapy (see full description below). This model structure was primarily selected as it acknowledges underlying patient heterogeneity, and supports the incorporation of external data with longer follow-up to inform the relationship between response and long-term survival outcomes. The use of a response-based model is also consistent with the modelling approach taken in the appraisal of bortezomib as an induction therapy for patients eligible for SCT (TA311), the only other therapy to be assessed by NICE in the newly diagnosed, transplant-eligible setting.(27) A comparison between the approach taken in this submission and the approach taken in TA311 is provided in Table 46 below.

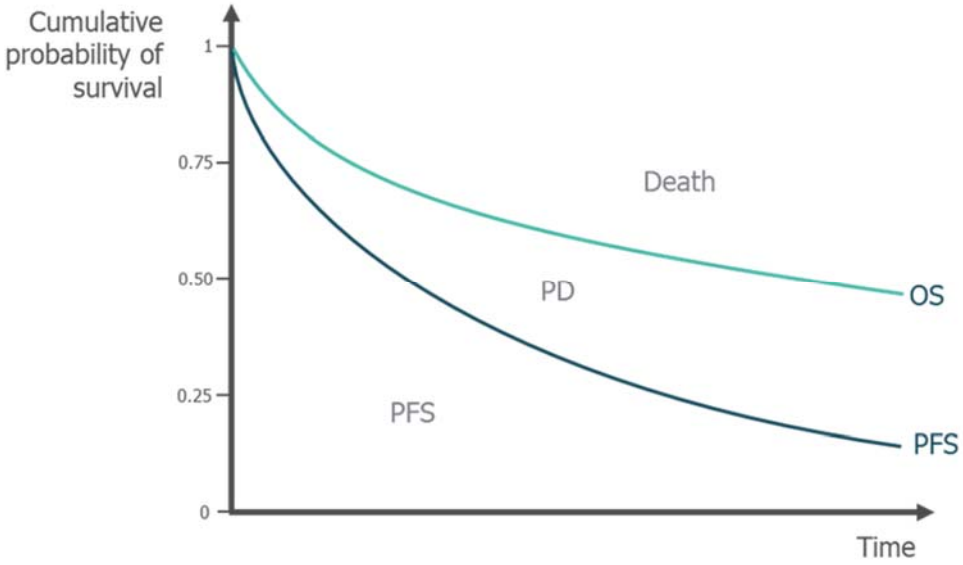
Rather than use a state-transition modelling approach, as per TA311, the response-based approach to modelling survival in this submission has been conducted within a 3-state partitioned survival modelling framework. The partitioned survival approach provides greater flexibility for incorporating external sources of survival, from which only summary data may be available, and given the similarity between the treatment arms in terms of the treatment pathway following disease progression, there is limited advantage in being able to model the subsequent lines of therapy as individual health states. By directly modelling study-observed events, the partitioned survival modelling approach is also likely to provide estimates of survival that closely match the observed survival from the original studies. Uncertainty in long-term survival can also be explored via the use of alternative distributions to extrapolate survival.

The partitioned survival model used in this submission consists of three health states: Progression-free (PF), Progressed disease (PD) and Death. The occupancy of health states over time was derived from the survival curves (PFS and OS), as described below and shown in Figure 27:

- The proportion of patients occupying the PF state was calculated as the proportion alive and progression-free (based on PFS curve)
- The proportion of patients occupying the PD state was calculated as the proportion alive (based on OS curve) minus the proportion of patients alive and progression-free (based on PFS curve)

- The proportion of patients occupying the death state was calculated as the proportion who had died (based on OS curve)

Figure 27: Partitioned survival model structure



Key: OS = overall survival; PD = progressed disease; PFS = progression-free survival.

To incorporate response into the analysis, PFS and OS were both modelled to be dependent on whether patients had achieved a response to treatment, with separate survival inputs used for 'responders' and 'non-responders'. In the model, response has been based on MRD negativity, which, as described in Section B.2.3.1, is a more sensitive measure of response than traditional assessments (e.g. based on CR), and several studies have been published demonstrating the benefits of achieving MRD-negative status following induction therapy and ASCT, in terms of prolonged OS and PFS.(61, 63)

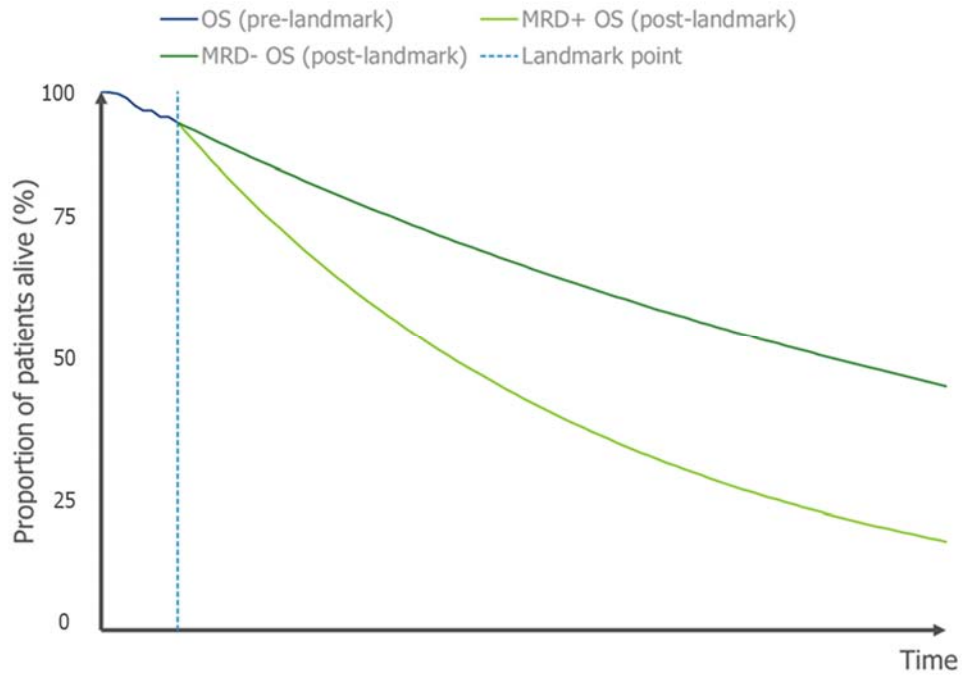
In Part 1 of the CASSIOPEIA trial, MRD negativity was assessed post-induction and after consolidation therapy (100 days post-ASCT) as a secondary outcome (refer to Section B.2.6).(43) Data from the CASSIOPEIA trial on post-consolidation MRD negativity have been used to inform the levels of response achieved in each treatment arm in the model (refer to Section B.3.3.1). The use of post-consolidation response rather than post-induction response ensures that the impact of ASCT and consolidation therapy is captured in the efficacy assessment.

As described in Section B.2.6.3, in order to mitigate against the effect of immortal time bias (patients needed to survive to experience the event), a 'landmark' approach was taken in the model whereby survival was only modelled to be dependent on response after a specific 'landmark' timepoint. The post-consolidation response assessment timepoint in CASSIOPEIA (100 days post-ASCT) was chosen as the landmark point in the model to align with the MRD negativity data used to establish whether patients had achieved a response. Survival (PFS and OS) before and after the landmark point were modelled as follows (refer also to Figure 28):

- **Pre-landmark point:** PFS and OS based on the observed survival data from the CASSIOPEIA trial (ITT population, PHA; median follow-up = 29.2 months) from the time of randomisation
- **Post-landmark point:** PFS and OS based on the extrapolation of survival data from the time of post-consolidation response, with PFS and OS modelled separately for patients who achieve

MRD-negative status (MRD-) and those who do not (MRD+) at the post-consolidation assessment timepoint (refer to Section B.3.3.2 for the survival inputs used)

Figure 28: Response-based approach incorporating landmark analysis



Key: MRD = minimal residual disease; OS = overall survival; PFS = progression-free survival.
The diagram uses OS as an example. The same is also done in the model for PFS.

The mean time from the start of induction therapy to 100 days post-ASCT in CASSIOPEIA was approximately 37 weeks in both treatment arms, and a model cycle of 4 weeks was chosen in order to align with the 28-day treatment cycles for both DBTd and BTd (see Section B.3.2.3).(44) The pre-landmark period therefore corresponded to cycles 0–8 of the model with the post-landmark period covering the remainder of the model time horizon from cycle 9 onwards.

Features of the economic analysis

A summary of the features of the CUA presented as part of this submission compared to those included in TA311, the only other NICE appraisal of a treatment for newly diagnosed patients with MM who are eligible for ASCT, is presented in Table 46.

Table 46: Features of the economic analysis

Factor	Previous appraisals	Current appraisal	
	TA311	Chosen values	Justification
Time horizon	Lifetime (30 years)	Lifetime (40 years)	Consistent with the NICE reference case
Model structure	Response-based model, with survival outcomes based on post-induction response (CR, PR, NR) State-transition model framework with health states for 2nd line, 3rd line and further lines. Transition probabilities for transitions to later lines of therapy were not based on response or the treatment received at 1st line	Response-based model, with survival outcomes based on post-consolidation MRD negativity (MRD- or MRD+) Partitioned survival model in which the cost of subsequent therapies is included in the PD health state	In both TA311 and this submission, a response-based model, using data from external sources to inform the relationship between response and long-term survival outcomes, was used due to the immaturity of survival data from the 1st line trials. Whereas TA311 aimed to capture the benefit of BTd derived from the proportion of patients who underwent transplantation, in CASSIOPEIA, the proportion of patients who completed transplantation were similar between treatment arms. The benefit of DBTd in this submission is thus derived from greater depth of response achieved post-consolidation, leading to improved long-term survival outcomes. MRD status has been used in this submission as it is a more sensitive measure of response than traditional assessments (e.g. based on CR), and several studies have been published demonstrating the benefits of achieving MRD negativity following induction therapy and ASCT, in terms of prolonged OS and PFS.(61, 63) Post-consolidation response has been used (rather than post-induction response) in order to capture the impact of ASCT and consolidation therapy on depth of response and therefore long-term survival outcomes. Given the similarity between the treatment arms in terms of the treatment pathway following disease progression, the ability to model subsequent lines of therapy as individual health states was not considered to be necessary. A partitioned survival model was therefore utilised. This approach provides greater flexibility for incorporating external sources of survival, produces reliable estimates of survival versus the observed trial data, and allows for uncertainty in long-term survival estimates to be explored.

Treatment waning effect?	None The probability of death in each cycle was dependent on post-induction response and was assumed to be constant with respect to time. In the model, the effect of treatment was modelled in terms differences in post-induction response and via the use of transition probabilities for progression from post-SCT that were treatment-dependent as well as response-dependent. No treatment waning effect on progression post-SCT was included in the model.	None OS and PFS (post-landmark) are modelled to be dependent on post-consolidation response and, for DBTd, survival inputs are based on the application of HRs (OS and PFS) to the BTd arm. In the model, the long-term effect of treatment was therefore modelled in terms of differences in post-consolidation response and via the application of HRs to the BTd arm. In the base case analysis, the same HRs were applied for the entire duration of the remaining time in the model.	No treatment waning effect was applied in the base case analysis as there is no evidence to suggest if, or when, the treatment effect of daratumumab on survival would wane over time. As well as not being included in TA311, treatment waning was not considered in the previous NICE appraisals of daratumumab at later lines of therapy (TA573 and TA510). Evidence of a treatment effect with daratumumab in both MRD-negative and MRD-positive patients was demonstrated by the landmark analysis of the CASSIOPEIA trial (refer to Section B.2.6.3). A treatment effect with daratumumab regardless of MRD response is [redacted] Superior survival outcomes for DBTd patients is driven by deeper responses, and reflect the unique mechanism of action of daratumumab, which is to modulate the immune system to better fight the disease.(118, 119) To fully explore uncertainty, and the possibility that the treatment effect of daratumumab may wane over time, scenario analyses have been conducted in which the HRs for DBTd versus BTd (PFS and OS) are set to equal one (i.e. no treatment effect) at specified timepoints in the model (5- and 10-years). This is considered an extremely conservative approach, not supported by clinical evidence, however, provides a useful upper bound to characterise the uncertainty in the daratumumab treatment effect in the long-term.
Source of utilities	Various sources were used for utility values for induction therapy, ASCT and post-ASCT, including van Agthoven et al. (2004), Segeren thesis and Beusterien et al. (2010) Van Agthoven et al. (2004) was used for utility values for subsequent lines of therapy ('2nd line and 3rd line' and 'further lines')	EQ-5D-5L data from CASSIOPEIA were used to derive utility values for the PF health state (including separate values for induction therapy, post-induction to post-consolidation response, and PF post-consolidation). The utility value used in TA311 for the '2nd line and 3rd line' health states from van	For consistency with the patient population and source of efficacy inputs used in the model, utility values derived from the CASSIOPEIA EQ-5D-5L data were preferred in the base case analysis. As patients are expected to spend a greater period of time in 2nd and 3rd line, compared to 4th line, the value of 0.69 from TA311 was used for the PD health state utility in the model. Scenario analyses have also been conducted in which lower utility values were used for PD and in which alternative sources were used for all health state utility values (refer to Section B.3.8.3).

	• 1st line: start of treatment to post-induction response = 0.57 • 1st line: post-induction to post-SCT response = 0.65 • 1st line (SCT) = various from 0.59 (3 months) to 0.75 (18+ months) • 1st line (non-SCT) = 0.83 for CR, 0.76 for PR and 0.65 for NR • 2nd line and 3rd line = 0.69 • Further lines = 0.644	Agthoven et al. (2004) were used for the PD health state • PF induction therapy = 0.57 • PF post-induction to post-consolidation response = 0.68 • PF post-consolidation = 0.73 • PD = 0.69	
Source of costs	NHS reference costs and the British National Formulary Costs included: • Drug acquisition and administration for 1st line and subsequent therapies • Concomitant medications (e.g. prophylaxis) • ASCT costs • Monitoring costs • Management of adverse events (grade 3 and above)	NHS reference costs and the British National Formulary Costs included: • Drug acquisition and administration for 1st line and subsequent therapies • Concomitant medications (e.g. prophylaxis) • ASCT costs • Monitoring costs • Management of adverse events (grade 3 and above, with incidence ≥5%, plus any grade nausea and upper respiratory tract infections) • End-of-life cost	Cost assumptions used in the model (administration costs, incidence of adverse events, monitoring costs, end-of-life cost) have been based on previous appraisals in MM, including previous daratumumab appraisals (TA573 and TA510). The cost of ASCT has been calculated using the approach used in TA311.

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; CR = complete response; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EQ-5D-5L = EuroQol-5D, 5 levels; HR; hazard ratio; MRD; minimal residual disease; NR = no response; OS = overall survival; PD = progressed disease; PF = progression-free; PFS = progression-free survival; PR = partial response.

B.3.2.3 Intervention technology and comparators

Intervention

The intervention included in the CUA was DBTd as an induction therapy prior to ASCT and a consolidation therapy post-ASCT. The treatment protocol included in the model for induction therapy, ASCT and consolidation therapy in the DBTd arm is consistent with that which was followed in the CASSIOPEIA trial (refer to Section B.3.5.1 for full details), and the SmPC-recommended posology for daratumumab in this setting:(5, 44)

- Four 28-day cycles of DBTd as induction therapy
- High-dose chemotherapy and ASCT
- Two 28-day cycles of DBTd as consolidation therapy

In the CASSIOPEIA trial, not all patients completed the full number of cycles of induction or consolidation therapy. In order to reflect the extent of treatment exposure in the CASSIOPEIA trial, the number of cycles of induction and consolidation therapy received by patients in the model was based on data from the trial (see Table 47). The use of CASSIOPEIA to determine the extent of treatment exposure in the model is consistent with the use of the trial data to also derive efficacy inputs for the model (refer to Section B.3.3).

In the base case analysis, the cost of DBTd was calculated assuming that patients would receive the SC formulation of daratumumab (refer to Section B.3.5.1). Non-inferiority between the weight-based IV formulation of daratumumab (which was used in CASSIOPEIA) and the SC formulation of daratumumab has been demonstrated as part of the COLUMBA (MMY3012) trial, which was the primary source of evidence for the license extension granted in June 2020.(104, 120) As such, it is considered appropriate to use the efficacy data from CASSIOPEIA for DBTd in the model whilst reflecting the cost to the NHS of the SC formulation.

Comparators

As described in Section B.1.3.4, BTd is the primary comparator in this submission. Bortezomib-based regimens (BTd or Bd) are the only treatments recommended by NICE for this indication, and clinical expert opinion indicates triplet-therapy with BTd is considered to represent SOC as an induction therapy for newly diagnosed MM patients eligible for ASCT.(1, 27) The CUA has therefore been conducted to assess the cost-effectiveness of DBTd versus BTd.

As with DBTd, the treatment protocol included in the model for the BTd arm was based on the CASSIOPEIA trial (with full details provided in Section B.3.5.1):

- Four 28-day cycles of BTd as induction therapy
- High-dose chemotherapy and ASCT
- Two 28-day cycles of BTd as consolidation therapy

The use of four cycles of induction therapy is consistent with the bortezomib SmPC, which also recommends that patients with at least a partial response also receive two additional cycles of BTd.(113) Whilst consolidation therapy following ASCT is not NICE approved nor routinely funded in UK clinical practice, feedback from a recent advisory board meeting involving three UK clinicians was that the use of six cycles of BTd induction in clinical practice was common, with the additional two cycles often used whilst patients are being scheduled in for stem cell collection and transplant.(1). To ensure consistency with the efficacy inputs used in the model, which are based on post-consolidation assessments from CASSIOPEIA, and to better reflect the existing cost to the NHS of BTd induction, the

cost of two cycles of BTd consolidation therapy was also included in the model. As per DBTd, it was assumed that not all patients would receive the full number of cycles of induction or consolidation therapy in the BTd arm, based on data from the CASSIOPEIA trial (Table 47).(43)

In CASSIOPEIA, thalidomide was administered at a dose of 100 mg once daily (QD) in all cycles and dexamethasone was administered at a dose of 40 mg for Cycles 1–2 and on Days 1 and 2 of Cycles 3–4, with 20 mg dexamethasone administered on subsequent days of Cycles 3–4.(44) The 100 mg QD dose of thalidomide was administered to mitigate the risk of heterogeneity in daily dosing during the trial and UK clinical experts confirmed that the 100 mg QD dose is used to manage adverse events associated with thalidomide, such as peripheral neuropathy, and represents standard clinical practice in England.(1, 121) This is however different to the recommended posology in the bortezomib SmPC, which specifies that thalidomide is initially administered at a dose of 50 mg QD for Days 1–14 of Cycle 1, and if tolerated can be increased to 100 mg QD on Days 15–28 of Cycle 1 and further increased to 200 mg QD for all days in subsequent cycles; and that dexamethasone is administered at a dose of 40 mg on all days of all cycles.(113)

A scenario analysis has therefore been conducted in which the bortezomib SmPC-recommended posology for thalidomide and dexamethasone is utilised in the calculation of drug acquisition costs for BTd instead of the posology specified in the CASSIOPEIA trial protocol (refer to Section 6.7.3 for the results of scenario analyses). The efficacy of BTd in this scenario was assumed to be same as the base case analysis, and was therefore based on data from the CASSIOPEIA trial. That the modified BTd dose (100 mg thalidomide QD) has efficacy that is similar to, and certainly no worse than, the SmPC-recommended dose (up to maximum of 200 mg thalidomide QD) has been demonstrated in naïve and matching-adjusted indirect comparisons between patients receiving the two doses in the CASSIOPEIA and PETHEMA trials, respectively.(116) In the MAIC, modified BTd was seen to have similar or better efficacy compared with the BTd label.(116) Refer to Section B.2.13 for further details.

Stem-cell transplantation and consolidation therapy

In the model, the number of cycles of induction and consolidation therapy received by patients in the DBTd and BTd arms, and the proportion of patients assumed to receive high-dose chemotherapy and ASCT, were based on data from the CASSIOPEIA trial (Table 47). In order to align with the efficacy inputs used in the model, these inputs were derived from the ITT population of CASSIOPEIA.

The proportion of patients receiving ASCT and the number of cycles of induction and consolidation therapy received only directly affected the costs that were applied in the model (refer to Section B.3.5.1). The efficacy benefits of induction, ASCT and consolidation therapy were implicitly captured in the model via the use of post-consolidation response from the ITT population of CASSIOPEIA, which includes response for all patients irrespective of whether ASCT was received or whether all cycles of induction and consolidation therapy were completed.

As the proportion of patients receiving induction therapy, ASCT and consolidation therapy are specified in the model using the data from CASSIOPEIA, the total cost of treatment was applied as a single, lump-sum cost in the first cycle of the model (i.e. when all patients are still alive).

Table 47: Proportion of patients receiving induction, ASCT and consolidation therapy(43, 44)

	Proportion of patients			
Treatment	**Induction: 1 cycle**	**Induction: 2 cycles**	**Induction: 3 cycles**	**Induction: 4 cycles**
DBTd	98.7%	97.6%	96.7%	95.4%
BTd	99.3%	97.8%	97.0%	94.5%
Treatment	**ASCT[a]**	**Consolidation: 1 cycle**	**Consolidation: 2 cycles**	
DBTd	90.1%	85.8%	85.5%	
BTd	89.3%	82.7%	80.6%	

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

[a] The proportion of patients receiving ASCT in the model is based on the proportion of patients who completed transplantation in CASSIOPEIA.

B.3.3 *Clinical parameters and variables*

The clinical inputs used in the model were primarily derived from the CASSIOPEIA trial, which is the primary source of evidence for DBTd as a treatment for adult patients with newly diagnosed MM who are eligible for ASCT (refer to Section B.2.2). However, due to the immaturity of survival data from CASSIOPEIA (particularly for OS), external data with longer follow-up were also used to inform the relationship between MRD status and OS/progression in the model (refer to Section B.3.3.2).(61, 63)

B.3.3.1 Induction and consolidation: response and survival

Survival inputs pre-landmark

Survival for DBTd and BTd before the landmark point (model cycles 0–8) was modelled using the observed PFS and OS data from the respective arms of the CASSIOPEIA trial. The data used were based on the ITT population from the PHA (median follow-up = 29.2 months) (refer to Figure 12 for PFS and Figure 16 for OS in Section B.2.6.2).(65)

Post-consolidation response

As described in Section B.3.2.2, survival after the landmark point was modelled to be dependent on MRD status at the time of the post-consolidation response assessment, with separate survival inputs used for patients with MRD-negative and MRD-positive status. The proportion of patients achieving post-consolidation MRD negativity in the model was based on data from the ITT population of CASSIOPEIA, using a sensitivity threshold of 10^{-5} and measured using MFC (Table 48).

The post-consolidation timepoint was utilised in the model as this captures the benefits of induction, ASCT and consolidation therapy on response and was the only timepoint in the CASSIOPEIA trial (other than post-induction) at which MRD negativity was assessed.(81) The post-consolidation assessment was prespecified in CASSIOPEIA and was also the timepoint used for the primary efficacy endpoint in the trial.(81)

Table 48: MRD negativity at post-consolidation assessment(80, 81)

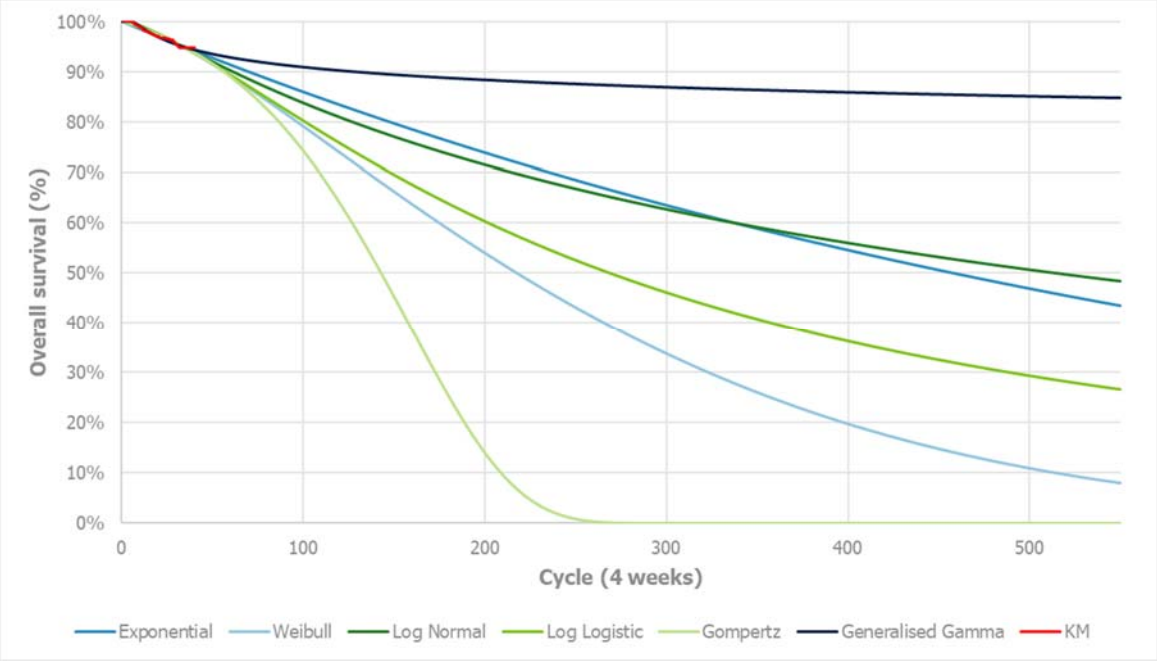
	BTd	DBTd
MRD-negative (%) (95% CI)	43.5 (39.3, 47.8)	63.7 (59.5, 67.8)
MRD-positive (%)	56.5	36.3
Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; CI = confidence interval; MRD = minimal residual disease.		

B.3.3.2 Survival inputs for post-consolidation

As described in Section B.3.2.2, a response-based model was developed in order to best utilise the data available from CASSIOPEIA and external sources to model long-term outcomes for newly diagnosed MM patients, and to reflect the aims of induction, ASCT and consolidation therapy, which is to attain a deep and durable response.

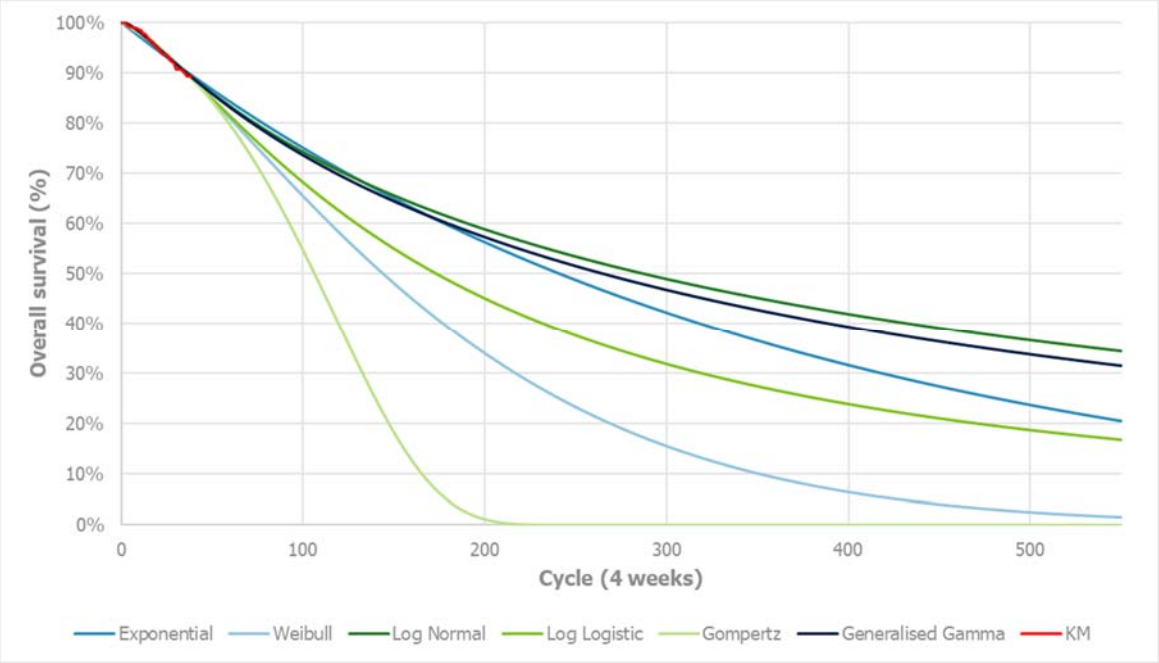
Survival outcomes were initially explored by directly extrapolating the OS data for DBTd and BTd from CASSIOPEIA, rather than modelling survival based on response. However, there was wide variation in the OS predicted by the different extrapolations (see Figure 29 for DBTd and Figure 30 for BTd), which would translate into high levels of uncertainty in the cost-effectiveness results.

Figure 29: Extrapolation of OS for DBTd (ITT population, median follow-up = 29.2 months)



Key: DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention-to-treat; KM = Kaplan-Meier; OS = overall survival.

Figure 30: Extrapolation of OS for BTd (ITT population, median follow-up = 29.2 months)



Key: BTd = bortezomib, thalidomide and dexamethasone; ITT = intention-to-treat; KM = Kaplan-Meier; OS = overall survival.

Approaches for modelling survival based on response were therefore explored. For each approach, consideration was given to both internal validity (i.e. how well the predicted survival fit the observed data from CASSIOPEIA) and external validity in terms of the clinical plausibility of long-term survival predictions.

SLR/meta-analysis for survival outcomes and MRD negativity

The prognostic significance of MRD status, and its relationship with long-term survival outcomes (PFS and OS) in patients with newly diagnosed MM has been established through SLR/meta-analysis.(61) In June 2019, an expanded SLR/meta-analysis was conducted which explored the prognostic utility of MRD for PFS/OS across a range of different disease settings (newly diagnosed transplant-eligible and ineligible MM, and relapsed/refractory MM), MRD sensitivity thresholds (10^{-4}, 10^{-5}, 10^{-6}), cytogenetic subgroups (high risk versus standard risk) and method of MRD assessment (MFC versus NGS versus PCR).(63) Importantly, this meta-analysis in a large cohort of MM patients confirmed that MRD negativity is significantly associated with extended PFS and OS, including those with newly diagnosed MM who are eligible for ASCT.

Extrapolation of survival outcomes for DBTd and BTd based on post-consolidation MRD status were initially explored using pooled Kaplan-Meier data from the SLR/meta-analysis. However, the survival outcomes predicted by the model substantially underestimated OS when compared to the observed trial data for both DBTd and BTd, and real-world evidence of outcomes from the PHE datasets (refer to Section B.2.9.5; and discussed further below). This is likely because the SLR/meta-analysis included a number of older trials which do not capture the shift in outcomes for MM patients due to the introduction of novel agents as well as trials with a range of MRD sensitivity thresholds, including 10^{-4}. Both relative and absolute survival estimates from the SLR/meta-analysis are therefore likely to underestimate outcomes for patients in current clinical practice who achieve MRD negativity at higher sensitivity thresholds (i.e. 10^{-5} or 10^{-6}).

In order to improve the internal validity of the model survival outcomes, alternative approaches were explored which leveraged survival data directly from the CASSIOPEIA trial, complemented with external data from the expanded SLR/meta-analysis.

Approach to modelling survival in the base case analysis: use of CASSIOPEIA and expanded SLR/meta-analysis survival data

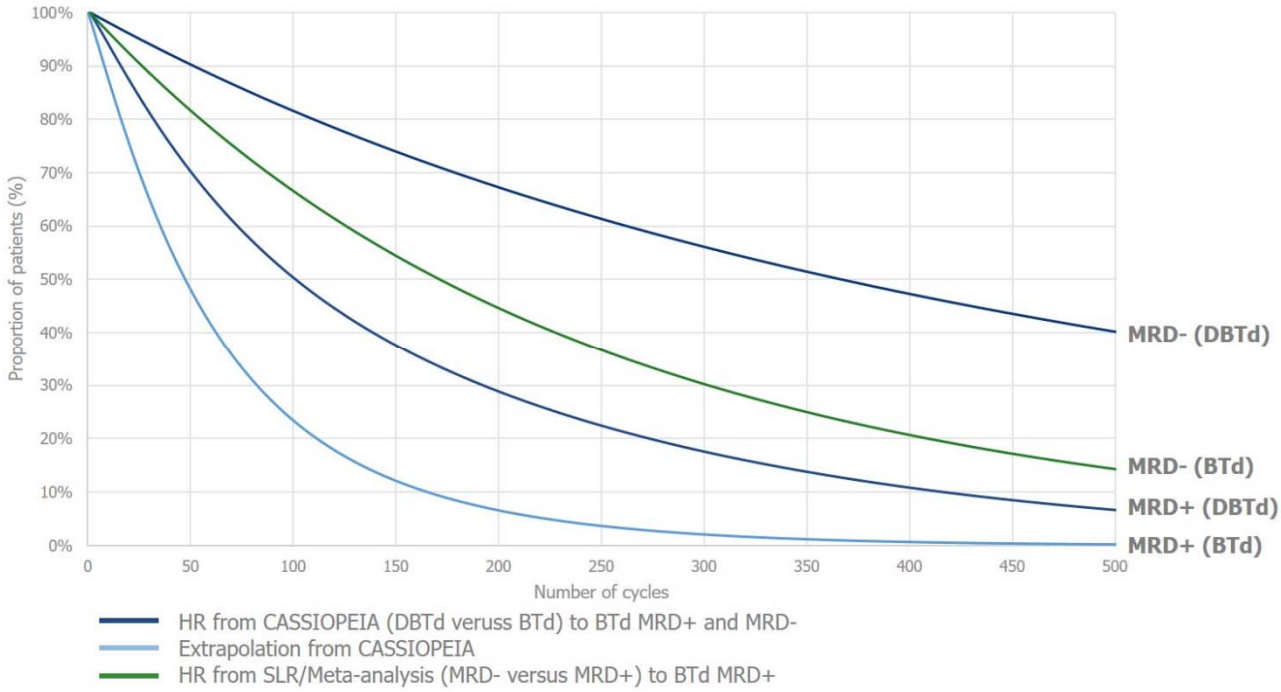
The CUA presented in this submission makes use of survival data from CASSIOPEIA and also further analyses of the expanded SLR/meta-analysis to inform the relationship between MRD negativity and survival. As described above, the inclusion of data from CASSIOPEIA was considered necessary to ensure the internal validity of survival outcomes predicted by the model when compared to the observed outcomes from the CASSIOPEIA trial. The further analysis of the expanded SLR/meta-analysis for MRD was conducted to incorporate the latest (May 2019) data cut from CASSIOPEIA and include those trials where MRD negativity was specifically assessed at 100 days post-ASCT (n=9 studies for OS and n=15 studies for PFS in patients newly-diagnosed MM who are eligible for ASCT). This analysis excluded the DBTd arm from CASSIOPEIA so as to evaluate the impact of MRD status on survival outcomes in NDTE MM based on existing standard of care treatments, and without including the treatment effect associated with daratumumab. Details of the SLR methodology and the list of studies included in the meta-analyses used to inform the model are presented in Appendix M.

In the CUA, survival was modelled to be dependent on both post-consolidation MRD status (MRD-positive or MRD-negative) and also treatment arm (DBTd or BTd). The inclusion of a treatment effect in addition to MRD response is supported by results of the landmark analysis of the CASSIOPEIA trial, which demonstrated improved survival in the DBTd arm versus BTd for both MRD-negative and MRD-positive patients (refer to Section B.2.6.3). As previously noted, a PFS treatment effect regardless of MRD response is [redacted].The superior outcomes for DBTd treated patients reflect deeper responses (higher proportion of patients achieving MRD negativity at sensitivity 10^{-6} and higher conventional response for MRD-positive patients measured by IMWG criteria) due to the unique mechanism of action of daratumumab, which is to modulate the immune system to better fight the disease.(118, 119)

In the base case analysis, the following approach for modelling post-landmark survival was taken (Figure 31), with the same approach used for modelling both OS and PFS:

- **BTd MRD+**: for patients in the BTd arm with MRD-positive response at the post-consolidation assessment point, OS and PFS were modelled by extrapolating IPD directly from the CASSIOPEIA trial (PHA; median follow-up = 29.2 months). The majority of patients treated with BTd were MRD-positive (56.5%), and a higher number of events for this cohort relative to any other MRD subgroup provided the most mature source of data from CASSIOPEIA to extrapolate from
- **BTd MRD-**: for patients in the BTd arm with MRD-negative response at the post-consolidation assessment point, OS and PFS were modelled via the application of a HR (MRD- versus MRD+) to the BTd MRD-positive survival curve. The HRs used in the base case were derived from the statistical analysis performed on the expanded SLR/meta-analysis described above
- **DBTd MRD+ and DBTd MRD-**: in the DBTd arm, OS and PFS for MRD-positive and MRD-negative patients were modelled via the application of HRs to the corresponding BTd survival curves (DBTd MRD+ versus BTd MRD+ and DBTd MRD- versus BTd MRD-). The HRs used in the base case were derived from the landmark analysis of the CASSIOPEIA trial (refer to Section B.2.6.3)

Figure 31: Approach to modelling survival by response and treatment arm in the base case analysis



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MRD = minimal residual disease; SLR = systematic literature review.

Alternative approaches using data from both CASSIOPEIA and the expanded SLR/meta-analysis have been explored in scenario analyses (see description later on in this section). However, the approach described above for the base case analysis is considered to make best use of the data available from both internal/external sources of evidence and also best reflect the survival benefit observed in the DBTd arm of CASSIOPEIA.

Full details of how survival inputs for the base case analysis have been derived are provided below.

Extrapolation of survival and application of hazard ratios

Where extrapolation of PFS and OS was required (i.e. for BTd MRD+), survival analyses were performed in accordance with the guidance provided in the NICE DSU Technical Support Document (TSD) 14.(122) The full range of parametric distributions were explored (exponential, Weibull, loglogistic, lognormal, Gompertz, and generalised gamma), with each model assessed in terms of goodness-of-fit statistics (Akaike information criterion [AIC] and the Bayesian information criteria [BIC]), visual inspection of the survival curves to the observed data from the CASSIOPEIA trial, and clinical plausibility of long-term survival predictions.

Given that the observed data for the BTd MRD-positive subgroup are still relatively immature, the clinical plausibility of long-term extrapolations was considered to be an important factor in selecting curves for the base case analysis.

The appropriateness of modelling survival via the use of HRs is dependent on the assumption of proportional hazards. This was assessed and confirmed for each outcome and MRD subgroup for which HRs were used in the model via inspection of log-cumulative hazard plots (refer Appendix N).

In order to explore uncertainty in the long-term treatment effect of daratumumab, scenario analyses have been conducted in which the HRs for DBTd versus BTd (PFS and OS) are set to equal one (i.e. no treatment effect) at specified timepoints in the model (5- and 10-years), to reflect the possibility that

the treatment effect of daratumumab may wane over time. No treatment waning effect was applied in the base case analysis and there is no evidence to suggest if and when the treatment effect of daratumumab on survival (for MRD-positive or MRD-negative patients) would wane over time. As well as not being included in TA311, treatment waning was also not considered in the previous NICE appraisals of daratumumab at later lines of therapy (TA573 and TA510).

Progression-free survival

Extrapolation of PFS for BTd patients with a post-consolidation MRD+ response was performed using the data from the landmark analysis of the CASSIOPEIA trial (refer to Section B.2.6.3), i.e. PFS was extrapolated from the time of post-consolidation response assessment.

Goodness-of-fit statistics for each parametric distribution explored are presented in Table 49 and the extrapolated curves are presented in Figure 32. The distribution associated with the lowest AIC and BIC values was the exponential, although other distributions, such as the Weibull and loglogistic had similar AIC and BIC values. On visual inspection of the survival curves, each of the distributions provided a reasonable fit to the initial slope of the Kaplan-Meier curve. Due to a lower number of events, and higher level of censoring, limited reliance was placed on visual fit towards the tail end of the Kaplan-Meier curve where there is greatest uncertainty.

The choice of distribution for the base case has largely been informed by comparing the clinical plausibility of survival estimates predicted by each model (Table 50). Clinical expert feedback from an advisory board meeting involving three UK clinicians indicated that for patients who still have residual disease (i.e. MRD-positive), less than 10% would be progression-free at 10 years, and none would be progression-free at 20 years.(1) That PFS rates would be less than 10% after 10 years was also supported by a UK clinician from whom feedback was obtained separately to the advisory board.(123) This clinician also noted that between 20–30% of MRD-positive patients would be expected to be progression-free after 5 years.(123) The Weibull and generalised gamma distributions both produced similar estimates of PFS at all timepoints modelled, which were within the ranges expected by the clinicians. The exponential distribution, which was associated with the lowest AIC and BIC values, predicted survival rates which were at the upper end of clinician estimates.

For the base case analysis, the Weibull distribution was used for the extrapolation of BTd MRD+ PFS, as this predicted plausible estimates of long-term survival when compared to the clinician's expectations. In order to explore uncertainty and provide an alternative estimate of long-term PFS, a scenario analyses has also been conducted using the exponential distribution to extrapolate BTd MRD+ PFS. The exponential distribution was associated with the best statistical fit in terms of AIC and BIC and provides a more optimistic estimate of long-term PFS whilst remaining within the bounds of clinical plausibility (refer to Section B.3.8.3 for the results of scenario analyses).

Table 49: Goodness-of-fit statistics for BTd MRD+ PFS (landmark analysis) survival models

Survival model	AIC	BIC
Exponential	**863.73**	**867.30**
Weibull	865.00	872.14
Lognormal	870.89	878.04
Loglogistic	864.99	872.13
Gompertz	865.47	872.62
Generalised Gamma	866.97	877.68

Key: AIC = Akaike's information criterion; BIC = Bayesian information criterion; BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.

Table 50: Comparison of predicted survival rates for BTd MRD+ PFS (landmark analysis) survival models

Survival model	PFS survival rates		
	5 years	10 years	20 years
Clinician estimate	20–30%[a]	<10%	<1%
Exponential	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]
Loglogistic	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]
Generalised Gamma	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.
[a] Feedback from UK clinician, not part of the clinical advisory board meeting for DBTd.(123)

Figure 32: Extrapolation of PFS for BTd MRD+ (landmark analysis)

Key: BTd = bortezomib, thalidomide and dexamethasone; KM = Kaplan-Meier; MRD = minimal residual disease; PFS = progression-free survival.

BTd MRD- and DBTd MRD+/- progression-free survival

PFS for the other patients included in the model (BTd MRD- and DBTd MRD-/+) were based on the application of HRs, as described in Table 51. In order to derive the survival probabilities for the overall cohort (MRD- and MRD+ combined), the MRD- and MRD+ PFS were weighted by the proportion of patients with MRD- and MRD+ status at the post-consolidation assessment timepoint (i.e. at the landmark point) (refer to Table 48).

The clinical plausibility of PFS predicted by the model, when also considering survival for MRD- patients and the DBTd cohort after applying these HRs, is discussed in a subsequent section.

Table 51: Hazard ratios for modelling PFS

Intervention and post-consolidation MRD response	HR (95% CI)	Application and source
BTd MRD-	[redacted]	HR (MRD- versus MRD+) applied to BTd MRD+ and derived from the expanded SLR/meta-analysis – transplant-eligible studies reporting survival from 100 days post-consolidation and excluding the DBTd arm from CASSIOPEIA
DBTd MRD+	[redacted]	HR (DBTd MRD+ versus BTd MRD+) applied to BTd MRD+ and derived from CASSIOPEIA landmark analysis
DBTd MRD-	[redacted]	HR (DBTd MRD- versus BTd MRD-) applied to BTd MRD- and derived from CASSIOPEIA landmark analysis

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MRD = minimal residual disease; PFS: progression-free survival.

Overall survival

BTd MRD+ overall survival

The approach used for the extrapolation of OS for BTd patients with a post-consolidation MRD+ response was the same as that used for PFS i.e. OS was extrapolated from the time of post-consolidation response assessment using data from the CASSIOPEIA landmark analysis.

Goodness-of-fit statistics for each parametric distribution explored are presented in Table 52 and the extrapolated curves are presented in Figure 33. The distributions with the lowest AIC and BIC values were the Gompertz and exponential, respectively. As with PFS, each of the distributions produced survival curves with a reasonable visual fit when compared to the initial slope of the Kaplan-Meier curve. The long-term extrapolations with OS were however more varied, with the exponential and Gompertz providing very different predictions of long-term OS (Table 53).

Feedback from the clinical advisory board meeting was that at 10 years, an OS rate of approximately 44% would be expected for NDTE patients with MRD-positive response based on cancer survival data reported in England from the Office for National Statistics (ONS) for persons aged 55–64 years at diagnosis.(1, 124) The 10-year survival rates predicted by the models were all higher than 44%, with the exponential distribution being associated with most conservative estimate of 10-year survival (52%). Beyond 10 years, the other distributions were all associated with more optimistic predictions of long-term OS than the exponential (Table 53 and Figure 33). Several of these extrapolations were also considered to predict implausibly high estimates of OS when compared to age- and sex-matched general population mortality (Figure 33). With the exception of the Weibull and exponential distributions,

the model estimates of OS exceeded survival in the general population within 30 years of the model time horizon.

For the base case analysis, the exponential distribution, which was associated with the best statistical fit according to BIC, was used for the extrapolation of BTd MRD+ OS, as this was the only distribution that provided a clinically plausible estimate of long-term OS. In the absence of other plausible estimates of long-term OS, scenario analyses were not conducted to explore alternative distributions for the extrapolation of BTd MRD+ OS. The choice of the exponential distribution for the base case was the preferred curve choice from all three UK clinicians attending the advisory board meeting.(1) The use of the exponential distribution was also supported by feedback from the UK clinician whose feedback was sought independently of the advisory board.(123) This clinician noted that 5-year OS rates are expected to be no higher than 70%.(123)

Table 52: Goodness-of-fit statistics for BTd MRD+ OS (landmark analysis) survival models

Survival model	AIC	BIC
Exponential	404.29	**407.96**
Weibull	405.48	412.82
Lognormal	403.70	411.04
Loglogistic	405.11	412.45
Gompertz	**403.36**	410.70
Generalised Gamma	405.44	416.45

Key: AIC = Akaike's information criterion; BIC = Bayesian information criterion; BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival.

Table 53: Comparison of predicted survival rates for BTd MRD+ OS (landmark analysis) survival models

Survival model	OS survival rates			
	5 years	10 years	20 years	30 years
Clinician estimate	≤70%[a]	44%	-	-
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	[redacted]
Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Generalised Gamma	[redacted]	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival.
[a] Feedback from UK clinician, not part of the clinical advisory board meeting for DBTd.(123)

Figure 33: Extrapolation of OS for BTd MRD+ (landmark analysis)

Key: BTd = bortezomib, thalidomide and dexamethasone; GPM = general population mortality; KM = Kaplan-Meier; MRD = minimal residual disease; OS = overall survival.

BTd MRD- and DBTd MRD+/- overall survival

OS for the other patients included in the model (BTd MRD- and DBTd MRD-/+) were based on the application of HRs, as described in Table 54. As for PFS, in order to derive the survival probabilities for the overall cohort (MRD- and MRD+ combined), the MRD- and MRD+ OS were weighted by the proportion of patients with MRD- and MRD+ status at the post-consolidation assessment timepoint (i.e. at the landmark point) (refer to Table 48).

The clinical plausibility of OS predicted by the model, when also considering survival for MRD- patients and the DBTd cohort after applying these HRs, is discussed in a subsequent section.

Table 54: Hazard ratios for modelling overall survival

Intervention and post-consolidation MRD response	HR (95% CI)	Application and source
BTd MRD-	[redacted]	HR (MRD- versus MRD+) applied to BTd MRD+ and derived from the expanded SLR/meta-analysis – transplant-eligible studies reporting survival from 100 days post-consolidation and excluding the DBTd arm from CASSIOPEIA
DBTd MRD+	[redacted]	HR (DBTd MRD+ versus BTd MRD+) applied to BTd MRD+ and derived from CASSIOPEIA landmark analysis
DBTd MRD-	[redacted]	HR (DBTd MRD- versus BTd MRD-) applied to BTd MRD- and derived from CASSIOPEIA landmark analysis

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; HR = hazard ratio; MRD = minimal residual disease.

General population mortality

The risk of mortality for patients with MM is expected to be higher than those of the general population when matched for age and gender. To ensure that OS predicted by the model for the overall cohort (MRD- and MRD+ combined) did not exceed that of the general population, age- and gender-matched general population mortality (based on life tables from England, 2016–2018) was used in any cycle where the predicted probability of death was lower than general population mortality.

Validation of survival (MRD+ and MRD- combined) in the base case analysis

The PFS and OS outcomes predicted by the model for the overall cohort (i.e. MRD- and MRD+ combined, weighted by the proportion of patients achieving post-consolidation MRD negativity), are presented in Figure 34 (for DBTd) and Figure 35 (for BTd). A comparison of survival rates from the model with the rates reported from the CASSIOPEIA trial are also presented in Table 55. For OS, the model provided a good visual fit to the Kaplan-Meier data from CASSIOPIEA with the proportion of patients predicted to be alive in the model generally consistent with the observed data from the trial. The proportion of patients predicted to be alive and progression-free in the model was slightly, but consistently, lower than the observed PFS data from CASSIOPEIA. However, the underestimation was consistent across both treatment arms and is not expected to materially impact the relative treatment effect or model results. The use of an alternative extrapolation for BTd MRD+ PFS, which is associated with a higher estimate of long-term PFS (exponential), has also been explored in scenario analyses. However, the use of the exponential distribution resulted in survival estimates for the first 36 months that were very similar to those seen using the Weibull, with differences in PFS extrapolations between the different distributions only becoming evident in the long term.

Public Health England linked datasets

As described in Section B.2.9.5, Janssen commissioned a real-world evidence study utilising the PHE linked datasets to investigate PFS/OS for NDTE MM patients. In Table 55, survival estimates are also provided from analyses conducted using routine patient data made available through the National Cancer Registration and Analysis Service (NCRAS) from PHE. Specifically, data from multiple, linked datasets were used to identify a cohort of patients in England with newly diagnosed MM who received ASCT, and report key characteristics, including clinical outcomes (OS and PFS), information on treatments received, and patient demographics, as well as information on healthcare resource utilisation and prognostics[IV]. All incidence primary diagnoses of newly diagnosed MM among patients aged ≥18 years old in England that were captured from the 1st January 2015 were considered for the analyses[V].

In total, [redacted] patients who were newly diagnosed with MM in England between 1st January 2015 and 31st December 2018 (inclusive) and who had received ASCT were included in the analysis, with follow-up to 31st December 2019. Patients who were receiving treatment as part of the Cancer Drugs Fund (CDF) in England were excluded from the analysis cohort with the exception of daratumumab patients that did not receive a CDF regimen at a prior or subsequent treatment line. The average age of patients included in the analysis was similar to patients in the CASSIOPEIA trial (see Table 56 for comparison versus the BTd arm), although a proportion of patients over the age of 65 years old were included in the PHE dataset cohort ([redacted]). Comparisons of disease stage and patient fitness between the PHE cohort and CASSIOPEIA are however limited, as [redacted]. Data from the SACT dataset could be linked to [redacted], which provided information on the treatments received and survival outcomes of these patients up to the 31st December 2019. BTd was the most commonly used regimen at first-line for patients who received ASCT ([redacted]), followed by BCd ([redacted]) and Bd ([redacted]). Survival rates at 12, 18 and 24 months are reported from the analysis for the time from the initiation of first-line therapy to death or censoring (OS) and for the time from the initiation of first-line therapy to death, censoring or the start of a new treatment line (which was considered to be a reasonable proxy for PFS), for patients receiving BTd and for all patients, regardless of first-line therapy (Table 55).

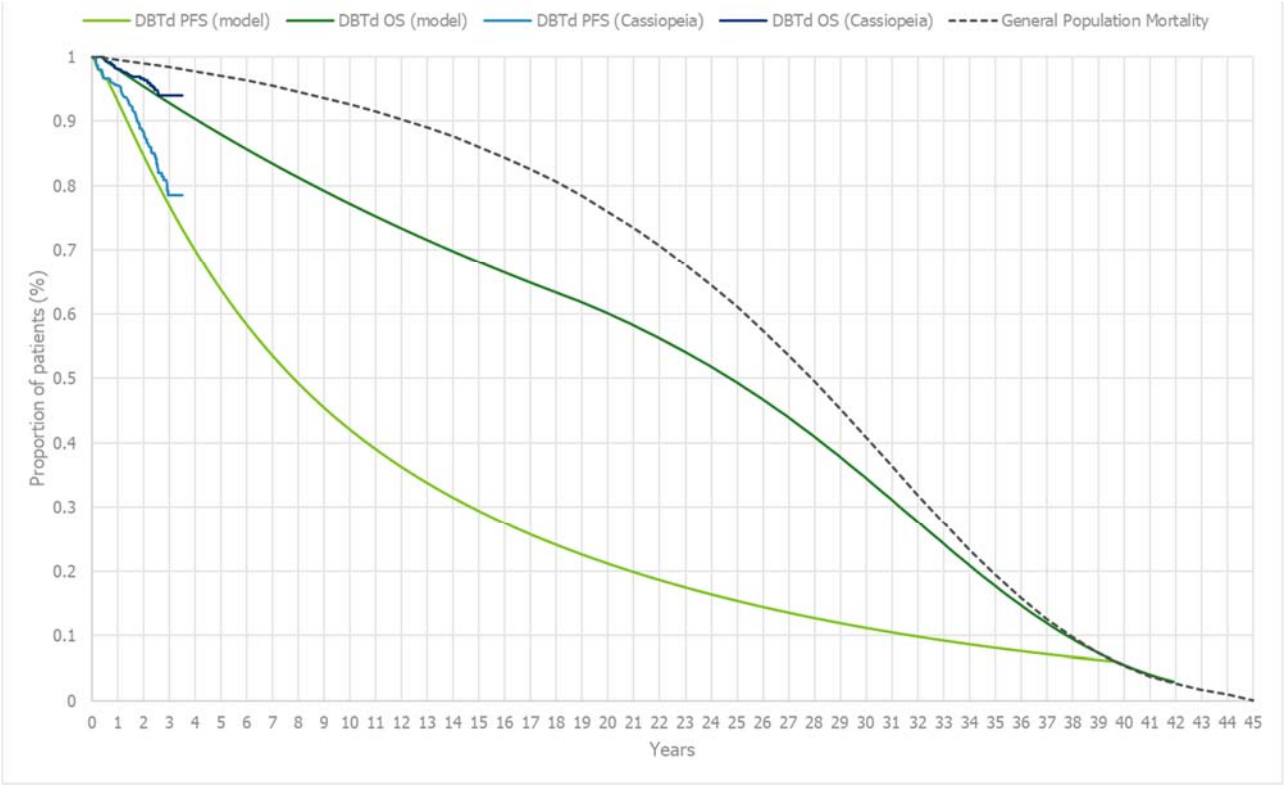
As shown in Table 55, the OS predicted by the model for the BTd arm was consistent with the survival rates from the PHE cohort (for BTd and all first-line therapy), with similar rates reported in the PHE cohort and the CASSIOPEIA trial. By contrast, the model predictions for PFS were generally higher than the survival rates reported in the PHE cohort, despite consistently underestimating PFS compared to the CASSIOPEIA trial. Differences in outcomes between CASSIOPEIA and the PHE cohort are likely related to the impact on survival of consolidation and maintenance treatment which are not funded by the NHS in England. Reassuringly, the model predicts similar BTd OS rates to both the CASSIOPEIA trial and PHE cohort, and at least predicts BTd PFS rates that are within the range provided by outcomes from the CASSIOPEIA trial and PHE cohort. That PFS predicted by the model for BTd is higher than PFS from the PHE cohort suggests that the use of efficacy data from CASSIOPEIA for BTd

[IV] The datasets that informed the analyses included: quality-assured and standardised diagnostic and pathological data submitted by NHS service providers to form the cancer registry; the Hospital Episode Statistics dataset (HES), which describes secondary care, including inpatient, outpatient and accident and emergency admissions; the Systemic Anti-Cancer Therapy dataset (SACT), which contains cancer-specific systemic treatment information; the Radiotherapy Dataset (RTDS), which describes cancer-specific radiotherapy treatments; and routine death registration data provided by the Office for National Statistics (ONS)

[V] A diagnosis of MM was defined according to the International Classification of Disease of Oncology 3rd edition (ICD-O-3) morphology code 9732 (multiple myeloma, myelomatosis, plasma cell myeloma and myeloma not otherwise specified), and this definition was utilised based on the recommendation of a PHE pathologist

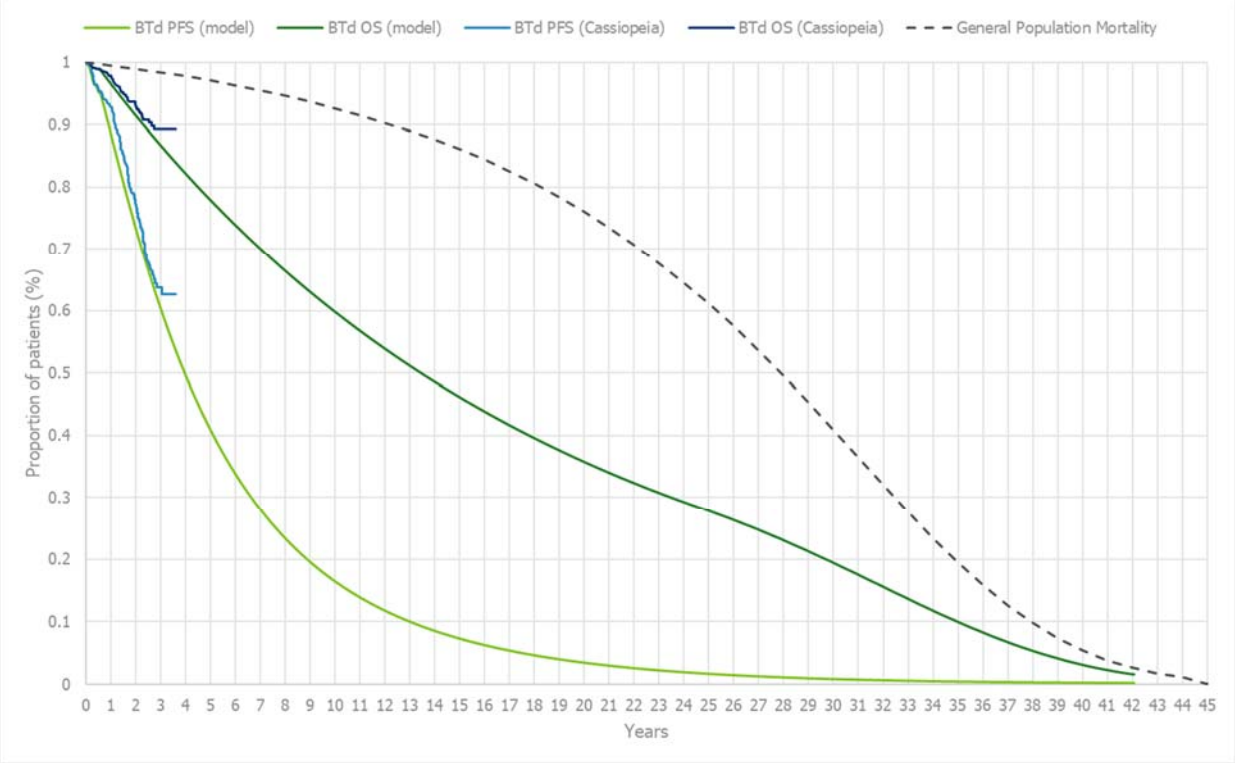
will bias the results of the CUA in favour of BTd when considering how BTd is typically used in UK clinical practice (i.e. without consolidation therapy).

Figure 34: Comparison of OS and PFS for DBTd predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined)



Key: DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival; PFS = progression-free survival.

Figure 35: Comparison of OS and PFS for BTd predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined)



Key: BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival; PFS = progression-free survival.

Table 55: Survival rates predicted by the model compared to those observed in CASSIOPEIA (May 2019 data cut; PHA) and the PHE dataset

Outcome	% alive (95% CI)				% alive and progression free (95% CI)[a]			
Time (months)	12	18	24	36	12	18	24	36
DBTd								
Model	98.1%	96.9%	95.4%	92.9%	93.2%	89.3%	84.8%	77.1%
CASSIOPEIA	98.1% (96.6, 99.0)	97.2% (95.4, 98.3)	96.6% (94.7, 97.9)	-	95.4% (93.3, 96.9)	92.5% (89.9, 94.5)	88.4% (85.3, 90.9)	-
BTd								
Model	96.7%	94.3%	91.6%	86.8%	88.8%	81.6%	73.6%	60.5%
CASSIOPEIA	97.8% (96.1, 98.7)	95.1% (92.9, 96.7)	93.2% (90.6, 95.0)	-	92.9% (90.3, 94.8)	85.3% (82.0, 88.1)	77.4% (73.4, 80.8)	-
PHE cohort All first-line treatments (ASCT-positive)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PHE cohort BTd as first-line treatment (ASCT-positive)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Key = BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; PHA = Post-hoc Interim Analysis; PHE = Public Health England.
[a] For the PHE cohort, the survival rates are based on survival from the initiation of first-line therapy to death, censoring or the start of a new treatment line. It was assumed that start of a new treatment line was a reasonable proxy for progression.

Table 56: Comparison of patient characteristics from the Public Health England dataset (at diagnosis) and BTd arm of the CASSIOPEIA trial (at baseline)

	PHE dataset (n=[redacted])	CASSIOPEIA BTd (n=542)
Mean age (SD)	[redacted]	56.7 (7.03)
Median age (range)	[redacted]	58.0 (26–65)
Female, n (%)	[redacted]	223 (41.1)
Age category, n (%)		
<50 years	[redacted]	90 (16.6)
50-64 years	[redacted]	452 (83.4)[a]
65 years and older	[redacted]	-
ECOG score, n (%)		
0	[redacted]	257 (47.4)
1	[redacted]	230 (42.4)
2	[redacted]	55 (10.1)
3	[redacted]	-
4	[redacted]	-
Null	[redacted]	-
ISS stage, n (%)		
I	[redacted]	228 (42.1)
II	[redacted]	233 (43.0)
III	[redacted]	81 (14.9)
Null	[redacted]	-

Key: BTd = bortezomib, thalidomide and dexamethasone; ECOG = Eastern Cooperative Oncology Group; ISS = International Staging System; PHE = Public Health England; SD = standard deviation.
[a] ≥50-65 years in CASSIOPEIA

Scenario analyses for modelling response-based survival

Alternative approaches for modelling response-based survival using data from the CASSIOPEIA trial and statistical analysis performed on the expanded SLR/meta-analysis were explored as scenario analyses (refer to Section B.3.8.3 for the results of scenario analyses). The two alternative approaches were as follows:

Scenario A: *As per the base case analysis, with the exception that*: survival for patients in the BTd arm with MRD- response post-consolidation was also modelled by extrapolating IPD directly from the CASSIOPEIA trial (as per BTd MRD+; refer Appendix O for the choice of extrapolations in this scenario). Survival for patients in the DBTd arm was modelled via the application of HRs from the CASSIOPEIA trial, as per the base case analysis. As such, in this scenario, only data from the CASSIOPEIA trial were used to determine survival.

This approach does not leverage external data with longer follow-up to inform long-term survival predictions. Therefore, the use of the HRs from the statistical analysis performed on the

expanded SLR/meta-analysis for BTd MRD- survival was preferred in the base case analysis. The aim of this scenario was to explore the impact on the ICER of not using the external data at all.

Scenario B: *As per the base case analysis, with the exception that*: survival for patients in the DBTd arm with MRD- response post-consolidation was modelled via the application of the HR from the statistical analysis performed on the expanded SLR/meta-analysis (HR for MRD- versus MRD+) to the DBTd MRD+ survival curve.

This approach does not capture the treatment effect of daratumumab for patients who achieve MRD negativity, as demonstrated in CASSIOPEIA and [redacted].2.6.3 The use of the HR (DBTd MRD- versus BTd MRD-) from the CASSIOPEIA landmark analysis was therefore preferred in the base case analysis. The aim of this scenario was to explore the impact on the ICER of not including a daratumumab treatment effect for MRD- patients.

B.3.3.3 Adverse events

In the model a proportion of patients were assumed to experience adverse events following treatment with induction, ASCT and consolidation therapy based on data from the CASSIOPEIA trial (Table 57).

The adverse events included in the model were those Grade 3 and 4 events that were reported in at least 5% of patients in the CASSIOPEIA trial (Part 1: induction, ASCT, and consolidation), and also nausea and upper respiratory tract infections (URTI) (any grade). Grade 1 and 2 events were not included in the model as these are unlikely to be associated with considerable health-related costs or changes in patient HRQoL. Any-grade nausea and URTI were however included in the model as being events of clinical importance, as per clinician feedback provided as part of NICE TA510.(125) The inclusion rule that events must have occurred in at least 5% of patients in the CASSIOPEIA trial was selected so as to capture adverse events that would impact patients consistently enough to have validity in a real-world setting where adverse events are monitored in a less strict manner compared with a clinical trial setting.

The change in utility and cost associated with each adverse event are presented in Section B.3.4.4 and B.3.5.3, respectively. As the proportion of patients experiencing adverse events following treatment with induction, ASCT and consolidation therapy is based on data from CASSIOPEIA, the cost and disutility of adverse events were applied in the first cycle of the model (i.e. when all patients are still alive)

Table 57: Incidence of adverse events included in the model (induction, ASCT and consolidation)(44)

Adverse event	DBTd	BTd	Source
Neutropenia	27.61%	14.68%	CASSIOPEIA CSR Part 1 (Grade 3 or 4 Treatment-emergent adverse events during induction/ASCT/consolidation occurring in at least 5% of patients).
Lymphopenia	16.98%	9.67%	
Thrombocytopenia	11.01%	7.43%	
Febrile neutropenia	6.72%	5.20%	
Stomatitis	12.69%	16.36%	

Peripheral sensory neuropathy	8.77%	8.55%	Nausea and upper respiratory tract infection included based on clinical importance [all grades]
Nausea	30.22%	24.16%	
Upper respiratory tract infection	6.16%	3.35%	
Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; CSR = Clinical Study Report.			

B.3.4 *Measurement and valuation of health effects*

B.3.4.1 Health-related quality-of-life data from clinical trials

Utility values for the induction, ASCT and post-consolidation periods were derived using EQ-5D-5L data from the CASSIOPEIA trial, which were collected at the following timepoints: baseline, Cycle 4 Day 28, and Day 100 post-ASCT.(44) EQ-5D utility was seen to be similar between the two treatment arms at each timepoint (refer to Section B.2.6.4 and Appendix L), and so the utility values used in the model were based on pooled EQ-5D data across treatment arms, with the same values applied to both model cohorts.(44)

The utility values used in the model are presented in Table 58. These were derived using the cross-walk method reported by van Hout et al. (2013) to map EQ-5D-5L dimension scores to utilities using the UK EQ-5D-3L value set.(126)

Table 58: Mean EQ-5D utility during induction/ASCT/consolidation phase of CASSIOPEIA

	Baseline, mean (SD)	Cycle 4 Day 28, mean (SD)	Day 100 post-ASCT, mean (SD)
Pooled DBTd and BTd	0.57 (0.31)	0.68 (0.22)	0.73 (0.17)
Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EQ-5D-5L = EuroQol-5D, 5 levels; SD = standard deviation.			

The utility values from the trial were applied in the model as follows:

- 'Baseline' utility from the trial was applied to patients entering the model and for the duration of induction therapy, based on mean duration in CASSIOPEIA (Table 59; model cycles 0–3 in both treatment arms)
- 'Cycle 4 Day 28' utility from the trial was applied once patients had completed induction therapy and was assumed to be maintained until the time of the post-consolidation response assessment (Table 59; model cycles 4–8 in both treatment arms). The higher utility value in this period compared to the induction therapy period reflects the expectation that patients would experience some benefit from having received induction therapy, but would no longer experience adverse events associated with induction therapy
- 'Day 100 post-ASCT' utility from the trial was applied for the remainder of the time horizon for patients who remained progression-free

Table 59: Duration for health-state utility values (induction and post-induction to post-consolidation response) based on CASSIOPEIA(44)

	Duration of induction therapy, weeks	Duration from completion of induction therapy to response assessment, weeks[a]
DBTd	15.65	21.63
BTd	15.42	21.42

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

[a] Based on the mean gap between induction and SCT in CASSIOPEIA + 100 days.

For the PD health state, utility in the model was based on data from the published literature, including sources used in previous UK HTA of induction therapy (i.e. SMC ID 927/13 and NICE TA311).(27, 117) In the bortezomib HTA, utility data were mainly derived from van Agthoven et al. (2004), with values of 0.69 (2nd and 3rd lines) and 0.64 (further line treatment) used for later lines of therapy following 1st line disease progression.(62) As patients are expected to spend a greater period of time in 2nd and 3rd line, compared to 4th line, the value of 0.69 was used for the PD health state utility in the model. That anti-cancer therapies are now also available at 4th line (refer to Section B.1.3.4), supports the use of the higher utility value from TA311. This value also has reasonable face validity when compared to the values used from CASSIOPEIA for the PF health states in that it is: a) lower than the value used for patients in the PF health state following the post-consolidation assessment timepoint, which is consistent with the expectation that HRQoL would be reduced for patients with relapsed/refractory MM, and b) it is similar to the Cycle 4 Day 28 value from CASSIOPEIA, when patients would have just completed anti-cancer therapy.

A summary of the utility values used in the base case analysis are presented in Table 60. Several scenario analyses were also conducted in which lower utility values were used for PD and in which alternative sources were used for all health state utility values (Table 61; refer to Section B.3.8.3 for the results of scenario analyses).

Table 60: Health-state utility values included in the model (base case)

	Value	Source/justification
PF (induction therapy)	0.57[a]	CASSIOPEIA EQ-5D-5L at baseline (DBTd arm) with utility derived using van Hout et al. (2012)(126)
PF (post-induction to post-consolidation response)	0.68[a]	CASSIOPEIA EQ-5D-5L at Cycle 4 day 28 (DBTd arm) with utility derived using van Hout et al. (2012)(126)
PF (post-consolidation)	0.73	CASSIOPEIA EQ-5D-5L at Day 100 post ASCT (DBTd arm) with utility derived using van Hout et al. (2012)(126)
PD	0.69	Average of van Agthoven et al. (2004) utility values from baseline to 18-months post-treatment which were used for 2nd line and 3rd line health state utility values in TA311 (27, 62, 117)

Key: ASCT = autologous stem cell transplant; CSR = Clinical Study Report; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EQ-5D-5L = EuroQol-5D, 5 levels; PF = progression free; PD: progressed disease.

Table 61: Health-state utility values included in the model (scenario analyses)

	Value	Source/justification
Scenario: lower utility value for PD (TA311)		
PF	-	As per base case
PD	0.644	From van Agthoven et al. (2004)(62) and used in TA311(27) Utility for 'Further lines'
Scenario: lower utility value for PD (TA510)		
PF	-	As per base case
PD	0.57	From Palumbo et al. (2013)(127) and used in TA510(125) Utility for PD after 4th line treatment
Scenario: TA311 utility values from van Agthoven et al. (2004) and Segeren thesis		
PF (induction therapy)	0.57	From Segeren thesis (not reported in van Agthoven et al. (2004)) and used in TA311(27) Utility for 'From start of treatment until post-induction response'
PF (post-induction to post-consolidation response)	0.65	From Segeren thesis (not reported in van Agthoven et al. (2004)) and used in TA311(27) Utility for 'From post-induction response to post-SCT response'
PF (post-consolidation)	0.75	From van Agthoven et al. (2004)(62) and used in TA311(27) Utility for 'SCT patients 18+ months after SCT'
PD	0.69	From van Agthoven et al. (2004)(62) and used in TA311(27) Utility for '2nd and 3rd line treatments'. Average of van Agthoven et al. (2004) utility values from baseline to 18-months post-treatment

Key: EQ-5D-3L = EuroQol-5D, 3 levels; PF = progression free; PD: progressed disease; SCT = stem cell transplant.

B.3.4.2 Mapping

HRQoL data were collected in the CASSIOPEIA trial using the EQ-5D-5L.(44) In accordance with the NICE position statement in the use of EQ-5D-5L to derive utility values, the EQ-5D-5L descriptive scores from CASSIOPEIA were mapped onto the 3L UK value set using the mapping function developed by van Hout et al. (2012) and the tool available on the EuroQol website.(126, 128, 129) The utility values presented in Table 58, which are used in the model, are derived from this mapping exercise.

B.3.4.3 Health-related quality-of-life studies

A SLR of humanistic burden was conducted to identify evidence on HRQoL, patient-reported outcomes and utilities in patients with newly diagnosed MM (refer Appendix H). In total, the review identified nineteen primary studies (nine from the original review and ten from an update of the original SLR) reporting on HRQoL and other patient-reported outcomes in patients with transplant-eligible newly diagnosed MM. Only one study (Abonour et al. [2018]) was identified that reported utilities in patients with transplant-eligible newly diagnosed MM.(130) This study reported EQ-5D index values from patients in a US registry who received ASCT and did or did not receive maintenance therapy (n=244 any maintenance; n=169 lenalidomide maintenance only; n=137 no maintenance).(130) Patients who received consolidation therapy prior to maintenance were however excluded from the analysis.(130)

As described above in Section B.3.4.1, utility values used in the model were based on data collected as part of the CASSIOPEIA trial and those used in previous UK HTA for induction therapies (i.e. ID 927/13 and NICE TA311).(27, 117) With the availability of EQ-5D data from the CASSIOPEIA trial, with which to derive utility values for induction, ASCT and post-consolidation, the use of EQ-5D data from Abonour et al. (2018) was not explored in the CUA. The utility values reported from Abonour et al. (2018) in the no maintenance group (0.75 at baseline, 0.79 at pre-ASCT, 0.83 during follow-up from 100 days post ASCT, and 0.79 at disease progression) were consistently higher than those derived from the CASSIOPEIA trial and those used in previous UK HTA for induction therapies (see below).(130) Furthermore, when compared to UK population norms for utility individuals aged 55–64 years (0.80), the values from Abonour et al. (2018) are implausibly high when considering the symptom burden associated with MM.(130, 131) Utility values from the CASSIOPEIA trial were therefore preferred in the CUA for consistency with the patient population and source of efficacy inputs used in the model, and also the utility that might be expected of patients with MM in the UK.

B.3.4.4 Adverse reactions

Decrements in utility were applied in the model for the proportion of patients who experienced adverse events associated with induction therapy, ASCT and consolidation. The utility decrements used in the model were primarily based on those used in previous UK HTA of daratumumab (Table 62).(125, 132)

It was assumed that the loss of utility associated with adverse events would not last for the entire duration of induction, ASCT and consolidation therapy (~37 weeks in both treatment arms). The utility values applied were therefore adjusted such that the duration of disutility was assumed to be 28 days (equivalent to one cycle of induction therapy), as per the assumption used in NICE TA510.(125) Taking into account the proportion of patients experiencing each adverse event in each treatment arm (Table 57), the total disutility across all events included in the model was 0.02 for DBTd and 0.01 for BTd.

Table 62: Duration and utility decrements associated with adverse events included in the model

Adverse event	Duration (days)	Utility	Total decrement	Source
Neutropenia	28.00	-0.15	0.02	Based on TA573/TA510 (Brown 2013/Partial
Lymphopenia	28.00	-0.07	0.01	

Thrombocytopenia	28.00	-0.31	0.03	Review TA171)(125, 132)
Febrile neutropenia	28.00	-0.39	0.04	
Stomatitis	28.00	-0.15	0.02	Lloyd et al. (2006)(133)
Peripheral sensory neuropathy	28.00	-0.07	0.01	Based on TA573/TA510 (Brown 2013/Partial Review TA171)(125, 132)
Nausea	28.00	-0.10	0.01	Lloyd et al. (2006)(133)
Upper respiratory tract infection	28.00	-0.19	0.02	Based on TA573/TA510 (Brown 2013/Partial Review TA171)(125, 132)

B.3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis

A summary of utility values included in the base case analysis are presented in Table 63.

In the model, the health state utility values were also age-adjusted using the population norm values for EQ-5D as reported in Janssen et al. (2014).(134)

Table 63: Summary of utility values for cost-effectiveness analysis

State	Mean utility value	95% CI	Reference in submission	Source/justification
Health state utility values				
PF (induction therapy)	0.57	0.55–0.59	Section B.3.4.1	Based on EQ-5D-5L data from CASSIOPEIA (pooled across treatment arms), with utilities derived using the mapping function from van Hout et al. (2012)(126, 135)
PF (post-induction to post-consolidation response)	0.68	0.66–0.69		
PF (post-consolidation)	0.73	0.72–0.74		
PD	0.69	-		Based on utility values used in TA311 (for 2nd and 3rd line).(27) Alternative PD utility values are explored in scenario analyses
Adverse event utility decrements				
Neutropenia	0.02	-	Section B.3.4.4	Based on TA573/TA510 (Brown 2013/Partial Review TA171)(125, 132) Adjusted for assumed duration of 28 days
Lymphopenia	0.01	-		
Thrombocytopenia	0.03	-		
Febrile neutropenia	0.04	-		
Stomatitis	0.02	-		Lloyd et al. (2006)(133) Adjusted for assumed duration of 28 days
Peripheral sensory neuropathy	0.01	-		Based on TA573/TA510 (Brown 2013/Partial Review TA171)(125, 132) Adjusted for assumed duration of 28 days
Nausea	0.01	-		Lloyd et al. (2006)(133) Adjusted for assumed duration of 28 days
Upper respiratory tract infection	0.02	-		Based on TA573/TA510 (Brown 2013/Partial Review TA171)(125, 132) Adjusted for assumed duration of 28 days

Key: CI = confidence interval; EQ-5D-5L = EuroQol-5D, 5 levels; PD = progressed disease; PF = progression-free

B.3.5 *Cost and healthcare resource use identification, measurement and valuation*

The analysis was conducted from the perspective of the NHS in England and therefore included only costs that would be incurred by the health system. Appropriate sources of unit costs, such as NHS reference costs 2018–19, British National Formulary (BNF) and drugs and pharmaceutical electronic market information tool (eMIT) were used for cost inputs in the model.

The following cost types were included in the model: drug acquisition and administration costs for induction/consolidation therapy and subsequent therapies, cost of concomitant medication for induction/consolidation therapies, ASCT costs, costs associated with monitoring, and costs associated with the management of adverse events.

As part of the economic SLR (refer to Appendices G and I), no cost or resource use studies that were specific to the UK were identified in either the original or updated searches.

B.3.5.1 Intervention and comparators' costs and resource use

Drug acquisition costs – induction and consolidation

The dosing regimens for induction and consolidation therapies included in the model are presented in Table 64. These were based on treatment protocols specified for the CASSIOPEIA trial, with the exception that for DBTd the dose of daratumumab was based on the SC formulation (1,800 mg fixed dose) that is now available as a result of a license extension in June 2020. A scenario analysis was also conducted in which the weight-based dose and IV formulation of daratumumab (16 mg/kg), which was used in CASSIOPEIA, was used in the cost calculations for DBTd. In this scenario, only the acquisition and administration costs of daratumumab was changed.

The unit costs and total costs per administration associated with the individual therapies included within the induction and consolidation regimens are presented in Table 65 (daratumumab and bortezomib) and Table 66 (thalidomide and dexamethasone). The cost per administration for bortezomib (BSA-based dosing) were calculated using the mean BSA (1.88 m^2) of patients included in the CASSIOPEIA trial (Table 6), with the mean weight (75.67 kg) from CASSIOPEIA also used in the scenario for the IV formulation of daratumumab (weight-based dosing).(44) In the base case analysis, it was assumed that there would be no vial sharing and so the number of vials required per administration was rounded up to the nearest whole integer.

In the cost-effectiveness analyses presented in this submission, the cost per vial of bortezomib is based on the list price. However, the cost of bortezomib may vary in different regions because of negotiated procurement discounts and use of generic versions of the drug.

The total cost of induction and consolidation therapy applied in the model was £85,797.13 for DBTd and £20,194.21 for BTd (Table 67). The total costs presented in Table 67 take into account that not all patients are modelled to receive all four cycles of induction therapy and all two cycles of consolidation therapy, as described in Section B.3.2.3.

Concomitant medication costs

The cost of concomitant medications were also included in the model based on the recommendations provided in the SmPCs for daratumumab, bortezomib and thalidomide (Table 68).(5, 113, 136)

Specifically, for patients receiving thalidomide the cost of prophylaxis for thromboembolic events (consisting of low molecular weight heparin and aspirin) was included in the model. In line with the minimum recommended duration of prophylactic treatment, it was assumed that patients would be administered prophylaxis for 5 months.(136) For patients receiving daratumumab, both pre-infusion medications (antipyretics and antihistamine) and post-infusion medications (oral corticosteroids) were included in the model, as per the recommendations for daratumumab as part of a combination therapy.(5) For patients receiving bortezomib, the cost of antiviral treatment for the prevention of herpes zoster reactivation was included in the model based on recommendations from the bortezomib SmPC.(113)

The total cost of concomitant medications was applied as a single cost in the first cycle of the model, along with the main treatment costs. Concomitant medications for treatments received at subsequent lines of therapy were not included in the model.

Stem-cell transplantation costs

The costs associated with ASCT were applied in the model for the proportion of patients who were modelled to receive ASCT following induction therapy (see Table 47 for the proportions used in the base case analysis for each treatment arm). The costs included in the model were those related to stem cell mobilisation and harvesting, ablation (with high-dose chemotherapy), the transplant procedure itself, and post-transplant treatment to facilitate stem cell engraftment and reduce the duration of neutropenia, which is consistent with the costs of ASCT included in TA311.(27) In the CASSIOPEIA trial, a proportion of patients in each treatment arm (110/543 [20.3%] DBTd arm and 39/542 [7.2%] BTd arm; all randomised patients) also received plerixafor to mobilise stem cells, in addition to cyclophosphamide/G-CSF.(44) The cost of plerixafor has therefore been included in the ASCT cost calculations, with the proportion of patients receiving plerixafor based on the CASSIOPEIA trial. The other ASCT costs were the same in each treatment arm.

The cost of ASCT included in the model is presented in Table 69.

Table 64: Summary of dosing regimens for first-line treatment included in the model

Treatment	Treatment Phase	Duration	Drugs per cycle	Total Administrations	Source/Justification
DBTd[a]	Induction	4 cycles of 28 days	Daratumumab – 1,800 mg QW for 2 cycles, Q2W for 2 cycles	11.69	CASSIOPEIA protocol (with the exception of daratumumab dose, which is based on the SC formulation) As described in Table 47, only a proportion of patients are modelled to receive the full number of cycles of induction and consolidation therapy (based on CASSIOPEIA)(81), which is reflected in the total number of administrations
			Bortezomib – 1.3 mg/m^2 BW for 2 weeks	15.54	
			Thalidomide – 100 mg QD	108.75	
			Dexamethasone – 40 mg BW for cycles 1 and 2	15.71	
			Dexamethasone – 40 mg BW in week 1, 20 mg BW in weeks 2 and 3 for cycles 3 and 4	11.52	
	Consolidation	2 cycles of 28 days	Daratumumab – 1,800 mg Q2W	3.43	
			Bortezomib – 1.3 mg/m^2 BW for 2 weeks	6.85	
			Thalidomide – 100 mg QD	47.96	
			Dexamethasone – 20 mg BW for weeks 1–3	10.28	
BTd	Induction	4 cycles of 28 days	Bortezomib – 1.3 mg/m^2 BW for 2 weeks	15.54	
			Thalidomide – 100 mg QD	108.80	
			Dexamethasone – 40 mg BW for 2 cycles	15.76	

			Dexamethasone – 40 mg BW week 1, 20 mg BW weeks 2, 3, for 2 cycles	11.49	
	Consolidation	2 cycles of 28 days	Bortezomib – 1.3 mg/m^2 BW for 2 weeks	6.53	
			Thalidomide – 100 mg QD	45.72	
			Dexamethasone – 20 mg BW for weeks 1–3	9.80	

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; BW = bi-weekly; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; IV = intravenous; QD = daily; QW = every week; Q2W = every 2 weeks; SC = subcutaneous.
[a] In the scenario analysis using the IV formulation of daratumumab, the dose of daratumumab was 16 mg/kg per administration.

Table 65: Drug acquisition costs for first-line treatment (daratumumab and bortezomib)

Treatment	Dose per administration	Total dose per administration	Vial Size	Cost per Vial	Vials per administration[b]	Cost per administration
Daratumumab	1,800 mg	1,800 mg	1,800 mg	£4,320.00	1	£4,320.00
Daratumumab (IV)[a]	16 mg/kg	1,210.72 mg	100 mg	£360.00	13	£4,680.00
			400 mg	£1,440.00	-	
Bortezomib	1.3 mg/m²	2.45 mg	3.5 mg	£762.38	1	£762.38

Key: IV = intravenous.
Unit costs were derived from the BNF online.
[a] IV formulation of daratumumab was used in a scenario analysis only.
[b] The cost per administration for daratumumab in the scenario analysis is the same when calculated as 13x 100 mg vials or 3x 400 mg vials and 1x 100 mg vials.

Table 66: Drug acquisition costs for first-line treatment (thalidomide and dexamethasone)

Treatment	Unit Size (mg)	Pack Size	Cost per Pack	Cost per mg
Thalidomide	50	28	£298.48	£0.21
Dexamethasone	8	50	£25.17	£0.06
Unit costs were derived from the BNF online for thalidomide and eMIT for dexamethasone.				

Table 67: Summary of drug acquisition costs for first-line treatment – induction and consolidation

Treatment	Drug costs per cycle	Total regimen costs per cycle
DBTd		
Daratumumab (SC)	£65,317.13	£85,797.13
Bortezomib	£17,067.20	
Thalidomide	£3,341.00	
Dexamethasone	£71.80	
BTd		
Bortezomib	£16,828.62	£20,194.21
Thalidomide	£3,294.29	
Dexamethasone	£71.29	
Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; SC = subcutaneous.		

Table 68: Concomitant medication costs

Treatment	Dosing Regimen	Unit size	Units per pack	Cost per pack	Cost per unit	Total cost	Source
Thalidomide concomitant medication (prophylaxis for thrombosis)							
Low molecular weight heparin	Dalteparin 5000 units QD SC	5000	10	£8.84	£0.88	£80.98	**For calculation of total costs:** low molecular weight heparin and aspirin assumed to be taken by 60% and 40% of patients, respectively, based on inputs used in NICE TA311.(69) Duration of prophylaxis assumed to be 5 months based on the minimum recommended time in the Thalidomide Celgene SmPC.(136) **Dosing regimen:** CASSIOPEIA trial(81) **Drug costs**: Heparin: eMIT (5 mL solution for injection) Aspirin: eMIT (75 mg tablets)
Aspirin	75 mg QD	75 mg	28	£0.12	£0.00		
Daratumumab concomitant medications							
Antipyretic: oral paracetamol pre-infusion	1000 mg per administration of daratumumab	500 mg	32	£0.16	£0.01	£22.24	**Dosing regimen**: Darzalex SmPC, as per with TA573(117) **Drug costs**: Paracetamol: eMIT (500 mg soluble tablet) Diphenhydramine: MIMS 2020 Methylprednisone: BNF online (accessed April 2020); Drug
Antihistamine: oral/IV diphenhydramine pre-infusion	50 mg per administration of daratumumab	25 mg	20	£3.16	£0.16		
Corticosteroid: oral methylprednisolone post infusion	20 mg per administration of daratumumab	16 mg	30	£17.17	£0.57		

							Tariff Price (PART VIIIA Category C#)
Bortezomib concomitant medication							
Antiviral: aciclovir daily	400 mg QD	200 mg	25	£0.52	£0.02	£10.86 (DBTd arm)[a] £10.76 (BTd arm)[a]	**Dosing regimen**: Velcade SmPC(113) **Drug costs**: eMIT (200 mg dispersible tablet)
Key: IV = intravenous; QD = daily; SC = subcutaneous; SmPC = Summary of Product Characteristics. [a] Total cost differs between treatment arms due to the different duration of induction and consolidation therapy in each model cohort.							

Table 69: Stem-cell transplantation costs

Description	Intervention	Cost	Source	Administration Cost	Source
Mobilisation	Cyclophosphamide (1.5 g/m^2 BSA)	£48.96[a]	eMIT (500 mg powder for solution)	£254.14	NHS Reference Costs 2018-19. SB12Z DCRDN: Deliver Simple Parenteral Chemotherapy at First Attendance
	GCSF: Lenograstim 19.2 MU/m^2 daily (assumed 5 days)	£625.40[b]	BNF online (1 vial of 33.6 MU)	£493.70	NHS Reference Costs 2018-19. N10AF: Specialist Nursing, Cancer Related, Adult, Face to face (5 administrations)
	Plerixafor 20 mg daily (assumed 4 days) for a proportion of patients only[c]	£19,531.08	BNF online (1 vial of 24 mg)	£394.96	NHS Reference Costs 2018-19. N10AF: Specialist Nursing, Cancer Related, Adult, Face to face (4 administrations)
Harvest	Peripheral blood stem cell harvest	-	-	£1,132.57	NHS Reference Costs 2018-19. SA34Z DC: Peripheral Blood Stem Cell Harvest
Ablation	High dose melphalan 200 mg/m^2 (75%)	£824.22[d]	BNF online (50 mg)	£254.14	NHS Reference Costs 2018-19. SB12Z DCRDN: Deliver Simple Parenteral Chemotherapy at First Attendance
	Immediate dose melphalan 140 mg/m^2 (25%)	£206.06[c]	BNF online (50 mg)	-	-

Transplant	ASCT	-	-	£16,768.33	NHS Reference Costs 2018-19. SA26A EL: Peripheral Blood Stem Cell Transplant, Autologous, 19 years and over
Post-transplant	GCSF: Lenograstim 19.2 MU/m^2 daily (assumed 14 days)	£1,751.12[b]	BNF online (1 vial of 33.6 MU)	-	-
Total costs	£22,358.64 (not including the cost of plerixafor for a proportion of patients) **Weighted cost for DBTd arm:** £24,171.71 (based on 90.1% patients receiving ASCT and 20.3% patients receiving plerixafor) **Weighted cost for BTd arm:** £21,399.81 (based on 89.3% patients receiving ASCT and 7.2% patients receiving plerixafor)				

Key: ASCT = autologous stem cell transplant; BNF = British National Formulary; BCd = bortezomib, cyclophosphamide; BSA = body surface area; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; GCSF = granulocyte-colony stimulating factor; IV = intravenous; MU = mega unit.

[a] Based on a unit cost of £8.16 for cyclophosphamide 500 mg powder for solution and BSA of 1.88 m^2 from CASSIOPEIA.

[b] Based on a unit cost of £62.54 for lenograstim 33.6 MU vial and BSA of 1.88 m^2 from CASSIOPEIA.

[c] Based on a unit cost of £4,882.77 (excl. VAT) for plerixafor 24 mg vial. The cost of plerixafor is applied to a proportion of patients (20.3% DBTd arm and 7.2% BTd arm) based on data from CASSIOPEIA. For the cost of 2nd ASCT as a subsequent therapy (see sections below), it was assumed that the proportion of patients requiring plerixafor following BCd induction would be the same as the BTd arm.

[d] Based on a unit cost of £137.37 for melphalan 50 mg powder and solvent for solution for injection and BSA of 1.88 m^2 from CASSIOPEIA.

Subsequent therapies

As shown in Section B.1.3.4, several treatments have been recommended by NICE for patients with relapsed/refractory MM. To reflect that patients are expected to receive treatment following progression from induction, ASCT and consolidation, the model includes the cost of subsequent therapies. In accordance with the NICE position statement on the inclusion of therapies recommended via the Cancer Drugs Fund, only those treatments that have been recommended for routine funding by NICE, and not via the Cancer Drugs Fund, have been considered as subsequent therapies in the base case analysis.(137) A scenario analysis has also been conducted in which regimens recommended via the Cancer Drugs Fund have been included in the model (refer to Section B.3.8.3 for the results of scenario analyses).

The possibility that some patients may receive a 2nd ASCT was also included in the cost of subsequent therapies. Feedback from a recent advisory board with three UK clinicians indicated significant regional variation in the proportion of patients likely to receive a 2nd ASCT with different criteria used to assess eligibility. In the economic model it is estimated that between 8–10% of patients will receive a 2nd ASCT following progression after first-line treatment which was at the lower end of the range indicated by clinicians.(1) With patients receiving either DBTd or BTd as first-line induction therapy, it was assumed that patients would receive BCd as induction therapy for the 2nd ASCT (3 cycles of 21 days, as per GMMG-M5).(94, 95) A higher proportion of patients in the DBTd arm were assumed to receive a 2nd ASCT compared to those in the BTd arm (10% versus 8%) due to the deeper responses and prolonged period of remission achieved with DBTd relative to BTd.

In the model, which consists of only two health states (PF and PD), the cost of subsequent therapies across all lines of therapy (2nd line to 4th line) has been included as a single, per-cycle cost which is applied in all cycles for patients in the PD health state. An advantage of this approach (i.e. applying a per-cycle cost) compared to applying the cost as a single, lump-sum cost is that the impact of annual discounting of costs in the model and the impact of deaths on the number of patients receiving subsequent treatment is captured to some extent. In order to calculate this per-cycle cost, the total cost of treatment at each line of therapy was first calculated. As the majority of subsequent therapies are ‘treat to progression’, the total cost of treatment was based on median TTP/PFS reported from clinical trials for each regimen. As such, this approach to costing subsequent therapies takes into account the high attrition observed between lines of therapy. The per-cycle cost of subsequent therapies (across all lines) was then calculated as the sum of the total cost of treatment at each line (2nd line, 3rd line and 4th line) divided by the mean time spent in the model in the PD health state. As part of the calculation, it was assumed that all surviving patients would go on to receive subsequent therapy at 2nd line, 3rd line and 4th line following progression.

To include the cost of 2nd ASCT (fixed duration and not ‘treat to progression’), the treatment cost per cycle was calculated as the total cost of BCd induction therapy and ASCT divided by the median PFS for 2nd ASCT. A similar approach was also taken for Bd at 2nd line (8 cycles of 21 days) and panobinostat in combination with Bd (PBd) (8 or 16 cycles of 21 days) when provided at 3rd line, as neither of these regimens are recommended as ‘treat to progression’.(113, 138) Given that median PFS for PBd at 4th line is only expected to be 5.4 months, it was considered more reasonable to assume that patients would receive PBd until progression in 4th line, rather than receive the full 8 or 16 cycles.(139)

The proportion of patients receiving treatment with each subsequent therapy (by line of therapy) is presented in Table 70, and median TTP/PFS for each regimen is presented in Table 71.

Feedback from the clinical advisory board meeting indicated significant regional variation in the choice of treatment at 2nd line, which is therefore expected to vary in UK clinical practice. For the purposes of the model it was assumed that patients not receiving a 2nd ASCT would be equally distributed to either Bd or lenalidomide in combination with dexamethasone (Ld). Feedback from the clinicians at the advisory board was that carfilzomib in combination with dexamethasone (Cd) is rarely used in clinical practice, given the recommendation from NICE that it should only be used in patients who have not received prior treatment with bortezomib (TA457), and the expectation that transplant-eligible patients would receive a bortezomib-containing induction therapy (BTd, BCd or Bd).(1) It is expected that patients would be less likely to receive a given therapy (e.g. lenalidomide or panobinostat) if they had already received this treatment at an earlier line of therapy. The distribution of patients to treatments received in 3rd line (Ld or PBd) and 4th line (PBd or pomalidomide in combination with dexamethasone [Pd]) was there based on the treatments received at earlier lines of therapy.

With the exception of the slight difference in the proportion of patients receiving a 2nd ASCT (10% in the DBTd arm and 8% in the BTd arm), the distribution of patients to subsequent therapies is expected to be largely the same regardless of whether patients receive DBTd or BTd induction therapy at 1st line. However, this is only true when treatments recommended by NICE via the Cancer Drugs Fund are not taken into consideration (refer to Table 70). For example, patients who receive daratumumab as DBTd in 1st line would be less likely to then be treated with daratumumab in combination with Bd (DBd) (2nd line) or daratumumab monotherapy (4th line) at later lines of therapy than those who receive BTd.

The dosing regimens of subsequent therapies included in the model are presented in Table 72. These were based on dosing schedules outlined in the respective SmPCs or pivotal trials for each regimen. The unit costs and total costs per administration associated with the individual therapies included within the subsequent treatment regimens are presented in Table 73 (ixazomib, lenalidomide, pomalidomide and panobinostat) and Table 74 (cyclophosphamide; BSA-based dosing). The average cost per model cycle for Ld, Bd, BCd, PBd, Pd, and ixazomib in combination with Ld (ILd) are presented in Table 75 and the average cost per model cycle for DBd and daratumumab monotherapy are presented in Table 76. In calculating the cost of subsequent therapies in the model, it should be noted that:

- NICE recommendations for Ld, PBd, Pd and ILd are subject to the manufacturers providing the relevant treatments (lenalidomide, panobinostat, pomalidomide and ixazomib) in accordance with the terms of a confidential commercial arrangement. In the cost-effectiveness analyses provided in this submission, these treatments have all been included at list price.
- For DBd and daratumumab monotherapy (relevant for the 'including treatments recommended via Cancer Drugs Fund' scenario only), the number of daratumumab administrations per model cycle is not constant over time. An average cost per cycle until disease progression was calculated for each of these regimens assuming daratumumab administration as a SC injection (Table 76). When provided at later lines of therapy, daratumumab has been included at list price in the model in line with the other subsequent therapies.

Table 70: Distribution of patients to subsequent therapies (by line of therapy)

Treatment arm	Base case				Scenario: including treatments recommended via the CDF				
Line:	**2nd line**								
Subsequent therapy:	**Cd**	**Ld**	**Bd**	**BCd + 2nd ASCT**	**Cd**	**Ld**	**Bd**	**BCd + 2nd ASCT[a]**	**DBd**
DBTd	0%	45%	45%	10%	0%	45%	45%	10%	0%
BTd	0%	46%	46%	8%	0%	20%	0%	0%	80%
Line:	**3rd line**								
Subsequent therapy:	**Ld**	**PBd**	**-**	**-**	**Ld**	**PBd**	**ILd**	**-**	**-**
DBTd	55%	45%	-	-	0%	45%	55%	-	-
BTd	54%	46%	-	-	0%	20%	80%	-	-
Line:	**4th line**								
Subsequent therapy:	**Pd**	**PBd**	**-**	**-**	**Pd**	**PBd**	**D**	**-**	**-**
DBTd	100%	0%	-	-	100%	0%	0%	-	-
BTd	100%	0%	-	-	80%	0%	20%	-	-

Key: ASCT = autologous stem cell transplant; BCd = bortezomib, cyclophosphamide and dexamethasone; Bd = bortezomib and dexamethasone; BTd = bortezomib, thalidomide and dexamethasone; Cd = carfilzomib and dexamethasone; CDF = Cancer Drugs Fund; D = daratumumab; DBd = daratumumab, bortezomib, and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ILd = ixazomib, lenalidomide and dexamethasone; Ld = lenalidomide and dexamethasone; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone.

[a] Feedback from a recent advisory board with three UK clinicians indicated that the proportion of patients likely to receive a 2nd ASCT would be reduced with the availability of DBd at 2nd line.

Table 71: Median TTP/PFS for subsequent lines of treatment

Subsequent treatment	Median TTP/PFS (months)	Source
2nd line		
Ld	17.1	Based on median TTP from 1 prior therapy subgroup from Pooled MM-009 and MM-010 (Stadtmauer 2009)(140)
Bd	8.02	Based on median TTP in second-line patients from CASTOR (NICE TA573 manufacturer submission)(132)
BCd + 2nd ASCT	40.9	Based on median PFS from GMMG-MM5 trial (BCd arm with lenalidomide maintenance for 2 years)(94) (Note: based on induction therapy and ASCT in newly diagnosed patients and not 2nd ASCT)
DBd[a]	27.63	Based on median TTP in second-line patients from CASTOR (NICE TA573 manufacturer submission)(132)
3rd line		
Ld	14.1	Based on median TTP after 2 or 3 previous lines of therapy from TOURMALINE-MM1 (NICE TA505 manufacturer submission)(141)
PBd	12.68	Based on median TTP after at least 2 therapies from PANORAMA-1 (Richardson 2016)(142)
ILd[a]	28.8	Based on median TTP after 2 or 3 previous lines of therapy from TOMALINE-MM1 (NICE TA505 manufacturer submission) (141)
4th line		
Pd	4.7	Based on median TTP after at least 2 therapies from MM-003 (NICE TA427 manufacturer submission)(143)
PBd	5.4	Based on median PFS (ITT) from PANORAMA-2 (Richardson 2013)(139)
Daratumumab monotherapy[a]	4.0	Based on median PFS from GEN501/MMY-002 (Usmani 2016)(144)

Key: ASCT = autologous stem cell transplant; BCd = bortezomib, cyclophosphamide and dexamethasone; Bd = bortezomib and dexamethasone; CDF = Cancer Drugs Fund; DBd = daratumumab, bortezomib, and dexamethasone; ILd = ixazomib, lenalidomide and dexamethasone; ITT = intention-to-treat; Ld = lenalidomide and dexamethasone; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone; PFS = progression-free survival; TTP = time to progression.
[a] Recommended via CDF; scenario only

Table 72: Summary of dosing regimens for subsequent treatments

Treatment	Duration	Drugs per cycle	Average administrations per model cycle	Average total dose per model cycle	Source
Ld	Cycles of 28 days, until disease progression	Lenalidomide 25 mg QD for 3 weeks	21	525 mg	Revlimid SmPC(145)
		Dexamethasone 40 mg QW	4	160 mg	
Bd	Cycles of 21 days, maximum 8 cycles	Bortezomib – 1.3 mg/m^2 BW for 2 weeks	5.33	7 mg/m²	Velcade SmPC(113)
		Dexamethasone 20 mg 4 times weekly for 2 weeks	10.67	213 mg	
BCd + 2nd ASCT	Cycles of 28 days, 3 cycles	Bortezomib – 1.3 mg/m^2 BW for 2 weeks	4	10 mg/m²	GMMG-MM5 trial(94, 95)
		Cyclophosphamide - 900 mg/m^2 on day 1	1	1,695 mg/m²	
		Dexamethasone 40 mg BW for 3 weeks	6	240 mg	
PBd	Cycles of 21 days, maximum 16 cycles It was assumed that PBd 4th line would only be received until	Panobinostat 20 mg 3 times weekly for 2 weeks, for all cycles	9	180 mg	Farydak SmPC(138)
		Bortezomib – 1.3 mg/m^2 BW for 2 weeks, for cycles 1–8	5.33	7 mg/m²	

	disease progression[a]	Dexamethasone 20 mg 4 times weekly for 2 weeks, for cycles 1–8	10.67	213 mg	
		Bortezomib – 1.3 mg/m^2 QW for 2 weeks, for cycles 9–16	2.67	3 mg/m²	
		Dexamethasone 20 mg BW for 2 weeks, for cycles 9–16	5.33	107 mg	
Pd	Cycles of 28 days, until disease progression	Pomalidomide 4 mg QD for 3 weeks	21	84 mg	Imnovid SmPC(146)
		Dexamethasone 40 mg QW	4	160 mg	
DBd (Recommended via CDF; scenario only)	8 cycles of 21 days, followed by 28-day cycles of daratumumab monotherapy until disease progression	Daratumumab 1,800 mg QW for 3 cycles, then once per cycle	Number of administrations is not the same in each model cycle. See Table 76 for cost of DBd across different model cycles		CASTOR trial(69)
		Bortezomib – 1.3 mg/m^2 BW for 2 weeks			
		Dexamethasone 20 mg on days 1, 2, 4, 5, 8, 9, 11 and 12 of each cycle			
ILd	Cycles of 28 days, until	Ixazomib 4 mg QW for 3 weeks	3	12 mg	Ninlaro SmPC(147)

(Recommended via CDF; scenario only)	disease progression	Lenalidomide 25 mg QD for 3 weeks	21	525 mg	
		Dexamethasone 40 mg QW	4	160 mg	
Daratumumab monotherapy (Recommended via CDF; scenario only)	Cycles of 28 days, until disease progression	Daratumumab 1,800 mg QW for 2 cycles	4	7,200 mg	Darzalex SmPC(5)
		Daratumumab 16 mg/kg Q2W for 4 cycles	2	3,600 mg	
		Daratumumab 16 mg/kg Q4W until progression	1	1,800 mg	

Key: ASCT = autologous stem cell transplant; BCd = bortezomib, cyclophosphamide and dexamethasone; Bd = bortezomib and dexamethasone; BW = bi-weekly; CDF = Cancer Drugs Fund; DBd = daratumumab, bortezomib and dexamethasone; ILd = ixazomib, lenalidomide and dexamethasone; Ld = lenalidomide and dexamethasone; Pd = pomalidomide and dexamethasone; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone; SmPC = Summary of Product Characteristics; TTP = time-to-progression; QD = daily; Q2W = once every two weeks; Q4W = once every four weeks; QW = weekly.

[a] Given that median PFS for PBd at 4th line is only expected to be 5.4 months, it was considered more reasonable to assume that patients would receive PBd until progression in 4th line, rather than receive the full 8 or 16 cycles.(139) The cost of PBd in 4th line is based on the dosing schedule for the first 8 cycles.

Table 73: Drug acquisition costs for subsequent treatments (ixazomib, lenalidomide, pomalidomide and panobinostat)

Treatment	Unit Size (mg)	Pack Size	Cost per Pack	Cost per mg
Ixazomib	4	3	£6,336.00	£528.00
Lenalidomide	25	21	£4,368.00	£8.32
Pomalidomide	4	21	£8,884.00	£105.76
Panobinostat	20	6	£4,656.00	£38.80

Unit costs were derived from the BNF online

Table 74: Drug acquisition costs for subsequent treatments (cyclophosphamide)

Treatment	Dose per administration	Total dose per administration	Vial Size	Cost per vial	Vials per administration	Cost per administration
Cyclophosphamide	900 mg/m^2	1,694.70 mg	500.00	£8.16	4	£32.64
Unit costs were derived from eMIT						

Table 75: Summary of drug acquisition costs for subsequent treatments – drug costs per cycle (Ld, Bd, BCd, PBd, Pd, and ILd)

Treatment	Drug costs per cycle		Total regimen costs per cycle	
Ld				
Lenalidomide	£4,368.00		£4,378.07	
Dexamethasone	£10.07			
Bd				
Bortezomib	£4,066.03		£4,079.45	
Dexamethasone	£13.42			
BCd				
Bortezomib	£3,049.52		£9,291.79	
Cyclophosphamide	£32.64			
Dexamethasone	£15.10			
PBd				
	Cycles 1–8[a]	Cycles 9–16	Cycles 1–8[a]	Cycles 9–16
Panobinostat	£6,984.00	£6,984.00	£11,063.45	£9,023.73
Bortezomib	£4,066.03	£2,033.01		
Dexamethasone	£13.42	£6.71		

Pd		
Pomalidomide	£8,884.00	£8,894.07
Dexamethasone	£10.07	
ILd (Recommended via CDF; scenario only)		
Ixazomib	£6,336.00	£10,714.07
Lenalidomide	£4,368.00	
Dexamethasone	£10.07	

Key: ASCT = autologous stem cell transplant; BCd = bortezomib; cyclophosphamide and dexamethasone; Cd = carfilzomib and dexamethasone; CDF = Cancer Drugs Fund; ILd = ixazomib, lenalidomide and dexamethasone; Ld = lenalidomide and dexamethasone; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone.

[a] The cost of PBd in 4th line is based on the dosing schedule for the first 8 cycles.

Table 76: Summary of drug acquisition costs for subsequent treatments – drug cost per cycle (DBd and daratumumab monotherapy)

Treatment	Cycle from start of treatment	Daratumumab	Bortezomib	Dexamethasone	Total drug cost per cycle	Median TTP	Total cost until progression	Average cost per cycle
DBd (Recommended via CDF; scenario only)	Cycles 1–2	£17,280.00	£4,574.28	£15.10	£21,869.38	27.63	£178,398.30	£6,456.69
	Cycles 3–6	£8,640.00	£3,811.90	£12.59	£10,304.49			
	Cycles 7+ (Median TTP – 6)	£4,320.00	£0.00 / £0.00	£0.00	£4,320.00			
Daratumumab monotherapy (Recommended via CDF; scenario only)	Cycles 1–2	£17,280.00	-	-	£17,280.00	4.00	£69,120.00	£17,280.00
	Cycles 3–6	£8,640.00	-	-	£8,640.00			
	Cycles 7+ (Median TTP – 6)	£4,320.00	-	-	£4,320.00			

Key: CDF = Cancer Drugs Fund; DBd = daratumumab, bortezomib, and dexamethasone; TTP = time to progression.

Administration costs

The cost of administration was included for both first-line treatment (induction and consolidation) and subsequent therapies (Table 77). In line with the assumptions used in NICE TA573: for oral chemotherapy regimens (i.e. thalidomide and dexamethasone) a one-off cost was applied on treatment initiation, whereas for therapies administered via SC injection (i.e. daratumumab and bortezomib), a cost was applied for each administration.(132) On days where both bortezomib and daratumumab are administered together, it was assumed that this would be performed together by the same nurse and so the cost of only one administration was included in the analysis. The cost of a blood test prior to the first administration of daratumumab was also included in the cost of administration for DBTd.

In the scenario analysis using the IV formulation of daratumumab, the cost of daratumumab administration was based on the cost of delivering complex chemotherapy, including prolonged infusion. On days where both bortezomib and daratumumab are administered together, only the (higher) cost of the IV infusion was applied.

Table 77: Administration costs

Drug	Parameter	Cost	Source
Daratumumab (SC)	Subcutaneous administration	£98.74	NHS Reference Costs 2018-19. N10AF: Specialist Nursing, Cancer Related, Adult, Face to face
	Blood test (prior to first administration)	£2.79	NHS Reference Costs 2018-19. DAPS05: Haematology
Daratumumab (IV scenario)	First administration	£385.28	NHS Reference Costs 2018-19. SB14Z DCRDN: Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance
	Subsequent administration	£223.00	NHS Reference Costs 2018-19. SB15Z Outpatient: Deliver Subsequent Elements of a Chemotherapy Cycle
	Blood test (prior to first administration)	£2.79	NHS Reference Costs 2018-19. DAPS05: Haematology
Bortezomib	Subcutaneous administration	£98.74	NHS Reference Costs 2018-19. N10AF: Specialist Nursing, Cancer Related, Adult, Face to face
Cyclophosphamide	IV administration	£223.00	NHS Reference Costs 2018-19. SB15Z Outpatient: Deliver Subsequent Elements of a Chemotherapy Cycle
Oral Chemotherapies	First administration only	£185.71	NHS Reference Costs 2018-19. SB11Z Outpatient: Deliver Exclusively Oral Chemotherapy
Key: IV = intravenous; SC = subcutaneous.			

B.3.5.2 Health-state unit costs and resource use

Monitoring costs

Ongoing monitoring costs were included in the model, with the frequency of monitoring visits and tests dependent on whether patients were receiving active anti-cancer therapy (Table 78). For the PF health state, it was assumed that patients would receive 'on treatment' monitoring from the start of induction therapy until the time of the post-consolidation response assessment (i.e. for cycles 0–9), with the 'off treatment' monitoring costs applied in all subsequent model cycles until disease progression. In the PD health state, the 'on treatment' monitoring costs were applied in all model cycles, as it was assumed that patients would receive anti-cancer therapy for the majority of their lifetime.

The type and frequency of monitoring visits and tests were based on those used in NICE TA573.(132)

Table 78: Monitoring costs

Item	Frequency per cycle		Unit cost	Source
	On Treatment	Off Treatment		
Haematologist visit	0.92	0.32	£168.02	NHS Reference Costs 2018-19. WF01A: Clinical Haematology (303). Non-Admitted Face-to-Face Attendance, Follow-up
Full blood count	0.84	2.56	£2.79	NHS Reference Costs 2018-19. DAPS05: Haematology
Biochemistry	0.76	1.32	£1.10	NHS Reference Costs 2018-19. DAPS04: Clinical Biochemistry
Protein electrophoresis	0.52	0.72	£1.10	
Immunoglobulin	0.48	0.76	£1.10	
Urinary light chain excretion	0.20	0.20	£1.10	
Total cost per 28 days	**£159.08**	**£64.21**	-	Calculated

End-of-life cost

A one-off cost representing the cost of terminal care was applied in the model for the proportion of patients that died in each cycle. The cost applied in the model (£8,103.30) was derived from the cost used in NICE TA573, inflated to 2018–2019 using the Pay & Price Index to 2015-16 and the NHSCII Pay & Price Index to 2018-19.(132, 148)

B.3.5.3 Adverse reaction unit costs and resource use

The cost of managing adverse events experienced by patients receiving induction, ASCT and consolidation therapy was included in the model. The costs per event were based on NHS reference costs 2018–19 (or inflated to 2018–2019) and are presented in Table 79. These costs were applied to the proportion of patients experiencing each event in each of the treatment arms

in the model (Table 57) and were applied in the first cycle of the model. The total cost across all events included in the model was £1,771.09 for DBTd and £1,279.08 for BTd.

Table 79: Adverse event costs

Adverse Event	Costs	Source
Neutropenia	£1,417.51	NHS Reference Costs 2018-19. Weighted average of SA08G-SA08J: Other haematological or splenic disorders, with CC score 0-6+, non-elective long stay and short stay
Lymphopenia	£1,417.51	NHS Reference Costs 2018-19. Weighted average of SA08G-SA08J: Other haematological or splenic disorders, with CC score 0-6+, non-elective long stay and short stay
Thrombocytopenia	£1,660.53	NHS Reference Costs 2018-19. Weighted average of SA12G-SA12K: Thrombocytopenia with CC score 0-8+, non-elective long stay and short stay
Febrile neutropenia	£7,369.65	Inflated from TA510, based on NHS Reference Costs 2011-12. PA45Z: Febrile neutropenia with malignancy, using the Pay & Price Index to 2015-16 and the NHSCII Pay & Price Index to 2018-19(148, 149)
Stomatitis	£853.18	NHS Reference Costs 2018-19. Weighted average of CB02A-CB02F: Non-malignant, ear, nose, mouth, throat or neck disorders, with interventions (with CC score 1-4) and without interventions (with CC score 0-5+), non-elective long stay and short stay
Peripheral sensory neuropathy	£945.04	NHS Reference Costs 2018-19. Weighted average of WH08A and WH08B: Unspecified pain with CC score 0 and 1+, non-elective long stay and short stay
Nausea	£771.93	Inflated from TA510, based on NHS Reference Costs 2014-15. WA21Z: other procedures or healthcare problems, using the NHSCII Pay & Price Index to 2018-19(148, 149)
Upper respiratory tract infection	£598.73	NHS Reference Costs 2018-19. Weighted average of DZ19H-ZD19N: Other respiratory disorders with single intervention (with CC score 0-4) and without intervention (with CC score 0-11+), non-elective long stay and short stay

B.3.5.4 Miscellaneous unit costs and resource use

No additional costs were included in the CUA.

B.3.6 *Summary of base-case analysis inputs and assumptions*

B.3.6.1 Summary of base-case analysis inputs

A summary of inputs used in the base case analysis is presented in Table 80.

Table 80: Summary of variables applied in the economic model

Variable	Value		Reference to section in submission
Model settings			
Discount rate (costs and benefits)	3.5%		Section B.3.2
Time horizon	Lifetime (100 – starting age)		
Patient baseline characteristics			
Mean age	56.6 years		Section B.3.2.1
Mean body weight	75.67 kg		
Mean BSA	1.88 m^2		
% Male	58.5%		
Induction, ASCT and consolidation therapy	**DBTd**	**BTd**	Section B.3.2.3
% completing 1, 2, 3 and 4 cycles of induction therapy	98.7% 97.6% 96.7% 95.4%	99.3% 97.8% 97.0% 94.5%	
% receiving ASCT	90.1%	89.3%	
% completing 1 and 2 cycles of consolidation therapy	85.8% 85.5%	82.7% 80.6%	
Clinical inputs	**DBTd**	**BTd**	Section B.3.3.1
MRD negativity at post-consolidation assessment	63.7%	43.5%	
Survival inputs	**PFS**	**OS**	Section B.3.3.2
Extrapolation for BTd MRD+	Weibull	Exponential	
HR for MRD- versus MRD+	[redacted]	[redacted]	
HR for DBTd versus BTd (MRD-)	[redacted]	[redacted]	
HR for DBTd versus BTd (MRD+)	[redacted]	[redacted]	
Adverse events	**DBTd**	**BTd**	Section B.3.3.3
Neutropenia incidence	27.61%	14.68%	
Lymphopenia	16.98%	9.67%	

Thrombocytopenia	11.01%	7.43%	
Febrile neutropenia	6.72%	5.20%	
Stomatitis	12.69%	16.36%	
Peripheral sensory neuropathy	8.77%	8.55%	
Nausea	30.22%	24.16%	
Upper respiratory tract infection	6.16%	3.35%	
Utility inputs			
PF (induction therapy)	0.57		Section B.3.4.1
PF (post-induction to post-consolidation response)	0.68		
PF (post-consolidation)	0.73		
PD	0.69		
Adverse event disutility			
Neutropenia	0.02		Section B.3.4.4
Lymphopenia	0.01		
Thrombocytopenia	0.03		
Febrile neutropenia	0.04		
Stomatitis	0.02		
Peripheral sensory neuropathy	0.01		
Nausea	0.01		
Upper respiratory tract infection	0.02		
Cost inputs			
Daratumumab SC, cost per vial (1,800 mg)	£4,320.00		Section B.3.5.1
Bortezomib, cost per vial (3.5 mg)	£762.38		
Thalidomide, cost per pack	£298.48		
Dexamethasone, cost per pack	£25.17		
Lenalidomide, cost per pack	£4,368.00		

Cyclophosphamide, cost per vial	£8.16		
Pomalidomide, cost per pack	£8,884.00		
Panobinostat, cost per pack	£4,656.00		
Subsequent therapies	**DBTd**	**BTd**	
Ld – 2nd line	45%	46%	
Bd – 2nd line	45%	46%	
BCd + 2nd ASCT – 2nd line	10%	8%	Section B.3.5.1
Ld – 3rd line	55%	54%	
PBd – 3rd line	45%	46%	
Pd – 4th line	100%	100%	
Concomitant medication costs			
Low molecular weight heparin, cost per pack	£8.84		
Aspirin, cost per pack	£0.12		
Antipyretic: oral paracetamol, cost per pack	£0.16		Section B.3.5.1
Antihistamine: oral/IV diphenhydramine, cost per pack	£3.16		
Corticosteroid: oral methylprednisolone, cost per pack	£17.17		
Antiviral: aciclovir, cost per pack	£0.52		
	DBTd	**BTd**	
ASCT cost	£24,171.71	£21,399.81	Section B.3.5.1
Administration costs			
Subcutaneous administration	£98.74		
Oral administration	£185.71		Section B.3.5.1
IV administration	£223.00		
Blood test for daratumumab	£2.79		
Monitoring costs			Section B.3.5.2

Haematologist visit	£168.02		
Full blood count	£2.79		
Biochemistry	£1.10		
Protein electrophoresis	£1.10		
Immunoglobulin	£1.10		
Urinary light chain excretion	£1.10		
Adverse event costs			Section B.3.5.3
Neutropenia	£1,417.51		
Lymphopenia	£1,417.51		
Thrombocytopenia	£1,660.53		
Febrile neutropenia	£7,369.65		
Stomatitis	£853.18		
Peripheral sensory neuropathy	£945.04		
Nausea	£771.93		
Upper respiratory tract infection	£598.73		

Key: ASCT = autologous stem cell transplant; BCd = bortezomib, cyclophosphamide and dexamethasone; BSA = body surface area; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; IV = intravenous; Ld = lenalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone; PD = progressed disease; PF = progression free; PFS = progression-free survival; SC = subcutaneous.

B.3.6.2 Assumptions

A list of the key assumptions used in the base case analysis is provided in Table 81 alongside a description of scenarios conducted to explore the impact of these assumptions on the cost-effectiveness results. The results of these scenario analyses are presented in B.3.8.3.

Table 81: Assumptions used in the cost-utility analysis

Parameter	Assumption (base case)	Justification	Addressed in scenario analysis; rationale for scenario analysis
Approach to modelling survival	The model utilised a response-based approach in which survival (PFS and OS) after the post-consolidation assessment timepoint was modelled to be dependent on whether patients attained MRD negativity. The approach chosen for the base case analysis included the direct extrapolation of data from CASSIOPEIA (for BTd MRD+), the use of HRs from statistical analysis performed on the expanded SLR/meta-analysis exploring the relationship between MRD negativity and survival (for BTd MRD-), and the use of HRs from the CASSIOPEIA landmark analysis (for DBTd MRD+/-).	Direct extrapolation of data from the overall treatment arms of CASSIOPEIA (not based on response) resulted in a wide variation in long-term survival depending on the parametric model chosen (refer to Section B.3.3.2). A response-based model was therefore used. The use of a response-based model allows for the inclusion of external data in the model to more robustly model long-term outcomes. The attainment of a deep and durable response is a primary aim of induction, ASCT and consolidation therapy for patients with newly diagnosed MM, and the relationship between MRD negativity and long-term survival outcomes has been established.(61) The base case modelling approach leverages external data with longer follow-up to inform long-term survival outcomes for patients who attain post-consolidation MRD negativity in the BTd arm. The use of trial data directly from CASSIOPEIA for BTd MRD+, which represents the most mature data from the trial (i.e. highest proportion of events occurring), helped improve internal validity of modelled survival outcomes when compared to the observed data from CASSIOPEIA. Data from the CASSIOPEIA trial are utilised for the DBTd arm in order to model the observed treatment effect for daratumumab versus BTd in both MRD-positive and MRD-negative patients (see CASSIOPEIA landmark analysis in Section B.2.6.3).	Scenarios were conducted to explore the impact of external data and the daratumumab treatment effect. **Scenario 1A:** BTd MRD- survival is modelled by directly extrapolating CASSIOPEIA data rather than applying the HR from the statistical analysis performed on the expanded SLR/meta-analysis. By removing the use of external data, this scenario explores the impact of external data on the ICER. **Scenario 1B:** DBTd MRD- survival is modelled via the application of the HRs from the SLR/meta-analysis, rather than the HR from the CASSIOPEIA landmark analysis. By removing the daratumumab treatment effect for MRD- patients seen in CASSIOPEIA, this scenario explores the impact of the daratumumab treatment effect on the ICER.
Extrapolation of PFS and	Weibull distribution for the extrapolation of PFS and exponential distribution for the extrapolation of OS	The choice of parametric distribution for the base case analysis was made based on consideration of: statistical fit, visual fit when compared to the observed data from the CASSIOPEIA trial, and the clinical plausibility of	Alternative and more optimistic extrapolations of BTd MRD+ OS were not explored in scenario analyses given that these

OS for BTd MRD+		long-term survival estimates (refer to Section B.3.3.2). Given that the observed data for the BTd MRD-positive subgroup are still relatively immature, the clinical plausibility of long-term extrapolations was considered to be a critical factor in selecting curves for the base case analysis. Based on clinician feedback on the long-term PFS and OS that may be expected of BTd MRD-positive patients in clinical practice in England, the Weibull distribution was chosen for the extrapolation of PFS and the exponential distribution was chosen for the extrapolation of OS. Survival outcomes predicted by the model (MRD+ and MRD- combined) were also validated against the observed data from CASSIOPEIA and data from Public Health England on patients receiving first-line therapy and ASCT in England (refer to Section B.3.3.2).	predicted long-term OS rates that were higher than those expected based on clinician feedback and higher than general population mortality. The use of the exponential distribution in the base case also resulted in OS model predictions for MRD+ and MRD- combined that were consistent with the CASSIOPEIA trial and real-world outcomes reported from the PHE datasets. An alternative extrapolation for PFS, which provided a higher estimate of long-term PFS compared to the base case, was considered to be plausible and was explored in scenario analyses. **Scenario 2:** Exponential distribution used for the extrapolation of BTd MRD+ PFS
Application of daratumumab treatment effect	OS and PFS in the DBTd arm were modelled via the application of HRs for DBTd versus BTd from the CASSIOPEIA landmark analysis No treatment waning effect was included in the base case analysis, with the HRs from the landmark analysis applied for the duration of model time horizon (post-landmark)	In the landmark analysis for CASSIOPEIA, DBTd was associated with improvements in OS and PFS versus BTd in both MRD-positive and MRD-negative patients (refer to Section B.2.6.3). The application of HRs from the landmark analysis allows for the inclusion of this treatment effect in the modelled survival for DBTd. Evidence of a daratumumab treatment effect regardless of MRD response is [redacted]	Scenario 1B (described above) explores the impact of the daratumumab treatment effect on the ICER. To reflect the possibility that the treatment effect of daratumumab may wane over time, several scenario analyses have been conducted in which the HRs for DBTd versus BTd (PFS and OS) are set to equal one (i.e. no treatment effect) at a specified timepoint in the model. These

		Improvements in survival outcomes for MRD-negative patients (assessed using sensitivity threshold 10^{-5}) is considered plausible, as daratumumab-treated patients may achieve an even deeper level of response. In CASSIOPEIA, DBTd almost doubled the rate of MRD negativity compared to BTd at the higher sensitivity threshold of 10^{-6} (using NGS) (refer to Section B.2.6.1).(44) For MRD-positive patients, the improved survival outcomes in the DBTd arm is likely a reflection of the deeper conventional responses (i.e. according to IMWG criteria) achieved by daratumumab-treated patients (refer to Section B.2.6.1). In the absence of evidence to suggest that the treatment effect of daratumumab on survival (for MRD-positive and MRD-negative patients) would wane over time, no treatment waning effect was applied in the base case analysis. Prolonged survival benefit for patients treated with daratumumab is considered plausible given the unique mechanism of action of daratumumab, which is to modulate the immune system to better fight the disease. Treatment waning has not been included in previous NICE appraisals of daratumumab at later lines of therapy (TA573 and TA510) and was not included in TA311.	include scenarios in which the treatment effect is assumed to wane for both MRD-positive and MRD-negative patients, or for MRD-negative patients only. The timepoint from which the treatment waning effect was applied was also varied. These timepoints were arbitrarily chosen in the absence of evidence to suggest if and when the treatment effect would wane over time. **Scenario 3A:** No additional treatment effect of DBTd after 5 years (MRD+ and MRD-) **Scenario 3B:** No additional treatment effect of DBTd after 10 years (MRD+ and MRD-) **Scenario 3C:** No additional treatment effect of DBTd after 5 years (MRD- only) **Scenario 3D:** No additional treatment effect of DBTd after 10 years (MRD- only)
Daratumumab formulation	The cost of daratumumab was based on the fixed dose of 1,800 mg administered via SC injection, with efficacy for DBTd based on CASSIOPEIA (weight-based dose and IV infusion)	A licence extension for a SC formulation of daratumumab was received in June 2020 and is expected to be used by the majority of patients in clinical practice. Non-inferiority between the weight-based IV formulation of daratumumab (which was used in CASSIOPEIA) and the SC formulation of daratumumab has been demonstrated as part of the COLUMBA (MMY3012) trial.(104)	In CASSIOPEIA, which represents the primary source of clinical evidence for the analysis, daratumumab was administered as a weight-based dose via IV infusion. **Scenario 4:** Drug acquisition and administration costs for daratumumab were based on the IV formulation and weight-based dose (16 mg/kg) used in

			CASSIOPEIA. Only the acquisition and administration costs of daratumumab were adjusted in this scenario.
Subsequent treatments	Subsequent treatments (2nd, 3rd and 4th line) were included in the model based on those treatments that had been recommended by NICE for relapsed/refractory MM. The distribution of patients across the available treatments at each line of therapy were based on market share estimates and took into consideration the expectation that patients are likely to only receive a given therapy (e.g. daratumumab, lenalidomide or panobinostat) once in the treatment pathway. The treatment pathway (when only considering therapies recommended for routine funding by NICE) is expected to be the same in each of the model cohorts with the exception that slightly more patients receiving DBTd are expected to receive a 2nd ASCT due to the deeper responses achieved and prolonged period of remission relative to BTd.	The cost of subsequent therapies were included in the model to reflect the reality that patients with MM experience relapsed and refractory disease. In accordance with the NICE position statement on the inclusion of therapies recommended via the Cancer Drugs Fund, only those treatments that have been recommended for routine funding by NICE, and not via the Cancer Drugs Fund, have been considered as subsequent therapies in the base case analysis.(137)	Several therapies are recommended by NICE for use via the Cancer Drugs Fund, including DBd at 2nd line and daratumumab monotherapy at 4th line. Should these regimens become available for routine funding, then the treatment pathway would be expected to be markedly different for patients receiving DBTd or BTd at first line. This is based on the expectation that patients who receive daratumumab at first line would be less likely to receive daratumumab at later lines. **Scenario 5:** A scenario analysis has also been conducted in which regimens recommended via the Cancer Drugs Fund have been included in the model
Dosing of BTd	Cost of BTd was based on the CASSIOPEIA trial protocol	For consistency with the source of clinical inputs included in the model, the dosing of BTd was based on the CASSIOPEIA trial protocol. The dosing of BTd included in the model is also considered to be consistent with how BTd is typically administered in clinical practice.	The recommended posology for BTd in the bortezomib SmPC is different to that used in the CASSIOPEIA trial (refer to Section B.3.2.3).(113) A scenario was conducted in which the bortezomib SmPC posology was used to calculate the cost of BTd in the model.

			Scenario 6: Cost of BTd was based on the bortezomib SmPC recommended posology
Vial sharing	No vial sharing was assumed	In the base case analysis it was assumed that vials would not be shared or pooled across administrations. As such, drug wastage was assumed if the amount of drug required for a single dose was not an exact multiple of vial size.	With certain drugs administered in a hospital-based setting, there is the potential for vial sharing in clinical practice. A scenario was also conducted in which vial sharing was assumed to occur. **Scenario 7:** Vial sharing was assumed
Utility values	Utility values for PF (induction, ASCT and consolidation) were based on EQ-5D data from the CASSIOPEIA trial For PD, utility was based on the utility value used in TA311 for 2nd and 3rd lines from van Agthoven et al. (2004) (0.69), rather than utility for 'further lines' (0.644).(27, 62)	For consistency with the source of clinical inputs included in the model and the relevance of data from the CASSIOPEIA trial to the patient population of interest for this submission, the utility values used in the base case analysis were based on EQ-5D data from the CASSIOPEIA trial. As patients are expected to spend a greater period of time in 2nd and 3rd line, compared to 4th line, the higher value from TA311 (for 2nd and 3rd lines) was used for the PD health state utility in the model.	Utility values from the published literature (van Agthoven et al. [2004]) were used in the NICE submission for bortezomib as induction therapy, and other sources of utility values are now available.(27, 62) To explore the impact of using alternative utility values, scenario analyses have been conducted in which utility values from van Agthoven et al. (2004)) have been used instead of those derived from CASSIOPEIA, and in which lower utility values for PD have been explored. **Scenario 8A:** Lower utility value from van Agthoven et al. (2004) (0.644) was used for the PD health state **Scenario 8B:** Lower utility value from Palumbo et al. (2013) (PD in 4th line: 0.57) was used for the PD health state

			Scenario 8C: Utility values from van Agthoven et al. (2004) were used for all health states

Key: ASCT = autologous stem cell transplant; BCd = bortezomib, cyclophosphamide and dexamethasone; BIC = Bayesian information criterion; BTd = bortezomib, thalidomide and dexamethasone; CI: confidence interval; CUA = cost-utility analysis; DBd = daratumumab, bortezomib, and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; EQ-5D = EuroQol-5D; HR = hazard ratio; HTA = Health Technology Assessment; ICER = incremental cost-effectiveness ratio; IMWG = International Myeloma Working Group; IV = intravenous; Ld = lenalidomide and dexamethasone; MM = multiple myeloma; MRD = minimal residual disease; NGS = next generation sequencing; PBd = panobinostat, bortezomib and dexamethasone; Pd = pomalidomide and dexamethasone; PD = progressed disease; PFS = progression-free survival; PHE = Public Health England; OS = overall survival; SC = subcutaneous; SLR = systematic literature review; SmPC = Summary of Product Characteristics.

B.3.7 *Base-case results*

B.3.7.1 Base-case incremental cost-effectiveness analysis results

The deterministic base case results for DBTd versus BTd are presented in Table 82.

Compared to BTd, DBTd was associated with an increased number of life years (3.32) and QALYs ([redacted]), but also higher total costs ([redacted]). In the base case analysis, the ICER at list price versus BTd was [redacted] per QALY gained. [redacted]
2[redacted]

Disaggregated results from the base case analysis are presented in Appendix J for:

- Costs by cost category for 1st line treatment (induction, ASCT and consolidation therapy costs)
- Costs by health state (PF and PD, and end-of-life costs)
- QALYs by health state (PF and PD)

The difference in costs between treatment arms during induction, ASCT and consolidation was primarily due to differences in drug acquisition costs between DBTd and BTd induction therapies (i.e. due to the cost of daratumumab). The other sources of 1st line treatment costs applied in the model (e.g. administration, ASCT, monitoring, concomitant medication, adverse events) were broadly similar or the same between the treatment arms. The costs accrued in the PD health state, which includes subsequent therapy costs, and end-of-life costs were also similar between treatment arms. The difference in total costs between DBTd and BTd in the model were therefore largely attributable to the difference in drug acquisition costs in 1st line.

The difference in QALYs between treatment arms was primarily due to the difference in QALYs accrued in the PF health state ([redacted] for DBTd versus [redacted] for BTd). That the benefits of DBTd treatment are realised in the model as an increase in time spent in the PF health state, as well as an increase in QALYs overall, is consistent with the aims of 1st line treatment, which are to delay progression and achieve sustained remission.

Clinical outcomes (mean PFS and OS) are also presented in Appendix J. The survival rates predicted by the model, compared to observed data from CASSIOPEIA and the PHE linked datasets, are presented in Section B.3.3.2.

Table 82: Deterministic base case results

Intervention	Total costs	Total LYs	Total QALYs	Inc. costs	Inc. LYs	Inc. QALYs	ICER
DBTd	[redacted]	14.66	[redacted]				
BTd	[redacted]	11.34	[redacted]	[redacted]	3.32	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ICER = incremental cost-effectiveness ratio; LY = life year; QALY = quality-adjusted life year.

B.3.8 *Sensitivity analyses*

B.3.8.1 Probabilistic sensitivity analysis

A probabilistic sensitivity analysis (PSA) was conducted in order to assess the impact of parameter uncertainty on the results of the CUA. The PSA was run for 5,000 iterations and in each iteration model inputs for all parameters were randomly drawn from specified distributions (e.g. gamma for costs; beta for proportions and lognormal for HRs). Where possible the standard error or standard deviation associated with the mean value was used to define the distribution, otherwise it was assumed that the standard error would be 20% of the mean value. The inputs and distributions used in the PSA are summarised in Appendix P.

The average incremental cost-effectiveness results from the PSA are presented in Table 83. Taking into account the combined parameter uncertainty in the model, the ICERs for DBTd versus BTd were seen to be similar (albeit marginally higher) to those reported in the deterministic base case.

A scatter plot showing the results of each iteration from the PSA on the cost-effectiveness plane are presented in

Figure 36, and the cost-effectiveness acceptability curve is presented in

Figure 37. In the vast majority of PSA iterations, the DBTd arm was associated with a greater number of QALYs than BTd and the incremental costs remained relatively stable across the different iterations (

Figure 36). At willingness to pay-thresholds of £20,000 and £30,000 per QALY gained, the probability of DBTd being the more cost-effective treatment option was [redacted]% and [redacted]%, respectively.

Table 83: Average probabilistic cost-effectiveness results

Comparison versus	Inc. costs	Inc. QALYs	ICER
BTd	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year.

Figure 36: Cost-effectiveness plane for DBTd versus BTd

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; QALY = quality-adjusted life year.

Figure 37: Cost-effectiveness acceptability curve

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

B.3.8.2 Deterministic sensitivity analysis

Deterministic sensitivity analysis (DSA) was conducted by varying the input for each parameter in the model by ±20% of their mean value, whilst keeping all other inputs the same. For those parameters where 95% confidence intervals were available, the upper and lower limits of the confidence intervals were used instead to vary the model input. The inputs used in the DSA are presented in Appendix P.

As shown in Figure 38, the parameters with the greatest impact on the ICER were the HRs used to determine OS for patients in the DBTd arm. These HRs were based on the landmark analysis of the CASSIOPEIA trial (refer to Section B.2.6.3), and so represent the primary source of evidence for the relative efficacy of DBTd versus BTd in MRD-positive and MRD-negative patients. Scenario analyses exploring alternative assumptions and inputs relating to the daratumumab treatment effect have also been conducted and are presented in Section B.3.8.3.

As discussed in Section B.2.6.3, a daratumumab treatment effect for both MRD-positive and MRD-negative patients was demonstrated in the landmark analysis of PFS and OS from CASSIOPEIA, and is supported by [redacted] This is explained by the significantly deeper levels of response achieved by patients treated with daratumumab – as seen in CASSIOPEIA, where DBTd almost doubled the rate of MRD negativity compared to BTd at the higher sensitivity threshold of 10^{-6} (refer to Section B.2.6.1). Similarly, for MRD-positive patients, the improved survival outcomes in the DBTd arm is driven by deeper conventional response according to IMWG criteria achieved by daratumumab-treated patients (refer to Section B.2.6.1).

With the exception of the parameters relating to OS for DBTd and BTd (HRs and the exponential rate), the increase in the ICER from the base case was less than [redacted] per QALY gained for all other parameters varied in the DSA.

Figure 38: ICER tornado plot from deterministic sensitivity analyses – top 10 parameters

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; ICER = incremental cost-effectiveness ratio; MRD = minimal residual disease; OS = overall survival; PFS = progression-free survival.

B.3.8.3 Scenario analysis

The results of scenario analyses are presented in Table 84.

Across several scenarios, the ICER was similar to that seen in the base case analysis (e.g. scenarios for vial sharing, alternative utility inputs and BTd dosing based on the bortezomib SmPC). The use of the IV formulation of daratumumab resulted in an increase in the ICER compared to the base case, due to the higher drug acquisition and administration costs for DBTd (assuming no vial sharing). With the SC formulation offering a more convenient route of administration for patients, as well as a lower cost and an increase in NHS capacity, it is expected that the majority of patients would receive treatment via SC injection as per the base case analysis.

The scenarios exploring different approaches to modelling response-based survival (Scenarios 1A and 1B) or alternative PFS extrapolations (Scenario 2) were associated with higher ICERs compared to the base case analysis. With the exception of Scenario 1B, the increases in the ICER were however relatively modest (less than [redacted] per QALY gained). In Scenario 1B, OS and PFS for MRD- patients in the DBTd arm were based solely on the benefits of achieving MRD negativity via the use of the HRs from the statistical analysis performed on the expanded SLR/meta-analysis and not data from the CASSIOPEIA landmark analysis. However, as noted in Section B.3.3.2, this scenario fails to capture the treatment effect with daratumumab that has been observed in the CASSIOPEIA trial and across other daratumumab trials (refer to Section B.2.6.3), and is therefore considered to represent a highly conservative estimate of survival in the DBTd arm. The PFS and OS predicted by the model in this scenario (and Scenario 1A) are presented in Appendix O, with both PFS and OS from CASSOPEIA being consistently and considerably underestimated in the DBTd model cohort using the Scenario 1B approach.

The scenarios exploring treatment waning for daratumumab for both MRD-positive and MRD-negative patients (Scenario 3A, after 5 years; Scenario 3B, after 10 years), and MRD-negative patients only (Scenario 3C, after 5 years; Scenario 3D, after 10 years), were also associated with higher ICERs compared with the base case analysis. However, these scenarios are considered to be highly conservative and are not supported by the existing clinical evidence. These scenarios have been presented in order to fully explore uncertainty in the daratumumab treatment effect in the long term.

In the scenario exploring the impact of including drugs recommended by NICE via the Cancer Drugs Fund as subsequent therapies, DBTd was seen to dominate BTd, being associated with lower total costs and higher total QALYs. In this scenario, the costs accrued in the PD health state are considerably lower in the DBTd arm, reflecting the expected use of daratumumab at later lines of therapy in the BTd arm.

Table 84: Summary of results from scenario analyses

Scenario	Inc. costs	Inc. QALYs	ICER (£ per QALY)
Base case	[redacted]	[redacted]	[redacted]
1A: Approach to modelling BTd MRD- (Extrapolation of BTd MRD- from CASSIOPEIA using Weibull for PFS and Weibull for OS)	[redacted]	[redacted]	[redacted]
1B: Approach to modelling DBTd MRD- (Using HR for MRD- versus MRD+ from SLR/meta-analysis)	[redacted]	[redacted]	[redacted]

2: Extrapolation of BTd MRD+ PFS (Exponential)	[redacted]	[redacted]	[redacted]
3A: No additional treatment effect of DBTd after 5 years (MRD+ and MRD-)	[redacted]	[redacted]	[redacted]
3B: No additional treatment effect of DBTd after 10 years (MRD+ and MRD-)	[redacted]	[redacted]	[redacted]
3C: No additional treatment effect of DBTd after 5 years (MRD- only)[a]	[redacted]	[redacted]	[redacted]
3D: No additional treatment effect of DBTd after 10 years (MRD- only)[a]	[redacted]	[redacted]	[redacted]
4: Daratumumab IV formulation	[redacted]	[redacted]	[redacted]
5: Inclusion of subsequent therapies recommended via the Cancer Drugs Fund	[redacted]	[redacted]	[redacted]
6: Dosing for BTd (based on bortezomib SmPC)	[redacted]	[redacted]	[redacted]
7: With vial sharing	[redacted]	[redacted]	[redacted]
8A: PD utility = 0.644 from van Agthoven et al. (2004) (TA311)	[redacted]	[redacted]	[redacted]
8B: PD utility = 0.57 from Palumbo et al. (2013) (TA510)	[redacted]	[redacted]	[redacted]
8C: Utility values from van Agthoven et al. (2004)	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; ICER = incremental cost-effectiveness ratio; Inc. = incremental; IV = intravenous; MRD = minimal residual disease; OS = overall survival; PD = progressed disease; PFS = progression-free survival; QALY = quality-adjusted life year; SLR = systematic literature review; SmPC = Summary of Product Characteristics.

[a] In this scenario, the treatment effect is still applied across the entire model time horizon for MRD-positive patients.
Note: Refer to Table 81 for further information regarding each scenario.

B.3.8.4 Summary of sensitivity analyses results

As shown in the results of the PSA and extensive scenario analyses, the results of the CUA are relatively robust to uncertainty in parameter inputs and alternative inputs/assumptions. The scenarios with the greatest impact on the ICER were those relating to the treatment effect for DBTd versus BTd. This is consistent with the findings from the DSA which identified the HRs for the daratumumab treatment effect on OS as key model drivers.

As may be expected, the more pessimistic assumptions explored in the scenario analyses with regards to the daratumumab treatment effect resulted in ICERs that were higher than those in the base case analysis. Across all other scenarios, the ICER versus BTd was similar to the base case analysis and average probabilistic results, and was in the region of £30,000 per QALY gained.

B.3.9 *Subgroup analysis*

No cost-effectiveness analyses were conducted in subgroups.

B.3.10 *Validation*

Feedback on the plausibility of the survival inputs used in the model was originally obtained from one clinician in the UK. An advisory board was also held in August 2020 from which feedback was obtained from an additional three clinicians in the UK. At the advisory board, feedback was sought on the MM treatment pathway in the UK and the generalisability of the CASSIOPEIA trial data, as well as the plausibility of the survival inputs used in the model (as described below).

B.3.10.1 Validation of cost-effectiveness analysis

Validation and clinical plausibility of survival outcomes

Long-term PFS and OS extrapolations from the model were assessed using a combination of statistical goodness of fit criteria, visual inspection, real-world evidence of outcomes for UK patients, and clinical expert opinion on the plausibility of long-term extrapolations.

As described in Section B.3.3.2, the choice of extrapolation for BTd MRD+ PFS and OS was informed by feedback from UK clinicians on the long-term survival outcomes expected in clinical practice. The final (weighted) survival outcomes predicted by the model, which incorporate both MRD-positive and MRD-negative patients, were then compared against:

- Observed data from the CASSIOPEIA trial (DBTd and BTd)
- Data from PHE datasets on real-world survival outcomes for [redacted] patients in England who received first-line therapy and ASCT (and [redacted] patients who received first-line BTd and ASCT) (refer to Section B.3.3.2)

The model was seen to closely predict OS when compared to the CASSIOPEIA trial for both DBTd and BTd, and model predictions for BTd OS were also consistent with those reported in the PHE cohort. The proportion of patients in the model who were predicted to be alive and progression free was, however, slightly but consistently underestimated in both treatment arms when compared to CASSIOPEIA. Given that this underestimation occurred in both the DBTd and BTd model cohorts, and to a similar extent based on inspection of the different survival curves, the impact on the incremental cost-effectiveness results is expected to be limited. In contrast, the PFS estimates from the model were generally higher than the survival rates reported in the PHE cohort, with the model therefore predicting PFS survival rates that, reassuringly, were within the range provided by outcomes from the CASSIOPEIA trial and PHE cohort. The validity of survival outcomes predicted by the model with respect to clinical practice in England is further discussed below.

Validity of the model compared to clinical practice in England

For consistency with the primary source of evidence available for daratumumab in this indication, the inputs and assumptions used in the model were based on the trial design of CASSIOPEIA and the data that has been reported from it. By using the CASSIOPEIA trial as the basis for the model, the cost-effectiveness analyses are (by extension) also subject to the points raised in Section B.2.13 about the generalisability of CASSIOPEIA to clinical practice in England. For example, in using OS and PFS data from CASSIOPEIA, the survival outcomes predicted by the model will be based on patients (in both treatment arms) approximately 50% of whom will have received maintenance therapy in Part 2. However, due to the limited study follow-up (PHA; median follow-up = 29.2 months), and the eight-week dosing schedule of daratumumab maintenance, the exposure to maintenance treatment for either treatment group is limited. Also,

whilst absolute survival outcomes for both DBTd and BTd may as a result be better than expected in clinical practice in England, the relative treatment benefit for DBTd versus BTd that is incorporated in the model is unaffected by maintenance treatment.

Furthermore, the efficacy and cost of BTd consolidation therapy is included in the model, as per the CASSIOPEIA trial protocol. Consolidation with BTd is not however routinely used in clinical practice in England. It is therefore expected that the survival outcomes predicted by the model for BTd may overestimate survival for patients currently treated in England, and thus underestimate the relative treatment benefit of DBTd plus consolidation therapy in clinical practice. As noted above and in Section B.3.3.2, external validation of the predicted model outcomes against CASSIOPEIA and PHE datasets show consistency for OS and variation for PFS. The use of an alternative survival distribution to extrapolate PFS in the model (for BTd MRD+ patients) was explored as a scenario analysis, with the Weibull selected to provide a more optimistic estimate of long-term PFS when compared to the base case. In this scenario, DBTd was still associated with an ICER [redacted] per QALY gained versus BTd (see Section B.3.8.3).

Internal validation

The model programming was checked by an analyst who was not involved in the original development of the model using a validation checklist similar that reported in the published literature.(150) This involved a quality control check of the formulae used in the model and stress testing of the model to ensure that it behaves as expected when extreme values are used.

B.3.11 *Interpretation and conclusions of economic evidence*

Currently, there is no cure for MM. The primary goal of therapy is therefore to induce remission and delay disease progression. With each relapse, it becomes more challenging to induce a deep and durable response to treatment, with high attrition rates between lines of therapy highlighting the need to treat patients with the most efficacious regimens first. Despite several new treatments having been approved in later lines during the past decade, there has been limited progress in the development of new effective regimens for the management of NDTE MM patients with no new licenced therapy approved since BTd in 2013. Currently all patients eventually relapse leading to poorer prognosis, highlighting the high level of unmet need that still exists.

The economic analysis presented in this submission is robust, makes best use of available data, and captures the treatment effect of daratumumab over and above the attainment of MRD negativity. The daratumumab treatment effect has been consistently observed across the clinical development plan from daratumumab monotherapy in the relapsed/refractory setting to daratumumab combination therapy in newly diagnosed MM.(1) Due to immaturity of the survival data in the NDTE setting of CASSIOPEIA, a response-based modelling approach was taken which leveraged external data with longer follow-up to inform the relationship between MRD status and long-term survival outcomes. Indeed, the association between MRD status and PFS and OS for NDTE MM has already been established following an expanded SLR/meta-analysis which identified the strong prognostic value of MRD assessment.(63) The use of a response-based modelling approach is also consistent with the modelling approach taken in the appraisal of bortezomib as an induction therapy for patients eligible for SCT (TA311), the only other induction therapy to be assessed by NICE.(27)

The cost-effectiveness of DBTd as a treatment for adult patients with newly diagnosed MM who are eligible for ASCT was assessed via response-based CUA from the perspective of the NHS in

England. The comparator included in the CUA was BTd, which was recommended by NICE in TA311, and which represents SOC induction therapy for the majority of patients in clinical practice in England and is the intervention that DBTd would be expected to displace.(27) Direct evidence for the clinical efficacy and safety of DBTd versus BTd in the relevant patient population is available from the CASSIOPEIA trial and data from this trial is used in the economic evaluation.

For a minority of patients where thalidomide is not considered suitable (e.g. due to challenging thrombosis, or baseline neuropathy/neurotoxicity), BCd may be administered in clinical practice instead with the doublet therapy, Bd, rarely used. A comprehensive and robust SLR was carried out to identify clinical evidence on all relevant comparators (refer to Appendix D), however major differences in study design and outcomes assessed meant that it was not possible to incorporate BCd or Bd in a standard evidence network. The feasibility of conducting an MAIC based on response was explored, however data on post-consolidation MRD negativity was only reported from the CASSIOPEIA trial meaning a response-based MAIC to inform the CUA was not feasible. Instead, results from two MAICs and a naïve comparison of PFS and OS outcomes from the PHE linked datasets indicate that the efficacy of BCd and Bd are no better than BTd (refer to Section B.2.9). The costs across the three regimens are also expected to be similar, given the relatively low cost of thalidomide and cyclophosphamide (refer to Section B.3.5.1). As such, a pragmatic approach may be taken to economic analysis. That is, if DBTd is considered cost-effective versus BTd, the daratumumab quadruplet combination is also likely to be cost-effective versus BCd and Bd.

Model extrapolations have been assessed based on consideration of statistical/visual fit, real-world evidence of outcomes for UK patients, and clinical expert opinion. Whilst complexity of the CASSIOPEIA trial design, which included re-randomisation to maintenance therapy for Part 2, introduces some challenges regarding the generalisability of absolute survival outcomes, prespecified statistical analysis for PFS using the IPW methodology has demonstrated that the relative treatment effect modelled in this economic evaluation is robust and not subject to bias. This finding was later confirmed with longer study follow-up (median follow-up = 29.2 months) to support the EMA regulatory approval. Whilst Part 2 of the study remains blinded, Janssen does not have access to individual patient-level data and no further statistical analysis of PFS/OS can be performed.

The results of the CUA found DBTd to represent a cost-effective use of NHS resources in England, being associated with an ICER at list price of [redacted] per QALY gained versus BTd. [redacted] B.3.7.1 [redacted] The significant clinical benefits of introducing daratumumab to the front-line setting was also demonstrated by the incremental life-years (3.32) and QALYs ([redacted]) gained versus BTd in the CUA. The model results are considered to be robust, and the inputs and assumptions used in the model have been tested and explored via the use of extensive scenario and sensitivity analyses. In the PSA, the probability that DBTd would be cost-effective versus BTd at willingness-to-pay thresholds of £20,000 and £30,000 per QALY gained was estimated to be 18.7% and 51.8%, respectively.

As well as first-line treatment with DBTd and BTd, the model includes the expected costs of subsequent therapies expected to be received by patients in England. In accordance with the NICE position statement on the inclusion of therapies recommended via the Cancer Drugs Fund, several regimens have not been included as subsequent therapies in the base case analysis, including DBd (2nd line) and daratumumab monotherapy (4th line).(137) Should these subsequent

therapies become routinely funded in England at a later point, then the cost-effectiveness results for DBTd as a first-line therapy within the treatment pathway would be expected to improve (as shown in the scenario including drugs recommended on the Cancer Drugs Fund; refer to Section B.3.8.3). For example, patients treated with DBTd as induction/consolidation treatment at front-line are less likely to receive daratumumab at subsequent treatment lines (e.g. DBd at second-line or daratumumab monotherapy at fourth-line). Therefore, the cost of introducing daratumumab as a fixed duration therapy at front-line would be partly offset by fewer patients receiving daratumumab later on in the treatment pathway, as a treat-to-progression regimen.

In summary, the results of the CUA suggest that the use of daratumumab, in combination with bortezomib and dexamethasone, as an induction and consolidation therapy for newly-diagnosed adult patients with MM who are eligible for ASCT, would represent a cost-effective treatment strategy, being associated with an ICER of [redacted] versus BTd. DBTd addresses the unmet need for a safe and effective quadruplet therapy for NDTE MM patients that can drive deep responses and prolong remission whilst maintaining HRQoL. As a highly innovative and effective therapy, the use of DBTd earlier on in the MM treatment pathway would represent a step-change in the management of patients who are eligible for ASCT. Indeed, for those patients treated with DBTd who achieve MRD negativity, there is hope of long-term disease control and functional cure, with the mortality rate resembling that of the general population. Furthermore, with a fixed treatment duration of six cycles, DBTd offers a sustained period of treatment-free remission with good quality of life which is highly valued by both patients and carers. The positive effect that treatment with DBTd could have on informal carers in terms of reduced anxiety/depression and the ability to return to work is not captured in the economic analysis. Similarly, the psychological impact of achieving sustained remission, in terms of the sense of hope that patients and carers may experience in place of the fear of relapse, is not intrinsically captured as part of the QALY framework.

References

1. Janssen. [Data on File] Clinical Advisory Board Meeting Minutes. August 2020.
2. Janssen. [Data on File] Public Health England datasets. Standing cohort study of newly diagnosed multiple myeloma cancer patients in England (January 2015 to December 2018). (July 2020 Report).
3. Khagi Y, Mark TM. Potential role of daratumumab in the treatment of multiple myeloma. OncoTargets and therapy. 2014;7:1095-100.
4. de Weers M, Tai YT, van der Veer MS, Bakker JM, Vink T, Jacobs DC, et al. Daratumumab, a novel therapeutic human CD38 monoclonal antibody, induces killing of multiple myeloma and other hematological tumors. Journal of immunology (Baltimore, Md : 1950). 2011;186(3):1840-8.
5. Electronic Medicines Compendium (EMC). Darzalex 1,800 mg solution for injection: Summary of Product Characteristics 2020 [Available from: https://www.medicines.org.uk/emc/product/11488/smpc.
6. Electronic Medicines Compendium (EMC). Darzalex 20mg/ml concentrate for solution for infusion: Summary of Product Characteristics 2020. 2020 [Available from: https://www.medicines.org.uk/emc/product/7250/smpc.
7. Food and Drug Administration (FDA). Orphan Drug Designations and Approvals. Daratumumab. Available at: https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=397113 [
8. European Medicines Agency (EMA). Public Summary of Opinion on Orphan Designation: Daratumumab for the treatment of plasma cell myeloma. 2013.
9. Kyle RA RS. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. Leukemia. 2009;23(1):3-9.
10. Röllig C KS, Bornhäuser M. Multiple myeloma. The Lancet. 2015;385(9983):2197-208.
11. Kumar SK VR, Noga SJ, Berg D, Brent L, Dollar L, et al. Treating Multiple Myeloma Patients With Oral Therapies. Clinical Lymphoma, Myeloma and Leukemia. 2017;17(5):P243-51.
12. Palumbo A SO, Kyle R, Miguel JS, Orlowski RZ, Moreau P, et al. International Myeloma Working Group guidelines for the management of multiple myeloma patients ineligible for standard high-dose chemotherapy with autologous stem cell transplantation. Leukemia. 2009;23(10):1716-30.
13. Bianchi G AK. Understanding biology to tackle the disease: Multiple myeloma from bench to bedside, and back. CA: a cancer journal for clinicians 2014;23(10):422-44.
14. Palumbo A, Anderson K. Multiple myeloma. The New England journal of medicine. 2011;364(11):1046-60.
15. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(26):2863-9.
16. Richardson P, Mitsiades C, Schlossman R, Ghobrial I, Hideshima T, Chauhan D, et al. The treatment of relapsed and refractory multiple myeloma. Hematology American Society of Hematology Education Program. 2007:317-23.
17. Brenner H, Gondos A, Pulte D. Recent major improvement in long-term survival of younger patients with multiple myeloma. Blood. 2008;111(5):2521-6.
18. Kumar SK, Rajkumar SV, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, et al. Improved survival in multiple myeloma and the impact of novel therapies. Blood. 2008;111(5):2516-20.
19. Pratt G, Jenner M, Owen R, Snowden JA, Ashcroft J, Yong K, et al. Updates to the guidelines for the diagnosis and management of multiple myeloma. British journal of haematology. 2014;167(1):131-3.
20. Usmani S, Ahmadi T, Ng Y, Lam A, Desai A, Potluri R, et al. Analysis of Real-World Data on Overall Survival in Multiple Myeloma Patients With ≥3 Prior Lines of Therapy Including a Proteasome Inhibitor (PI) and an Immunomodulatory Drug (IMiD), or Double Refractory to a PI and an IMiD. The oncologist. 2016;21(11):1355-61.
21. Yong K, Delforge M, Driessen C, Fink L, Flinois A, Gonzalez-McQuire S, et al. Multiple myeloma: patient outcomes in real-world practice. British journal of haematology. 2016;175(2):252-64.
22. Voorhees P, Fonseca R, Cote S, Lam A, Ukropec J, Maiese E. Characterization of frontline treatment patterns and the proportion of patients reaching subsequent lines of therapy in transplant eligible patients with newly diagnosed multiple myeloma: . HemaSphere. 2019;3:274-5.

23. Hajek R. Strategies for the Treatment of Multiple Myeloma in 2013: Moving Toward the Cure 2013. Available from: https://www.intechopen.com/books/multiple-myeloma-a-quick-reflection-on-the-fast-progress/strategies-for-the-treatment-of-multiple-myeloma-in-2013-moving-toward-the-cure.
24. Cancer Research UK (CRUK). Myeloma Statistics 2017. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma.
25. National Institute for Health and Care Excellence (NICE). NG35. Myeloma: diagnosis and management. (2016) Available at: https://www.nice.org.uk/guidance/ng35.
26. Moreau P, San Miguel J, Sonneveld P, Mateos MV, Zamagni E, Avet-Loiseau H, et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2017;28(suppl_4):iv52-iv61.
27. National Institute for Health and Care Excellence (NICE). TA311: Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation. Available at: https://www.nice.org.uk/guidance/ta311 [
28. Johnsen AT TD, Petersen MA, Pedersen L, Groenvold M. Health related quality of life in a nationally representative sample of haematological patients. European Journal of Haematology. 2009;83(2):139-48.
29. Jordan K, Proskorovsky I, Lewis P, Ishak J, Payne K, Lordan N, et al. Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2014;22(2):417-26.
30. pH-Associates. A top-line summary of the existing evidence related to patient experience, health-related quality of life burden, and patient priorities in multiple myeloma: Targeted Literature Review. 2017 MCO-DARA-047 (part A).
31. Hulin C, Hansen T, Heron L, Pughe R, Streetly M, Plate A, et al. Living with the burden of relapse in multiple myeloma from the patient and physician perspective. Leukemia research. 2017;59:75-84.
32. Rizzo M, Xu Y, Panjabi S, Iheanacho I. A systematic literature review of the humanistic burden of multiple myeloma. Value in Health. 2014;17(7):A537.
33. Acaster S, Gaugris S, Velikova G, Yong K, Lloyd A. Impact of the treatment-free interval on health-related quality of life in patients with multiple myeloma: a UK cross-sectional survey. Supportive Care in Cancer. 2013;21(2):599-607.
34. He et al. 2020. Patient Perceptions Regarding Multiple Myeloma and Its Treatment: Qualitative Evidence From Interviews With Newly Diagnosed and Relapsed/Refractory Patients in the United Kingdom, France, and Germany. Presented at virtual ISPOR; May 18–20, 2020.
35. Myeloma UK. Measuring Patient Preferences: An exploratory study to determine how patient preferences data could be used in health technology assessment (HTA) Project report. Available at: https://www.myeloma.org.uk/wp-content/uploads/2019/07/NICE-Patient-Preferences-Report.pdf.
36. Kurtin S, Lilleby K, Spong J. Caregivers of multiple myeloma survivors. Clinical journal of oncology nursing. 2013;17(1):25-32.
37. Molassiotis A, Wilson B, Blair S, Howe T, Cavet J. Unmet supportive care needs, psychological well-being and quality of life in patients living with multiple myeloma and their partners. Psycho-oncology. 2011;20(1):88-97.
38. Rajkumar SV. Treatment of multiple myeloma. Nature reviews Clinical oncology. 2011;8(8):479-91.
39. Engelhardt M, Terpos E, Kleber M, Gay F, Wäsch R, Morgan G, et al. European Myeloma Network recommendations on the evaluation and treatment of newly diagnosed patients with multiple myeloma. Haematologica. 2014;99(2):232-42.
40. National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology. Multiple Myeloma. Version 4.2020. Available at: https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf. 2020.
41. Harousseau JL. Induction therapy in multiple myeloma. Hematology American Society of Hematology Education Program. 2008:306-12.
42. Gay F, Engelhardt M, Terpos E, Wäsch R, Giaccone L, Auner HW, et al. From transplant to novel cellular therapies in multiple myeloma: European Myeloma Network guidelines and future perspectives. Haematologica. 2018;103(2):197-211.
43. Moreau P, Attal M, Hulin C, Arnulf B, Belhadj K, Benboubker L, et al. Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell

transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label, phase 3 study. Lancet (London, England). 2019;394(10192):29-38.
44. Janssen. [Data on File] MMY3006. Clinical Study Report: Part 1. 2019.
45. Janssen. [Data on File] MMY3006. Post-hoc Interim Analysis. May 1, 2019.
46. Janssen. [Data on File] MMY3006 (CASSIOPEIA). Clinical Protocol. 2015.
47. Rajkumar SV, Harousseau JL, Durie B, Anderson KC, Dimopoulos M, Kyle R, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. Blood. 2011;117(18):4691-5.
48. Durie BG, Harousseau JL, Miguel JS, Bladé J, Barlogie B, Anderson K, et al. International uniform response criteria for multiple myeloma. Leukemia. 2006;20(9):1467-73.
49. Thoennissen GB, Gorlich D, Bacher U, Aufenberg T, Husken AC, Hansmeier AA, et al. Autologous Stem Cell Transplantation in Multiple Myeloma in the Era of Novel Drug Induction: A Retrospective Single-Center Analysis. Acta haematologica. 2017;137(3):163-72.
50. Paquin AR, Kumar SK, Buadi FK, Gertz MA, Lacy MQ, Dispenzieri A, et al. Overall survival of transplant eligible patients with newly diagnosed multiple myeloma: comparative effectiveness analysis of modern induction regimens on outcome. Blood cancer journal. 2018;8(12):125.
51. Lahuerta JJ, Mateos MV, Martinez-Lopez J, Rosinol L, Sureda A, de la Rubia J, et al. Influence of pre- and post-transplantation responses on outcome of patients with multiple myeloma: sequential improvement of response and achievement of complete response are associated with longer survival. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(35):5775-82.
52. van de Velde HJ, Liu X, Chen G, Cakana A, Deraedt W, Bayssas M. Complete response correlates with long-term survival and progression-free survival in high-dose therapy in multiple myeloma. Haematologica. 2007;92(10):1399-406.
53. Harousseau JL, Attal M, Avet-Loiseau H. The role of complete response in multiple myeloma. Blood. 2009;114(15):3139-46.
54. Chanan-Khan AA, Giralt S. Importance of achieving a complete response in multiple myeloma, and the impact of novel agents. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(15):2612-24.
55. Gay F, Larocca A, Wijermans P, Cavallo F, Rossi D, Schaafsma R, et al. Complete response correlates with long-term progression-free and overall survival in elderly myeloma treated with novel agents: analysis of 1175 patients. Blood. 2011;117(11):3025-31.
56. Bladé J, Esteve J, Rives S, Martínez C, Rovira M, Urbano-Ispizua A, et al. High-dose therapy autotransplantation/intensification vs continued standard chemotherapy in multiple myeloma in first remission. Results of a non-randomized study from a single institution. Bone Marrow Transplantation. 2000;26(8):845-9.
57. Kapoor P, Kumar SK, Dispenzieri A, Lacy MQ, Buadi F, Dingli D, et al. Importance of achieving stringent complete response after autologous stem-cell transplantation in multiple myeloma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31(36):4529-35.
58. García de Veas Silva JL, Bermudo Guitarte C, Menéndez Valladares P, Rojas Noboa JC, Kestler K, Duro Millán R. Prognostic Value of Serum Free Light Chains Measurements in Multiple Myeloma Patients. PLoS One. 2016;11(11):e0166841-e.
59. Palumbo A, Rajkumar SV, San Miguel JF, Larocca A, Niesvizky R, Morgan G, et al. International Myeloma Working Group consensus statement for the management, treatment, and supportive care of patients with myeloma not eligible for standard autologous stem-cell transplantation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(6):587-600.
60. Paiva B, van Dongen JJ, Orfao A. New criteria for response assessment: role of minimal residual disease in multiple myeloma. Blood. 2015;125(20):3059-68.
61. Munshi NC, Avet-Loiseau H, Rawstron AC, Owen RG, Child JA, Thakurta A, et al. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. JAMA oncology. 2017;3(1):28-35.
62. van Agthoven M, Segeren CM, Buijt I, Uyl-de Groot CA, van der Holt B, Lokhorst HM, et al. A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study. European journal of cancer (Oxford, England : 1990). 2004;40(8):1159-69.

63. Munshi, N. et al. Expanded Meta-Analysis Confirms the Association Between MRD and Long-term Survival Outcomes in Multiple Myeloma (MM). Poster presented at American Society of Hematology (ASH). 2019
64. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. The Lancet Oncology. 2016;17(8):e328-e46.
65. European Medicines Agency. Darzalex Assessment Report Variation. Available at: https://www.ema.europa.eu/en/documents/variation-report/darzalex-h-c-4077-ii-0030-epar-assessment-report-variation_.pdf.
66. Higgins J, Sterne J, Savovic J, Page M, Hrobjartsson A, Bourtron I, et al. A revised tool for assessing risk of bias in randomized trials. Cochrane Database of Systematic Reviews. 2016;10(Suppl 1):29-31.
67. Hervé Avet-Loiseau IUdC, Toulouse, France. Minimal Residual Disease assessment in Multiple Myeloma by Next-Generation Sequencing.
68. Lokhnygina Y, Helterbrand JD. Cox regression methods for two-stage randomization designs. Biometrics. 2007;63(2):422-8.
69. Spencer A, Lentzsch S, Weisel K, Avet-Loiseau H, Mark TM, Spicka I, et al. Daratumumab plus bortezomib and dexamethasone versus bortezomib and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of CASTOR. Haematologica. 2018;103(12):2079-87.
70. Spencer A, et al. Daratumumab, Bortezomib, and Dexamethasone (DVd) Versus Bortezomib and Dexamethasone (Vd) in Relapsed or Refractory Multiple Myeloma (RRMM): Updated Efficacy and Safety Analysis of CASTOR. Poster 3145. Presented at ASH 2017.
71. Mateos M-V, et al. Oral presentation at: 58th American Society of Hematology (ASH) Annual Meeting and Exposition; December 3-6 2016; San Diego, CA, USA. .
72. Palumbo A, Chanan-Khan A, Weisel K, Nooka AK, Masszi T, Beksac M, et al. Daratumumab, bortezomib, and dexamethasone for multiple myeloma. New England Journal of Medicine. 2016;375(8):754-66.
73. Weisel et al. Efficacy and Safety of Daratumumab, Bortezomib and Dexamethasone (D-Vd) Versus Bortezomib and Dexamethasone (Vd) in First Relapse Patients (pts) With Multiple Myeloma (MM): Four-Year Update of CASTOR. ASH Poster 3192. December 7-10, 2019. Orlando, Florida. .
74. Mateos et al. Efficacy and Safety of Daratumumab, Bortezomib and Dexamethasone (D-Vd) Versus Bortezomib and Dexamethasone (Vd) in First Relapse Patients With Multiple Myeloma (MM): Update of CASTOR. ASH Poster 3270. December 1-4, 2018. San Diego, California. .
75. Harousseau JL, Attal M, Avet-Loiseau H, Marit G, Caillot D, Mohty M, et al. Bortezomib plus dexamethasone is superior to vincristine plus doxorubicin plus dexamethasone as induction treatment prior to autologous stem-cell transplantation in newly diagnosed multiple myeloma: results of the IFM 2005-01 phase III trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(30):4621-9.
76. Goldschmidt H, Ashcroft J, Szabo Z, Garderet L. Navigating the treatment landscape in multiple myeloma: which combinations to use and when? Annals of hematology. 2019;98(1):1-18.
77. Mellors PW, Binder M, Buadi FK, Lacy MQ, Gertz MA, Dispenzieri A, et al. Time to plateau as a predictor of survival in newly diagnosed multiple myeloma. American Journal of Hematology. 2018;93(7):889-94.
78. Lehners N, Becker N, Benner A, Pritsch M, Löpprich M, Mai EK, et al. Analysis of long-term survival in multiple myeloma after first-line autologous stem cell transplantation: impact of clinical risk factors and sustained response. Cancer Med. 2018;7(2):307-16.
79. Hoering A, Durie B, Wang H, Crowley J. End points and statistical considerations in immuno-oncology trials: impact on multiple myeloma. Future Oncol. 2017;13(13):1181-93.
80. Roussel M, Moreau P, Attal M, Eisenmann JC. Improvement in health-related quality of life (HRQoL) for newly diagnosed multiple myeloma (NDMM) transplant eligible patients treated with daratumumab, bortezomib, thalidomide, and dexamethasone (D-VTd) vs VTd alone: CASSIOPEIA. European Hematology Association (EHA)2019.
81. Janssen. [Data on File] Unanchored Matching-Adjusted Indirect Comparison of D-VTd/VTd (CASSIOPEIA) vs. Vd (IFM 2005-01) and VCd (GMMG-MM5) for the Treatment of Patients with Newly Diagnosed Multiple Myeloma who are Eligible for Autologous Stem Cell Transplant. 8 November 2019.
82. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.

83. Kvam AK, Fayers PM, Wisloff F. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. European journal of haematology. 2011;87(4):330-7.
84. Bedard G, Zeng L, Zhang L, Lauzon N, Holden L, Tsao M, et al. Minimal important differences in the EORTC QLQ-C30 in patients with advanced cancer. Asia-Pacific journal of clinical oncology. 2014;10(2):109-17.
85. Kaufman JL, Dimopoulos MA, Leiba M, Morton J, Ho PJ, Kim K, et al. Efficacy and safety of daratumumab, lenalidomide, and dexamethasone (D-Rd) in relapsed or refractory multiple myeloma (RRMM): Updated subgroup analysis of POLLUX based on cytogenetic risk. American Society of Clinical Oncology; 2019.
86. Weisel KC, Sonneveld P, Mateos M-V, Hungria VTM, Spencer A, Estell J, et al. Efficacy and Safety of Daratumumab, Bortezomib, and Dexamethasone (D-Vd) Versus Bortezomib and Dexamethasone (Vd) in First Relapse Patients (pts) with Multiple Myeloma (MM): Four-Year Update of Castor. Blood. 2019;134(Supplement_1):3192-.
87. Avet-Loiseau H, Moreau P, Attal M, Hulin C, Arnulf B, Corre J, et al. Efficacy of daratumumab (DARA) + bortezomib/thalidomide/dexamethasone (D-VTd) in transplant-eligible newly diagnosed multiple myeloma (TE NDMM) based on minimal residual disease (MRD) status: Analysis of the CASSIOPEIA trial. Journal of Clinical Oncology. 2019;37(15_suppl):8017-.
88. Moreau P, Avet-Loiseau H, Facon T, Attal M, Tiab M, Hulin C, et al. Bortezomib plus dexamethasone versus reduced-dose bortezomib, thalidomide plus dexamethasone as induction treatment prior to autologous stem cell transplantation in newly diagnosed multiple myeloma. Blood. 2011:blood-2011-05-355081.
89. Cavo M, Tacchetti P, Patriarca F, Petrucci MT, Pantani L, Galli M, et al. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. The Lancet. 2010;376(9758):2075-85.
90. Ludwig H, Viterbo L, Greil R, Masszi T, Spicka I, Shpilberg O, et al. Randomized phase II study of bortezomib, thalidomide, and dexamethasone with or without cyclophosphamide as induction therapy in previously untreated multiple myeloma. Journal of Clinical Oncology. 2013;31(2):247-55.
91. Rosinol L, Oriol A, Teruel AI, Hernandez D, Lopez-Jimenez J, de la Rubia J, et al. Superiority of bortezomib, thalidomide, and dexamethasone (VTD) as induction pretransplantation therapy in multiple myeloma: a randomized phase 3 PETHEMA/GEM study. Blood. 2012;120(8):1589-96.
92. Moreau P, Hulin C, Macro M, Caillot D, Chaleteix C, Roussel M, et al. VTD is superior to VCD prior to intensive therapy in multiple myeloma: results of the prospective IFM2013-04 trial. Blood. 2016;127(21):2569-74.
93. Sunami K, Matsumoto M, Fuchida S-i, Omoto E, Takamatsu H, Adachi Y, et al. Bortezomib-based strategy with autologous stem cell transplantation for newly diagnosed multiple myeloma: a phase II study by the Japan Study Group for Cell Therapy and Transplantation (JSCT-MM12). International journal of clinical oncology. 2019;24(8):966-75.
94. Goldschmidt H, Mai EK, Dürig J, Scheid C, Weisel KC, Kunz C, et al. Response-Adapted Lenalidomide Maintenance in Newly Diagnosed, Transplant-Eligible Multiple Myeloma: Results from the Multicenter Phase III GMMG-MM5 Trial. Blood. 2017;130(Suppl 1):400-.
95. Mai EK, Bertsch U, Dürig J, Kunz C, Haenel M, Blau IW, et al. Phase III trial of bortezomib, cyclophosphamide and dexamethasone (VCD) versus bortezomib, doxorubicin and dexamethasone (PAd) in newly diagnosed myeloma. Leukemia. 2015;29(8):1721-9.
96. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):940-7.
97. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. 2016. Available at http://www.nicedsu.org.uk. 2016.
98. Kumar S, Flinn I, Richardson PG, Hari P, Callander N, Noga SJ, et al. Randomized, multicenter, phase 2 study (EVOLUTION) of combinations of bortezomib, dexamethasone, cyclophosphamide, and lenalidomide in previously untreated multiple myeloma. Blood, The Journal of the American Society of Hematology. 2012;119(19):4375-82.
99. Gregersen H, Do T, Kristensen IB, Frølund UC, Andersen NF, Nielsen LK, et al. A randomized placebo-controlled phase II study of clarithromycin or placebo combined with VCD

induction therapy prior to high-dose melphalan with stem cell support in patients with newly diagnosed multiple myeloma. Experimental hematology & oncology. 2018;7:18.
100. El-Ghammaz AM, Abdelwahed E. Bortezomib-based induction improves progression-free survival of myeloma patients harboring 17p deletion and/or t(4;14) and overcomes their adverse prognosis. Annals of hematology. 2016;95(8):1315-21.
101. Kumar L, kumar Chellapuram S, Sahoo R, Gupta R. VRd versus VCd as induction therapy for newly diagnosed multiple myeloma: A Phase III, randomized study. Clinical Lymphoma, Myeloma and Leukemia. 2019;19(10):e361.
102. Tanaka K, Toyota S, Akiyama M, Wakimoto N, Nakamura Y, Najima Y, et al. Efficacy and Safety of a Weekly Cyclophosphamide-Bortezomib-Dexamethasone Regimen as Induction Therapy Prior to Autologous Stem Cell Transplantation in Japanese Patients with Newly Diagnosed Multiple Myeloma: A Phase 2 Multicenter Trial. Acta haematologica. 2019;141(2):111-8.
103. Higgins JP, Sterne JA, Savovic J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. Cochrane database of systematic reviews. 2016;10(Suppl 1):29-31.
104. Mateos M-V, Nahi H, Legiec W, Grosicki S, Vorobyev V, Spicka I, et al. Efficacy and safety of the randomized, open-label, non-inferiority, phase 3 study of subcutaneous (SC) versus intravenous (IV) daratumumab (DARA) administration in patients (pts) with relapsed or refractory multiple myeloma (RRMM): COLUMBA. American Society of Clinical Oncology; 2019.
105. ClinicalTrials.gov. NCT02541383. A Study to Evaluate Daratumumab in Transplant Eligible Participants With Previously Untreated Multiple Myeloma (Cassiopeia). Available at: https://clinicaltrials.gov/ct2/show/NCT02541383.
106. ClinicalTrials.gov. NCT03710603. A Study Comparing Daratumumab, VELCADE (Bortezomib), Lenalidomide and Dexamethasone Compared to VELCADE, Lenalidomide and Dexamethasone in Subjects With Previously Untreated Multiple Myeloma (Perseus). Available at: https://clinicaltrials.gov/ct2/show/NCT03710603.
107. ClinicalTrials.gov. NCT02874742. A Study Comparing Daratumumab, Lenalidomide, Bortezomib, and Dexamethasone (D-RVd) Versus Lenalidomide, Bortezomib, and Dexamethasone (RVd) in Subjects With Newly Diagnosed Multiple Myeloma. Available at: https://clinicaltrials.gov/ct2/show/NCT02874742.
108. ClinicalTrials.gov. NCT02951819. A Study to Evaluate Dara-CyBorD in Previously Untreated and Relapsed Subjects With Multiple Myeloma. Available at: https://clinicaltrials.gov/ct2/show/NCT02951819.
109. Prideaux SM, Conway O'Brien E, Chevassut TJ. The genetic architecture of multiple myeloma. Advances in hematology. 2014;2014:864058.
110. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. Nature reviews Cancer. 2012;12(5):335-48.
111. Barlogie B, Mitchell A, van Rhee F, Epstein J, Morgan GJ, Crowley J. Curing myeloma at last: defining criteria and providing the evidence. Blood. 2014;124(20):3043-51.
112. IQVIA. [Data on File] Harmony Multiple Myeloma Therapy Monitor. Executive Summary - UK Wave 22 (May/Jun 2020). 2020.
113. European Medicines Agency. Velcade. EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/velcade.
114. Mohty M, Terpos E, Mateos M, M C. Multiple myeloma (MM) treatment in real-world clinical practice: a focus on induction regimens prior to autologous stem cell transplantation (auto-SCT) from the prospective, multinational, non-interventional EMMOS study. EBMT. 2019.
115. Mohty M, Terpos E, Mateos MV, Cavo M, Lejniece S, Beksac M, et al. Multiple Myeloma Treatment in Real-world Clinical Practice: Results of a Prospective, Multinational, Noninterventional Study. Clinical lymphoma, myeloma & leukemia. 2018;18(10):e401-e19.
116. Sonneveld P. Impact of Modified Thalidomide Dosing In Bortezomib/Thalidomide/Dexamethasone for Patients With Newly Diagnosed Multiple Myeloma Who Are Transplant-eligible: A Matching-adjusted Indirect Comparison. ASH. 2019;Poster 4739.
117. Scottish Medicines Consortium (SMC). ID 927/13: In combination with dexamethasone, or with dexamethasone and thalidomide, for the induction treatment of adult patients with previously untreated multiple myeloma who are eligible for high-dose chemotherapy with haematopoietic stem cell transplantation. Available at: https://www.scottishmedicines.org.uk/medicines-advice/bortezomib-velcade-fullsubmission-92713/.
118. Facon T, Kumar SK, Plesner T, Orlowski RZ, Moreau P, Bahlis N, et al. Phase 3 Randomized Study of Daratumumab Plus Lenalidomide and Dexamethasone (D-Rd) Versus Lenalidomide and

Dexamethasone (Rd) in Patients with Newly Diagnosed Multiple Myeloma (NDMM) Ineligible for Transplant (MAIA). Blood. 2018;132(Supplement 1):LBA-2-LBA-.
119. Mateos M-V, Cavo M, Bladé J, Dimopoulos MA, Suzuki K, Jakubowiak A, et al. Daratumumab Plus Bortezomib, Melphalan, and Prednisone Versus Bortezomib, Melphalan, and Prednisone in Patients with Transplant-Ineligible Newly Diagnosed Multiple Myeloma: Overall Survival in Alcyone. Blood. 2019;134(Supplement_1):859-.
120. European Medicines Agency. Darzalex Assessment Report Extension. Available at: https://www.ema.europa.eu/en/documents/variation-report/darzalex-h-c-4077-x-0032-epar-assessment-report-extension_en.pdf.
121. Richardson PG, Delforge M, Beksac M, Wen P, Jongen JL, Sezer O, et al. Management of treatment-emergent peripheral neuropathy in multiple myeloma. Leukemia. 2012;26(4):595-608.
122. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available at: http://nicedsu.org.uk/.
123. Janssen. Personal Communication with Consultant Haematologist in the UK. May 2020.
124. Office for National Statistics (ONS). Cancer survival in England - adults diagnosed (2013-2017). Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed.
125. National Institute for Health and Care Excellence (NICE). TA510: Daratumumab monotherapy for treating relapsed and refractory multiple myeloma. Available at: https://www.nice.org.uk/guidance/ta510.
126. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(5):708-15.
127. Palumbo A, Davies F, Lee D, Dhanasiri S, Facon T, Zaki M, et al. Quality of life weights (utilities) in refractory or relapsed and refractory multiple myeloma (RRMM) patients using EORTC-8D and EQ-5D. Lymphoma and myeloma 2013 an international congress on hematologic malignancies. New York, United States of America. 24-26 October 2013. Poster P-03.
128. National Institute for Health and Care Excellence (NICE). Position statement on use of the EQ-5D-5L value set for England (updated October 2019). Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l.
129. EuroQol. EQ-5D-5L Valuation. Crosswalk Index Value Calculator. Available at: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/.
130. Abonour R, Wagner L, Durie BG, Jagannath S, Narang M, Terebelo HR, et al. Impact of post-transplantation maintenance therapy on health-related quality of life in patients with multiple myeloma: data from the Connect® MM Registry. Annals of hematology. 2018;97(12):2425-36.
131. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. University of York. Centre for Health Economics Discussion Paper 172. November 1999.
132. National Institute for Health and Care Excellence (NICE). TA573: Daratumumab with bortezomib and dexamethasone for previously treated multiple myeloma. Available at: https://www.nice.org.uk/guidance/TA573.
133. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. British journal of cancer. 2006;95(6):683-90.
134. Janssen B, Szende A. Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J, editors. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht: Springer Netherlands; 2014. p. 19-30.
135. Janssen. [Data on File] MMY3006 (CASSIOPEIA) Clinical Study Report: Subgroup Appendix. 28 Mar 2019.
136. European Medicines Agency. Thalidomide Celgene. EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/thalidomide-celgene.
137. National Institute for Health and Care Excellence (NICE). Position statement: consideration of products recommended for use in the Cancer Drugs Fund as comparators, or in a treatment sequence, in the appraisal of a new cancer product. Available at: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/cancer-drugs-fund/CDF-comparator-position-statement.pdf.

138. European Medicines Agency. Farydak. EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/farydak.
139. Richardson PG, Schlossman RL, Alsina M, Weber DM, Coutre SE, Gasparetto C, et al. PANORAMA 2: panobinostat in combination with bortezomib and dexamethasone in patients with relapsed and bortezomib-refractory myeloma. Blood. 2013;122(14):2331-7.
140. Stadtmauer EA, Weber DM, Niesvizky R, Belch A, Prince MH, San Miguel JF, et al. Lenalidomide in combination with dexamethasone at first relapse in comparison with its use as later salvage therapy in relapsed or refractory multiple myeloma. European Journal of Haematology. 2009;82(6):426-32.
141. National Institute for Health and Care Excellence (NICE). TA505: Ixazomib with lenalidomide and dexamethasone for treating relapsed or refractory multiple myeloma. Available at: https://www.nice.org.uk/guidance/TA505.
142. Richardson PG, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, et al. Panobinostat plus bortezomib and dexamethasone in previously treated multiple myeloma: outcomes by prior treatment. Blood. 2016;127(6):713-21.
143. National Institute for Health and Care Excellence (NICE). TA427: Pomalidomide for multiple myeloma previously treated with lenalidomide and bortezomib. Available at: https://www.nice.org.uk/guidance/TA427.
144. Usmani SZ, Weiss BM, Plesner T, Bahlis NJ, Belch A, Lonial S, et al. Clinical efficacy of daratumumab monotherapy in patients with heavily pretreated relapsed or refractory multiple myeloma. Blood. 2016;128(1):37-44.
145. European Medicines Agency. Revlimid. EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/revlimid.
146. European Medicines Agency. Imnovid (previously Pomalidomide Celgene). EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/imnovid-previously-pomalidomide-celgene.
147. European Medicines Agency. Ninlaro. EPAR - Product Information. Summary of Product Characteristics. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/ninlaro.
148. Curtis L, Burns A. Unit Costs of Health and Social Care 2019, Personal Social Services Research Unit, University of Kent, Canterbury. Available at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2019/.
149. National Institute for Health and Care Excellence (NICE). GID-TA10449: Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]. Available at: https://www.nice.org.uk/guidance/indevelopment/gid-ta10449.
150. Büyükkaramikli NC, Rutten-van Mölken M, Severens JL, Al M. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. PharmacoEconomics. 2019;37(11):1391-408.

Appendices

Appendix C: Summary of product characteristics (SmPC) and European public assessment report (EPAR)

Appendix D: Identification, selection and synthesis of clinical evidence

Appendix E: Subgroup analysis

Appendix F: Adverse reactions

Appendix G: Published cost-effectiveness studies

Appendix H: Health-related quality-of-life studies

Appendix I: Cost and healthcare resource identification, measurement and valuation

Appendix J: Clinical outcomes and disaggregated results from the model

Appendix K: Checklist of confidential information

Appendix L: Additional trial data from CASSIOPEIA

Appendix M: SLR/meta-analysis of MRD negativity and survival outcomes

Appendix N: Log-cumulative hazard plots

Appendix O: Survival models and outcomes for scenario analyses

Appendix P: Sensitivity analysis inputs

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Clarification questions

September 2020

File name	Version	Contains confidential information	Date
ID1510_Company Clarification Response_ACIC	1.0	Yes	09/10/2020

Notes for company

Highlighting in the template

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.

Section A: Clarification on effectiveness data

Decision problem

A1. The decision problem table (Company Submission (CS), Table 1) is incomplete. Please provide the full table according to the NICE template.

The full decision problem table according to the NICE template is provided in Table 1 below:

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	People with previously untreated MM who are eligible for ASCT	Adult patients with newly diagnosed MM who are eligible for ASCT	This population is considered to be in line with the full marketing authorisation for this indication
Intervention	DBTd	As per the final scope	Not applicable
Comparator(s)	• Bortezomib with dexamethasone (Bd) or with dexamethasone and thalidomide (BTd) • Bortezomib with cyclophosphamide and dexamethasone (BCd) (off-label) • Cyclophosphamide with thalidomide and dexamethasone (CTd) (off-label)	• Bd • BTd • BCd (off-label)	Janssen does not consider CTd a relevant comparator to DBTd in this indication following clinical expert feedback that CTd is rarely used as an induction therapy for NDTE MM patients in England.[1] Real-world evidence supports limited CTd usage, with steady decline in prescribing and less than 2% of NDTE MM patients in England treated with CTd since 2018.[2] Furthermore, CTd is not recommended by NICE, or recognised by international or European clinical practice guidelines.
Outcomes	The outcome measures to be considered include: • Overall survival • Progression-free survival • Response rate • Minimal residual disease-negative status • Proportion of people undergoing high dose chemotherapy and autologous stem cell transplantation • Adverse effects of treatment • Health-related quality of life	As per the final scope	Not applicable
Key: ASCT = autologous stem cell transplant; Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, dexamethasone and thalidomide; CTd = cyclophosphamide, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MM = multiple myeloma; NDTE = newly diagnosed transplant-eligible.			

Clinical effectiveness SLR methods

A2. The company submission states, “the inclusion criteria of the systematic literature review were limited to randomised controlled trials (RCTs)” (CS Appendix Table 9, page 18), with “non-RCT” being an exclusion criterion (CS Appendix Table 12). Please explain why the single arm studies JSCT-MM12 and Tanaka et al. 2019 were included, as stated in CS Appendix Tables 11 & 15 and CS Appendix pages 36 & 64.

The CS correctly states that the clinical SLR was limited to RCT evidence. As the original SLR (conducted, May 2018) failed to identity any direct or indirect evidence necessary to form a network for comparison of DBTd against Bd or BCd, for the SLR update, it was deemed relevant to include details of any single-arm Bd or BCd studies identified (but not originally reported) in addition to comparative RCTs.

Data analysis methods

A3. The company submission states the number of participants with a post-consolidation evaluation was n=459 for daratumumab in combination with bortezomib, thalidomide and dexamethasone (DBTd) and n=436 for bortezomib, thalidomide and dexamethasone (BTd) (CS Figure 7). This implies that 84 (15%) of DBTd and 106 (19%) of BTd patients had missing data for the post consolidation response outcomes stringent complete response (sCR) (CS Table 13) and minimum residual disease (MRD) negative status (CS Appendix Table 65) when analysed by intention to treat (ITT).

(a) Please clarify how the missing response and MRD negative status outcomes data were imputed to achieve the ITT analysis population for these outcomes.

(b) Were any sensitivity analyses conducted with different imputation methods to test the impact of missing data on outcomes?

As per the Statistical Analysis Plan (SAP) for CASSIOPEIA Part 1, post-consolidation sCR rate was defined as the percentage of ITT subjects who achieved or maintained sCR status within 30 days of Day 100 post-ASCT. If the patient did not have any post baseline disease assessments, they were categorised as post-

consolidation “not evaluable”. The number of not evaluable response assessments 100 days post-ASCT were 15 (2.8%) and 10 (1.8%) on the BTd and DBTd arms respectively, as shown in CS Appendix L (Table 63).

Post-consolidation MRD negative rate was defined as the proportion of subjects who had negative MRD at Day 100 post-ASCT. For those with missing Day 100 post-ASCT assessment, Cycle 4 Day 28 assessment was carried forward. For analysis purposes, patients in the ITT population with missing MRD results were assumed to be MRD positive. This represented [redacted] and [redacted] patients in the DBTd and BTd arms from CASSIOPEIA respectively.

No sensitivity analysis was conducted to test the impact of missing data on outcomes.

A4. The ERG have noted that the sample sizes for health-related quality of life (HRQoL) outcomes reported in CS Figures 20 to 23 are smaller than the number of patients who had a post-consolidation assessment (as indicated in CS Figure 7), after accounting for rates of compliance with the HRQoL instruments (CS Table 26).

(a) Please explain the missing data.

(b) The analysis of HRQoL appears to be based on available cases. Why was an ITT analysis not conducted for HRQoL outcomes?

(c) Were any sensitivity analyses conducted to investigate the impact of missing data on HRQoL outcomes?

As per the SAP for CASSIOPEIA Part 1, for each PRO endpoint, a mixed effects model with repeated measures analysis was conducted to estimate the change from baseline at each time point between DBTd and BTd.

ITT subjects who had a baseline value and at least one post-baseline value were included in the analysis. The difference in the sample sizes noted by the ERG therefore relate to patients who either did not have a baseline value recorded or at least one post-baseline assessment.

An additional PRO analysis was conducted to describe the missing data pattern over the timeframe of the mixed model analyses. The denominator for this analysis was the total number of subjects in the ITT population. Refer to Table 2 and Table 3 below.

Table 2: Missing data pattern of EQ-5D-5L assessments

Note: PRO assessments mapping to an analysis visit window are summarized (O = Missing, X = Present).

Source: Dara3006\stat\T7_1.sas
MMY3006 PRO ANALYSIS (CLINICAL CUT-OFF 19JUN2018) 01APR2019 11:03 1

Table 3: Missing data pattern of EORTC QLQ-C30 assessments

Note: PRO assessments mapping to an analysis visit window are summarized (O = Missing, X = Present).

Source: Dara3006\stat\T7_3.sas
MMY3006 PRO ANALYSIS (CLINICAL CUT-OFF 19JUN2018) 01APR2019 11:03 1

To further explore the impact of missing data, a pattern mixture model was conducted. A total of 100 iterations were performed on each of the EQ-5D-5L VAS and Utility Scores, and all EORTC QLQ-C30 subscales. In each iteration, a Markov chain Monte Carlo (MCMC) method was used to first impute missing data up to a monotone missing data pattern. Next, a propensity score method was employed to impute the remaining missing data. The complete data from each iteration was analysed using the same mixed-effects repeated measures model. Finally, LS mean estimates from each model were combined to yield a single estimate summarising the entire imputation process. The results from this analysis are presented below:

Table 1.7.2: Mixed Model for Change in EQ-5D-5L VAS and Utility Score: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_2.sas

Table 1.7.2: Mixed Model for Change in EQ-5D-5L VAS and Utility Score: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.
Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

Table 1.7.4: Mixed Model for Change in EORTC QLQ-C30 Subscale Scores: Average of 100 Imputations
Analysis Population: Intent-to-treat Subjects

Note: In each iteration, a Markov chain Monte Carlo (MCMC) method is used to first impute missing data up to a monotone missing data pattern. A propensity score method is then used to impute the remaining missing data.
LS Means are derived based on a mixed effects model with repeated measures, in which the dependent variable is change from baseline in score, and independent variables are baseline, visit, treatment, visit by treatment interaction, and randomization stratification factors - Site (HOVON vs. IFM), ISS staging (I, II, III), and cytogenic risk (Standard vs. High) as fixed effects and individual subject as random effect.

Source: Dara3006\stat\T7_4.sas

A5. Please clarify which stratification factors were used in the hazard ratio analyses for progression free survival (PFS) (CS Tables 17 & 19), TTP (CS Table 22) and overall survival (OS) (CS Tables 23 & 24). Please also clarify whether these included the randomisation stratification factors; ISS staging, cytogenic risks and site affiliation.

As per the SAP for CASSIOPEIA Part 1, PFS results (along with TTP and OS) were not stratified by the 3 randomisation factors or any other stratification factors. This was due to the expected small number of events for the time-to-event (TTE) analysis at the end of study follow-up for Part 1.

A6. The inverse probability weighting (IPW) adjustment to PFS accounting for potential bias due to the second randomisation is reported in CS Appendix section L.7.

(a) Please clarify why OS is not included.

(b) Please consider a sensitivity analysis using the weighted Kaplan-Meier method with time-dependent weights as opposed to fixed weights (Miyahara, 2010).

As per the SAP for CASSIOPEIA Part 1, due to the anticipated low number of events and immaturity of survival data, no alpha spend was allocated to the statistical analysis of OS. As referred in Section B.2.6.2 of the CS, to protect the integrity of CASSIOPEIA Part 2, the IPW analysis for PFS was conducted by a sequestered group independent from the Janssen study team. A similar IPW analysis for OS was not requested or published as part of the EMA regulatory process and results from such analysis have not been shared with the Janssen study team.

As Part 2 of the study remains blinded, Janssen does not have access to the patient-level data necessary to perform the requested sensitivity analysis for PFS.

A7. The CS states there were two versions of the international staging system (ISS) (CS Tables 6 & 7).

(a) Please clarify which version of the ISS was used:

- (i) as the stratification factor in the statistical tests
- (ii) in the subgroup analyses (CS Appendix Figures 36 to 40)
- (iii) in the MAIC analyses (CS section B.2.9.3)
- (iv) in the comparison of PHE and CASSIOPEIA data (CS Table 56)

As per the SAP for CASSIOPEIA Part 1, the original version of the ISS stage (as opposed to the revised version, R-ISS) was used as a stratification factor within the permuted block randomisation and as a stratification factor in the statistical tests for the primary endpoint (sCR). As noted in our response to question A5, statistical analysis for the TTE endpoints in Part 1 were not stratified by ISS (or any other factor) due to the expected low number of events.

The revised ISS (R-ISS) criteria was first published in 2015, after initiation of the CASSIOPEIA study and after publishing the results from the IFM-2005-01 and GMMG-MM5. Therefore, the R-ISS staging criteria could not have been used in the IFM-2005-01 and GMMG-MM5 trials and matching of patient baseline characteristics was instead performed using the original (ISS) definition.

In terms of the PHE datasets, the original ISS was specified up to and including version 8 of the Cancer Outcomes and Services Dataset (COSD), with R-ISS only introduced following release of version 9, effective from 1st April 2020. The follow-up for the PHE analysis presented in the CS was to 31st December 2019, and therefore reflects the original ISS definition consistent with the results reported from both CASSIOPEIA and the MAICs.

(b) The ERG noted that recalculation of ISS altered the baseline risk categories for both DBTd and BTd (CS Tables 6 & 7). Please explain the implications of this for the results of analyses that included ISS as a stratification factor.

Recalculation of staging at baseline for DBTd and BTd using the R-ISS suggests a poorer overall prognosis for patients recruited into the daratumumab treatment arm of the study. Whilst the proportion of patients classified as grade III was comparable between arms (DBTd: 9.2%; BTd: 9.3%), the daratumumab arm included a greater proportion of patients classified as grade II (DBTd: 71.6%; BTd: 63.7%) and fewer classified as grade 1 (DBTd: 19.3%; BTd: 27.0%) compared to BTd.

Results are not available using R-ISS as a stratification factor (instead of ISS) however, as referred to in Section B.2.13 of the CS, the impact of this imbalance is likely to bias clinical efficacy results against DBTd in favour of BTd. In other words, the direction of any unresolved selection bias following randomisation using ISS (instead of R-ISS) is against DBTd. In this regard, the relative efficacy of DBTd

versus BTd reported in CASSIOPEIA in terms of rate of response (measured by conventional IMWG and MRD), and treatment effect in terms of the reported hazard ratios for PFS/PFS2/TTP/OS, are likely to represent a conservative estimate.

A8. The CS Table 7 shows eight patients were missing from the DBTd arm and two were missing from the BTd arm when the revised ISS was calculated. Please explain this discrepancy.

The revised ISS (R-ISS) introduced in 2015 includes an assessment of chromosomal abnormality (CA) detected by interphase fluorescent in situ hybridization (iFISH) and lactate dehydrogenase (LDH), in addition to B2-microglobulin and serum albumin (required for standard ISS assessment).[3]

The difference in the number of patients for the post hoc calculation of R-ISS compared to the ITT population noted in Table 7 of the CS included 8 patients on the DBTd arm missing an LDH test result at baseline compared to 1 patient on the BTd arm, which also included 1 patient without a baseline assessment for CA (i.e. 2 patients in total on the BTd arm missing R-ISS assessment).

Landmark analysis

A9. The number of patients at risk at baseline in the landmark analysis Kaplan-Meier graphs (CS Figures 17 and 18) differ from the sample sizes quoted in the economic model for the BTd arm (CASSIOPEIA RESPONSE MRD tab).

(a) Please clarify which figures are correct. If there are data missing from the landmark analyses please explain why.

(b) Please provide Kaplan-Meier data used for the DBTd OS and PFS landmark analysis, in the same format as for the BTd arm in the CASSIOPEIA RESPONSE MRD tab of the economic model.

Janssen confirm that the number of patients at risk have been misrepresented in the model. The numbers at risk at baseline in the landmark analyses for the BTd arm should be as per the CS for both PFS (MRD-, n=[redacted]; MRD+, n=[redacted]) and OS (MRD-, n=[redacted]; MRD+, n=[redacted]) (Figure 1 and Figure 2 below). Janssen apologise for the confusion and can confirm that no data are missing from the landmark analyses. Janssen also confirm that amending the numbers in the economic model has no impact on the results presented as part of the CS.

The Kaplan-Meier data used for the landmark analysis of DBTd OS and PFS are presented in the Excel spreadsheet ('DBTd Landmark Analysis_KM Data') provided alongside this response.

Figure 1: Landmark analysis: PFS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.

Figure 2: Landmark analysis: OS from post-consolidation assessment by treatment arm and MRD status at the time of the post-consolidation assessment (median follow-up = 29.2 months)

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; GPM = general population mortality; MRD = minimal residual disease; OS = overall survival.

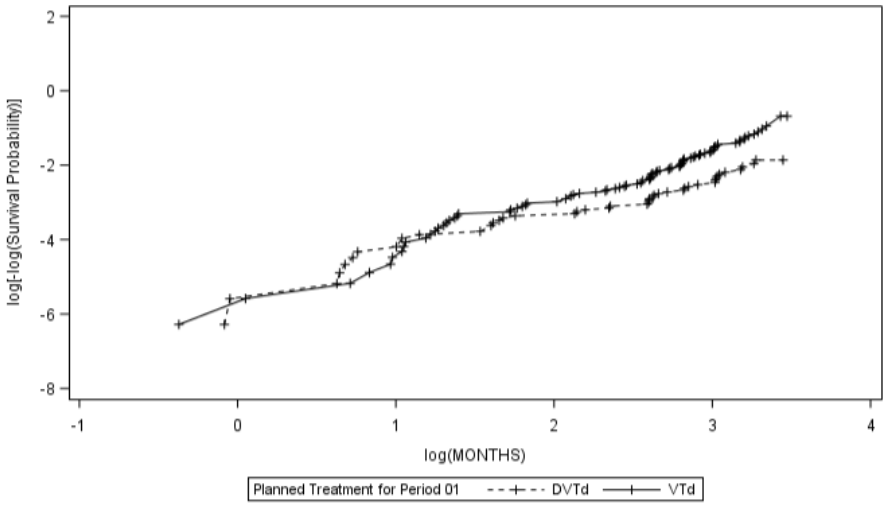
A10. The numbers at risk in the landmark analysis Kaplan-Meier graphs (CS Figures 17 and 18) are identical for OS and PFS in the MRD- group but differ between OS and PFS in the MRD+ group. Please clarify if this is because deaths or progression occurred only in the MRD+ group prior to day zero of the landmark analysis.

Janssen confirm that the above interpretation is correct. The numbers at risk in the landmark analyses are identical for PFS and OS in the MRD- group but differ between PFS and OS in the MRD+ group because all death or progression events that occurred pre-landmark analysis, occurred in MRD+ patients.

A11. Please explain how the assumption of proportional hazards was tested for the Cox analyses and survival analyses reported in CS sections B.2.6.2 and B.2.6.3.

To investigate the hazard ratio across different phases of treatment, the proportional hazards (PH) assumption was tested for PFS analysis by log-log plot (Figure 3).

Figure 3: Log-log plot for PFS (ITT population, median follow-up = 18.8 months)[4]

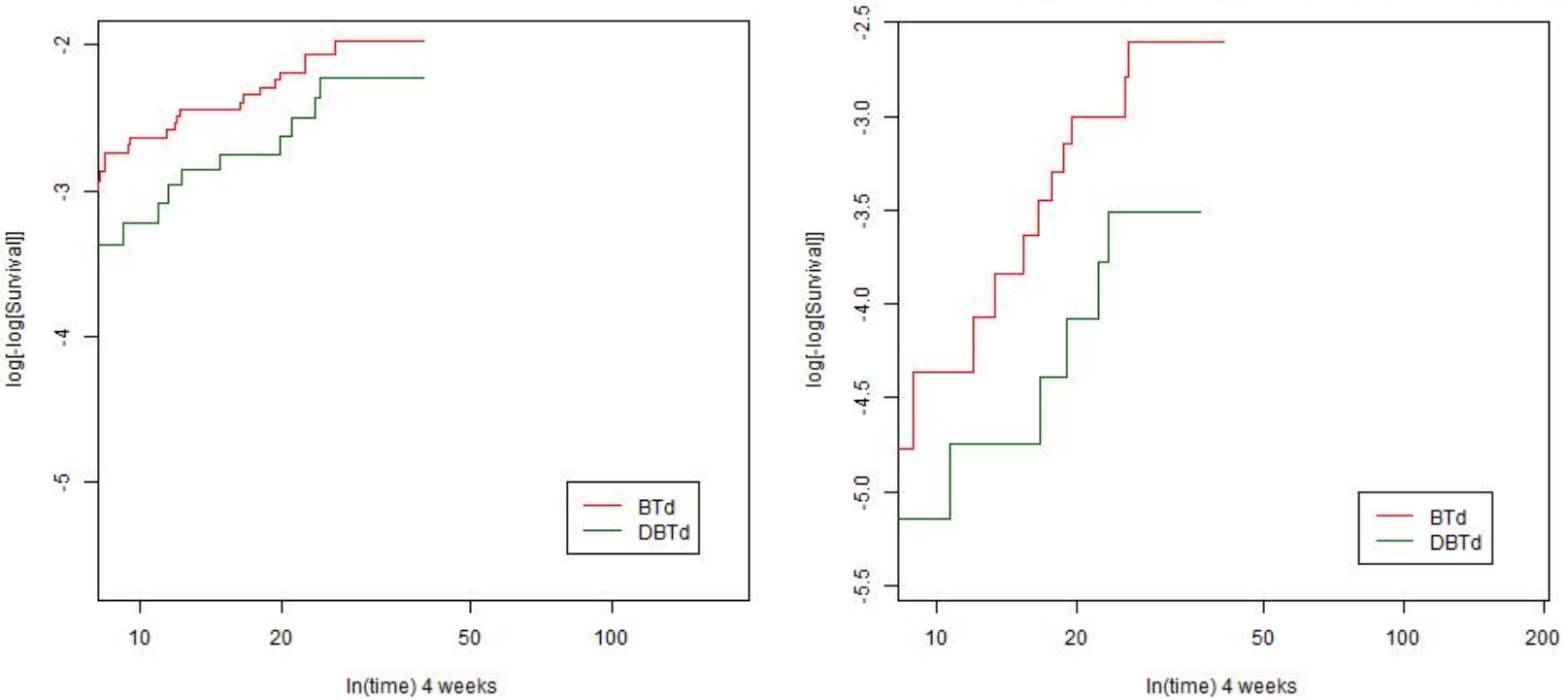


Key: VTd = bortezomib (VELCADE) + thalidomide + dexamethasone; DVTd = daratumumab + bortezomib (VELCADE) + thalidomide + dexamethasone.
[GEFPFS01D_3006.RTF] [JNJ-54767414\MMY3006\DBR_PART1_CSR\RE_PART1_CSR\PROD\GEFPFS01D_3006.SAS] 27NOV2018, 09:29

The two curves overlap at early timepoints indicating violation of the PH assumptions. To obtain further insight on PFS benefit across different treatment phases, a piecewise hazard ratio by study phase was conducted with the point estimates for HRs in each of the different treatment phases indicating a benefit for DBTd compared to BTd (refer to Section B.2.6.2 of the CS). As per the SAP for CASSIOPEIA Part 1, the PH assumption was not tested for the OS results reported in CS Section B.2.6.2.

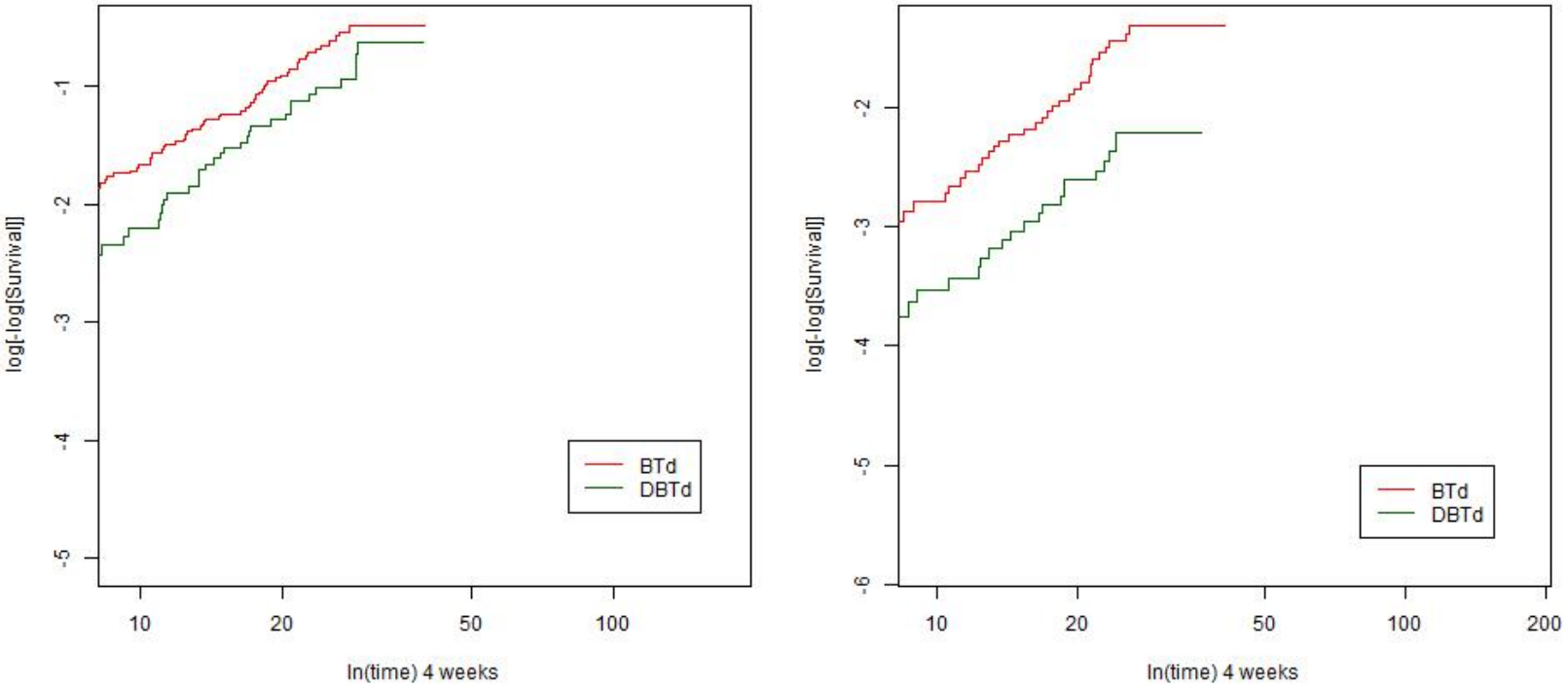
As presented in Section B.2.6.3 of the CS, the landmark analyses used Cox proportional hazard models to determine the treatment effect for DBTd and BTd in terms of OS and PFS in both the MRD+ and MRD- groups. The validity of the proportional hazards assumption between the treatment arms was tested by visual examination of the log-cumulative hazard plots where convergence (or crossing) of the two curves was considered to be evidence of a violation of the PH assumption, as recommended in Technical Support Document (TSD) 14. As shown in Figure 4 and Figure 5 (Appendix N of the CS), the curves remain parallel for both OS and PFS in the MRD+ and MRD- groups, indicating no violations of the proportional hazards assumption.

Figure 4: Log-cumulative hazard plot for OS (BTd vs DBTd) for MRD+ (left) and MRD- (right)



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD: minimal residual disease; OS: overall survival.

Figure 5: Log-cumulative hazard plot for PFS (BTd vs DBTd) for MRD+ (left) and MRD- (right)



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD: minimal residual disease; PFS = progression-free survival.

Matched-adjusted indirect comparisons (MAICs)

A12. The population baseline characteristics for the studies included in the MAICs are listed in CS Table 32. Please clarify if there are any prognostic factors for multiple myeloma (MM) that are not included in this table?

Clinical expert feedback received by Janssen indicates that all important prognostic factors for MM are included in the baseline characteristics listed in Table 32 of the CS, and were therefore considered for inclusion in the MAIC analysis[i]. Extramedullary disease was noted as not being reported (it requires whole body MRI imaging which was not routine clinical practice when IFM-2005-01 and GMMG-MM5 were conducted) however clinical expert feedback obtained by Janssen suggests that its omission is not anticipated to impact the reliability of the MAIC as it only affects a small proportion of patients (approximately 10% in clinical practice).[i]

A13. Please provide the median OS and PFS for each of the IFM 2005-01 and GMMG-MM5 studies following digitisation by the Guyot method.

Below are presented the published statistics for the PFS and OS curves for bortezomib-dexamethasone (Bd) and bortezomib-cyclophosphamide-dexamethasone (BCd) and the ones obtained using the re-created individual patient data (IPD) by the Guyot method including overlay plots. The GMMG-MM5 trial did not report the median progression-free survival (PFS) and overall survival (OS) for BCd; instead, it reported the 3-year rates and that is what was compared with the re-created IPD. Red curves on the IFM 2005-01 overlay plots and pink curves on the GMMG-MM5 overlay plots represent the re-created IPD. These show a close reproduction of the published data. Potential limitations causing deviations between the published and re-created medians include the lack of published number of patients at risk in each interval by the IFM 2005-01 trial. Similarly, while the GMMG-MM5 reported the number of patients at risk in each interval, it did not report the number of events in each arm. These are important inputs

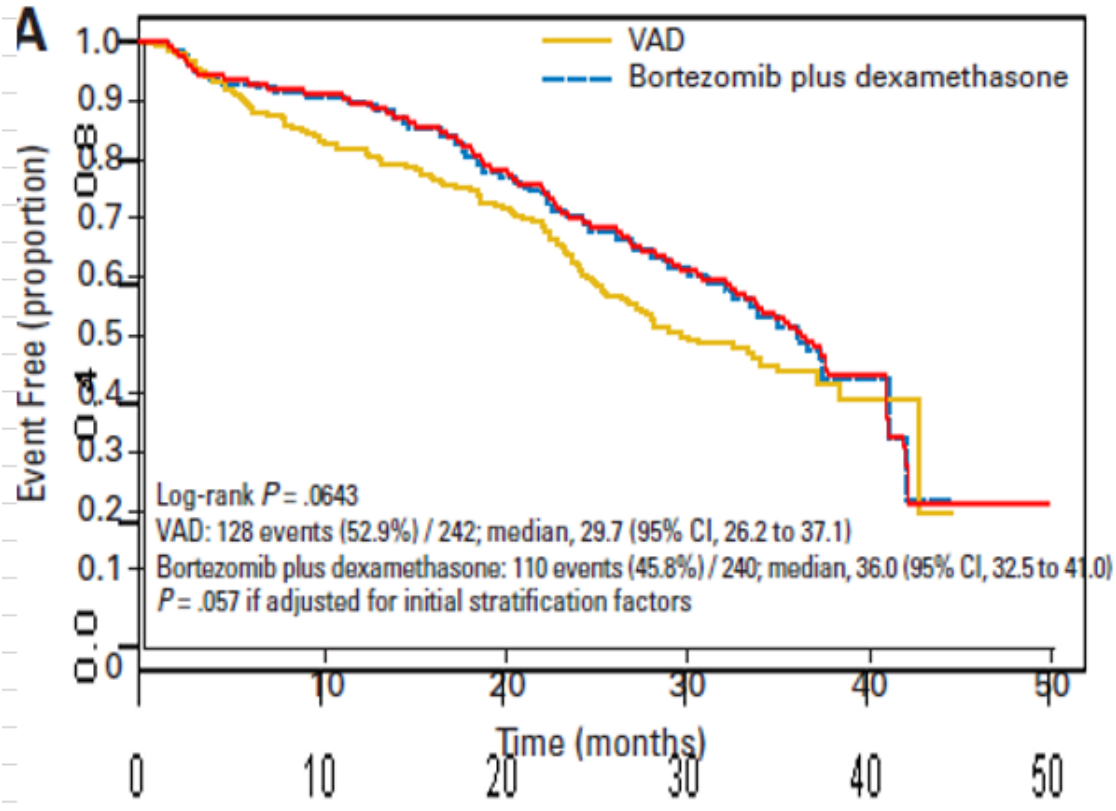
[i] Telephone/email correspondence between Dr Karthik Ramasamy and Janssen was received between 30th September and 1st October 2020.

for the Guyot method that could further improve the re-creation of the IPD if they were published.

PFS - Bd (IFM 2005-01)

Bd	N	Median (95 % CI)
Published	240	36.0 (32.5; 41)
IPD (Guyot method)	240	36.7 (33.3; 41.1)

Abbreviations: IPD=Individual patient data; CI=Confidence interval

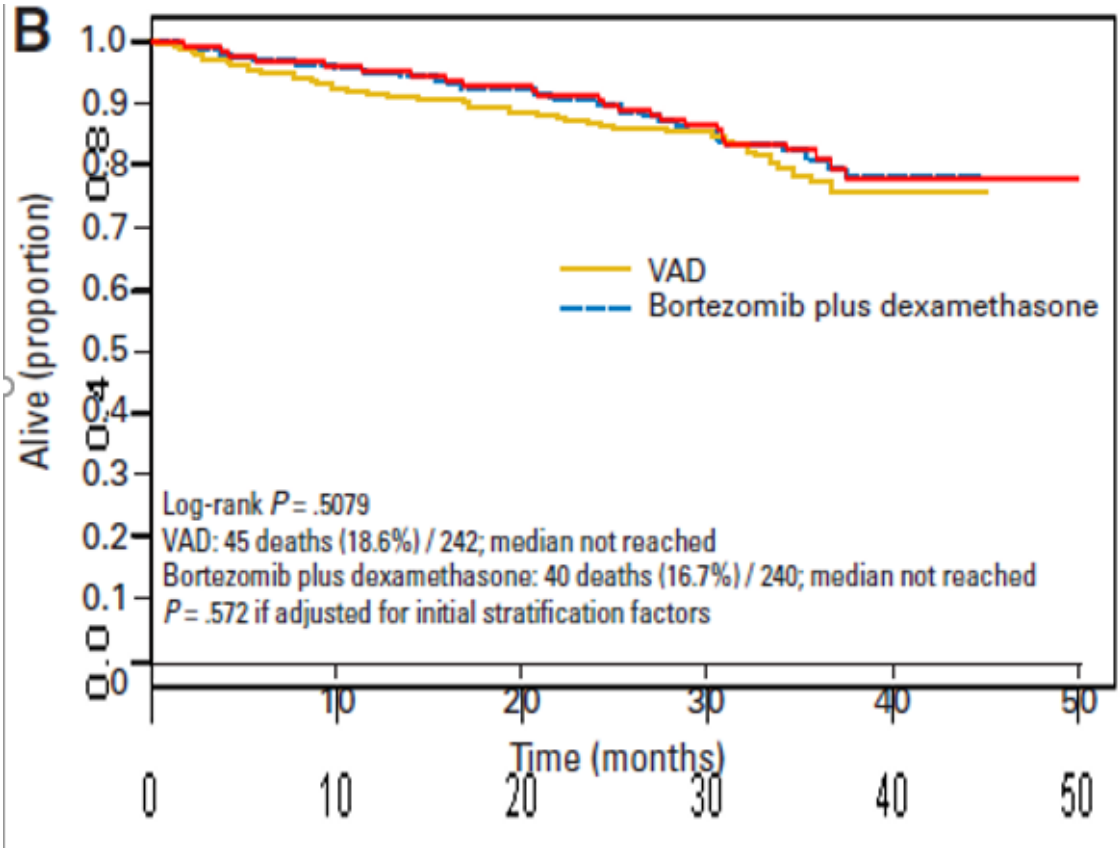


Source: Harousseau et al. 2010

OS- Bd (IFM 2005-01)

Bd	N	Median (95% CI)
Published	240	NE (NE; NE)
IPD (Guyot method)	240	NE (NE; NE)

Abbreviations: NE=Not Estimable; IPD=Individual patient data; CI=Confidence interval

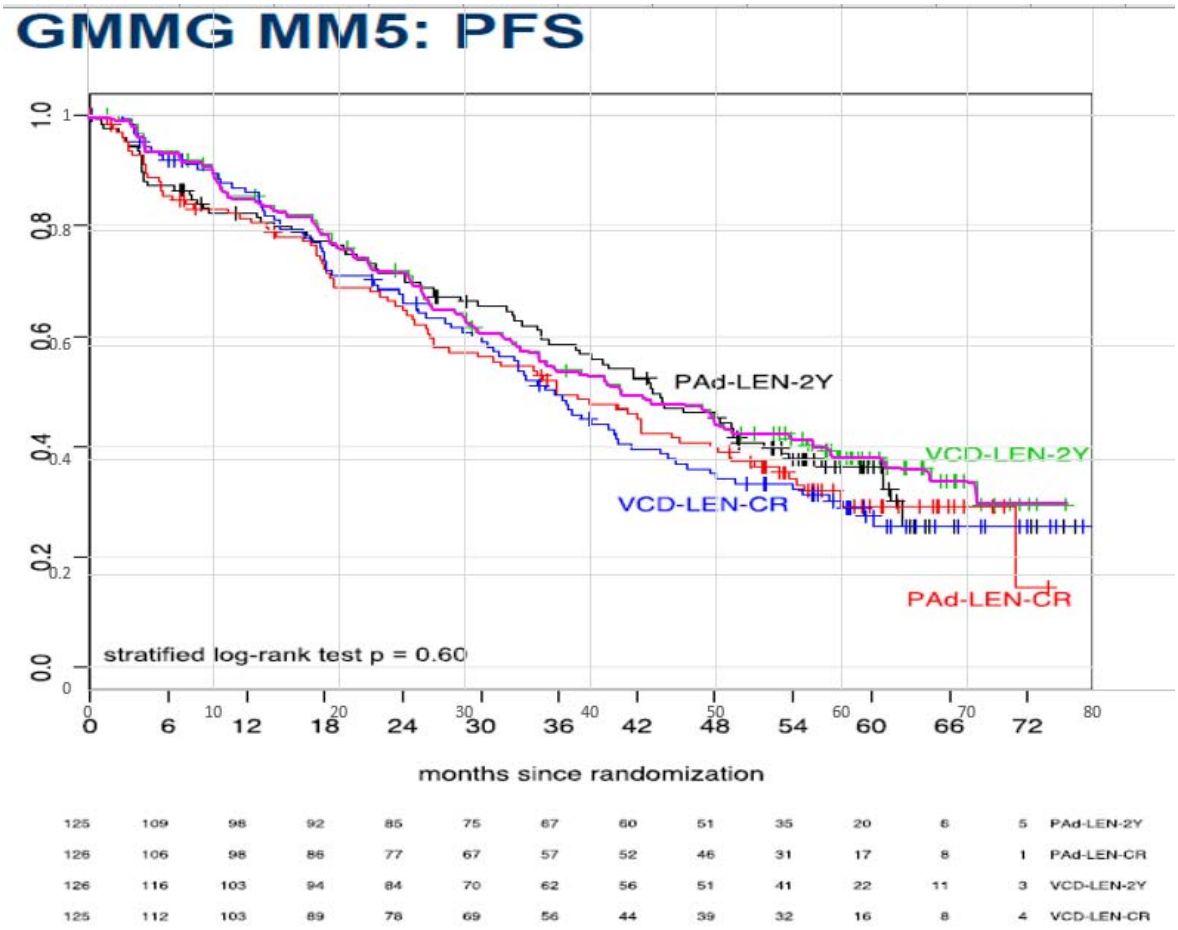


Source: Harousseau et al. 2010

PFS - BCd (GMMG-MM5)

BCd	N	Median	3-year rate
Published	126	NR	54%
IPD (Guyot method)	126	44.8	57%

Abbreviations: IPD=Individual patient data; NR=Not reported

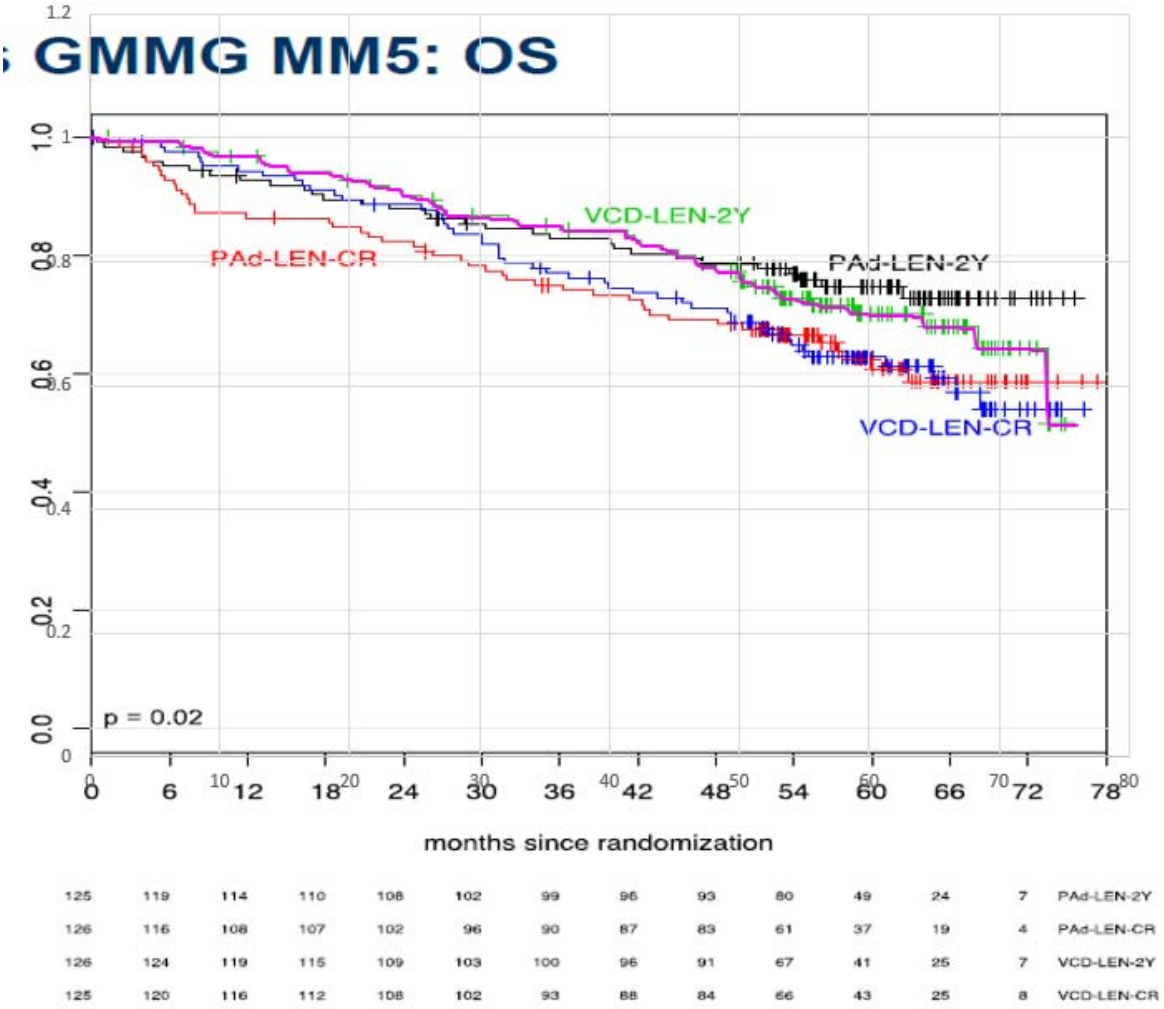


Source: Goldschmidt et al. 2017

OS - BCd (GMMG-MM5)

BCd	N	Median	3-year rate
Published	126	NR	85.2%
IPD (Guyot method)	126	NE	85.9%

Abbreviations: NR=Not reported; NE=Not Estimable; IPD=Individual patient data



Source: Goldschmidt et al. 2017

A14. Please explain the process for acquiring and extracting the Public Health England (PHE) data that are reported in "Janssen [Data on File] PHE Results tables.xlsx". Please clarify if the extracted data were checked for accuracy against the PHE source, and if so how?

The process for acquiring and extracting the PHE data reported in the CS is broadly provided in the PHE final report with the key points summarised below.

The PHE analysis was a descriptive, non-interventional study that used routine population-level data available through Public Health England's National Cancer Registration and Analysis Service (NCRAS) to identify, and subsequently track

outcomes for, a standing cohort of newly diagnosed MM patients.[5] Patients were selected into the cohort if they met the following criteria:

- Resident in England on the date of diagnosis;
- Aged 18 years or above on the date of diagnosis;
- Have an incident primary diagnosis of NDMM, defined according to the International Classification of Disease of Oncology 3rd edition (ICD-O-3) morphology code 9732 (multiple myeloma, myelomatosis, plasma cell myeloma and myeloma not otherwise specified). This definition was applied based on the recommendation of a Public Health England (PHE) pathologist.
- At least one cohort-relevant diagnosis between 01/01/2015 and 31/12/2018, inclusive.

Given the inclusion criteria above, patients or their corresponding tumour(s) were omitted from the cohort if:

- No recorded date of diagnosis;
- Patient age was missing at diagnosis, or the patient was aged <18 or >122 years at the first cohort-relevant diagnosis;
- There were known data quality issues with patient vital status, such as a diagnosis occurring after the date of death;
- The patient was flagged as being in receipt of a CDF-listed drug indicated for myeloma but not funded by Janssen at the time of analysis.
- The tumour was diagnosed via death certificate only.

Data related to patient baseline demographics and disease characteristics were obtained directly from the English national cancer registry. Patients were defined as HDT-ASCT eligible if they were documented as having received an autologous or

allogeneic peripheral blood stem cell transplant (identified using OPCS Classification of Interventions and Procedures version 4 (OPCS-4) codes X334 and X336, respectively) between their first cohort-relevant diagnosis and the end of follow-up.

For patients with linkage to SACT, lines of therapy were derived according to an algorithm outlined in the project protocol. Briefly, the algorithm selected on regimens that (i) contained at least one drug specifically indicated for the treatment of myeloma and (ii) were recorded in SACT as being delivered to a patient for the treatment of a C90 tumour. Changes in line were then defined according to pre-specified changes in the composition of drugs/drug classes between consecutive regimens and treatment-free interval duration.

Overall survival (OS) and progression-free survival (PFS) were estimated using dates of death contained within the English national cancer registry via the Office for National Statistics (ONS). For OS, time-at-risk was defined from the start of first-line therapy through to the date of death or censoring, where censoring was defined as the date of embarkation (moving outside of England) or the end of follow-up, whichever occurs first. Time-at-risk for PFS was defined from the start of first-line treatment through to the date of disease progression. Progression is typically defined by tumour growth, increased invasiveness or metastasis. As with lines of therapy, such information is not routinely available within the cancer registry. Accordingly, disease progression was defined as a change in treatment line or death, whichever occurs first.

The analysis was conducted by an experienced Senior Analyst working at PHE, familiar with the NCRAS linked datasets. Quality assurance of coding and extracted results was subsequently performed by a PHE lead for haematology, with a strong clinical background in myeloma. Data extraction was carried out within the NCRAS Cancer Analysis System using SQL, and descriptive analysis was completed using a combination of SQL and Stata.

A15. Please explain why serum lactate dehydrogenase (LDH) was not included in the matching with bortezomib, cyclophosphamide and dexamethasone (BCd). The proportion of patients with ≤ULN or >ULN serum LDH is unbalanced between GMMG-

MM5 and the other studies (CS Table 32). Please provide a sensitivity analysis including serum LDH in the matching.

The proportion of patients with LDH ≤ULN (82.5%) or >ULN (17.5%) in GMMG-MM5 was matched in both the base-case and sensitivity analysis 1 in the MAIC versus BCd. The classification of ≤ULN and >ULN represents a binary outcome with the total number of patients in GMMG-MM5 summing to 100%. By matching for the proportion of patients >ULN, ≤ULN was, by definition, also matched.

In CASSIOPEIA LDH was based on local lab, whereas in the GMMG-MM5 trial, it was not reported whether LDH was based on local or central lab. In CASSIOPEIA, patient-dependent cut-offs of 213 U/L or 225 U/L were used for defining the ULN for LDH. The cut-off used to define ULN in GMMG-MM5 trial was not reported in the publication. Therefore, considering that there might be potential differences in the definitions of the ULN between studies, a second sensitivity analysis (i.e. sensitivity analysis 2), excluding LDH was performed.

A16. There is little variation in the MAIC base case, sensitivity analyses, and naïve comparison. The ERG disagrees that there is a clear rationale for the use of the MAIC over the simulated treatment comparison (STC) methodology. There is a lack of overlap in certain characteristics such as renal insufficiency and creatinine, and the distribution of weights in CS Appendix D.1.7 shows a large number of subjects with zero weights. Please provide a STC as a scenario analysis.

There were only 2 and 1 patients in the BTd and DBTd arms respectively with renal insufficiency in CASSIOPEIA (May 1st, 2019 CASSIOPEIA data cut). Therefore, this baseline characteristic cannot be adjusted for in either an MAIC or STC. Creatinine and LDH were matched in the base-case MAIC scenario for the comparison with GMMG-MM5.

The implementation of STC requires derivation of a predictive equation using parametric survival methodology. The development of an equation would require, in general, at least 8 events per baseline covariate added to that equation. Adjusting for all 12 covariates (many of which are categorical with more than 2 categories per covariate)

included in the base-case MAIC for the GMMG-MM5 trial would therefore require at least 96 events per outcome. Thus, an STC may not be feasible for some of the comparisons due to the immaturity of the survival outcomes in CASSIOPEIA. In addition, the implementation of an unanchored STC would require simulation of comparator-like trial data (since pseudo-IPD must be used for predicting OS and PFS in comparator- like population). This is because the efficacy outcomes of interest are non-linear (i.e. OS and PFS are survival outcomes) and the impact of performing an unanchored indirect comparison on a different scale than that of the linear predictor (which is the case here with survival outcome) is introducing extra complexities and the impact of these on the bias are not yet fully known (see NICE DSU TSD 18, sections 2.3.2 and 2.3.3). Consequently, estimation of the standard errors of the effect estimates using bootstrapping techniques would be required.

Given this, and whilst acknowledging the limitations of the MAIC methodology, Janssen does not consider the STC as a suitable alternative method.

A17. Please provide the R code and input data for the MAIC analyses.

The analysis for the MAICs was conducted in SAS 9.4. Programs used for the comparison with Bd and BCd are provided in the attached folder, “SAS code for MAIC vs. IFM2005-01 and GMMG-MM5”.

Regrettably, Janssen is unable to provide the requested individual patient-level data from CASSIOPEIA Part 1 used as input for the MAIC due to company policy.

Section B: Clarification on cost-effectiveness data

SLR of MRD status on survival outcomes

B1. Please explain the relationship between the following 3 systematic literature reviews (SLRs). Please explain how the methods differed between these SLRs and why:

- The SLR reported in CS section B.3.3.2 and CS Appendix M
- The SLR reported in the abstract by Munshi et al 2019 [reference 63]
- The SLR reported in the journal paper by Munshi et al. 2017 [reference 61]

The original SLR for MRD was reported by Munshi et al. 2017 and included both transplant eligible and ineligible studies (controlled studies, randomised controlled studies or patient cohort studies). An updated and expanded SLR of the evidence supporting the prognostic utility of MRD in myeloma was conducted by Munshi et al. 2019 which considered both randomised controlled trials and observational studies, including newly diagnosed patients and relapsed/refractory MM.

The SLR reported in CS section B.3.3.2 and CS Appendix M refers to the most recent 2019 SLR. As referred in the CS, for the purposes of the DBTd submission, additional screening was performed to update the meta-analysis results to only include studies that met the following criteria:

- Studies where transplant was performed
- Studies where MRD was measured at 100 days post-SCT
- Studies representing standard of care

In addition, the analysis was updated with the latest data cut of CASSIOPEIA representing a median follow-up of 29.2 months (BTd arm only).

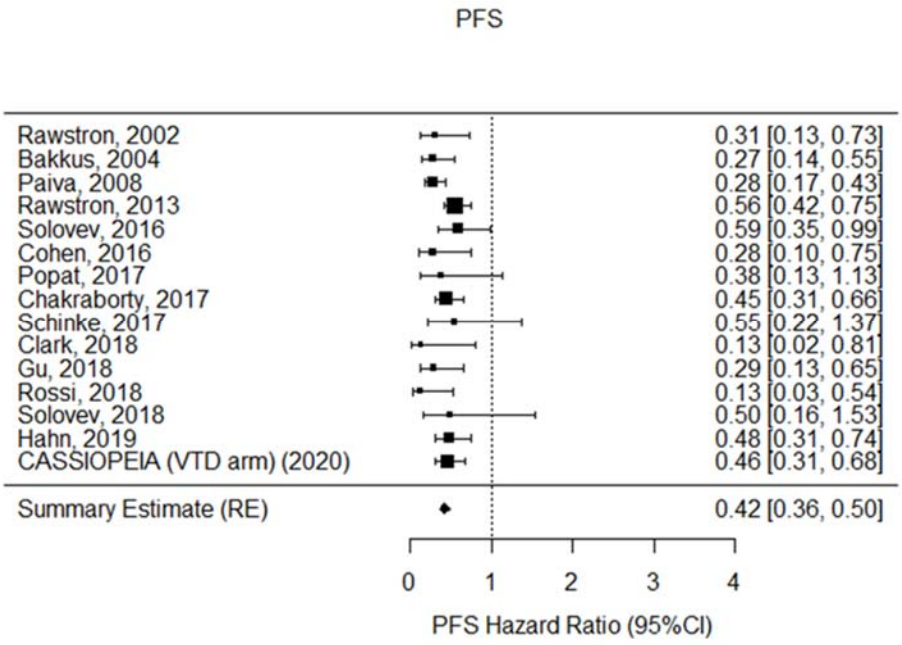
B2. Priority question. CS section B.3.3.2 states "the SLR/meta-analysis included a number of older trials which do not capture the shift in outcomes for MM patients due to the introduction of novel agents as well as trials with a range of MRD sensitivity thresholds, including 10^{-4}."

(a) Please clarify which are these "older trials"? Please test the influence of these trials in a sensitivity analysis.

(b) The date range for studies included in the expanded SLR goes back to 2002 (CS Appendix Tables 85 and 86). This date range is the same as for the original SLR reported by Munshi et al 2017 [reference 61]. The ERG note that expanding the SLR has not solved the problem of including unrepresentative older trials. Please explain this.

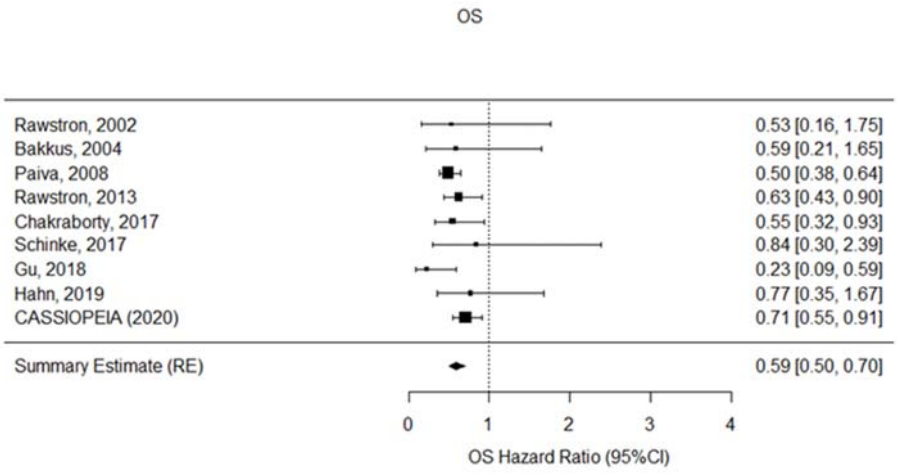
To address potential chronicity bias due to the inclusion of older studies, Janssen qualitatively assessed if there is any trend in the reported PFS/OS HRs when ordered by publication date. The forest plots in Figure 6 and Figure 7 present the reported HRs for each study from older to more recent with no clear trend observed.

Figure 6: PFS HRs by MRD status, by publication year



Note: MRD HR <1 favours MRD-negative status

Figure 7: OS HRs by MRD status, by publication year



Note: MRD HR <1 favours MRD-negative status

To further investigate the potential impact of older trials, a sensitivity analysis was conducted to exclude studies older than 2015 (arbitrary 5-year cut-off). The analysis for PFS included 11 studies[6,7,8,9,10,11,12,13,14,15,16], and resulted in a HR of 0.44 (95% CI: 0.37-0.53, SE = 0.10, p<0.001). The analysis for OS included five studies[10,11,14,15,16], and resulted in a HR of 0.63 (95% CI: 0.49-0.82, SE=0.13, p<0.001). Both OS and PFS results are in line with the main analysis, which indicates that the chronicity bias does not substantially impact the outcomes.

Finally, Janssen also investigated whether older trials may introduce bias due to the use of a lower sensitivity threshold (10^{-4}). For this, analysis was limited to only include studies where MRD-negativity was defined at the threshold of 10^{-5} (no study with the threshold of 10^{-6} was available). The PFS meta-analysis included five studies[10,13,14,15,16], and resulted in a HR of 0.45 (95%CI: 0.35-0.59, SE=0.13, p<0.001). The analysis for OS included four studies[10,14,15,16] and resulted in a HR of 0.62 (95%CI: 0.40-0.96, SE=0.22, p=0.03).

In conclusion, the impact of chronicity bias for the studies included in the MRD meta-analysis appears limited in terms of relative effects. The impact of older studies or studies with lower MRD sensitivity threshold on absolute survival outcomes, however, remains unclear.

B3. <u>Priority question</u>. The searches in the SLR of MRD status on survival outcomes are 15 months out of date (this is inconsistent with the other SLRs provided in CS Appendices D, G, H, I which were updated more recently). Please update the searches or provide a clear justification that all relevant studies have been identified.

Due to time constraints, Janssen has initially prioritised screening all RCTs identified in the clinical SLR update which were published after May 2019 (refer to CS, Appendix D). A list of the studies screened is provided in Table 4 below.

Table 4: List of publications identified in the SLR in newly diagnosed transplant-eligible MM published after May, 2019

Publication	Trial ID	Was the trial previously captured in the original SLR?	Is MRD reported?	Does the publication present a HR or a KM curve reporting the survival endpoints stratified by MRD?
Moreau 2019[17]	CASSIOPEIA MMY3006	Yes		
Voorhees 2020[18]	GRIFFIN	No	Yes	No
Luoma 2019[19]	FMG-MM02 (NCT01790737)	No	Yes	No (MRD-negative stratified into sustained negative and unsustained negative. Patient numbers are not reported so the curves cannot be pooled to produce a single MRD-negative curve).
Rosiñol 2019[20]	GEM2012MENOS 65	Yes		
Sunami 2019[21]	JSCT-MM12	No	No	No
Tanaka 2019[22]	NR	No	No	No
van de Donk 2018[23]	HOVON-50 (NTR238)	No	No	No
Horvath 2019[24]	VCAT study	No	Yes	No
Gregersen 2018[25]	CLAIM (NCT02573935)	No	No	No
Hulin 2019[26]	CASSIOPEIA MMY3006	Yes		
Avet-Loiseau 2019 [27]	CASSIOPEIA MMY3006	Yes		
Moreau 2019[28]	CASSIOPEIA MMY3006	Yes		
Yong 2019[29]	Cardamon study	No	Yes	No
Jackson 2019[30]	Myeloma XI Trial	No	No	No
Roussel 2019[31]	IFM 2014-03	No	No	No
Voorhees 2019[32]	GRIFFIN	No	Yes	No

Moreau 2019[33]	CASSIOPEIA MMY3006	Yes		
Shuang 2019[34]	NCT02577783	No	No	No
Kumar 2019[35]	REF/2016/08/012 008	No	Yes	No
Sonneveld 2019[36]	CASSIOPEIA MMY3006	Yes		
Pawlyn 2019[37]	Pawlyn 2019	No	No	No
Gay 2019[38]	FORTE	No	Yes	No
Scheid 2019[39]	HOVON-65/GMMG-HD4	No	No	No
Olivia 2019[40]	FORTE	No	Yes	No

None of the identified RCT studies reported survival outcomes by MRD status. Janssen intend updating the search results to include eligible non-RCT studies published after May 2019. Results from this review will however not be available before 31st October 2020 based on a preliminary search which provided around 300 hits plus manual search in this year's conference proceedings.

B4. Priority question. CS Appendix Table 84 does not report any selection criteria specifically for patients with newly diagnosed transplant eligible multiple myeloma (NDTE MM). Please explain how studies on these patients were identified.

Table 5 describes the complete list of eligibility criteria for the MRD SLR in MM in all disease settings and further selection criteria relevant to the CS.

Table 5: *MRD SLR eligibility criteria*

	Inclusion Criteria	**Exclusion Criteria**
SLR on survival outcomes by MRD in MM		
Population	Patients with MM	Patients without a primary diagnosis of MM
Intervention/comparator	Any treatment	Allo-SCT
Outcomes	OS and/or PFS stratified by MRD status (using any MRD definition) Any PRO, TTP or PFS2 reported by MRD status	Survival data that cannot be extracted or is not available MRD measured in peripheral blood (PB)

		MRD assessed by PET-CT
Study Design	RCTs and non-RCT study design	Economic models, case reports, comments and editorials, animal/in-vitro studies
Date Limit	• No date limit applied on indexed databases search • Conference abstract and other materials (grey literature): 3 years (2016-2019): EHA, ASH, ISPOR, ASCO. • SLRs: 5 years (2014-2019)	• Conference abstract or other materials (grey literature) published before 2016 • SLRs published before 2014
Language	English language	Non-English language
Additional selection criteria		
Population	Newly diagnosed transplant-eligible patients	Relapsed/refractory patients, transplant-ineligible patients.
Intervention	• Transplant • Treatments representing standard of care (SoC)	• Transplant was not performed • Studies with D-VTd (CASSIOPEIA D-VTd arm)
Outcomes	MRD measured at 100 days post-SCT	MRD measured at a different timepoint

B5. <u>Priority question</u>. Please provide a list of all studies that were excluded at full-text screening with the reason(s) for exclusion. Please include:

- The studies included in the 677 excluded publications referred to in CS Appendix Figure 61
- The studies that were excluded because no hazard ratio or Kaplan-Meier plot was available for overall survival (these studies were reported in 57 publications, according to the text at the start of CS Appendix section M.2)
- Any further excluded studies (see question B6)

Please see the excel table attached for a list of all studies excluded at full-text screening. Note that some studies that did not have hazard ratio or Kaplan-Meier curves

by MRD status were nevertheless included if they reported on the same trial as another included publication and provided background information (e.g., patient characteristics or details of MRD assessment protocol).

B6. Priority question. CS Appendix M.2 states that 45 studies provided hazard ratios or Kaplan-Meier plots which could be utilised in the meta-analysis. However, only 15 studies were used for analysis of PFS, of which 9 were used for analysis of OS (CS Appendix Tables 85 and 86). Please explain why 30 of the 45 studies are not accounted for. Please list these 30 excluded studies in the response to question B5).

The 45 studies include a mix of newly diagnosed transplant-eligible, ineligible, and relapsed/refractory patient populations with various treatments and MRD measured at different time points. Further, 30 studies were excluded based on the additional selection criteria described in question B4. Table 6 lists those studies.

Table 6. Excluded studies that were used in the primary analysis but excluded from the analysis submitted to NICE

Study	Study ID	Reason for exclusion
Ferrero, 2015[41]	GIMEMA VEL-03-096	MRD not measured at 100 days post-transplant
Korde, 2015[42]	NCT01402284	No transplant
Korthals, 2012[43]	NA	MRD not measured at 100 days post-transplant
Paiva, 2015[44]	NA	Not newly-diagnosed transplant-eligible population
Martinez-Lopez, 2017[45]	PETHEMA/GEM2010MAS65 (NCT01237249)	Not newly-diagnosed transplant-eligible population
Perrot, 2018[46]	IFM2009	MRD not measured at 100 days post-transplant
Bahlis, 2019 [47]	MAIA	Not newly-diagnosed transplant-eligible population
De Tute, 2016 [48]	NCRI Myeloma XI trial	Not newly-diagnosed transplant-eligible population
Oliva, 2017[49]	EMN02/HO95 // NCT01208766	MRD not measured at 100 days post-transplant
Sanchez-Vega, 2016[50]	GEM2005 / GEM05MENOS65 / NCT00461747	Not newly-diagnosed transplant-eligible population
Mateos, 2019 [51]	ALCYONE / NCT02195479	Not newly-diagnosed transplant-eligible population

Martinez-Sanchez, 2008[52]	NA	Not newly-diagnosed transplant-eligible population
Putkonen, 2010[53]	NA	MRD not measured at 100 days post-transplant
Avet-Loiseau, 2018 [54]	CASTOR / NCT02136134	Not newly-diagnosed transplant-eligible population
Avet-Loiseau, 2018[54]	POLLUX / NCT02076009	Not newly-diagnosed transplant-eligible population
Swedin, 1998[55]	NA	MRD not measured at 100 days post-transplant
Fukumoto, 2016[56]	NA	Not newly-diagnosed transplant-eligible population
Silvennoinen, 2014[57]	NCT00861250	Not newly-diagnosed transplant-eligible population
Flores-Montero, 2017[58]	NA	Not newly-diagnosed transplant-eligible population
Paiva, 2011[59]	GEM05MAS65 / NCT00443235	Not newly-diagnosed transplant-eligible population
Li, 2019[60]	NCT02086942 / NCT02248428	Not newly-diagnosed transplant-eligible population
Gambella, 2019[61]	NCT01091831 / NCT01208766	MRD not measured at 100 days post-transplant
Alonso, 2019[62]	NA	Not newly-diagnosed transplant-eligible population
Rasche, 2018[63]	NA	Not newly-diagnosed transplant-eligible population
Facon, 2019[64]	CLARION / NCT01818752	Not newly-diagnosed transplant-eligible population
Shah, 2018[65]	NA	MRD not measured at 100 days post-transplant
Yong, 2018[66]	MUK-five	Not newly-diagnosed transplant-eligible population
Austin, 2018[67]	NA	MRD not measured at 100 days post-transplant
Paiva, 2020[68]	GEM2012MENOS65 / NCT01916252	MRD not measured at 100 days post-transplant
Ludwig, 2015 [69]	NCT00531453	MRD not measured at 100 days post-transplant

B7. <u>Priority question</u>. No information has been provided on the validity of the included studies. Please provide a risk of bias assessment for each included

study, using the same criteria as reported for the CASSIOPEIA study in CS section B.2.5.

Janssen intend performing a risk of bias assessment for studies included in the MRD meta-analysis using the ROBINS-I tool developed by Cochrane.[70] This tool is specific for non-randomised studies, however, we consider it suitable for this review as MRD assessment is always conducted on unrandomized population, as it implies dividing the patients within a treatment arm based on their response to the treatment. In addition, MRD analysis is often conducted post-hoc on bone marrow samples and/or on a population pooled from different treatment arms.

Results from this assessment will be available, Friday 16th October.

Meta-analysis of MRD status on survival outcomes

B8. <u>Priority question</u>. CS Appendix M.2 provides no information on the methods employed for the meta-analysis, other than that a random effects model was preferred based on an assessment of heterogeneity. Please provide a full description of the meta-analysis method. Please clarify if any parts of the method were the same as those reported by Munshi et al. [reference 61]?

The meta-analysis referred in CS Appendix M.2 was performed by fitting a random effects model to obtain a pooled effect estimate of the hazard ratio for MRD negativity. There is evidence in favour of using a random effects model over a fixed effects model for meta-analyses where studies are not uniform in design and population (irrespective of heterogeneity) (Riley, 2011).[71] In addition, a random effects model was previously used in a meta-analysis in the newly diagnosed transplant-eligible indication.[72] Therefore, a random effects model was more justified than fixed effects. The statistical significance level was set at $p<0.05$.

Heterogeneity in design and population among the studies eligible for meta-analysis was assessed using the I^2 test, with a maximum likelihood (ML) estimator. Refer to question B10 below for further details.

The methodology applied in the meta-analysis differs from the one applied in Munshi et al. 2017. In the CS, a frequentist meta-analysis on mean outcomes (HRs) was used, while in Munshi 2017, a Cox regression was fitted on the pooled curves generated based on a simulated IPD. The methodology used in the CS repeats the methodology used in Munshi et al. 2020 manuscript, which has been reviewed and accepted for publication in Blood Advances. The methodology involved conducting a meta-analysis by fitting a random effect model to obtain the pooled estimate, generating the pooled KM curves based on the simulated IPD, and conducting subgroup analysis to address the potential bias caused by the differences in disease setting, eligibility for MRD assessment by conventional response, MRD assay and sensitivity and time of MRD assessment.

B9. Priority question. Please provide a tabulation of the study characteristics and population characteristics for the included studies. Please clarify how well the studies compare with CASSIOPEIA in terms of their designs and population characteristics?

Table 7 below summarizes the study characteristics. The majority of the studies used for the meta-analysis were non-randomized. However, we do not expect that this reduces the quality of the meta-analysis by introducing any additional bias as the initial randomisation is not relevant for MRD analysis (please refer to question B7 for more details).

CASSIOPEIA's patient population is comparable in terms of age, gender distribution, and type of measurable disease. The ISS score is comparable to most studies, although there are outliers (e.g., Rossi, 2018: ISS I – 83%). The high-risk population percentage is higher in many studies, although this parameter is often not reported. An analysis comparing the hazard ratio for MRD-negativity in standard and high cytogenetic risk populations was conducted by Munshi et al. 2019 and did not show substantial differences in MRD-negativity benefit between those populations with the PFS hazard ratio for MRD- 0.37 (95%CI: 0.18-0.73) and 0.44 (95%CI: 0.34-0.56) in standard risk and high-risk populations, respectively.[73] The OS hazard ratio for MRD- was 0.65 (95%CI: 0.55-0.76) and 0.63 (95%CI: 0.43-0.91) in standard risk and high-risk

populations, respectively.[73] Based on these findings, Janssen does not expect baseline cytogenetics to have a significant impact on the MRD hazard ratio.

Table 7. Study characteristics

Author	Trial ID	Study design	Phase	Number of patients	Age (median)	Male (%)	ISS	Type of measurable disease	Cytogenetic profile (high risk, %)
	CASSIOPEIA (VTd arm)	RCT	III	1,085 MRD measured in 1,085	59	59%	I – 40% II – 45% III – 15%	IgG – 59% IgA – 17% Other – 3%	16%
Rawstron, 2013[74]	MRC Myeloma IX Study	RCT	III	1,111 MRD measured in 397	59	62%	I - 25% II - 34% III - 31% NA – 10%	IgG – 60% IgA – 22% Other – 17%	47%
Paiva, 2008[75]	PETHEMA/GEM2000 / NCT00560053	Prospective single-arm	III	295 MRD measured in 295	59	54%	I – 39% II – 41% III – 20%	IgG – 55% IgA – 27% Other – 18%	16%
Popat, 2017[Error! Bookmark not defined.]	PADIMAC	Prospective single-arm	II	153 MRD measured in 27	55	NA	NA	NA	33%
Cohen, 2016[Error! Bookmark not defined.]	NA	Prospective single-arm	II	40 MRD measured in 19	61	55%	I – 43% II – 10% III – 5% NA – 43%	IgG – 55% IgA – 23% Other – 23%	NA
Clark, 2018[Error! Bookmark not defined.]	NCT01215344	Prospective single-arm	II	32 MRD measured in 25	57	68%	I – 52% II – 44% III – 4%	IgG – 68% IgA – 12% Other – 20%	32%
Chakraborty, 2017[Error! Bookmark not defined.]	NA	Retrospective	NA	185 MRD measured in 185	61	54%	I – 17% II – 42% III – 24% NA – 17%	NA	100%
Schinke, 2017[Error! Bookmark not defined.]	TT3b-TT5a	NA	NA	883 MRD measured in 109	60	63%	I – 33% II – 40% III – 27%	IgG – 58% IgA – 21% Other – 21%	NA
Bakkus, 2004[76]	EBMT Phase III	RCT	NA	67 MRD measured in 60	54	58%	NA	IgG – 60% IgA – 35%	NA

								Other – 5%	
Rawstron, 2002 [77]	MRC Myeloma VII	RCT	NA	45 MRD measured in 45	55	53%	NA	NA	NA
Gu, 2018 Error! Bookmark not defined.	NA	Prospective single-arm	NA	104 MRD measured in: NA	54	67%	I – 37% II – 38% III – 25%	IgG – 56% IgA – 14% Other – 30%	59%
Rossi, 2018 Error! Bookmark not defined.	NA	Prospective single-arm	NA	30 MRD measured in 30	58	53%	I – 83% II – 7% III – 10%	NA	25%
Hahn, 2019 Error! Bookmark not defined.	STAMiNA / PRIMER, NCT01109004	RCT	III	437 MRD measured in 311	NA	NA	I – 33% II – 33% III – 29%	NA	NA
Solovev, 2018 Error! Bookmark not defined.	NA	Prospective single-arm	NA	70 MRD measured in 37	56	66%	I – 40% II – 27% III – 33%	NA	NA
Solovev, 2016 Error! Bookmark not defined.	NA	Prospective single-arm	NA	52 MRD measured in 52	54	37%	NA	NA	NA

ISS: International Staging System for Multiple Myeloma; NR: not reported

Table 8. Study characteristics (continued)

Author	Treatment	Eligibility for MRD assessment by response	MRD assay and sensitivity
	VTd or D-VTd (arm excluded from the analysis)	Any response	
Rawstron, 2013	CTD or CVAD + ASCT	Any response	MFC. 10^{-4}
Paiva, 2008	VBMCP/VBAD induction + ASCT + melphalan consolidation	Any response	MFC. 10^{-4}
Popat, 2017	PAD induction + ASCT	Only CR	MFC. 10^{-4}
Cohen, 2016	ASCT + Bortezomib consolidation	NA	MFC. 10^{-4}

Clark, 2018	VRD induction + ASCT	Any response	MFC. 10^{-4}
Chakraborty, 2017	ASCT (other therapies unknown)	Any response	MFC. 10^{-4}
Schinke, 2017	VTD and PACE chemo + ASCT + VRD maintenance	At least VGPR	NGS. 10^{-5}
Bakkus, 2004	Conventional chemo + ASCT	Any response	PCR. 10^{-4}
Rawstron, 2002	C-VAMP and HDT induction + ASCT	Any response	MFC. 10^{-4}
Gu, 2018	Bortezomib induction + ASCT + thalidomide/lenalidomide and/or IFN-a	At least VGPR	MFC. 10^{-5}
Rossi, 2018	PAD induction	At least VGPR	MFC. 10^{-4}
Hahn, 2019	ASCT + RVD consolidation + LEN maintenance	Any response	MFC. 10^{-5}
Solovev, 2018	Bortezomib induction + ASCT +/- LEN maintenance	Only CR	MFC. 10^{-5}
Solovev, 2016	ASCT + Bortezominb maintenance	CR	MFC. Sensitivity: NA

ASCT (Autologous Stem Cell Transplantation); C-VAMP (vincristine, amethopterine, methotrexate, and prednisone + cyclophosphamide); CTD (cyclophosphamide, thalidomide, and dexamethasone); CVAD (cyclophosphamide, vincristine, doxorubicin, and dexamethasone); HDT (high-dose chemotherapy); IMiD (Immunomodulatory drugs); LEN (lenalidomide); PACE (cisplatin, doxorubicin, cyclophosphamide, etoposide); PAD (bortezomib, doxorubicin and dexamethasone); PI (protease inhibitors); RVD (lenalidomide, bortezomib, and dexamethasone); VBMCP (vincristine, carmustine, melphalan, cyclophosphamide, and prednisone); VRD (bortezomib, lenalidomide, and dexamethasone); VTD (bortezomib, thalidomide, and dexamethasone).

B10. Priority question. Please explain how heterogeneity was assessed for each included study and provide the results of the heterogeneity assessment.

Heterogeneity was assessed using the ML estimator. The statistical significance of I^2 was assessed through the Chi-square test for residual heterogeneity (by definition, I^2 = 100% x (Q-df)/Q, where Q is a chi-square statistic and df is degrees of freedom). The non-significant Chi-square test suggests relative homogeneity of the studies. However, it should be noted that Chi-square test is very sensitive to sample size and tends to be insignificant in small samples.

In the PFS meta-analysis, the heterogeneity was estimated at [redacted]% (95%CI: [redacted]-[redacted]). The Chi-square test did not reach statistical significance, suggesting relative homogeneity of the studies. The Chi-square test resulted in Q=[redacted] (df = [redacted]), p=[redacted]. In the OS analysis, the heterogeneity was estimated at [redacted] % (95%CI: [redacted]-[redacted] %). The Chi-square test did not reach statistical significance: Q=[redacted] (df = [redacted]), p=[redacted].

B11. Priority question. Please explain how the validity (risk of bias) of individual studies was considered when conducting the meta-analysis.

Individual studies were assessed using the modified Strobe statement which focused on the quality and completeness of reporting rather than study design. Considering the limitations of MRD assessment as a secondary/post-hoc analysis on pooled populations and in unrandomized settings, Janssen considered them all sufficiently robust and of a decent quality for inclusion. However, as stated in our response to question B7, a formal risk of bias assessment will be conducted for included studies based on the ROBINS-I tool developed by Cochrane.

B12. Please explain whether sensitivity analyses were conducted to explore the impact of study heterogeneity and study validity on the meta-analysis results (questions B10 and B11). If so, please provide the results of these. If not, please justify why sensitivity analyses were not conducted.

Janssen has conducted several subgroup analyses to adjust for effect modifiers (variables that significantly impact the association of MRD and PFS/OS). The variables were selected based on qualitative evidence from the extracted

publications. The results for PFS and OS are presented in the tables below. Consistent with the main analysis, all subgroup analyses show that the survival is positively associated with MRD-negativity. Whilst a deterioration in hazard ratio is observed with an increase in sensitivity threshold, Janssen note that this is not consistent with the expanded SLR/meta-analysis reported by Munshi et al. 2019 based on a larger number of studies and may reflect selection bias.

Table 9. Subgroup analyses: PFS

Subgroup	Number of observations	Outcome
Base case	**n=15**	[redacted]
Sensitivity		
Sensitivity of 10^{-4}	n=9 [6,7,8,11,12,49,50,51,52]	[redacted]
Sensitivity of 10^{-5}	n=5 [10,13,14,15,16]	[redacted]
Sensitivity of 10^{-6}	Not enough observations	NA
Eligibility for MRD assessment by response criteria		
Only patients achieving CR	n=3 [6,9,13]	[redacted]
Patients achieving at least VGPR	n=3 [8,10,16]	[redacted]

Table 10. Subgroup analyses: OS

Subgroup	Number of observations	Outcome
Base case	**n=9**	[redacted]
Sensitivity		
Sensitivity of 10^{-4}	n=5 [11,49,50,51,52]	[redacted]
Sensitivity of 10^{-5}	n=4 [10,14,15,16]	[redacted]
Sensitivity of 10^{-6}	Not enough observations	NA
Eligibility for MRD assessment by response criteria		
Only patients achieving CR	No observations	NA
Patients achieving at least VGPR	n=2 [10,16]	[redacted]

B13. Please explain why fewer studies were included in the meta-analysis reported in CS Appendix M compared to the studies included in the NDTE MM subgroup of the Munshi et al. 2019 meta-analysis [reference 63].

Please refer to our response to question B1 for the additional screening criteria relevant to the CS.

B14. Please state the analysis software used for the meta-analysis reported in CS Appendix M and provide the statistical code and input data.

Statistical analyses were performed using the "metafor" R package Version 2 for frequentist meta-analyses embedded in the open-source JASP software.[78] The interface of JASP does not allow us to extract the source code. The input data is presented below; the analysis was conducted on log-transformed values.

Table 11: Input data for PFS MRD analysis

Author	HR TDE	HR TDE 95% LL	HR TDE 95% UL	HR TDE log	HR TDE SE log
Rawstron, 2013	0.557	0.415	0.746	-0.58609	0.149741
Paiva, 2008	0.275	0.16	0.4	-1.29098	0.233748
Popat, 2017 (arm 1)	0.38	0.13	1.15	-0.96129	0.562718
Cohen, 2016	0.277	0.102	0.752	-1.28371	0.50923
Clark, 2018	0.133	0.022	0.813	-2.01757	0.923329
Chakraborty, 2017	0.45	0.31	0.66	-0.79851	0.192772
Schinke, 2017	0.55	0.22	1.37	-0.60624	0.468501
Bakkus, 2004	0.274	0.136	0.552	-1.29426	0.357509
Rawstron, 2002	0.308	0.1296	0.732	-1.17766	0.441665
Gu, 2018	0.29	0.13	0.65	-1.23787	0.410571
Rossi, 2018	0.13	0.03	0.51	-2.07	0.71
CASSIOPEIA (BTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hahn, 2019	0.48	0.31053	0.741957	-0.73	0.22
Solovev, 2018	0.4966	0.2094	2	-0.7	0.58
Solovev, 2016	0.588	0.294	0.833	-0.53	0.27

Table 12 Input data for OS MRD analysis

Author	HR OS	HR OS 95% LL	HR OS 95% UL	HR OS log	HR OS SE log

Rawstron, 2013	0.6269	0.4344	0.9046	-0.46697	0.187124
Paiva, 2008	0.49505	0.42	0.711	-0.7031	0.13429
Chakraborty, 2017	0.55	0.32	0.92	-0.59784	0.269401
Schinke, 2017	0.8411	0.2962	2.388	-0.17304	0.532443
Bakkus, 2004	0.5873	0.209	1.651	-0.53222	0.527245
Rawstron, 2002	0.5345	0.1629	1.755	-0.62642	0.6064
Gu, 2018	0.23	0.09	0.59	-1.46968	0.479672
Hahn, 2019	0.77	0.354128	1.674255	-0.26136	0.396292
CASSIOPEIA (BTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Proportional hazards assumptions

B15. Priority question. CS Appendix N reports log-cumulative hazard plots for OS and PFS for MRD+ versus MRD- with BTd.

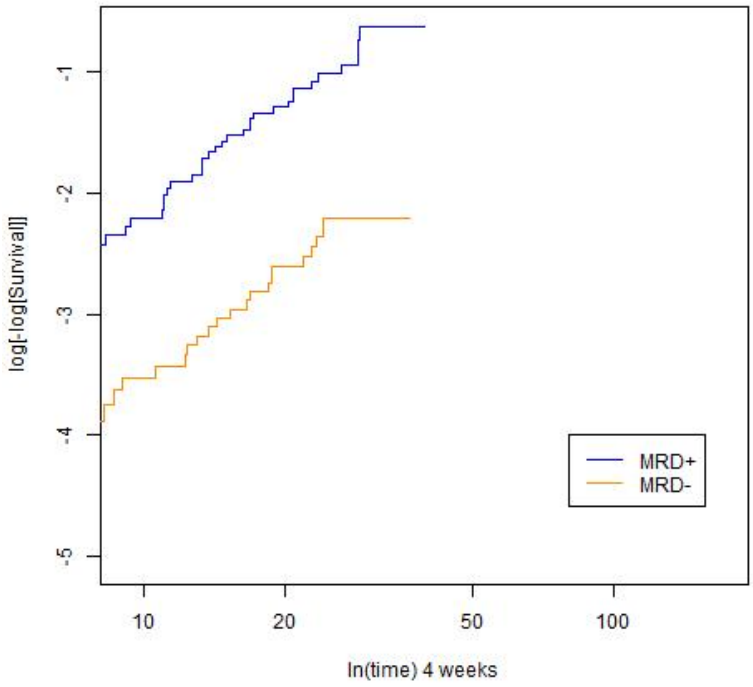
(a) Please clarify whether these analyses are based on CASSIOPEIA data only, or if they also include Kaplan-Meier data from other studies included in the expanded SLR and meta-analysis of the impact of MRD status on survival outcomes.

(b) Please provide similar graphs for the CASSIOPEIA DBTd arm.

Janssen confirm that the log-cumulative hazard plots for PFS and OS in the BTd arm (MRD+ versus MRD-) presented in Appendix N of the CS are based on the landmark analyses which were performed using individual patient data from the CASSIOPEIA trial; no Kaplan-Meier data from other studies were included.

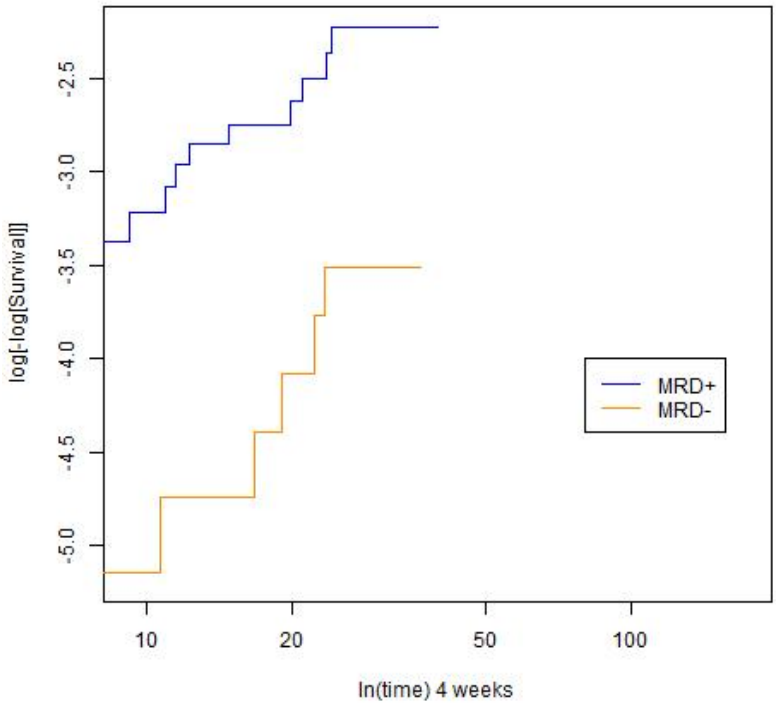
Similar analysis for the DBTd arm are presented in Figure 8 (PFS) and Figure 9 (OS) below.

Figure 8: Log-cumulative hazard plot for PFS in the DBTd arm (MRD+ versus MRD-)



Key: DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD: minimal residual disease; PFS: progression-free survival.

Figure 9: Log-cumulative hazard plot for OS in the DBTd arm (MRD+ versus MRD-)



Key: DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MRD: minimal residual disease; OS: overall survival.

Utilities

B16. The model includes 'age adjustment utilities', cited as Janssen 2013. These parameters are not mentioned in the CS. Please specify the source of these parameters and explain how they are used in the model.

As stated in Section B.3.4.5 of the CS, the health state utility values were age-adjusted using the population norm values for EQ-5D as reported in Janssen et al. (**2014**) (full reference: Janssen B, Szende A. Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J, editors. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht: Springer Netherlands; 2014. p. 19-30); this was erroneously cited in the model as Janssen et al. (**2013**). The EQ-5D index population norms using the UK-England value set are presented in Table 13. Age-specific adjustments in utility were applied via a multiplicative method, dividing the utility for current age versus the age at entry into the model (aligning with the mean age of sample informing health state utility values), and multiplying this adjustment factor by the health state utility value. As such, patient utility values will always be equal to or less than the health state utility value (assuming current age is equal or less than age at entry), with gradual reductions over time reflecting the reduction in the age adjusted utility values, presented in Table 13.

Table 13: Age adjusted utility values based on the EQ-5D index population norms (UK-England TTO value sets)

Age group	Value
18–24	0.929
25–34	0.919
35–44	0.893
45–54	0.855
55–64	0.810
65–74	0.773
75+	0.703

Key: EQ-5D = EuroQol-5 dimensions; TTO = time trade-off.

Costs

B17. The base case analysis assumes no vial sharing for daratumumab and bortezomib. The numbers of vials are rounded up based on only the mean body

weight or mean body surface area (BSA) for the patient population (CS Table 45). The ERG note that this may not give accurate cost estimates, as the mean number of vials per administration without vial sharing depends on the distributions of weight and BSA. Please revise the cost calculations based on the distributions of these characteristics within the CASSIOPEIA population.

As stated in Section B.3.5.1 of the CS, in the base case analysis, it was assumed that there would be no vial sharing or pooling across administrations, therefore the number of vials required per administration was rounded up to the nearest whole integer. In the base case analysis, the cost of daratumumab was based on the fixed dose (1,800 mg administered via subcutaneous injection) and was not weight-based. However, a scenario analysis (Scenario 4 as per the CS) was conducted in which the weight-based dose and IV formulation of daratumumab (16 mg/kg) was used. In this scenario analysis, the cost of daratumumab was calculated based on the mean weight of patients in CASSIOPEIA.

In order to re-calculate the cost of daratumumab for this scenario based on a distribution of the weight of patients in CASSIOPEIA, the categorised weight classes presented in the CASSIOPEIA CSR (<50, 50–64, 65–85 and >85 kg) were converted into point estimates using the total minimum and maximum values (range: 44.0, 142.5 kg) and the mean of the upper and lower bounds of the individual weight class (Table 14).

Table 14: Baseline weight of patients in CASSIOPEIA

Weight (kg) (category)	Weight (kg) (point estimate)	Value, n (%) (N=1,085)
<50	46.50	23 (2.12)
50–64	57.00	254 (23.41)
65–85	75.00	538 (49.59)
>85	114.25	270 (24.88)

The revised total cost per administration of daratumumab, based on a distribution of the weight of patients in CASSIOPEIA, is presented in Table 15.

Table 15: Revised drug acquisition cost calculations for daratumumab

Weight (kg) (point estimate)	Total dose per administration (mg)	Vials per administration	Cost per administration

46.5	744.00	8.00	£4,748.02
57	912.00	10.00	
75	1,200.00	12.00	
114.25	1,828.00	19.00	

The results of a scenario analysis, in which the weight-based dose and IV formulation of daratumumab (16 mg/kg) was used and based on a distribution of the weight of patients in CASSIOPEIA, are presented in Table 16.

Table 16: Impact of daratumumab IV formulation, based on a distribution of the weight of patients

Scenario	Inc. costs	Inc. QALYs	ICER (£ per QALY)
Base case: Daratumumab SC formulation	[redacted]	[redacted]	[redacted]
Scenario 4 (as per CS): Daratumumab IV formulation, based on mean weight of patients	[redacted]	[redacted]	[redacted]
Scenario 4 with daratumumab IV formulation, based on a distribution of the weight of patients	[redacted]	[redacted]	[redacted]

Key: ICER = incremental cost-effectiveness ratio; IV = intravenous; SC = subcutaneous; QALY = quality-adjusted life year.

It should be noted that it has not been possible to present a revised calculation for the cost of bortezomib based on a distribution of BSA as this information is not available in the CASSIOPEIA CSR.

B18. CS Table 76 cites the cost of daratumumab in the DBd regimen at 2L/3L/4L as £8,640 in cycles 3-6. However, the model estimates this cost at £6,480 (1.5 x £4,320). Please confirm which value is correct.

Janssen confirm that the cost presented in the economic model is correct and that in this instance, the CS is incorrect. Janssen apologise for this inconsistency; the correct cost of daratumumab in the DBd regimen at 2L/3L/4L is £6,480, as per the model (see Table 17 in response to question B19 below).

B19. The average cost per model cycle for the CDF DBTd regimen and daratumumab monotherapy are calculated by dividing the total costs until progression by the median TTP/PFS in months (27.63 and 4.00), rather than in 4

week cycles (25.42 and 3.68): (CS Table 76 and model sheet Treatment Costs). Please confirm these figures and correct if necessary.

Janssen confirm that the average cost per model cycle for the subsequent treatments, DBd and daratumumab monotherapy (recommended via CDF; scenario only), should have been calculated by dividing the total costs until disease progression by the median TTP/PFS in 4 week cycles not months, and apologise for this error. The corrected drug acquisition cost calculations for DBd and daratumumab monotherapy are presented in Table 17.

Table 17: Summary of drug acquisition costs for subsequent treatments – drug cost per cycle (DBd and daratumumab monotherapy)

Treatment	Cycle from start of treatment	Daratumumab	Bortezomib	Dexamethasone	Total drug cost per cycle	Median TTP	Total cost until progression	Average cost per cycle
DBd (Recommended via CDF; scenario only)	Cycles 1–2	£17,280.00	£4,574.28	£15.10	£21,869.38	25.42	£168,839.57	£6,642.69
	Cycles 3–6	£6,480.00	£3,811.90	£12.59	£10,304.49			
	Cycles 7+ (Median TTP – 6)	£4,320.00	£0.00	£0.00	£4,320.00			
Daratumumab monotherapy (Recommended via CDF; scenario only)	Cycles 1–2	£17,280.00	-	-	£17,280.00	3.68	£69,120.00	£18,784.29
	Cycles 3–6	£8,640.00	-	-	£8,640.00			
	Cycles 7+ (Median TTP – 6)	£4,320.00	-	-	£4,320.00			

Key: CDF = Cancer Drugs Fund; DBd = daratumumab, bortezomib, and dexamethasone; TTP = time to progression.

The results of a scenario analysis, in which the corrected cost per model cycles for subsequent therapies has been used and applied to scenario 5 (as per the CS), are presented in Table 18. It should be noted that this amendment does not impact the base case results as only those treatments that have been recommended for routine funding by NICE, and not via the CDF, have been considered as subsequent therapies in the base case analysis.

Table 18: Impact of the corrected cost per model cycle for subsequent therapies

Scenario	Inc. costs	Inc. QALYs	ICER (£ per QALY)
Base case	[redacted]	[redacted]	[redacted]
Scenario 5 (as per CS): Inclusion of subsequent therapies recommended via the CDF	[redacted]	[redacted]	[redacted]
Scenario 5 with corrected cost per model cycle for subsequent therapies	[redacted]	[redacted]	[redacted]

Key: CDF = Cancer Drugs Fund; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year.

B20. The duration of induction therapy for BTd is reported in CS Table 59 as 15.42 weeks. The ERG notes this is inconsistent with the value of 15.52 in the economic model which is based on 3.57 months and consistent with the value reported in Table 9 of the Clinical Study Report. Please explain this discrepancy.

Janssen confirm that the duration of therapy presented in the economic model is correct and that in this instance, the CS is incorrect. Janssen apologise for this inconsistency; the correct duration of induction therapy for BTd should be 15.52 weeks (3.57 months), as per the model and Table 9 of the CSR (see Table 19 below).

Note that this discrepancy does not impact the base case results.

Table 19: Duration for health-state utility values (induction and post-induction to post-consolidation response) based on CASSIOPEIA

	Duration of induction therapy, weeks	Duration from completion of induction therapy to response assessment, weeks[a]
DBTd	15.65	21.63
BTd	15.52	21.42

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

[a] Based on the mean gap between induction and SCT in CASSIOPEIA + 100 days.

Section C: Textual clarification and additional points

C1. Please confirm or provide the references for the ALCYONE and MAIA trials which are cited in CS section B.2.6.3. Are these references 119 and 118 respectively?

That is correct; the ALCYONE and MAIA trials cited in CS Section B.2.6.3 refer to a pooled analysis of the survival data from the studies referenced 119 and 118 respectively in the CS.

C2. Please provide missing footnote [a] for CS Table 60.

Janssen confirm that there should be no footnote [a] within Table 60 of the CS and apologise that the CS in this instance is incorrect.

References

[1] Janssen. [Data on File] Clinical Advisory Board Meeting Minutes. August 2020.
[2] Janssen. [Data on File] Public Health England datasets. Standing cohort study of newly diagnosed multiple myeloma cancer patients in England (January 2015 to December 2018). (July 2020 Report).
[3] Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(26):2863-9.
[4] Janssen. [Data on File] MMY3006. Clinical Study Report: Part 1. 2019.
[5] Henson et al. “Data Resource Profile: National Cancer Registration Dataset in England” International Journal of Epidemiology. 2019
[6] Popat, Rakesh, et al. "Outcomes of stratification to ASCT or not based on depth of response: results of a phase 2 trial assessing the impact of minimal residual disease (MRD) in multiple myeloma patients with deferred ASCT (PADIMAC)." *Blood* 130.Supplement 1 (2017): 1864-1864.
[7] Cohen, Oliver C., et al. "Bortezomib Consolidation Following Upfront ASCT for Multiple Myeloma Deepens Disease Response and MRD-Negative Rate without Compromising Response to Subsequent Bortezomib Salvage: Results of a Phase II Study." (2016): 4508-4508.
[8] Rossi, Giovanni, et al. "Minimal residual disease and log-reduction of plasma cells are associated with superior response after double autologous stem cell transplant in younger patients with multiple myeloma." *Cytometry Part B: Clinical Cytometry* 96.3 (2019): 195-200.
[9] Solovev, Maxim V., et al. "Maintenance Therapy after Autologous Haematopoietic Stem Cell Transplantation (auto-HSCT) in Multiple Myeloma Patients with and without Minimal Residual Disease (MRD)." (2016): 2260-2260.
[10] Gu, Jingli, et al. "Longitudinal flow cytometry identified “minimal residual disease”(MRD) evolution patterns for predicting the prognosis of patients with transplant-eligible multiple myeloma." *Biology of Blood and Marrow Transplantation* 24.12 (2018): 2568-2574.
[11] Chakraborty, Rajshekhar, et al. "Impact of post-transplant response and minimal residual disease on survival in myeloma with high-risk cytogenetics." *Biology of Blood and Marrow Transplantation* 23.4 (2017): 598-605.
[12] Clark, C. Amos, et al. "Prospective trial of minimal residual disease assessment by multiparametric flow cytometry for multiple myeloma in the era of bortezomib-based chemotherapy." *Bone marrow transplantation* 53.12 (2018): 1589-1592.
[13] Solovev, Maxim V., et al. "Efficacy of maintenance therapy following auto-HSCT depending on MRD status in patients with multiple myeloma." *Blood* 132.Supplement 1 (2018): 3432-3432.
[14] Hahn, Theresa E., et al. "Minimal residual disease (MRD) assessment before and after autologous hematopoietic cell transplantation (AutoHCT) and maintenance for multiple myeloma (MM): results of the prognostic immunophenotyping for myeloma response (PRIMeR) study." *Biology of Blood and Marrow Transplantation* 25.3 (2019): S4-S6.
[15] Janssen. [Data on File] MMY3006. Post-hoc Interim Analysis. May 1, 2019.
[16] Schinke, Carolina, et al. "The prognostic value of the depth of response in multiple myeloma depends on the time of assessment, risk status and molecular subtype." *Haematologica* 102.8 (2017): e313.

[17] Philippe Moreau, Michel Attal, Cyrille Hulin, Bertrand Arnulf, Karim Belhadj, et al. (2019). Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label, phase 3 study. Lancet 2019; 394: 29–38.

[18] Voorhees PM, Kaufman JL, Laubach JP, Sborov DW, Reeves B, Rodriguez C, Chari A, et al. (2020). Daratumumab, Lenalidomide, Bortezomib, & Dexamethasone for Transplant-eligible Newly Diagnosed Multiple Myeloma: GRIFFIN. Blood. pii: blood.2020005288. doi: 10.1182/blood.2020005288

[19] Sini Luoma, Pekka Anttila, Marjaana Säily, Tuija Lundan, et al. (2019). RVD induction and autologous stem cell transplantation followed by lenalidomide maintenance in newly diagnosed multiple myeloma: a phase 2 study of the Finnish Myeloma Group. Ann Hematol. 2019; 98(12): 2781–2792.

[20] Laura Rosiñol, Albert Oriol, Rafael Rios, Anna Sureda, María Jesús Blanchard, et al. (2019). Bortezomib, lenalidomide, and dexamethasone as induction therapy prior to autologous transplant in multiple myeloma. Blood. 134(16): 1337–1345.

[21] Kazutaka Sunami, Morio Matsumoto, Shin ichi Fuchida, Eijiro Omoto, et al. (2019). Bortezomib based strategy with autologous stem cell transplantation for newly diagnosed multiple myeloma: a phase II study by the Japan Study Group for Cell Therapy and Transplantation (JSCT MM12). Japan Society of Clinical Oncology 24, pages966–975

[22] Keisuke Tanaka, Shigeo Toyota, Megumi Akiyama, Naoki Wakimoto, Yuichi Nakamura, et al. (2019). Efficacy and Safety of a Weekly Cyclophosphamide-Bortezomib-Dexamethasone Regimen as Induction Therapy Prior to Autologous Stem Cell Transplantation in Japanese Patients with Newly Diagnosed Multiple Myeloma: A Phase 2 Multicenter Trial. Acta Haematol 2019;141:111–118

[23] van de Donk NW, van der Holt B, Minnema MC, Vellenga E, et al. (2018). Thalidomide before and after autologous stem cell transplantation in recently diagnosed multiple myeloma (HOVON-50): long-term results from the phase 3, randomised controlled trial. Lancet Haematol; 5(10):e479-e492.

[24] Noemi Horvath, Andrew Spencer, Melita Kenealy, Douglas Joshua, et al. (2019). Phase 3 study of subcutaneous bortezomib, thalidomide, and prednisolone consolidation after subcutaneous bortezomib-based induction and autologous stem cell transplantation in patients with previously untreated multiple myeloma: the VCAT study. Leuk Lymphoma. 60(9):2122-2133

[25] Gregersen H, Do T, Kristensen IB, FrÃ¸lund UC, Andersen NF, Nielsen LK, Andersen CL, Klausen TW, Vangsted AJ, Abildgaard N. (2018). A randomized placebo-controlled phase II study of clarithromycin or placebo combined with VCD induction therapy prior to high-dose melphalan with stem cell support in patients with newly diagnosed multiple myeloma. Exp Hematol Oncol. 13;7:18. doi: 10.1186/s40164-018-0110-0.

[26] Hulin C., Moreau P., Attal M., Belhadj K., Benboubker L., Caillot D., Facon T., et al. (2019). Stem cell (SC) yield and transplantation results from transplant-eligible newly diagnosed multiple myeloma (TE NDMM) patients (pts) receiving daratumumab (DARA) + bortezomib/thalidomide/dexamethasone (D-VTd) in the phase 3 CASSIOPEIA study.Journal of Clinical Oncology 37, no. 15_suppl (May 20, 2019) 8042-8042.

[27] Avet-Loiseau H., Moreau P., Attal M., Hulin C., Arnulf B., Corre J., Garderet L., Karlin L., Lambert J., MacRo M., Perrot A., Sonneveld P., Levin M.-D., Klein S., Chiu C., Pei L., De Boer C., Kampfenkel T., Wuilleme S., Béné M.-C. (2019). Efficacy of

daratumumab (DARA) + bortezomib/thalidomide/dexamethasone (D-VTd) in transplant-eligible newly diagnosed multiple myeloma (TE NDMM) based on minimal residual disease (MRD) status: Analysis of the CASSIOPEIA trial. Journal of Clinical Oncology 37, no. 15_suppl (May 20, 2019) 8017-8017.
[28] Moreau P., Attal M., Hulin C., Béné M.-C., Broijl A., Caillot D., Delforge M., Dejoie T., Facon T., Lambert J., Leleu X., MacRo M., Perrot A., Zweegman S., Ahmadi T., Chiu C., Pei L., Vermeulen J., Avet-Loiseau H., Sonneveld P. (2019). Phase 3 randomized study of daratumumab (DARA) + bortezomib/thalidomide/dexamethasone (D-VTd) vs VTd in transplant-eligible (TE) newly diagnosed multiple myeloma (NDMM): CASSIOPEIA Part 1 results. Journal of Clinical Oncology 37, no. 15_suppl (May 20, 2019) 8003-8003.
[29] Yong K., Popat R., Wilson W., Pang G., Jenner R., De Tute R.M., Ramasamy K., Streetly M, et al. (2019). Efficacy and safety of carfilzomib at 56mg/m with cyclophosphamide and dexamethasone (K56Cd) in newly diagnosed multiple myeloma patients followed by ASCT or K56Cd consolidation: Initial results of the phase 2 cardamon study. Blood (2019) 134 (Supplement_1): 861.
[30] Jackson G., Pawlyn C., Cairns D., Jones J.R., Kishore B., Garg M., Williams C., Karunanithi K., et al. (2019). Lenalidomide induction and maintenance maximizes outcome for newly diagnosed transplant eligible myeloma patients irrespective of risk status: Long-term follow-up of the myeloma Xi trial. Blood (2019) 134 (Supplement_1): 1910.
[31] Roussel M., Hebraud B., Hulin C., Leleu X., Facon T., et al. (2019). Twice weekly induction with ixazomib-lenalidomide-dexamethasone (IRD) combination followed by extended ird consolidation and lenalidomide maintenance in transplant eligible patients with newly diagnosed multiple myeloma (NDMM): A phase 2 study from the intergroupe francophone du myelome (IFM 2014-03). Blood (2019) 134 (Supplement_1): 3159.
[32] Voorhees P.M., Kaufman J.L., Laubach J.P., Sborov D.W., Reeves B., Rodriguez C., Chari A., et al. (2019). Depth of Response to Daratumumab (DARA), Lenalidomide, Bortezomib, and Dexamethasone (RVd) Improves over Time in Patients (pts) with Transplant-Eligible Newly Diagnosed Multiple Myeloma (NDMM): Griffin Study Update. Blood (2019) 134 (Supplement_1): 691.
[33] Moreau P., Zweegman S., Perrot A., Hulin C., Caillot D., Facon T., et al. (2019). Evaluation of the prognostic value of positron emission tomography-computed tomography (PET-CT) at diagnosis and follow-up in transplant-eligible newly diagnosed multiple myeloma (TE NDMM) patients treated in the phase 3 cassiopeia study: Results of the cassiopet companion study. Blood (2019) 134 (Supplement_1): 692.
[34] Shuang Y., Shang J., Yan L., Jin S., Wu D., Fu C. (2019). Bortezomib, Liposome Doxorubicin and Dexamethasone (PDd) is Superior in Safety and Not Inferior in Efficiency to Bortezomib, Doxorubicin and Dexamethasone (PAd) As Induction Therapy in New-Diagnosed Multiple Myeloma Patients. Clinical Lymphoma Myeloma and Leukemia Volume 19, Issue 10, Supplement, October 2019, Pages e204-e205
[35] Kumar L., Chellapuram S.K., Sahoo R., Gupta R. (2019). VRd versus VCd as induction therapy for newly diagnosed multiple myeloma: A Phase III, randomized study. Clinical Lymphoma, Myeloma & Leukemia, 2019-10-01, Volume 19, Issue 10, Pages e361-e361
[36] Sonneveld P., Attal M., Perrot A., Hulin C., Caillot D., Facon T., et al. (2019). Daratumumab Plus Bortezomib, Thalidomide, and Dexamethasone (D-VTd) in Transplant-eligible Newly Diagnosed Multiple Myeloma (NDMM): Subgroup Analysis

of High-risk Patients (Pts) in CASSIOPEIA. Clinical Lymphoma, Myeloma & Leukemia, 2019-10-01, Volume 19, Issue 10, Pages e2-e3
[37] Pawlyn C., Davies F., Cairns D., Striha A., Hockaday A., Kishore B., et al. (2019). Quadruplet KCRD (Carfilzomib, Cyclophosphamide, Lenalidomide and Dexamethasone) Induction for Newly Diagnosed Myeloma Patients. Clinical Lymphoma Myeloma and Leukemia Volume 19, Issue 10, Supplement, October 2019, Page e2
[38] Gay F.M., Foà R., Musto P., Cerrato C., Gamberi B., Zamagni E., et al. (2019). Updated efficacy data and MRD analysis according to risk status in newly diagnosed myeloma patients treated with carfilzomib + lenalidomide or cyclophosphamide (FORTE trial). Journal of Clinical Oncology (2018) 36:15 Supplement 1.
[39] Scheid C., Lokhorst H., Blau I.W., Van Der Holt B., Bertsch U., Zweegman S., et al. (2019). Bortezomib-based induction and maintenance overcomes the negative prognostic impact of renal impairment and del17p in transplant-eligible myeloma patients: Long term results from the phase iii hovon-65/gmmg-HD4 study after median 137 months follow up. Blood (2019) 134 Supplement 1.
[40] Oliva S., Genuardi E., Belotti A., Frascione P.M.M., Galli M., et al. (2019). Minimal residual disease evaluation by multiparameter flow cytometry and next generation sequencing in the forte trial for newly diagnosed multiple myeloma patients. Blood (2019) 134 Supplement 1.
[41] Ferrero, S., et al. "Long-term results of the GIMEMA VEL-03-096 trial in MM patients receiving VTD consolidation after ASCT: MRD kinetics' impact on survival." Leukemia 29.3 (2015): 689-695.
[42] Korde N, Roschewski M, Zingone A, et al. Treatment With Carfilzomib-Lenalidomide-Dexamethasone With Lenalidomide Extension in Patients With Smoldering or Newly Diagnosed Multiple Myeloma. *JAMA Oncol.* 2015;1(6):746-754.
[43] Korthals, Mark, et al. "The level of minimal residual disease in the bone marrow of patients with multiple myeloma before high-dose therapy and autologous blood stem cell transplantation is an independent predictive parameter." Biology of Blood and Marrow Transplantation 18.3 (2012): 423-431.
[44] Paiva, Bruno, et al. "The prognostic value of multiparameter flow cytometry minimal residual disease assessment in relapsed multiple myeloma." *Haematologica* 100.2 (2015): e53.
[45] Martinez-Lopez, J., et al. "Analytical and clinical validation of a novel in-house deep-sequencing method for minimal residual disease monitoring in a phase II trial for multiple myeloma." *Leukemia* 31.6 (2017): 1446-1449.
[46] Perrot, Aurore, et al. "Minimal residual disease negativity using deep sequencing is a major prognostic factor in multiple myeloma." *Blood* 132.23 (2018): 2456-2464.
[47] Bahlis N, et al. Daratumumab Plus Lenalidomide and Dexamethasone (D-Rd) Versus Lenalidomide and Dexamethasone (Rd) in Patients With Newly Diagnosed Multiple Myeloma (NDMM) Ineligible for Transplant: Updated Analysis of MAIA. *Blood.* 2019;134:1875.
[48] de Tute R, , , et al. Impact of minimal residual disease in transplant ineligible myeloma patients: results from the UK NCRI Myeloma XI trial. *Blood.* 2016;128(22):245-245.
[49] Oliva S, et al. Minimal residual disease by multiparameter cytometry in transplant eligible patients with newly diagnosed multiple myeloma: results from the EMN02/HO95 phase 3 trial. Haematologica. 2017;102.

[50] Sanchez-Vega B, et al. Prognostic Impact of Molecular Response Assessed By Next-Generation Sequencing in a Large Cohort of Multiple Myeloma Patients. Blood. 2016;128(22):3283-3283.
[51] Mateos MV, Dimopoulos MA, Cavo M, et al. Daratumumab plus Bortezomib, Melphalan, and Prednisone for Untreated Myeloma. N Engl J Med. 2018;378(6):518-528.
[52] Martinez-Sanchez P, Montejano L, Sarasquete ME, et al. Evaluation of minimal residual disease in multiple myeloma patients by fluorescent-polymerase chain reaction: the prognostic impact of achieving molecular response. Br J Haematol. 2008;142(5):766-774.
[53] Putkonen M, Kairisto V, Juvonen V, et al. Depth of response assessed by quantitative ASO-PCR predicts the outcome after stem cell transplantation in multiple myeloma. Eur J Haematol. 2010;85(5):416-423.
[54] Avet-Loiseau H, et al. Evaluation of sustained minimal residual disease (MRD) negativity in relapsed/refractory multiple myeloma (RRMM) patients (Pts) treated with daratumumab in combination with lenalidomide plus dexamethasone (D-Rd) or bortezomib plus dexamethasone (D-Vd): analysis of pollux and castor. Blood. 2018;132:3272-3272.
[55] Swedin A, Lenhoff S, Olofsson T, Thuresson B, Westin J. Clinical utility of immunoglobulin heavy chain gene rearrangement identification for tumour cell detection in multiple myeloma. Br J Haematol. 1998;103(4):1145-1151.
[56] Fukumoto K, Fujisawa M, Suehara Y, et al. Prognostic impact of immunophenotypic complete response in patients with multiple myeloma achieving better than complete response. Leuk Lymphoma. 2016;57(8):1786-1792.
[57] Silvennoinen R, Lundan T, Kairisto V, et al. Comparative analysis of minimal residual disease detection by multiparameter flow cytometry and enhanced ASO RQ-PCR in multiple myeloma. Blood Cancer J. 2014;4:e250.
[58] Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia. 2017;31(10):2094-2103.
[59] Paiva B, Martinez-Lopez J, Vidriales MB, et al. Comparison of immunofixation, serum free light chain, and immunophenotyping for response evaluation and prognostication in multiple myeloma. J Clin Oncol. 2011;29(12):1627-1633.
[60] Li H, Li F, Zhou X, et al. Achieving minimal residual disease-negative by multiparameter flow cytometry may ameliorate a poor prognosis in MM patients with high-risk cytogenetics: a retrospective single-center analysis. Ann Hematol. 2019;98(5):1185-1195.
[61] Gambella M, Omede P, Spada S, et al. Minimal residual disease by flow cytometry and allelic-specific oligonucleotide real-time quantitative polymerase chain reaction in patients with myeloma receiving lenalidomide maintenance: A pooled analysis. Cancer. 2019;125(5):750-760.
[62] Alonso R, Cedena MT, Gomez-Grande A, et al. Imaging and bone marrow assessments improve minimal residual disease prediction in multiple myeloma. Am J Hematol. 2019;94(8):853-861.
[63] Rasche L, Alapat D, Kumar M, et al. Combination of flow cytometry and functional imaging for monitoring of residual disease in myeloma. Leukemia. 2019;33(7):1713-1722.
[64] Facon T, Lee JH, Moreau P, et al. Carfilzomib or bortezomib with melphalan-prednisone for transplant-ineligible patients with newly diagnosed multiple myeloma. Blood. 2019;133(18):1953-1963.

[65] Shah GL. Depth of Response and Outcomes in Patients with Multiple Myeloma Undergoing Autologous Stem Cell Transplantation. Blood. 2018;132:4619-4619.
[66] Yong K, et al. Maintenance with Carfilzomib Following Carfilzomib, Cyclophosphamide and Dexamethasone at First Relapse or Primary Refractory Multiple Myeloma (MM) on the Phase 2 Muk Five Study: Effect on Minimal Residual Disease. Blood. 2018;132:802-802.
[67] Austin M. Sustained MRD negativity at 12 months post-ASCT predicts outcomes for myeloma patients: a real world study. EHA Library. Jun 16, 2018; 215626.
[68] Paiva, Bruno, et al. "Measurable residual disease by next-generation flow cytometry in multiple myeloma." *Journal of Clinical Oncology* 38.8 (2020): 784-792.
[69] Ludwig H, Greil R, Masszi T, et al. Bortezomib, thalidomide and dexamethasone, with or without cyclophosphamide, for patients with previously untreated multiple myeloma: 5-year follow-up. *Br J Haematol.* 2015;171(3):344-354.
[70] https://www.strobe-statement.org/?id=available-checklists
[71] Riley, Richard D., Julian PT Higgins, and Jonathan J. Deeks. "Interpretation of random effects meta-analyses." *Bmj* 342 (2011).
[72] Landgren, Ola, and S. Vincent Rajkumar. "New developments in diagnosis, prognosis, and assessment of response in multiple myeloma." (2016): 5428-5433.
[73] Munshi, N. et al. Expanded Meta-Analysis Confirms the Association Between MRD and Long-term Survival Outcomes in Multiple Myeloma (MM). Poster presented at American Society of Hematology (ASH). 2019
[74] Rawstron, Andy C., et al. "Minimal residual disease assessed by multiparameter flow cytometry in multiple myeloma: impact on outcome in the Medical Research Council Myeloma IX Study." *J Clin Oncol* 31.20 (2013): 2540-2547.
[75] Paiva, Bruno, et al. "Multiparameter flow cytometric remission is the most relevant prognostic factor for multiple myeloma patients who undergo autologous stem cell transplantation." *Blood, The Journal of the American Society of Hematology* 112.10 (2008): 4017-4023.
[76] Bakkus, Marleen HC, et al. "Post-transplantation tumour load in bone marrow, as assessed by quantitative ASO-PCR, is a prognostic parameter in multiple myeloma." *British journal of haematology* 126.5 (2004): 665-674.
[77] Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." *Blood, The Journal of the American Society of Hematology* 100.9 (2002): 3095-3100.
[78] https://jasp-stats.org/

Research to support the relevance of MRD in multiple myeloma: SLR update

Ingress-Health

By: Margarita Kulakova

Date: Nov 2020

Version: 1.0

Ingress Health NL Hofplein 20 3032 AC Rotterdam Netherlands +31 (0)10 3330 877 info@ingress-health.com
Ingress Health GER Alter Holzhafen 19 3966 Wismar Germany +49 (0) 3841 758 1014 www.ingress-health.com

SLR PROTOCOL

We conducted an SLR update for the purpose of the meta-analysis to evaluate the predictive utility of MRD detection in patients with MM for overall survival (OS) and progression-free survival (PFS). The initial search was conducted in November 2018 with a subsequent update in June 2019. The current update includes newer publications, as well as the materials of the latest conferences.

The SLR was updated from the following sources:

- MEDLINE (via PubMed)
- EMBASE
- Conference proceedings: ASH, ASCO, and EHA.

For database searches, we used the same search syntax as was used in the initial SLRs to ensure consistency. All searches were conducted within the data range from June 1, 2019 to present (October 24, 2020). Latest conference proceedings may not be indexed by Embase yet. Therefore, we conducted a hand search at the websites of all conferences to ensure that all relevant material were identified.

The search syntax for Pubmed is presented in the table below. The same syntax was adapted for Embase.

Table 1. MEDLINE Search Strategy

Category	No.	Search syntax	Hits
Disease terms	#1	("Multiple myeloma" [Mesh] OR Multiple myeloma [Title/Abstract] OR Kahler disease [Title/Abstract] OR Kahler's disease [Title/Abstract] OR Myelomatosis [Title/Abstract] OR Plasma cell myeloma [Title/Abstract])	53,040
Disease setting terms	#2	(newly diagnosed[Title/Abstract] OR newly-diagnosed[Title/Abstract] OR untreated[Title/Abstract] OR naïve[Title/Abstract])	305,582
Intervention terms	#3	(transplant*[tiab] OR SCT[tiab] OR ASCT[tiab] OR autoSCT[tiab] OR NDMM-TE[tiab])	489,132
MRD terms	#4	(MRD [All Fields] OR minimal residual disease [All Fields] OR "neoplasm, residual" [MeSH] OR (response [Title/Abstract] AND (flow cytometry [Title/Abstract] OR next generation flow [Title/Abstract] OR polymerase chain reaction [Title/Abstract] OR ASO-qPCR [Title/Abstract] OR next-generation sequencing [Title/Abstract])))	73,410
Language	#5	English[lang]	26,959,370
Time range	#6	("2019/06/01"[PDAT] : "2020/10/24"[PDAT])	2,025,488
Final term	#5	(#1 AND #2 AND #3 AND #4) AND #5 AND #6	81

Each abstract was reviewed by two independent investigators to determine its relevance for inclusion in the SLR. Disagreements between these investigators were resolved by a third investigator. All publications rejected were assigned a reason for exclusion.

The publications were included based on the eligibility criteria described in the table below.

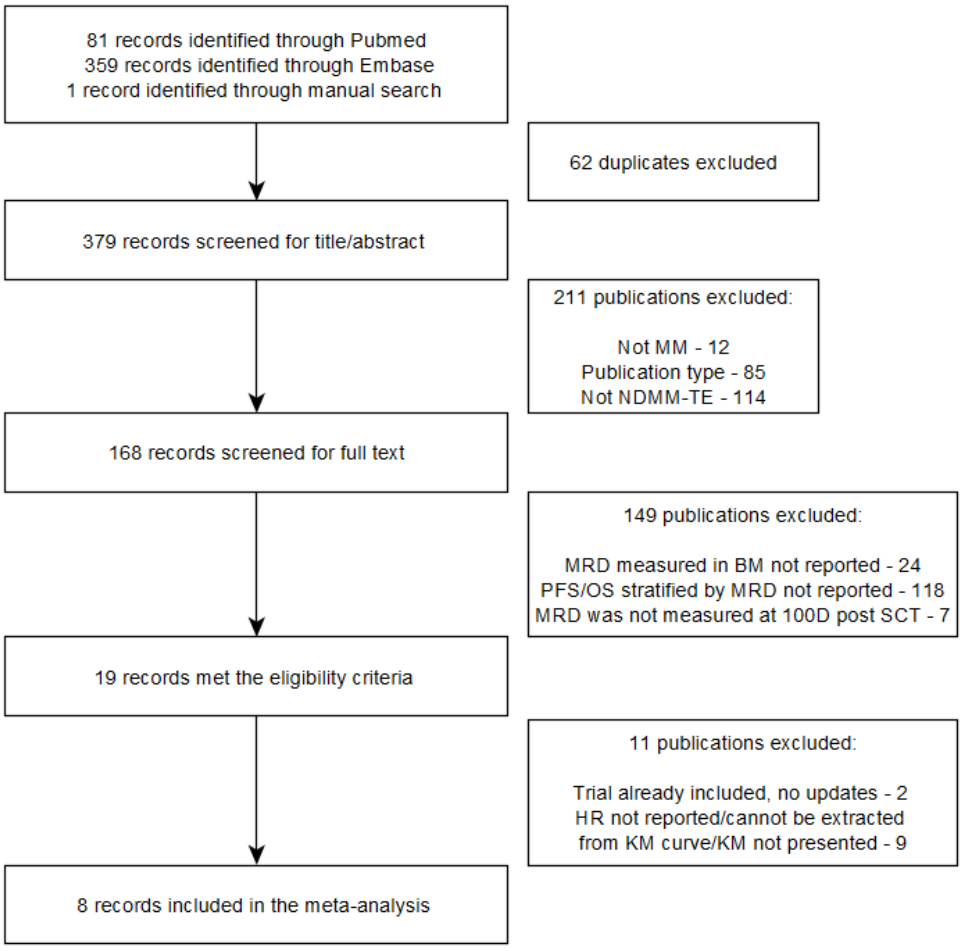
Table 2. MRD SLR eligibility criteria

	Inclusion Criteria	**Exclusion Criteria**
SLR on survival outcomes by MRD in MM		
Population	Patients with MM	Patients without a primary diagnosis of MM
Intervention/comparator	Any treatment	Allo-SCT
Outcomes	OS and/or PFS stratified by MRD status (using any MRD definition)	Survival data that cannot be extracted or is not available MRD measured in peripheral blood (PB) MRD assessed by PET-CT
Study Design	RCTs and non-RCT study design	Economic models, case reports, comments and editorials, animal/in-vitro studies
Date Limit	• No date limit applied on indexed databases search • Conference abstract and other materials (grey literature): 1.5 years (June 2019-Oct 2020): EHA, ASH, ISPOR, ASCO.	• Conference abstract or other materials (grey literature) published before 2019 and/or included in the previous search
Language	English language	Non-English language
Additional selection criteria		
Population	Newly diagnosed transplant-eligible patients	Relapsed/refractory patients, transplant-ineligible patients.
Intervention	• Transplant • Treatments representing standard of care (SoC)#	• Transplant was not performed • Studies with D-VTd (CASSIOPEIA D-VTd arm)
Outcomes	MRD measured at 100 days post-SCT	MRD measured at a different timepoint

RESULTS OF THE SLR

The figure below presents the PRISMA diagram. The total number of database hits was 440 publications. One additional publication relevant to the analysis was identified through a manual search.

Figure 1. PRISMA diagram



D: days. HR: hazard ratio. KM: Kaplan-Meier. MRD: minimal residual disease. NDMM-TE: newly diagnosed transplant-eligible multiple myeloma. SCT: autologous stem cell transplant.

19 publications met the eligibility criteria. Of those, 11 were excluded from the meta-analysis, as described in the table below.

Table 3. List of studies that met the eligibility criteria, but were not included in the meta-analysis

Publication	Study ID	Reason for exclusion from the meta-analysis
Moreau (2019)[1]	CASSIOPEIA	Trial already included. Data on file was used; the publication does not provide any new information.
Gay (2019)[2]	FORTE	HR for PFS or OS not reported; KM curve not presented.
Medina (2019)[3]	GEM2012MENOS65	Trial already included as Paiva (2020). The publication does not provide any new information.
Mookerjee (2019)[4]	NA	HR for PFS or OS not reported; KM curve not presented.
Salgado (2019)[5]	NA	HR for PFS or OS not reported; KM curve not presented.
Boncompagni (2019)[6]	NA	HR for PFS or OS not reported (only for relapse); KM curve not presented.
Yan (2019)[7]	BDH 2008/02	HR cannot be extracted from KM curve (patient numbers not reported).
Antonioli (2019)[8]	NA	Reports on the same study as Boncompagni (2019). HR for PFS or OS not reported (only for relapse); KM curve not presented.
Solovev (2019)[9]	NA	HR for PFS or OS not reported; KM curve not presented.
Kunacheewa (2019)[10]	NA	HR for PFS or OS not reported; KM curve not presented.
Patel (2020)[11]	NA	HR for PFS or OS specific for post-SCT patients not reported; KM curve for post-SCT patients cannot be extracted.

HR: hazard ration. KM: Kaplan-Meier. MRD: minimal residual disease. NA: not applicable. OS: overall survival. PFS: progression-free survival.

Eight publications were eligible for the inclusion in the meta-analysis, adding five PFS and two OS observations in the base case. Another two publications were included in the sensitivity analysis; in this study, only 83.5%/84.5%[1] patients underwent SCT, which may bias the outcomes. The assumptions and limitations associated with the study inclusion are presented in the table below.

Table 4. List of studies included in the meta-analysis (base case and sensitivity)

Publication	Study ID	Notes and limitations
Publications included in the base case		
Luoma (2019)[12]	A study of the Finnish Myeloma Group	-
Ribolla (2020)[13]	NA	-
Chan (2019)[14]	PADIMAC	The study has been included as Popat (2017). This publication contains an update (OS value)
Parrondo (2019)[15]	NA	Timing of MRD assessment was described as: "after ASCT." We assume that this corresponds to 100 days post-ASCT.
Paiva (2020)[16]	TOURMALINE-MM3	MRD was assessed at screening, which, according to TOURMALINE-MM3 study protocol, corresponded to ≥75 days post-SCT. This was assumed to be representative of 100 days post-SCT.
Garifullin (2019)[17]	NA	Timing of MRD assessment was described as: "after autoSCT". We assume that this corresponds to 100 days post-ASCT.
Publications included in the sensitivity analysis		
Alonso (2020)[18] Fernandez (2019)[19]	NA	ASCT was performed in 83.5% (or 84.5%) of the patients. The population was assumed representative of NDMM-TE, however, the inclusion of non-SCT patients may bias the results.

ASCT: autologous stem cell transplant. MRD: minimal residual disease. NA: not applicable. NDMM-TE: newly diagnosed transplant-eligible multiple myeloma.

The study characteristics of the final pool of studies (including the initial and the updated analyses) are described in the table below.

[1] This number is reported as 84.5% in Fernandez (2019).

Table 5. Study characteristics

Author	Trial ID	Study design	Number of patients	Age (median)	Male (%)	ISS	Type of measurable disease	Cytogenetic profile (high risk, %)
[Data on file][20]	CASSIOPEIA (VTd arm)	RCT	542 MRD measured in 542	58	59%	I – 42% II – 43% III – 15%	IgG – 58% IgA – 18% Other – 4%	16%
Rawstron, 2013[21]	MRC Myeloma IX Study	RCT	1,111 MRD measured in 397	59	62%	I - 25% II - 34% III - 31% NR – 10%	IgG – 60% IgA – 22% Other – 17%	47%
Paiva, 2008[22]	PETHEMA /GEM2000 / NCT00560053	Prospective single-arm	295 MRD measured in 295	59	54%	I – 39% II – 41% III – 20%	IgG – 55% IgA – 27% Other – 18%	16%
Popat, 2017[23] Chan (2019)[14]	PADIMAC	Prospective single-arm	153 MRD measured in 27	55	NR	NR	NR	33%
Cohen, 2016[24]	NR	Prospective single-arm	40 MRD measured in 19	61	55%	I – 43% II – 10% III – 5% NR – 43%	IgG – 55% IgA – 23% Other – 23%	NR
Clark, 2018[25]	NCT01215344	Prospective single-arm	32 MRD measured in 25	57	68%	I – 52% II – 44% III – 4%	IgG – 68% IgA – 12% Other – 20%	32%
Chakraborty, 2017[26]	NR	Retrospective	185 MRD measured in 185	61	54%	I – 17% II – 42% III – 24% NR – 17%	NR	100%
Schinke, 2017[27]	TT3b-TT5a	RCT	883 MRD measured in 109	60	63%	I – 33% II – 40% III – 27%	IgG – 58% IgA – 21% Other – 21%	NR
Bakkus, 2004[28]	EBMT Phase III	RCT	67 MRD measured in 60	54	58%	NR	IgG – 60% IgA – 35% Other – 5%	NR
Rawstron, 2002 [29]	MRC Myeloma VII	RCT	45 MRD measured in 45	55	53%	NR	NR	NR
Gu, 2018 [30]	NR	Prospective single-arm	104 MRD measured in: NR	54	67%	I – 37% II – 38% III – 25%	IgG – 56% IgA – 14% Other – 30%	59%
Rossi, 2018[31]	NR	Prospective single-arm	30 MRD measured in 30	58	53%	I – 83% II – 7% III – 10%	NR	25%

Hahn, 2019 [32]	STAMiNA / PRIMER, NCT01109004	RCT	437 MRD measured in 311	NR	NR	I – 33% II – 33% III – 29%	NR	NR
Solovev, 2018[33]	NR	Prospective single-arm	70 MRD measured in 37	56	66%	I – 40% II – 27% III – 33%	NR	NR
Solovev, 2016 [34]	NR	Prospective single-arm	52 MRD measured in 52	54	37%	NR	NR	NR
Luoma, 2019 [12]	A study of the Finnish Myeloma Group	RCT	80 MRD measured in 80	63	53%	I – 26% II – 55% III – 19%	IgG – 64% IgA – 20% Other – 16%	14%
Ribolla, 2020 [13]	NA	Prospective single-arm	83 MRD measured in 83	60	57%	NR	NR	24%
Parrondo, 2019 [15]	NA	Retrospective	110. MRD measured in 110	63	46%	I – NR II – NR III – 28%	NR	100%
Paiva, 2020 [16]	TOURMALINE-MM3	RCT	656. MRD measured in 582	58	63%	I – 37% II – 34% III – 29%	IgG – 58% IgA – 22% Other – 20%	18%
Garifullin, 2019 [17]	NA	Prospective single-arm	89. MRD measured in 39 post-transplant patients	58	NR	NR	NR	NR
Alonso, 2020 [18] Fernandez, 2019 [19]	NA	Retrospective	139. MRD measured in 139	59	52%	I – 33% II – 30% III – 36%	IgG – 60% IgA – 22% Other – 18%	37%

ISS: International Staging System for Multiple Myeloma; NR: not reported

Table 6. Study characteristics (continued)

Author	Treatment	Eligibility for MRD assessment by response	MRD assay and sensitivity
CASSIOPEIA (VTd arm)	VTd (D-VTd arm excluded from the analysis)	Any response	MFC. 10^{-5}
Rawstron, 2013	CTD or CVAD + ASCT	Any response	MFC. 10^{-4}
Paiva, 2008	VBMCP/VBAD induction + ASCT + melphalan consolidation	Any response	MFC. 10^{-4}
Popat, 2017	PAD induction + ASCT	Only CR (PFS)	MFC. 10^{-4}
Chan, 2019		VGPR (OS)	MFC. 10^{-4}
Cohen, 2016	ASCT + Bortezomib consolidation	NA	MFC. 10^{-4}
Clark, 2018	VRD induction + ASCT	Any response	MFC. 10^{-4}
Chakraborty, 2017	ASCT (other therapies unknown)	Any response	MFC. 10^{-4}
Schinke, 2017	VTD and PACE chemo + ASCT + VRD maintenance	At least VGPR	NGS. 10^{-5}
Bakkus, 2004	Conventional chemo + ASCT	Any response	PCR. 10^{-4}
Rawstron, 2002	C-VAMP and HDT induction + ASCT	Any response	MFC. 10^{-4}
Gu, 2018	Bortezomib induction + ASCT + thalidomide/lenalidomide and/or IFN-a	At least VGPR	MFC. 10^{-5}
Rossi, 2018	PAD induction	At least VGPR	MFC. 10^{-4}
Hahn, 2019	ASCT + RVD consolidation + LEN maintenance	Any response	MFC. 10^{-5}
Solovev, 2018	Bortezomib induction + ASCT +/- LEN maintenance	Only CR	MFC. 10^{-5}
Solovev, 2016	ASCT + Bortezomib maintenance	CR	MFC. Sensitivity: NR
Luoma, 2019	RVD induction + ASCT + LEN maintenance	CR	MFC. 10^{-4}
Ribolla, 2020	VTd induction + MEL conditioning + ASCT + KRd/Cy-Dex consolidation +/- LEN maintenance	NR	MFC. 10^{-5}
Parrondo, 2019	Carfilzomib/bortezomib/triplet induction + ASCT	NR	NR
Paiva, 2020	SoC induction + MEL conditioning + ASCT +/- IMiD	At least VGPR	MFC. 10^{-5}
Garifullin, 2019	VD/CVD/VMP/PAD/RD/VRD/ThalD/PomD/chemotherapy induction + MEL conditioning + ASCT	Any response	MFC. 10^{-4}
Alonso, 2020 Fernandez, 2019	VMP/CVD/BPV/VTD/Vel/Dex/VRd induction + ASCT (83.5%) + LEN maintenance	CR	NGS. 10^{-4}

ASCT (Autologous Stem Cell Transplantation); BPV (bendamustine, prednisone and bortezomib); C-VAMP (vincristine, amethopterine, methotrexate, and prednisone + cyclophosphamide); CTD (cyclophosphamide, thalidomide, and dexamethasone); CVAD (cyclophosphamide, vincristine, doxorubicin, and dexamethasone); CVD (bortezomib, cyclophosphamide and dexamethasone); Cy-Dex (cyclophosphamide plus dexamethasone); HDT (high-dose chemotherapy); IMiD (Immunomodulatory drugs); KRd (carfilzomib-lenalidomide-dexamethasone); LEN (lenalidomide); MEL (melphalan); PACE (cisplatin, doxorubicin, cyclophosphamide, etoposide); PAD (bortezomib, doxorubicin and dexamethasone); PI (protease inhibitors); RVD (lenalidomide, bortezomib, and dexamethasone); SoC (standard of care); VBMCP (vincristine, carmustine, melphalan, cyclophosphamide, and prednisone); VD (Bortezomib - Dexamethasone); Vel/Dex (Bortezomib/dexamethasone); VMP (bortezomib, melphalan, and prednisone); VRD (bortezomib, lenalidomide, and dexamethasone); VTD (bortezomib, thalidomide, and dexamethasone).

RESULTS OF THE META-ANALYSIS

The meta-analysis was repeated using the same methodology as the meta-analysis previously reported. A random effects model was fitted to obtain a pooled effect estimate of HR for MRD negativity. Heterogeneity in design and population among the studies eligible for meta-analysis was assessed with I^2 test using the maximum likelihood (ML) estimator. The statistical significance level was set at $p<0.05$.

The input data are presented in the Appendix. The analysis was conducted on log-transformed values.

The base case analysis for PFS was based on 20 observations and resulted in the HR of [redacted] The figure below presents the forest plot. The heterogeneity was estimated at [redacted] with a p-value of [redacted] based on a chi-square distribution.

Figure 2. Forest plot: PFS (base case)

The base case analysis for OS was based on 11 observations and resulted in the HR of [redacted]. The forest plot is presented in the figure below. The heterogeneity was estimated at [redacted], with a p-value of [redacted] based on a chi-square distribution.

Figure 3. Forest plot: OS (base case)

Subgroup analyses were repeated for the updated dataset, to adjust for effect modifiers (variables that significantly impact the association of MRD and PFS/OS). For PFS, two additional sensitivity analyses were conducted: one to include the study where only 83.5/84.5% of patients underwent ASCT (Alonso, 2020 / Fernandez, 2019) and another to replace the after ASCT assessment with post-ASCT+ consolidation in Parrondo (2019), as both timepoints may correspond to 100 days post-SCT.

The results for PFS and OS subgroup/sensitivity analyses are presented in Table 7 and Table 8 below. All subgroup analyses show that the survival is positively associated with MRD-negativity.

Table 7. Subgroup analyses: PFS

Subgroup	Number of observations	Outcome
Base case	**n=20**	[redacted]
Sensitivity		
Sensitivity of 10-4	n=11 [21,28,22,29,24,23,26,25,31,12,17]	[redacted]
Sensitivity of 10-5	n=7 [27,30,33,32,13,16,20]	[redacted]
Sensitivity of 10-6	No observations	NA
Eligibility for MRD assessment by response criteria		
Only patients achieving CR	n=4 [34,23,33,12]	[redacted]
Patients achieving at least VGPR	n=4 [16,27,30,31]	[redacted]
Other sensitivity analyses		
Alonso (2020)/Fernandez (2019) included	n=21	[redacted]

HR replaced with post-SCT and consolidation timepoint in Parrondo (2019). HR = 0.31 (95% CI: 0.09 – 0.99)	n=20	[redacted]

Table 8. Subgroup analyses: OS

Subgroup	Number of observations	Outcome
Base case	**n=11**	[redacted]
Sensitivity		
Sensitivity of 10-4	n=6 [29,28,22,21,26,14]	[redacted]
Sensitivity of 10-5	n=5 [27,30,32,20,13]	[redacted]
Sensitivity of 10-6	No observations	NA
Eligibility for MRD assessment by response criteria		
Only patients achieving CR	No observations	NA
Patients achieving at least VGPR	n=3 [14,27,30]	[redacted]

APPENDIX

Table 9. Input data for PFS MRD analysis

Author	HR PFS	HR PFS 95% LL	HR PFS 95% UL	HR PFS log	HR PFS SE log
Rawstron, 2013	0.557	0.415	0.746	-0.58609	0.149741
Paiva, 2008	0.275	0.16	0.4	-1.29098	0.233748
Popat, 2017	0.38	0.13	1.15	-0.96129	0.562718
Cohen, 2016	0.277	0.102	0.752	-1.28371	0.50923
Clark, 2018	0.133	0.022	0.813	-2.01757	0.923329
Chakraborty, 2017	0.45	0.31	0.66	-0.79851	0.192772
Schinke, 2017	0.55	0.22	1.37	-0.60624	0.468501
Bakkus, 2004	0.274	0.136	0.552	-1.29426	0.357509
Rawstron, 2002	0.308	0.1296	0.732	-1.17766	0.441665
Gu, 2018	0.29	0.13	0.65	-1.23787	0.410571
Rossi, 2018	0.13	0.03	0.51	-2.07	0.71
CASSIOPEIA (VTD)	0.4615	0.3143	0.6776	-0.77327	0.195972
Hahn, 2019	0.48	0.31053	0.741957	-0.73	0.22
Solovev, 2018	0.4966	0.2094	2	-0.7	0.58
Solovev, 2016	0.588	0.294	0.833	-0.53	0.27
Luoma, 2019	0.232	0.112	0.48	-1.46102	0.371247
Ribolla, 2020	0.20	0.07	0.54	-1.60944	0.521192
Parrondo, 2019	0.44	0.23	0.86	-0.82098	0.336442
Paiva, 2020	0.118	0.075	0.189	-2.13707	0.23578
Garifullin, 2019	0.326	0.827	11.39	-1.12086	0.669053

Table 10 Input data for OS MRD analysis

Author	HR OS	HR OS 95% LL	HR OS 95% UL	HR OS log	HR OS SE log
Rawstron, 2013	0.6269	0.4344	0.9046	-0.46697	0.187124
Paiva, 2008	0.49505	0.42	0.711	-0.7031	0.13429
Chakraborty, 2017	0.55	0.32	0.92	-0.59784	0.269401
Schinke, 2017	0.8411	0.2962	2.388	-0.17304	0.532443
Bakkus, 2004	0.5873	0.209	1.651	-0.53222	0.527245
Rawstron, 2002	0.5345	0.1629	1.755	-0.62642	0.6064
Gu, 2018	0.23	0.09	0.59	-1.46968	0.479672
Hahn, 2019	0.77	0.354128	1.674255	-0.26136	0.396292
CASSIOPEIA (VTD)	0.7092	0.5511	0.9126	-0.34362	0.128669
Ribolla, 2020	0.20	0.04	0.99	-1.60944	0.818578
Chan, 2019	0.34	0.10	1.20	-1.07881	0.633905

REFERENCES

[1] Moreau, Philippe, et al. "Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label, phase 3 study." *The Lancet* 394.10192 (2019): 29-38.

[2] Gay, Francesca, et al. "Integrative analysis of baseline prognostic features and achievement of minimal residual disease negativity as predictors of early relapse in transplant-eligible multiple myeloma patients." (2019): 605-605.

[3] Medina, Alejandro, et al. "Prognostic implications of MRD assessment in multiple myeloma patients: comparison of Next-Generation Sequencing and Next-Generation Flow." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e47.

[4] Mookerjee, Anjali, et al. "Post-transplant minimal residual disease assessment in Multiple myeloma." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e180.

[5] Salgado, Anna Beatriz, et al. "Minimal residual disease and quality sample evaluation by Next Generation Flow cytometry in multiple myeloma patients: a Brazilian experience." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e179.

[6] Boncompagni, Riccardo, et al. "Minimal Residual Disease (MRD) ratio before and after Autologous Stem Cell Transplantation (ASCT) in Multiple Myeloma." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e172.

[7] Yan, Y., et al. "PF593 THE IMPACT OF RESPONSE KINETICS FOLLOWING INITIAL THERAPY FOR MULTIPLE MYELOMA IN THE ERA OF NOVEL AGENTS." *HemaSphere* 3.S1 (2019): 249.

[8] Indexed via Embase, however, reference not found.

[9] Solovev, Maxim, et al. "The Value of Minimal Residual Disease as a Progression-Free Survival Predictor After Autologous Hematopoietic Stem Cell Transplantation in Multiple Myeloma Patients." *Clinical Lymphoma, Myeloma and Leukemia* 19 (2019): S324.

[10] Kunacheewa, Chutima, et al. "Correlation of minimal residual disease negativity (MRD-) with prognosis in high-risk myeloma (HRM)." (2019): 8024-8024.

[11] Patel, Dilan A., et al. "Minimal residual disease negativity and lenalidomide maintenance therapy are associated with superior survival outcomes in multiple myeloma." *Bone Marrow Transplantation* 55.6 (2020): 1137-1146.

[12] Luoma, Sini, et al. "RVD induction and autologous stem cell transplantation followed by lenalidomide maintenance in newly diagnosed multiple myeloma: a phase 2 study of the Finnish Myeloma Group." Annals of hematology 98.12 (2019): 2781-2792.

[13] Ribolla, Rossella et al. "REAL LIFE DATA OF PROGNOSTIC MINIMAL RESIDUAL DISEASE EVALUATION AFTER AUTOLOGOUS STEM CELL TRANSPLANTATION IN NEWLY DIAGNOSED MULTIPLE MYELOMA PATIENTS." EHA Library. Ribolla R. 06/12/20; 297949; PB2033

[14] Chan, Wei Yee, et al. "Outcomes of newly diagnosed multiple myeloma with deferred ASCT after achieving≥ VGPR following PAD induction in the Phase 2 PADIMAC study." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e288-e289.

[15] Parrondo, Ricardo, et al. "Carfilzomib-based induction improves progression free survival and rate of minimal residual disease negativity following autologous transplant in high risk multiple myeloma patients: a retrospective analysis." *Clinical Lymphoma, Myeloma and Leukemia* 19.10 (2019): e217.

[16] Paiva, Bruno, et al. "UNDETECTABLE MEASURABLE RESIDUAL DISEASE (MRD) AS ENDPOINT OF MAINTENANCE THERAPY: RESULTS FROM THE TOURMALINE-MM3 TRIAL OF IXAZOMIB VS PLACEBO MAINTENANCE IN NEWLY DIAGNOSED MULTIPLE MYELOMA (NDMM)". EHA Library. Paiva B. 06/12/20; 294849; EP932

[17] Garifullin, A., et al. "PB2359 THE ROLE OF AUTOLOGOUS STEM CELL TRANSPLANTATION IN INCREASING THE EFFICIENCY OF TREATMENT IN PATIENTS WITH MULTIPLE MYELOMA." *HemaSphere* 3.S1 (2019): 1051.

[18] Alonso, Rafael, et al. "Prolonged lenalidomide maintenance therapy improves the depth of response in multiple myeloma." *Blood Advances* 4.10 (2020): 2163-2171.

[19] Fernández, Rafael Alonso, et al. "Prolonged Lenalidomide Maintenace Therapy Improves the Quality and Deep of Response in Multiple Myeloma." Clinical Lymphoma, Myeloma and Leukemia 19.10 (2019): e285-e286.

[20] Data on file
[21] Rawstron, Andy C., et al. "Minimal residual disease assessed by multiparameter flow cytometry in multiple myeloma: impact on outcome in the Medical Research Council Myeloma IX Study." *J Clin Oncol* 31.20 (2013): 2540-2547.
[22] Paiva, Bruno, et al. "Multiparameter flow cytometric remission is the most relevant prognostic factor for multiple myeloma patients who undergo autologous stem cell transplantation." *Blood, The Journal of the American Society of Hematology* 112.10 (2008): 4017-4023.
[23] Popat, Rakesh, et al. "Outcomes of stratification to ASCT or not based on depth of response: results of a phase 2 trial assessing the impact of minimal residual disease (MRD) in multiple myeloma patients with deferred ASCT (PADIMAC)." *Blood* 130.Supplement 1 (2017): 1864-1864
[24] Cohen, Oliver C., et al. "Bortezomib Consolidation Following Upfront ASCT for Multiple Myeloma Deepens Disease Response and MRD-Negative Rate without Compromising Response to Subsequent Bortezomib Salvage: Results of a Phase II Study." (2016): 4508-4508.
[25] Clark, C. Amos, et al. "Prospective trial of minimal residual disease assessment by multiparametric flow cytometry for multiple myeloma in the era of bortezomib-based chemotherapy." *Bone marrow transplantation* 53.12 (2018): 1589-1592.
[26] Chakraborty, Rajshekhar, et al. "Impact of post-transplant response and minimal residual disease on survival in myeloma with high-risk cytogenetics." *Biology of Blood and Marrow Transplantation* 23.4 (2017): 598-605.
[27] Schinke, Carolina, et al. "The prognostic value of the depth of response in multiple myeloma depends on the time of assessment, risk status and molecular subtype." *Haematologica* 102.8 (2017): e313.
[28] Bakkus, Marleen HC, et al. "Post-transplantation tumour load in bone marrow, as assessed by quantitative ASO-PCR, is a prognostic parameter in multiple myeloma." *British journal of haematology* 126.5 (2004): 665-674.
[29] Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." *Blood, The Journal of the American Society of Hematology* 100.9 (2002): 3095-3100.
[30] Gu, Jingli, et al. "Longitudinal flow cytometry identified “minimal residual disease”(MRD) evolution patterns for predicting the prognosis of patients with transplant-eligible multiple myeloma." *Biology of Blood and Marrow Transplantation* 24.12 (2018): 2568-2574.
[31] Rossi, Giovanni, et al. "Minimal residual disease and log-reduction of plasma cells are associated with superior response after double autologous stem cell transplant in younger patients with multiple myeloma." *Cytometry Part B: Clinical Cytometry* 96.3 (2019): 195-200.
[32] Hahn, Theresa E., et al. "Minimal residual disease (MRD) assessment before and after autologous hematopoietic cell transplantation (AutoHCT) and maintenance for multiple myeloma (MM): results of the prognostic immunophenotyping for myeloma response (PRIMeR) study." *Biology of Blood and Marrow Transplantation* 25.3 (2019): S4-S6.
[33] Solovev, Maxim V., et al. "Efficacy of maintenance therapy following auto-HSCT depending on MRD status in patients with multiple myeloma." *Blood* 132.Supplement 1 (2018): 3432-3432.
[34] Solovev, Maxim V., et al. "Maintenance Therapy after Autologous Haematopoietic Stem Cell Transplantation (auto-HSCT) in Multiple Myeloma Patients with and without Minimal Residual Disease (MRD)." (2016): 2260-2260.

MRD analysis on newly diagnosed transplant-eligible patients diagnosed with multiple myeloma

Risk of bias assessment of the studies included in the meta-analysis

Summary

The risk of bias assessment was conducted using the ROBINS-I tool developed by Cochrane. This tool is specific for non-randomized studies, however, Janssen consider it suitable for this data as MRD assessment is always conducted on unrandomized population, as it implies dividing the patients within a treatment arm based on their response to the treatment. In addition, MRD analysis is often conducted post-hoc on bone marrow samples and/or on population pooled from different treatment arms.

The table below summarize the definitions used in ROBINS-I tool in relation to the analysis. The goal of the analysis is to obtain a pooled HR estimate to assess the effect of MRD-negative status on survival. Therefore, for the purpose of the assessment, MRD-negativity was considered an "intervention" and MRD-positive status as "comparator", as referred in ROBINS-I.

Participants	Newly diagnosed transplant-eligible patients with multiple myeloma
Experimental intervention	MRD-negative status
Comparator	MRD-positive status
Outcomes	PFS and OS Hazard ratio (HR)

The list of confounders that are relevant to the analysis are the following:

1. Eligibility for MRD assessment by conventional response

MRD can be assessed only in patients achieving at least CR, or only those achieving at least VGPR, or in all patients. HR $^{mrd-}$ measured in a group of patients achieving CR or better is less favourable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.

2. Definition of MRD-negativity based on the sensitivity threshold

MRD-negativity can be defined at sensitivity of 10[-4], 10[-5] or 10[-6]. Higher sensitivity threshold favors MRD-negativity.

3. MRD assay

MRD can be measured using next-generation sequencing (NGS), multiparameter flow cytometry (MFC), or polymerase chain reaction (PCR) assay. Adjustment for this variable is not always warranted, as it is highly correlated with sensitivity (as such, NGS assay usually has a sensitivity of 10[-6], while PCR only achieves 10[-4] - 10[-5]).

4. Treatment

While certain modern treatments result in higher rates of MRD-negativity than older treatments, it is still unknown whether they have an impact on the relative effect of MRD-negativity on survival. However, Janssen does not expect it to be an important source of bias in the analysis, as the study selection is limited to the transplant-eligible population.

Most studies were of moderate risk of bias. First, this was related to incomplete reporting of baseline characteristics or statistical methods, as many included publications are conference abstracts and do not contain all the necessary information. For example, a mix of patients with different conventional response may introduce a selection bias, or the use of adjusted Cox may change the HR outcome. Second, all the studies where HR was not reported and had to be estimated based on simulated individual patient data (IPD) and reconstructed KM curve were considered of moderate risk. It should be noted that the simulated KM curves were visually checked against the published curves and the number of progression/death and censoring events were validated, when possible. However, the exact magnitude or direction of the bias introduced by the manual curve extraction is still unknown. Another important source of bias was the selection of patients with any conventional response for the MRD assessment.

The table below summarises the risk of bias assessment. CASSIOPEIA study was not included in the assessment, as the analysis was conducted on in-house IPD.

Table 1 Summary of risk of risk of bias assessment

Publication	Risk of bias	Comments
Solovev (2016) [1]	Low	The authors reported that the compared groups were overall balanced for known prognostic factors. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. HR was reported in the publication, however, the method of HR estimation was not specified.
Solovev (2018) [2]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. However, the baseline characteristics were not well described.
Shinke (2017) [3]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was uniform in terms of known confounders: all patients achieved at least VGPR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. The authors conducted several subgroup analyses.
Bakkus (2004) [4]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.
Rawstron (2002) [5]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision.

		Both direction and magnitude of any potential bias were discussed in the publication.
Paiva (2008) [6]	Low	HR was reported, statistical methods were clearly described. Baseline characteristics were reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted multiple subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.
Popat (2017) [7]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. Potential sources of confounding were described but not addressed. However, the population was uniform in terms of known confounders: all patients achieved at least CR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment.
Cohen (2016) [8,9]	Moderate	HR was reported; however, the statistical methods were not clearly described. The population was not uniform in terms of the treatment received. The conventional responses are unknown. The authors did not conduct subgroup or interaction analyses, or discuss the limitations of the MRD analysis.
Clark (2018) [10]	Low	HR was reported, statistical methods were clearly described. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors did not conduct subgroup and interaction analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.
Chakraborty (2017) [11]	Low	HR was reported, statistical methods were clearly described. Background characteristics were clearly described. The population was uniform in terms of cytogenetic risk, but not conventional response. It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors conducted subgroup analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.
Gu (2018) [12]	Low	HR was reported, statistical methods were clearly described. Background characteristics were extensively reported. The population was uniform in terms of most of the confounders. The authors conducted subgroup analyses on MRD. Both direction and magnitude of any potential bias were discussed in the publication.
Hanh (2019) [13]	Moderate	HR was reported, but statistical methods were not clearly described. Baseline characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors did not conduct subgroup and interaction analyses for MRD, other than time of MRD assessment. Neither direction nor magnitude of a potential bias (other than time of assessment) were discussed in the publication.

Rossi (2018) [14]	Moderate	HR was reported. Background characteristics were reported. The population was uniform in terms of most of the confounders (conventional response, assay, sensitivity, treatment). The authors conducted subgroup analyses on time of MRD assessment only, however, the subgroups were not well defined. Both direction and magnitude of other potential bias were discussed in the publication, but not addressed in the analyses.
Rawstron (2013) [15]	Moderate	HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics of patients included in the MRD assessment were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.


Risk of bias domains
D1
D2
D3
D4
D5
D6
D7
Overall
Study
Solovev 2016
Solovev 2018
Shinke 2017
Bakkus 2004
Rawstron 2002
Paive 2008
Popat 2017
Cohen 2016
Clark 2018
Chakraborty 2017
Gu 2018
Hanh 2019
Rossi 2018
Rawstron 2013
Domains:
D1: Bias due to confounding.
D2: Bias due to selection of participants.
D3: Bias in classification of interventions.
D4: Bias due to deviations from intended interventions.
D5: Bias due to missing data.
D6: Bias in measurement of outcomes.
D7: Bias in selection of the reported result.
Judgement
Serious
Moderate
Low
No information

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool

Version 19 September 2016

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	Newly diagnosed transplant-eligible patients with multiple myeloma
Experimental intervention	MRD-negative status
Comparator	MRD-positive status
Outcomes	PFS and OS Hazard ratio (HR)

List the confounding domains relevant to all or most studies

Eligibility for MRD assessment by conventional response (MRD only assessed in patients achieving at least CR/at least VGPR/all patients), treatment, assay, sensitivity.

List co-interventions that could be different between intervention groups and that could impact on outcomes

Not applicable

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

NOT APPLICABLE

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
Definition of MRD-negativity based on the sensitivity threshold	MRD-negativity defined at sensitivity of 10[-4], 10[-5] or 10[-6]	Yes	Yes	Higher sensitivity threshold favors MRD-negativity
MRD assay	MRD measured with NGS, MFC, or PCR.	NA	No, this variable is highly correlated with sensitivity.	NA
Eligibility for MRD assessment by conventional response	MRD only assessed in patients achieving at least CR or at least VGPR or in all patients	Yes	Yes	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.
Treatment	Treatment, eligibility for transplant	NA. It is unknown whether treatments impact the effect of achieving MRD-negativity on survival.	No	The given selection of studies is homogenous enough in terms of treatment, as all patients are eligible for transplant.

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important.

NOT APPLICABLE

Risk of bias assessment. Solovev (2016)

1. **Solovev, Maxim V., et al. "Maintenance Therapy after Autologous Haematopoietic Stem Cell Transplantation (auto-HSCT) in Multiple Myeloma Patients with and without Minimal Residual Disease (MRD)." (2016): 2260-2260.**

   Outcome: PFS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PN	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	-	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	-	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low The authors reported that the compared groups were overall balanced for known prognostic factors.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low Criteria for MRD assessment were clearly described.	Low / Moderate / Serious / Critical / NI

Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	PN	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Moderate The choice of method and sensitivity threshold were not explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	The use of more sensitive assay or higher sensitivity threshold could result in a bigger difference in survival between MRD- and MRD+ groups.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI

If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI

5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients were measured for MRD. The authors report the eligibility criteria, and the sources and methods of selection of participants for MRD assessment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of survival endpoints was not reported. However, this is not expected to bias the results.	Low / Moderate / Serious / Critical / NI

Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable. Different definitions of PFS/OS (from MRD assessment vs. first diagnosis) can change t=0 of the survival curves.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR estimation method not described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	For example, the use of adjusted Cox could have changed the HR outcome. However, the method is not clearly described.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Low The authors reported that the compared groups were overall balanced for known prognostic factors. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. HR was reported in the publication, however, the method of HR estimation was not specified.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Solovev (2018)

Solovev, Maxim V., et al. "Efficacy of maintenance therapy following auto-HSCT depending on MRD status in patients with multiple myeloma." Blood 132.Supplement 1 (2018): 3432-3432.

Outcome: PFS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PN	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	PY	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate The authors did not report any adjustment for confounding variables. However, the population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Criteria for MRD assessment were clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI

4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI

5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients measured for MRD. The authors report the eligibility criteria, and the sources and methods of selection of participants for MRD assessment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Low The definition of survival endpoints was not reported. However, this is not expected to bias the results.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable. Different definitions of PFS/OS (from MRD assessment vs. first diagnosis) can change t=0 .	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	NI	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. However, the baseline characteristics were not well described.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Shinke (2017)

Schinke, Carolina, et al. "The prognostic value of the depth of response in multiple myeloma depends on the time of assessment, risk status and molecular subtype." Haematologica 102.8 (2017): e313.

Outcome: PFS and OS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PY	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Several subgroup analyses were conducted and sources of confounding were qualitatively described	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Criteria for MRD assessment were clearly described, all patients were assessed for MRD.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	The definition of MRD negativity was clearly described. The analysis was repeated using a different assay.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI

4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI

5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients measured for MRD. The authors report the eligibility criteria, and the sources and methods of selection of participants for MRD assessment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Moderate The authors did not give a clear definition of the response and the survival endpoints.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable. Different definitions of PFS/OS (from MRD assessment vs. first diagnosis) can change t=0 of the survival curves.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was uniform in terms of known confounders: all patients achieved at least VGPR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. The authors conducted several subgroup analyses.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Bakkus (2004)

Bakkus, Marleen HC, et al. "Post-transplantation tumour load in bone marrow, as assessed by quantitative ASO-PCR, is a prognostic parameter in multiple myeloma." British journal of haematology 126.5 (2004): 665-674.

Outcome: PFS and OS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Y	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Several subgroup analyses were conducted and sources of confounding were qualitatively described	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low Criteria for MRD assessment were clearly described.	Low / Moderate / Serious / Critical / NI

Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described. The choice of the assay and sensitivity were explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI

If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI

5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients measured for MRD. The authors report the eligibility criteria, and the sources and methods of selection of participants for MRD assessment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints.	Low / Moderate / Serious / Critical / NI

Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Rawstron (2002)

Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." Blood, The Journal of the American Society of Hematology 100.9 (2002): 3095-3100.

Outcome: PFS and OS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PY	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Multivariate regression analyses were conducted and sources of confounding were qualitatively described	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate MRD definition was pre-specified. However, the criteria for participation in MRD assessment were not clearly defined. The background patient characteristics were not reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low Criteria for MRD assessment were clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions	
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2	

4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI

5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients that were measured for MRD were included in the survival analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Moderate The authors gave a clear definition of the survival endpoints. HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Paiva (2008)

Paiva, Bruno, et al. "Multiparameter flow cytometric remission is the most relevant prognostic factor for multiple myeloma patients who undergo autologous stem cell transplantation." Blood, The Journal of the American Society of Hematology 112.10 (2008): 4017-4023.

Outcome: PFS and OS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Y	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Univariate regression analyses were conducted and sources of confounding were qualitatively described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Criteria for MRD assessment were clearly described. The background patient characteristics were reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described. The choice of the assay and sensitivity were explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions	
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2	

4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	PN (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	Y	Y / PY / PN / N / NI

5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Y	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	N	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	NI	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate Less patients were included in the survival analysis than measured for MRD. The reason was not explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. Statistical methods were clearly described. HR was reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Low HR was reported, statistical methods were clearly described. Baseline characteristics were reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted multiple subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Popat (2017)

Popat, Rakesh, et al. "Outcomes of stratification to ASCT or not based on depth of response: results of a phase 2 trial assessing the impact of minimal residual disease (MRD) in multiple myeloma patients with deferred ASCT (PADIMAC)." Blood 130.Supplement 1 (2017): 1864-1864.

Outcome: PFS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PN	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	PN	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate Potential sources of confounding were described but not addressed. However, the population was uniform in terms of known confounders: all patients achieved at least CR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate Criteria for MRD assessment were clearly described. The background patient characteristics were not reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions	
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2	

4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	PN (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	Y	Y / PY / PN / N / NI

5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Y	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	N	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	NI	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate Less patients were included in the survival analysis than measured for MRD. The reason was not explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Moderate The authors did not give a clear definition of the survival endpoints. Statistical methods were not clearly described. HR was not reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable. Different definitions of PFS/OS (from MRD assessment vs. first diagnosis) can change t=0 .	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. Potential sources of confounding were described but not addressed. However, the population was uniform in terms of known confounders: all patients achieved at least CR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Cohen (2016)

Cohen, Oliver C., et al. "Bortezomib Consolidation Following Upfront ASCT for Multiple Myeloma Deepens Disease Response and MRD-Negative Rate without Compromising Response to Subsequent Bortezomib Salvage: Results of a Phase II Study." (2016): 4508-4508.
Cohen, Oliver C., et al. "Bortezomib consolidation post-ASCT as frontline therapy for multiple myeloma deepens disease response and MRD-negative rate whilst maintaining QOL and response to re-treatment at relapse." British journal of haematology (2018).

Outcome: PFS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	N	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	N	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Serious Confounding was not discussed or addressed.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate Criteria for MRD assessment were clearly described. However, the population was not uniform in terms of conventional response or treatment.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Serious The definition of MRD negativity was not clearly described. The choice of the assay and sensitivity were not explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions	
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2	

4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	PN (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	Y	Y / PY / PN / N / NI

5.3 Were participants excluded due to missing data on other variables needed for the analysis?	PN	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	Y	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	PN	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low Less patients were included in the survival analysis than measured for MRD. The reasons were explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI

Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Moderate Statistical methods were not clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Moderate HR was reported; however, the statistical methods were not clearly described. The population was not uniform in terms of the treatment received. The conventional responses are unknown. The authors did not conduct subgroup or interaction analyses, or discuss the limitations of the MRD analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Clark (2018)

Clark, C. Amos, et al. "Prospective trial of minimal residual disease assessment by multiparametric flow cytometry for multiple myeloma in the era of bortezomib-based chemotherapy." Bone marrow transplantation 53.12 (2018): 1589-1592.

Outcome: PFS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PN	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	PN	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate Confounding in MRD analysis was discussed but was not addressed in any additional analyses.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Moderate The population was not uniform in terms of conventional response (mix of CR, VGPR and less than VGPR). This has not been addressed. Criteria for MRD assessment were not clearly described.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described. However, the choice of the assay and sensitivity were not explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions

If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	PY (out of those included in MRD assessment)	Y / PY / PN / N / NI

5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients assessed for MRD were included in the survival analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI

6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. Statistical methods were clearly described. HR was reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Low HR was reported, statistical methods were clearly described. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors did not conduct subgroup and interaction analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Chakraborty (2017)

Chakraborty, Rajshekhar, et al. "Impact of post-transplant response and minimal residual disease on survival in myeloma with high-risk cytogenetics." Biology of Blood and Marrow Transplantation 23.4 (2017): 598-605.

Outcome: PFS and OS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PY	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Confounding in MRD analysis was discussed and addressed in subgroup analyses.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low MRD definition was pre-specified. The background patient characteristics were reported. The population was uniform in terms of cytogenetic risk, but not uniform in terms of conventional response (mix of CR, VGPR and less than VGPR).	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described. The choice of the assay and sensitivity were explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions

If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Y (out of those included in MRD assessment)	Y / PY / PN / N / NI

5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients assessed for MRD were included in the survival analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI

6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. Statistical methods were clearly described. HR was reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Low HR was reported, statistical methods were clearly described. Background characteristics were clearly described. The population was uniform in terms of cytogenetic risk, but not conventional response. It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors conducted subgroup analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Gu (2018)

Gu, Jingli, et al. "Longitudinal flow cytometry identified “minimal residual disease”(MRD) evolution patterns for predicting the prognosis of patients with transplant-eligible multiple myeloma." Biology of Blood and Marrow Transplantation 24.12 (2018): 2568-2574.

Outcome: PFS and OS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants’ follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Y	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	PY	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI
1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Confounding in MRD analysis was discussed and addressed in additional analyses.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI
2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?	-	NA / Y / PY / PN / N / NI
2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?		NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI

Risk of bias judgement	Low Criteria for MRD assessment were clearly described. MRD definition was pre-specified. The background patient characteristics were reported. The population was uniform in terms of conventional response (at least VGPR).	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Y	Y / PY / PN / N / NI
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
Risk of bias judgement	Low The definition of MRD negativity was clearly described. The choice of the assay and sensitivity were explained.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions

If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI
4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	NI	Y / PY / PN / N / NI

5.2 Were participants excluded due to missing data on intervention status?	Y	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	PY	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	N	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	PN	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate The number of patients included in the survival analysis is unknown.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI

6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. Statistical methods were clearly described. HR was reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
Risk of bias judgement	Low	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

Risk of bias judgement	Low HR was reported, statistical methods were clearly described. Background characteristics were extensively reported. The population was uniform in terms of most of the confounders. The authors conducted subgroup analyses on MRD. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
Optional: What is the overall predicted direction of bias for this outcome?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Hanh (2019)

Hahn, Theresa E., et al. "Minimal residual disease (MRD) assessment before and after autologous hematopoietic cell transplantation (AutoHCT) and maintenance for multiple myeloma (MM): results of the prognostic immunophenotyping for myeloma response (PRIMeR) study." Biology of Blood and Marrow Transplantation 25.3 (2019): S4-S6.

Outcome: PFS and OS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PY	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	NI	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI

1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate Confounding in MRD analysis was not discussed or addressed. Only one additional analysis on time of MRD assessment was conducted.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI

2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention? 2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	-	NA / Y / PY / PN / N / NI NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate MRD definition was pre-specified. The background patient characteristics were not reported. However, the population was not uniform in terms of conventional response (mix of CR, VGPR and less than VGPR). This has not been addressed.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any response. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions

	3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
	Risk of bias judgement	Moderate The definition of MRD negativity was not clearly described. The choice of the assay and sensitivity were not explained.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions			
	If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI

	4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
	If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
	4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
	4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
	4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
	4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
	Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data

5.1 Were outcome data available for all, or nearly all, participants?	PY (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients assessed for MRD were included in the survival analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes

	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
	6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
	6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. HR was reported.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to measurement of outcomes?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result			
	Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI

	7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
	7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Moderate The method of HR estimation was not clearly described.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias			
	Risk of bias judgement	Moderate HR was reported, but statistical methods were not clearly described. Baseline characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors did not conduct subgroup and interaction analyses for MRD, other than time of MRD assessment. Neither direction nor magnitude of a potential bias (other than time of assessment) were discussed in the publication.	Low / Moderate / Serious / Critical / NI

	Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Rossi (2018)

Rossi, Giovanni, et al. "Minimal residual disease and log-reduction of plasma cells are associated with superior response after double autologous stem cell transplant in younger patients with multiple myeloma." Cytometry Part B: Clinical Cytometry 96.3 (2019): 195-200.

Outcome: PFS HR (reported in the publication)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	PY	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	PY	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI

1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low Several subgroup and interaction analyses were conducted and sources of confounding were qualitatively described	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI

	2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention? 2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	-	NA / Y / PY / PN / N / NI NA / Y / PY / PN / N / NI
	2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
	2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI
	Risk of bias judgement	Moderate. Criteria for MRD assessment were not clearly described.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions

	3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
	Risk of bias judgement	Moderate The definition of MRD negativity was not clearly described.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to classification of interventions?		Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions			
	If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI

	4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
	If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
	4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
	4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
	4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
	4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
	Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data

5.1 Were outcome data available for all, or nearly all, participants?	Y	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	PY	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	PY	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	N	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	NI	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate The number of patients measured for MRD who were included in MRD assessment is unknown.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	Not expected to influence the outcomes.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes

	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
	6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
	6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Moderate The authors did not give a clear definition of the response and the survival endpoints.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable. Different definitions of PFS/OS (from MRD assessment vs. first diagnosis) can change t=0 of the survival curves.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result			
	Is the reported effect estimate likely to be selected, on the basis of the results, from...		

	7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
	7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
	7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Low HR was reported, statistical methods were clearly described.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of the reported result?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias			
	Risk of bias judgement	Moderate HR was reported. Background characteristics were reported. The population was uniform in terms of most of the confounders (conventional response, assay, sensitivity, treatment). The authors conducted subgroup analyses on time of MRD assessment only, however, the subgroups were not well defined. Both direction and magnitude of other potential bias were discussed in the publication, but not addressed in the analyses.	Low / Moderate / Serious / Critical / NI

	Optional: What is the overall predicted direction of bias for this outcome?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Risk of bias assessment. Rawstron (2013)

Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." Blood, The Journal of the American Society of Hematology 100.9 (2002): 3095-3100.

Outcome: PFS and OS HR (estimated based on the simulated IPD and reconstructed KM)

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	Y	Y / PY / PN / N
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:		
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Not applicable	NA / Y / PY / PN / N / NI

1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	Not applicable	NA / Y / PY / PN / N / NI

Questions relating to baseline confounding only		
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Y	NA / Y / PY / PN / N / NI
1.5. **If Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y	NA / Y / PY / PN / N / NI
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Not applicable	NA / Y / PY / PN / N / NI
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable	NA / Y / PY / PN / N / NI

	1.8. **If Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable	NA / Y / PY / PN / N / NI
	Risk of bias judgement	Low Multivariate regression analyses were conducted and sources of confounding were qualitatively described	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to confounding?	-	Favours experimental / Favours comparator / Unpredictable

Bias in selection of participants into the study			
	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	N	Y / PY / PN / N / NI

2.2. **If Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention? 2.3 **If Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	-	NA / Y / PY / PN / N / NI NA / Y / PY / PN / N / NI
2.4. Do start of follow-up and start of intervention coincide for most participants?	Y	Y / PY / PN / N / NI
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Moderate MRD definition was pre-specified. The criteria for participation in MRD assessment were clearly defined. However, the background characteristics of patients included in MRD assessment were not reported.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to selection of participants into the study?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in classification of interventions

	3.1 Were intervention groups clearly defined?	PY	Y / PY / PN / N / NI
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Not applicable	Y / PY / PN / N / NI
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	NI	Y / PY / PN / N / NI
	Risk of bias judgement	Low Criteria for MRD assessment were clearly described.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to classification of interventions?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to deviations from intended interventions			
	If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Not applicable	Y / PY / PN / N / NI

	4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Not applicable	NA / Y / PY / PN / N / NI
	If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
	4.3. Were important co-interventions balanced across intervention groups?	Not applicable	Y / PY / PN / N / NI
	4.4. Was the intervention implemented successfully for most participants?	Not applicable	Y / PY / PN / N / NI
	4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable	Y / PY / PN / N / NI
	4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable	NA / Y / PY / PN / N / NI
	Risk of bias judgement	Not applicable	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Not applicable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias due to missing data

5.1 Were outcome data available for all, or nearly all, participants?	Y (out of those included in MRD assessment)	Y / PY / PN / N / NI
5.2 Were participants excluded due to missing data on intervention status?	N	Y / PY / PN / N / NI
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	N	Y / PY / PN / N / NI
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	-	NA / Y / PY / PN / N / NI
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	-	NA / Y / PY / PN / N / NI
Risk of bias judgement	Low All patients that were measured for MRD were included in the survival analysis.	Low / Moderate / Serious / Critical / NI
Optional: What is the predicted direction of bias due to missing data?	-	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in measurement of outcomes

	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Not applicable	Y / PY / PN / N / NI
	6.2 Were outcome assessors aware of the intervention received by study participants?	Not applicable	Y / PY / PN / N / NI
	6.3 Were the methods of outcome assessment comparable across intervention groups?	Y	Y / PY / PN / N / NI
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Low The authors gave a clear definition of the survival endpoints. HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to measurement of outcomes?	Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Bias in selection of the reported result			
	Is the reported effect estimate likely to be selected, on the basis of the results, from...		

	7.1. ... multiple outcome *measurements* within the outcome domain?	N	Y / PY / PN / N / NI
	7.2 ... multiple *analyses* of the intervention-outcome relationship?	PN	Y / PY / PN / N / NI
	7.3 ... different *subgroups*?	N	Y / PY / PN / N / NI
	Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias.	Low / Moderate / Serious / Critical / NI
	Optional: What is the predicted direction of bias due to selection of the reported result?	Unpredictable.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Overall bias

	Risk of bias judgement	Moderate HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics of patients included in the MRD assessment were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	Low / Moderate / Serious / Critical / NI
	Optional: What is the overall predicted direction of bias for this outcome?	HR mrd- measured in a group of patients achieving CR or better is less favorable than in patients achieving VGPR or better or any reponse. This is explained by a smaller difference in PFS between MRD+ and MRD- patients due to similar response being achieved.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

[1] Solovev, Maxim V., et al. "Maintenance Therapy after Autologous Haematopoietic Stem Cell Transplantation (auto-HSCT) in Multiple Myeloma Patients with and without Minimal Residual Disease (MRD)." (2016): 2260-2260.

2 Solovev, Maxim V., et al. "Efficacy of maintenance therapy following auto-HSCT depending on MRD status in patients with multiple myeloma." Blood 132.Supplement 1 (2018): 3432-3432.

3 Schinke, Carolina, et al. "The prognostic value of the depth of response in multiple myeloma depends on the time of assessment, risk status and molecular subtype." Haematologica 102.8 (2017): e313.

4 Bakkus, Marleen HC, et al. "Post-transplantation tumour load in bone marrow, as assessed by quantitative ASO-PCR, is a prognostic parameter in multiple myeloma." British journal of haematology 126.5 (2004): 665-674

5 Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." Blood, The Journal of the American Society of Hematology 100.9 (2002): 3095-3100.

6 Paiva, Bruno, et al. "Multiparameter flow cytometric remission is the most relevant prognostic factor for multiple myeloma patients who undergo autologous stem cell transplantation." Blood, The Journal of the American Society of Hematology 112.10 (2008): 4017-4023

7 Popat, Rakesh, et al. "Outcomes of stratification to ASCT or not based on depth of response: results of a phase 2 trial assessing the impact of minimal residual disease (MRD) in multiple myeloma patients with deferred ASCT (PADIMAC)." Blood 130.Supplement 1 (2017): 1864-1864

8 Cohen, Oliver C., et al. "Bortezomib Consolidation Following Upfront ASCT for Multiple Myeloma Deepens Disease Response and MRD-Negative Rate without Compromising Response to Subsequent Bortezomib Salvage: Results of a Phase II Study." (2016): 4508-4508.

9 Cohen, Oliver C., et al. "Bortezomib consolidation post-ASCT as frontline therapy for multiple myeloma deepens disease response and MRD-negative rate whilst maintaining QOL and response to re-treatment at relapse." British journal of haematology (2018).

10 Clark, C. Amos, et al. "Prospective trial of minimal residual disease assessment by multiparametric flow cytometry for multiple myeloma in the era of bortezomib-based chemotherapy." Bone marrow transplantation 53.12 (2018): 1589-1592.

11 Chakraborty, Rajshekhar, et al. "Impact of post-transplant response and minimal residual disease on survival in myeloma with high-risk cytogenetics." Biology of Blood and Marrow Transplantation 23.4 (2017): 598-605.

12 Gu, Jingli, et al. "Longitudinal flow cytometry identified “minimal residual disease”(MRD) evolution patterns for predicting the prognosis of patients with transplant-eligible multiple myeloma." Biology of Blood and Marrow Transplantation 24.12 (2018): 2568-2574.

13 Hahn, Theresa E., et al. "Minimal residual disease (MRD) assessment before and after autologous hematopoietic cell transplantation (AutoHCT) and maintenance for multiple myeloma (MM): results of the prognostic immunophenotyping for myeloma response (PRIMeR) study." Biology of Blood and Marrow Transplantation 25.3 (2019): S4-S6.

14 Rossi, Giovanni, et al. "Minimal residual disease and log-reduction of plasma cells are associated with superior response after double autologous stem cell transplant in younger patients with multiple myeloma." Cytometry Part B: Clinical Cytometry 96.3 (2019): 195-200.

15 Rawstron, Andy C., et al. "Flow cytometric disease monitoring in multiple myeloma: the relationship between normal and neoplastic plasma cells predicts outcome after transplantation." Blood, The Journal of the American Society of Hematology 100.9 (2002): 3095-3100.

NICE National Institute for Health and Care Excellence

Patient organisation submission

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

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	Myeloma UK
3. Job title or position	[redacted]
4a. Brief description of the organisation (including who funds it). How many members does it have?	Myeloma UK is the only organisation in the UK dealing exclusively with myeloma. Our broad and innovative range of services cover every aspect of myeloma from providing information and support, to improving standards of treatment and care through research and campaigning. We receive no government funding and rely almost entirely on the fundraising efforts of our supporters. We also receive some unrestricted educational grants and restricted project funding from a range of pharmaceutical companies. We are not a membership organisation.
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	The table below shows the audited 2019 income from the relevant manufacturers. Funding is received for a range of purposes and activities namely core grants, project specific work including clinical trials, and gifts, honoraria or sponsorship.

Name of Company	Grants and project specific funding	Gifts, Honoraria and Sponsorship	Total (£)
Celgene	110,000	12,337	122,337
Janssen-Cilag	20,000	327	20,327

If so, please state the name of manufacturer, amount, and purpose of funding.	
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?	The information included in this submission has been gathered from the myeloma patients and carers we engage with through our research and services programmes, including: - A combination of structured telephone interviews and questionnaire with myeloma patients about living with myeloma, their experience, and expectations of treatment. We specifically sought out the views of patients who have received high-dose therapy and stem cell transplantation (HDT-SCT) and possibly, have been treated with daratumumab at a later line of therapy. This interview/questionnaire focused on the Phase III CASSIOPEIA clinical trial data comparing daratumumab (Darzalex®), bortezomib (Velcade®), thalidomide and dexamethasone (DVTD) with bortezomib, thalidomide and dexamethasone (VTD). *(Patient Quotes are in italics)* - A Myeloma UK patient experience survey[1] of over 1,000 patients, conducted alongside the myeloma results of the National Cancer Patient Experience Survey in 2016. - A multi-criteria decision analysis study of 560 myeloma patients. The study, funded by Myeloma UK and run by the European Medicines Agency (EMA) and University of Groningen, explored patient preferences for different benefit and risk outcomes in myeloma treatment.

[1] Myeloma Patient Experience Report 2016 (Insights from the National Cancer Patient Experience Survey) https://www.myeloma.org.uk/wp-content/uploads/2020/05/Myeloma-UK-Full-Report-Cancer-Patient-Experience-Survey-2014.pdf Accessed: 31/08/2020

	It has also been informed by analysis of the experiences and views of patients, family members and carers gathered via our Myeloma Infoline, Patient and Family Myeloma Infodays and posts to our online Discussion Forum.
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?	**What is it like to live with myeloma?** *"The uncertainty of not knowing when it will come back but the certainty of knowing it will is particularly difficult."* Myeloma is a highly individual and complex cancer originating from abnormal plasma cells in the bone marrow. There is currently no cure, but treatment can halt its progress and improve quality of life. The complications of myeloma can be significant, debilitating and painful and include severe bone pain, bone destruction, kidney damage, fatigue and a depleted immune system which can lead to increased infections. Myeloma is also a relapsing and remitting cancer which evolves over time and becomes resistant to treatment. Most patients can be successfully retreated at relapse; however, remission is usually associated with diminishing duration and depth of response over time. **First remission is therefore widely held as the best opportunity to gain the deepest response with the longest period until disease progression**.[1] It is also the point in their disease where many patients will be able to build on existing better quality of life since the burden of treatment and illness will be less than for patients who are multiply relapsed. Treatment side effects and frequent hospital visits have a social and practical impact on patients' lives, including significant financial implications. Reduction in mobility over time and a perceived increase in reliance on carers and family members, also impacts on patients' sense of control.

	"I find the mental aspect most challenging. Not knowing what it's going to do even though you understand it is coming back. It clipped my wings. I live in a world of before and after. Certain things become insignificant. But I do not give in to it." **What do carers experience?** *"I feel angry that I'm not going to get the future I wanted, but the hardest thing to feel is how my life at the moment is in limbo"* A Myeloma UK study into the experiences of carers and family members found that looking after someone with myeloma has a significant emotional, social and practical impact: - 94% of carers are emotionally impacted and found the uncertainty of myeloma a major factor - 25% of those in work had been unable to work or had to retire early to care for the person with myeloma - 84% always put the needs of their relative or friend with myeloma before their own - Only 42%of carers were not given enough information at diagnosis about how myeloma may affect them Living with myeloma is therefore often extremely challenging physically and emotionally for patients, carers, and family members. *"I had to think of my husband. You are in this as a team, it is not an individual battle."*

Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	The role of induction therapy prior to HDT-SCT is to decrease tumour burden, thus deepening the response rate and increasing the likelihood of engraftment, while retaining the maximum possible tolerability and minimum possible toxicity on normal hematopoietic cells. Newly diagnosed patients who are eligible for an HDT-SCT are hoping for as long a remission period as possible post-transplant and induction therapy plays a large part in this process. VTD has become the standard triplet therapy used for induction/consolidation treatment in myeloma in the UK. The role of bortezomib is particularly significant in its demonstrated usefulness in high-risk patients.[23] Other induction treatment combinations may be used under certain circumstances, for example, if you are unable to take thalidomide. However, VTD was also proven superior to bortezomib, cyclophosphamide, and dexamethasone (VCD), thus highlighting the effect of combining an IMiD with a proteasome inhibitor and dexamethasone.[4] VTD is a very effective combination and we know that what patients value most is treatments that will keep their myeloma under control and deliver the longest possible remission. VTD is associated with side effects, with bortezomib particularly associated with peripheral neuropathy. *"I didn't enjoy being on the thalidomide. I was not keen on the Velcade either. It gave me gastrointestinal issues. The thalidomide I never felt quite there. I was walking around in a cloud."*

[2] Sonneveld P, Goldschmidt H, Rosiñol L, et al. Bortezomib-based versus nonbortezomib-based induction treatment before autologous stem-cell transplantation in patients with previously untreated multiple myeloma: a meta-analysis of phase III randomized, controlled trials. J Clin Oncol. 2013;31(26):3279-3287. doi:10.1200/JCO.2012.48.4626

[3] Avet-Loiseau H, Leleu X, Roussel M, et al. Bortezomib plus dexamethasone induction improves outcome of patients with t(4;14) myeloma but not outcome of patients with del(17p). J Clin Oncol. 2010;28(30):4630-4634. doi:10.1200/JCO.2010.28.3945

[4] Moreau P, Hulin C, Macro M, et al. VTD is superior to VCD prior to intensive therapy in multiple myeloma: results of the prospective IFM2013-04 trial. Blood. 2016;127(21):2569-2574. doi:10.1182/blood-2016-01-693580

	"I am usually quite relaxed and laidback about things, but it did affect my mood sometimes. However, I learned how to control it. I knew when a mood swing would come. I think daily walking and mindfulness helped me here." However, there are well established methods of treating side effects and for most patients VTD has a tolerable side effect profile. The different types of treatment benefit that are most valued by patients are set out below. Each of these benefits will be delivered to a greater or lesser extent by individual treatments currently available on the NHS. Myeloma patients and their carers place a very high value on treatments that: • **Prolong their life** • **Put their myeloma into remission for as long as possible** • **Allow them to enjoy normal day-to-day life.** The Myeloma UK, EMA and the University of Groningen study showed that, achieving a lasting remission from treatment was the most important factor for most (75%) participants. This was true across all patient groups regardless of demographic and clinical characteristics. Treatments with minimal negative impact on quality of life are very important, particularly those with as few side effects as possible and of low severity. That said, data shows that patients will accept even severe side effects if the treatment has a superior efficacy, suggesting that efficacy is the strongest driver of treatment choice. *"At the early stage of diagnosis and treatment, the most important thing for me was to get a degree of confidence that the treatment would be successful and give a good remission time. For me, I was happy to deal with the side effects (within reason)."*

	Finally, it is important to recognise that patients do not see the survival benefits of individual treatments in isolation. They want the best possible remission and quality of life at each stage of their myeloma and see gains in survival from one treatment as a "bridge" to further treatments coming down the line.
8. Is there an unmet need for patients with this condition?	As set out above, current standard induction treatments do exist. However, given that myeloma is such a heterogeneous cancer there is still a need for a range of treatment options with different mechanisms of action at each stage of the treatment pathway. DVTD would present the first quadruplet combination treatment in myeloma with studies showing the superior results of quadruplet treatment regimens vs triplet treatment regimens. As stated above the relapsing/remitting nature of myeloma is usually associated with diminishing duration and depth of response over time. Studies show that in myeloma the first remission is often the deepest and longest remission period for the patient. As this quadruplet is positioned as an induction treatment for newly diagnosed myeloma patients who are eligible for HDT-SCT this is the best opportunity to achieve the deepest and longest remission period. This would be particularly significant for younger/fitter patients who more likely to be working and to have dependents and therefore face particular challenges in living with myeloma. *"I relapsed with just about 2 and half years of remission. This was a huge disappointment as being younger I thought I might have got a longer remission period."*
Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	Patients value treatments which are effective and put their myeloma into remission for as long as possible. **Data from the Phase III CASSIOPEIA trial indicates that the addition of daratumumab to VTD at induction can induce deeper responses to the treatment and increase progression free survival.** The main outcome of the trial was stringent complete response (sCR) 100 days after HDT-SCT, representing a deeper level of response than conventional complete response (CR). Patients receiving DVTD had significantly deeper responses compared to patients receiving VTD.

	- 29% of DVTD patients achieved sCR compared to 20% of VTD patients achieving sCR (p=0.001). - 39% of DVTD patients achieved a CR or better compared to 26% of VTD patents achieving CR (p<0.0001). Achieving a sCR or CR is associated with improved outcomes for patients. The International Myeloma Working Group (IWMG) examined the clinical features of long-term survival as it correlates to the depth of disease response. The research project found that achieving CR at one year was associated with superior progression free survival (PFS) (median PFS 3.3 years vs. 2.6 years) as well as overall survival (OS) (median OS 8.5 years vs. 6.3 years). The data identifies CR as an important predictor of long-term survival for HDT-SCT eligible myeloma patients.[5] A secondary end point was the proportion of patients with minimal residual disease (MRD) negativity at 100 days post HDT-SCT. The CASSIOPEIA data showed a significant difference in number patients achieving minimal residual disease (MRD) negativity between both arms of the clinical trial. 64% of DVTD patients achieved MRD negativity vs 44% of VTD patients achieving MRD negativity (p<0.0001). MRD-negative status after treatment for newly diagnosed myeloma is also associated with longer PFS and longer OS.[6] *"The biggest difference in impact seems to be 64% of patients on DVTD delivering no detectable amounts of myeloma. Hoping that daratumumab is the main factor in that. DVTD does look like a much more effective combination."* Although median PFS was not reached the clinical trial data shows an improvement on PFS at 18 months with 93% of DVTD patients still in remission compared to 85% of VTD patients still in remission.

[5] Usmani SZ, Hoering A, Cavo M, et al. Clinical predictors of long-term survival in newly diagnosed transplant eligible multiple myeloma - an IMWG Research Project. Blood Cancer J. 2018;8(12):1-7. doi:10.1038/s41408-018-0155-7

[6] Munshi, NC et al. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. JAMA oncology 2017;3(1)28-35. doi:10.1001/jamaoncol.2016.3160

	Finally, the fixed duration of treatment with six cycles (four induction and two consolidation post HDT-SCT) can provide patients with a level of certainty that the treatment has an end point. Following this, there will hopefully be an extended and possibly treatment-free remission which is highly valued by patients. Considering that the first remission is likely to be the longest and deepest remission this is the best opportunity for patients to retain a relatively high quality of life (QOL). *"For an extra drug with a deeper response, increased remission and longer treatment free period I would have bitten your hand off. Achieving a complete response would be a big win."* The addition of daratumumab to VTD allows a deeper response, increased PFS, and improved MRD negativity. All of which are associated with improved outcomes for patients.
Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	**Side Effects** Patients value treatments with fewer side effects with low severity ratings which stop when treatment ends. However, in practice patients will accept varying levels of toxicity in a treatment if it delivers good survival benefit and depending on the stage of their myeloma. When discussing side effects with patients some were concerned about the level of toxicity that a quadruplet combination might bring. However, one patient did say: "*The number of drugs, 3 or 4 is irrelevant to me, it's the effectiveness of the treatment."* In analysing the clinical trial data, the overall incidence of serious adverse events was comparable with serious adverse events occurring in 251 patients (47%) in the DVTD group and 255 patients (47%) in the VTD group. However, more participants experienced grade 3/4 neutropenia, thrombocytopenia, and lymphopenia in the DVTD group. Further to this, although infections were more common in the DVTD group (65% versus 57% the VTD group), the rate of grade 3/4 infection was similar in both groups (22% in

	DVTD groups vs 20% VTD group). It should also be noted that 35% of patients receiving daratumumab experienced an infusion related reaction. Except for haematological events, no clinically meaningful differences in adverse events were observed between treatment groups. The incidence of infusion-related reaction was consistent with other daratumumab studies. Although the median stem-cell yield was smaller and more patients received plerixafor in the DVTD group, successful transplantation was not affected. Adding daratumumab to bortezomib, thalidomide, and dexamethasone did not increase overall toxicity. The dosing schedule used is typical of real-world practice, and adverse events were clinically manageable and consistent with the known toxicities of bortezomib, thalidomide, and dexamethasone as well as daratumumab. When discussing side effects with patients who had received VTD as induction therapy and daratumumab at a later line of treatment many said that there was a higher burden of side effects associated with VTD induction. The daratumumab was "kinder" and therefore would not add to the treatment burden already existing with VTD. *"I think the DVTD will be a much better induction treatment if it gives a good response. There are a lot of side effects on the VTD and the having the daratumumab is much better in comparison. The daratumumab is kind and well tolerated so it would be easy to manage with VTD."* **Administration** Further to this the addition of daratumumab to VTD would mean extra hospital visits to receive the daratumumab by IV infusion. This does mean taking time out of the day to attend hospital. For some patients there are cost/capability issues associated with this and it can place an additional burden on carers who have to accompany the patient to hospital. Oral treatments are often valued by patients, particularly those who are working and have dependents. The future ability to have daratumumab subcutaneously would be highly valued by patients. *"I honestly can't pin any side effect to the daratumumab. I did react to the first infusion but I knew that was likely to happen….. Because the dara is now an injection as opposed to infusion, I take my pre- meds before I leave for the hospital and I can be in and out in 15 minutes or so."*

	That said, our patient engagement has shown that there are also patients who welcome their treatment being delivered in the safety of a hospital environment and the opportunity to interact with clinical staff and other patients. Overwhelmingly, clinical efficacy and the opportunity of a good remission outweighs any disadvantages in the method of administration.
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.	Patients who would receive DVTD as an induction therapy followed by HDT-SCT are more likely to be younger/fitter. Many of whom are likely to be working and/or have dependents and therefore face particular challenges in living with myeloma. The CASSIOPEIA trial evaluated adults less than 65 years old with the median age of the patients in the trial at 58 years. With the data showing that DVTD can offer an improved depth of response and length of PFS this could have a significant impact on the QOL for this patient population. A survey conducted in 2019 by Jackson and Galinsky et al looked at productivity losses in patients with newly diagnosed multiple myeloma following stem cell transplantation. There were 115 eligible survey respondents, 77% were economically active at the time of diagnosis and highlighted return to work as an important factor affecting their quality of life; only 39% of respondents were economically active post HDT-SCT. Patients with myeloma aspire to engage in productive lives post-HDT-SCT, but most are unable to do so. Access to treatments extending remission and supporting engagement in a productive life can have a positive impact both for patients and wider society.[7]

[7] Jackson G, Galinsky J, Alderson DEC, et al. Productivity losses in patients with newly diagnosed multiple myeloma following stem cell transplantation and the impact of maintenance therapy. Eur J Haematol. 2019;103: 393–401. https://doi.org/10.1111/ejh.13298

	"I am still working. My husband runs a software company and I help out. Treatment does not impact on my ability to work. I don't do much on steroid days, but I keep going when I get an opportunity. I worked through everything, treatment, and pandemic. It gives me structure and I'm really thankful for work as I could have had a huge mental spiral without work." *"Treatment has had an impact though. I have been so tired at times I struggle with finding the energy to look after two children but have generally just pushed through! I have no choice – in some ways this is a good thing, it keeps me active and takes my mind off myeloma. I haven't time to feel sorry for myself."*
Equality	
12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	No
Other issues	
13. Are there any other issues that you would like the committee to consider?	No

14. In clinical practise, only induction therapy is administered to people who have untreated multiple myeloma prior to autologous stem cell transplantation. However, the clinical trial offered patients consolidation therapy in addition to induction therapy. Do you expect this to have an impact on generalisability of results? If so, please describe the anticipated effect and explain why.	Standard clinical practice is to give induction therapy for 4-6 cycles before HDT-SCT. With the fixed duration of treatment with DVTD fixed at 6 cycles (4 pre-HDT-SCT and 2 post-HDT-SCT) we do not expect this to have an impact on the generalisability of results. Further to this, from our engagement with patients we are aware that some patients have been given extra cycles of induction therapy if there paraprotein levels have not sufficiently lowered pre-HDT-SCT. Also, some patients have been given further rounds of induction treatment post HDT-SCT if the patient has not had a sufficient response.
Key messages	
15. In up to 5 bullet points, please summarise the key messages of your submission	

- Data from the Phase III CASSIOPEIA trial indicates that the addition of daratumumab to VTD at induction increases the likelihood of patients achieving a sCR/CR and MRD Negativity.
- Patients who achieve a deeper response, increased PFS, and improved MRD negativity are associated with improved outcomes.
- Adding daratumumab to bortezomib, thalidomide, and dexamethasone did not increase overall toxicity. The dosing schedule used is typical of real-world practice, and adverse events were clinically manageable and consistent with the known toxicities of bortezomib, thalidomide, and dexamethasone as well as daratumumab. Patients remarked that the addition of daratumumab to induction therapy would not add to the treatment burden already existing with VTD.
- Considering that the first remission is likely to be the longest and deepest remission this is the best opportunity for patients to retain a relatively high quality of life (QOL).
- Patients who would receive DVTD as an induction therapy followed by HDT-SCT are more likely to be younger/fitter. Accessing DVTD at induction therapy would give these patients a better opportunity to achieve a longer remission period.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

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.

……

NICE National Institute for Health and Care Excellence

Professional organisation submission

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

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 the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

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 13 pages.

About you	
1. Your name	**xxxxxxxxxx**
2. Name of organisation	UK Myeloma Forum

3. Job title or position	[redacted]
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☒ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5a. Brief description of the organisation (including who funds it).	UK Myeloma Forum is the only organisation that represents Physicians, Nursing staff, Pharmacists and Healthcare professional who are directly involved with providing clinical care or research for patients with myeloma. Membership is free by application and members of the executive are elected by the membership. It aims to improve the care of myeloma patients through the development and promotion of trials and provides education about myeloma to healthcare professionals.
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	UKMF has received an unrestricted educational grant from Janssen-Cilag (£6,000 per annum), and Celgene (BMS, £10,000 per annum). UKMF has also received unrestricted educational grants from other pharmaceutical companies.

If so, please state the name of manufacturer, amount, and purpose of funding.	
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No
The aim of treatment for this condition	
6. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	Myeloma is currently incurable. Most people diagnosed with myeloma will die as a result of complications of the disease. Symptoms and signs associated with active myeloma include bone pain, fractures secondary to bone deposits, fatigue, anaemia, recurrent infections, renal failure, high calcium levels and occasionally spinal cord compression. Treatment is primarily aimed at reducing these symptoms by controlling the disease. There is a direct association between how well the myeloma is controlled and the improvement in quality of life. Patients are clinically better if in complete response rather than partial response. Additional aims of treatment are to control the disease (and thereby symptoms) for as long as possible (i.e. lengthen the progression free survival / duration of response), lengthen life associated with the disease (i.e. increase overall survival) and prevent significant morbidity associated with progression of the disease.
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by	There are internationally agreed criteria for assessing response (International Myeloma Working Group Rajkumar et al. Blood 2011;117:4691-4695 These are based on the proportional reduction of serum paraprotein / serum free light chains (serological markers of myeloma), urine monoclonal protein and the bone marrow proportion of myeloma plasma cells. Generally, a Partial Response (PR) or better is considered clinically significant. Increasingly with more efficacious treatments the aim of the therapy is to achieve Complete Response (CR) or Very Good Partial Response (VGPR) for as many patients as possible. It is apparent in many studies that the greater the

x cm, or a reduction in disease activity by a certain amount.)	depth of response the longer the duration of the response (CR>VGPR>PR). Patients who achieve a CR have a longer survival than those who do not. Achieving minimal residual disease (MRD) is associated with an even longer duration of response and overall survival.
8. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Yes. Myeloma is incurable with current therapy for the majority of patients. First line therapy should be aimed at achieving the highest possible response rates and the deepest possible responses leading to the longest / most durable responses which thereby reduces the morbidity and mortality associated with the myeloma. Currently available first line therapies for transplant eligible patients are Bortezomib Thalidomide Dexamethasone (VTd) or Bortezomib Cyclophosphamide Dexamethasone (VCd) they are well tolerated and induce remission in a significant proportion of patients. A small group of patients do not respond to these therapies and patients will inevitably relapse often developing new myeloma related issues. There is therefore a clear unmet need to provide better treatments to induce a longer and more durable period of remission and limit or prevent myeloma associated complications.
What is the expected place of the technology in current practice?	
9. How is the condition currently treated in the NHS?	
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	NICE guideline TA331 recommends Bortezomib Thalidomide Dexamethasone (VTd) for up to 6 cycles prior to autologous stem cell transplantation (ASCT). This is considered the standard of care for patients eligible for transplantation. A small group of patients may receive VCd if there are concerns with the initial use of Thalidomide (pre-existing neurotoxicity or significant thrombotic risk). Maintenance treatment after ASCT is not available but is widely available in other countries.
• Is the pathway of care well defined? Does it	The pathway of care is well defined. Patients are assessed according to autologous transplant eligibility and then treated on the transplant eligible or non-eligible pathways. The only variation in practice is

vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	whether patients receive VTd or VCd (both for up to 6 cycles). Outside of clinical trials no other induction regimens are routinely given. Patients are referred to specialist centres for the delivery of ASCT following local pathways.
• What impact would the technology have on the current pathway of care?	Daratumumab is a well-tolerated treatment that is widely given in combination with Bortezomib and Dexamethasone (DVd) as 2nd line therapy (Cancer Drug Fund), or as monotherapy as 4th line therapy (Cancer Drug Fund). Clinicians have widespread experience of delivering this treatment and dealing with any associated toxicities. Daratumumab would be given in addition to VTd for up to 6 cycles prior to ASCT. It would easily fit into the current treatment algorithm and would be easily delivered.
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Daratumumab would be given in addition to VTd for up to 6 cycles prior to ASCT. It would easily fit into the current treatment algorithm and would be easily delivered.
• How does healthcare resource use differ between the technology and current care?	Patients currently attend chemotherapy day units for weekly treatment with subcutaneous Bortezomib. Daratumumab would be given on the same occasion and would require some additional nursing / pharmacy resources. Giving Daratumumab subcutaneously (rather than intravenously) would reduce the amount of time patients spend in hospital. Patients would not need to attend the chemotherapy day unit on more days as they would receive Bortezomib and Daratumumab at the same time.
• In what clinical setting should the technology be used? (For example,	Specialist clinics

primary or secondary care, specialist clinics.)	
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	None. Daratumumab is currently used at 2nd and 4th line (via the CDF). There is extensive UK experience of this drug. It is currently given in combination with Bortezomib and dexamethasone at 2nd line (CDF).
11. 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?	Data from the CASSIOPEIA trial indicates that the addition of Daratumumab to VTd induces deeper responses to treatment and increases progression free survival. The primary end point was stringent complete response (sCR), representing a deeper level of response than conventional complete response (CR). Daratumumab VTd was clinically superior to VTd (which is the standard of care): • 28.9% of DVTd patients achieved sCR compared to 20.3% of VTd patients • 39% of DVTd patients achieved a CR or better compared to 26% of VTd patients Achieving an sCR or CR is associated with improved outcomes for patients. Data from the International Myeloma Working Group (IMWG) assessing a large group of myeloma patients demonstrate that improved depth of response correlates with improved long term survival. Patients achieving a CR at 1 year were

	associated with superior PFS (median PFS 3.3 vs 2.6 years) as well as OS (median OS 8.5 years vs 6.3 years) (Usmani et al Blood Cancer Journal, Nov 23 2018 doi:10.1038/s41408-018-0155-7) To further characterise depth of response, minimal residual disease (MRD) was assessed: • 63.7% of DVTd patients achieved MRD negative status vs 43% of VTd patients Achieving MRD negative status after treatment for a newly diagnosed myeloma patients is associated with a longer PFS and OS (Munshi et al JAMA 3,1 (2017):28-35) Median PFS was not reached when CASSIOPEIA reported, although there is a trend in an improved PFS at 18 months (93% DVTd vs 85% VTd). This highlights the potential benefit of this therapy that is well tolerated and associated with deeper and more durable response. As mentioned before this will be associated with better quality of life.
• Do you expect the technology to increase health-related quality of life more than current care?	Yes. This is a well-tolerated regiment with limited and manageable side effect profile. There are no additional concerning adverse events with Daratumumab given in combination with Bortezomib Thalidomide and Dexamethasone vs Bortezomib Thalidomide Dexamethasone in the Phase III CASSIOPEIA trial.
12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	No
The use of the technology	

13. 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.)	Daratumumab is widely used. Healthcare professional will have experience of administration and dealing with potential complications. There will be additional health resources needed to deliver the addition of Daratumumab to the standard of care. Patients will need to spend more time on day units to receive Daratumumab. As Daratumumab will be delivered on the same day as Bortezomib the number of days at home or in hospital is unchanged. It is unlikely there will be added side effects with this new therapy.
14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Response is based on clinical response to treatment after between 2 and 4 cycles of induction treatment.
15. Do you consider that the use of the technology will result in any substantial health-	Yes. Quality of life is likely to be improved due to reduced myeloma associated complications.

related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	
16. 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?	Yes, this is the first in class monoclonal antibody to be licenced in multiple myeloma. Its use in first line treatment induces increased depth and durability of response reducing both morbidity and overall survival in what still remains a generally incurable but increasingly chronic disease.
• Is the technology a 'step-change' in the management of the condition?	Yes because it improves depth or response which correlates with improved survival. This will lead to reduced myeloma associated complications.
• Does the use of the technology address any particular unmet need of the patient population?	No

17. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	Daratumumab is well tolerated and unlikely to have any impact on quality of life. Significant infusion related events are unusual, manageable and are usually only associated with the first infusion. There are no other concerning side effects.
Sources of evidence	
18. Do the clinical trials on the technology reflect current UK clinical practice?	CASSIOPEIA gave a fixed duration of 4 cycles of DVTd prior to ASCT and then 2 cycles of DVTd consolidation afterwards. This is different to current practice where patients receive up to 6 cycles of VTd prior to ASCT and no consolidation or maintenance.
• If not, how could the results be extrapolated to the UK setting?	The CASSIOPEIA trial is broadly comparable with current practice with the total number of cycles of VTD the same (6 in total). The difference between induction (4 vs up to 6) and consolidation (2 vs none) should not prevent this trial being seen as comparable to current practice.
• What, in your view, are the most important outcomes, and were they measured in the trials?	Depth of response. sCR, CR and MRD were measured in this trial. Survival has been assessed using PFS. Toxicity was assessed and no concern has been highlighted.
• If surrogate outcome measures were used, do they adequately predict	sCR, CR and MRD were measured in this trial as surrogates for long term survival. There is a wealth of data to support depth of response correlating with long term survival. Median PFS was not reached in the CASSIOPEIA trial, but demonstrated a favourable PFS in favour of DVTd (hazard ratio 0·47, 95% CI 0·33-0·67, p<0·0001).

long-term clinical outcomes?	
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	No.
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No.
20. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance [TA311]?	No.
21. How do data on real-world experience compare with the trial data?	There is limited experience with Real World DVTd for 1st line patients. However, there is an enormous experience with using the combination of Daratumumab with Bortezomib at 1st relapse (CDF). Depth of response and tolerability exceed expectations based on current outcomes for newly diagnosed transplant eligible myeloma patients.

Equality	
22a. Are there any potential equality issues that should be taken into account when considering this treatment?	No
22b. Consider whether these issues are different from issues with current care and why.	No
Key messages	
24. In up to 5 bullet points, please summarise the key messages of your submission. • Comparator in the CASSIOPEIA trial is equivalent to current standard of care (VTd) • Daratumumab is well tolerated, there is widespread experience of its use given alone or in combination with Bortezomib • There are many unmet needs for myeloma patients. Improved outcome with DVTd helps address these • Depth of response (MRD, sCR and CR) are considered clinically meaningful outcomes as they correlate with long term survival • The reported outcomes for D-VTD in a phase 3 trial are internationally considered to set a new gold standard for the 1st line treatment of newly diagnosed transplant eligible myeloma	

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

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.

.....

CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR Systematic Reviews Programme on behalf of NICE

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Report version post Factual Accuracy Check

Produced by	Southampton Health Technology Assessments Centre (SHTAC)
Authors	Joanne Lord, Professor, Health Economics Neelam Kalita, Senior Research Fellow, Health Economics Lois Woods, Senior Research Assistant, Information Science and Evidence synthesis Lorna Hazell, Senior Research Assistant, Evidence synthesis David Scott, Principal Research Fellow, Statistics Geoff Frampton, Senior Research Fellow, Evidence Synthesis
Correspondence to	Dr Geoff Frampton Southampton Health Technology Assessments Centre (SHTAC) Wessex Institute Alpha House Enterprise Road, University of Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	5th November 2020 Updated 26th November 2020 after factual accuracy check.

Source of funding: This report was commissioned by the NIHR Systematic Reviews Programme as project number 131928.

Acknowledgements

We thank the following for providing clinical advice and comments on the draft report: Dr Jenny Bird, Consultant Haematologist, University Hospitals Bristol and Weston NHS Trust; Dr Matthew Jenner, Consultant Haematologist, University Hospitals Southampton NHS Foundation Trust; Dr Charlotte Kallmeyer, Consultant Haematologist, United Lincolnshire

Hospitals NHS Trust; Dr Simon Stern, Honorary Myeloma Consultant, Royal Marsden Hospital NHS Foundation Trust and Consultant Haematologist, Epsom and St Helier University Hospitals NHS Trust; Professor Kwee Yong, Professor of Clinical Haematology, University College London Hospitals NHS Foundation Trust. We also thank Dr Jonathan Shepherd (SHTAC) for acting as an internal editor for the draft report.

Declared competing interests of the authors and expert clinical advisors

None from the authors. None from Dr Bird and Dr Kallmeyer. Dr Jenner received honoraria from Janssen (manufacturer of daratumumab) for attending an advisory board regarding the role of minimal residual disease (MRD) in myeloma (September 2020) which was not linked to any particular therapy. He also received honoraria from Celgene (manufacturer of thalidomide) for attending advisory boards on CAR-T cell therapy in relapsed/refractory myeloma; and honoraria for an advisory board for minimum residual disease in myeloma (October 2020). Dr Stern received speaker fees from Celgene for 2 educational meetings in 2019-2020 (which discussed the management of relapsed myeloma and spinal disease imaging); and received an educational grant from Janssen-Cilag (manufacturer of bortezomib and daratumumab) to attend the International Myeloma Workshop in September 2019. Professor Yong received fees from Celgene and Janssen-Cilag for 2 educational meetings in the UK and China (providing lectures to healthcare professionals in January and July 2020).

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Systematic Reviews Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Lord J, Kalita N, Woods L, Hazell L, Scott, D, Frampton G. Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]. SHTAC. 2020.

Contributions of authors

Jo Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Neelam Kalita critically appraised the health

economic systematic review, critically appraised the economic evaluation, and drafted the report; Lois Woods critically appraised the clinical effectiveness systematic review, conducted bibliographic searches, and drafted the report; Lorna Hazell critically appraised the clinical effectiveness systematic review and drafted the report; David Scott critically appraised the clinical effectiveness systematic review and drafted the report. Geoff Frampton critically appraised the clinical effectiveness systematic review, drafted the report, project managed the report and is the project guarantor.

Word count: 44,238

Commercial in confidence (CIC) information in blue
Academic in confidence (AIC) information in yellow

Table of Contents

1 Executive Summary 12
1.1 Overview of key model outcomes 12
1.2 Overview of the ERG's key issues 12
1.3 The decision problem: summary of the ERG's key issues 14
1.4 The clinical effectiveness evidence: summary of the ERG's key issues 14
1.5 The cost effectiveness evidence: summary of the ERG's key issues 17
1.6 Other key issues: summary of the ERG's view 20
1.7 Summary of ERG's preferred assumptions and resulting ICERs 20
2 INTRODUCTION AND BACKGROUND 23
2.1 Introduction 23
2.2 Background 23
2.2.1 Background information on multiple myeloma 23
2.2.2 Background information on daratumumab 25
2.2.3 The current treatment pathway for multiple myeloma and the proposed place of daratumumab 27
2.3 Critique of the company's definition of the decision problem 30
3 CLINICAL EFFECTIVENESS 36
3.1 Critique of the methods of the company's systematic literature review 36
3.2 ERG critique of the included clinical effectiveness studies 38
3.2.1 Included and ongoing studies 38
3.2.2 Risk of bias assessment 44
3.2.3 Outcomes assessment 45
3.2.4 Approach to study statistics 50
3.2.5 Clinical effectiveness results 55
3.2.6 Response outcomes 55
3.2.7 Survival outcomes 57
3.2.8 Health-related quality of life 61
3.3 Safety outcomes 64
3.3.1 Treatment exposure 64
3.3.2 Treatment-emergent adverse events 64
3.3.3 Infusion-related reactions 66
3.4 Critique of studies included in the indirect treatment comparisons 67
3.4.1 Rationale for indirect comparisons 67
3.4.2 Identification, selection and feasibility assessment of studies for ITC 67
3.4.3 Clinical heterogeneity assessment 68
3.4.4 Similarity of treatment effects 69

3.4.5 Risk of bias assessment for studies included in the ITC 69
3.5 Critique of the ITC methods 70
3.5.1 Data inputs to the MAICs 70
3.5.2 Statistical methods for the MAICs 70
3.5.3 Comparison of DBTd and BTd against Bd 71
3.5.4 Comparison of DBTd and BTd against BCd 72
3.5.5 External validation of MAIC results against real-world data 73
3.5.6 Results of the indirect comparisons 74
3.5.7 Summary of the ERG's critique of the indirect comparisons 75
3.6 Systematic literature review and meta-analysis of the impact of MRD status on survival outcomes 75
3.6.1 Identification and selection of studies 76
3.6.2 Assessment of study validity 78
3.6.3 Assessment of study heterogeneity 78
3.6.4 Method of data synthesis 80
3.6.5 Results of the meta-analysis 81
3.6.6 ERG critique of the SLR and meta-analysis 81
3.7 Conclusions of the clinical effectiveness section 81
4 COST EFFECTIVENESS 84
4.1 ERG critique of the company's review of cost-effectiveness evidence 84
4.2 ERG summary and critique of the company's submitted economic 85
4.2.1 NICE reference case checklist 85
4.2.2 Model structure 86
4.2.3 Population 89
4.2.4 Interventions and comparators 90
4.2.5 Perspective, time horizon and discounting 92
4.2.6 Treatment effectiveness and extrapolation 92
4.2.7 Health related quality of life 103
4.2.8 Resources and costs 105
5 COST EFFECTIVENESS RESULTS 110
5.1 Company's cost effectiveness results 110
5.2 Company's sensitivity analyses 110
5.2.1 Probabilistic sensitivity analysis 110
5.2.2 Deterministic sensitivity analysis 111
5.2.3 Scenario analysis 111
5.3 Model validation and face validity check 113
5.3.1 Processes for model checking 113
5.3.2 Face validity 114

5.3.3 Internal validation: comparison of survival predictions with CASSIOPEIA data ... 114
5.3.4 External validation: comparison of survival predictions with PHE data ... 117
5.3.5 Cross validation: comparison with TA311 BTd extrapolation ... 118
6 ERG ADDITIONAL ANALYSES ... 119
6.1 Exploratory and sensitivity analyses undertaken by the ERG ... 119
6.2 Impact on the ICER of additional ERG analyses ... 122
6.3 ERG's preferred assumptions ... 124
6.4 Conclusions of the cost effectiveness section ... 125
7 END OF LIFE ... 128
8 INNOVATION, EQUALITY AND DIVERSITY ... 128
8.1 Innovation ... 128
8.2 Equality and diversity ... 129
9 References ... 130
10 Appendices ... 137
10.1 ERG critique of clinical effectiveness searches ... 137
10.2 Company and ERG risk of bias assessments ... 139
10.3 Overview of response rate outcomes assessed in the CASSIOPEIA trial ... 141
10.4 Revised sensitivity analysis produced by the ERG ... 143
10.4.1 Analysis conducted on the company's base case ... 143
10.4.2 Analysis conducted on the ERG's preferred base case ... 145

LIST OF TABLES

Table 1 Company's base case results, deterministic 12
Table 2 Summary of key issues 13
Table 3 Cumulative cost-effectiveness results for ERG's preferred assumptions, deterministic analysis 22
Table 4 Prognostic factors included in the original (ISS) and revised (R-ISS) versions of the International Staging System for multiple myeloma 24
Table 5 Summary of the decision problem 31
Table 6 ERG appraisal of the company's systematic review methods 36
Table 7 Summary of the trial characteristics for CASSIOPEIA 40
Table 8 Patients' baseline characteristics in the CASSIOPEIA trial 42
Table 9 Response outcomes at primary data cut (median follow up 18.8 months) 56
Table 10 Response outcomes at post-hoc data cut (median follow up 29.2 months) 57
Table 11 Time to response at primary data cut (median follow up 18.8 months) 57
Table 12 PFS and PFS2 at the primary and post-hoc analysis data cuts 58
Table 13 Progression rates at the primary and post-hoc analysis data cuts 59
Table 14 Mortality rates at the primary and post-hoc analysis data cuts 59
Table 15 Landmark analysis hazard ratios for OS and PFS by MRD status 60
Table 16 Change from baseline in EQ-5D-5L index and VAS scores 61
Table 17 Changes from baseline in EORTC QLQ-C30 scores 62
Table 18: Summary of treatment-emergent adverse events in the CASSIOPEIA trial 64
Table 19: Most frequent TEAEs of any grade (≥10% in either group) that differ by ≥5% between study groups 65
Table 20: Most frequent Grade 3 or 4 TEAEs (≥5% in either group) 65
Table 21 Factors included in the MAICs of DBTd and BTd versus Bd 71
Table 22 Factors included in MAICs of DBTd and BTd versus BCd 72
Table 23 Results of the naive comparison and MAIC (DBTd and BTd versus Bd) 74
Table 24 Results of the naive comparison and MAIC (DBTd and BTd versus BCd) 75
Table 25 Eligibility criteria for the SLR of MRD status on survival outcomes 77
Table 26 Clinical effectiveness issues identified by the ERG 81
Table 27 NICE reference case 85
Table 28 Proportion MRD negative at post consolidation assessment 96
Table 29: Overview of company's approach to OS and PFS extrapolation (base case) 97
Table 30: Hazard ratios used for modelling PFS and OS 100
Table 31 Utilities used in the base case model 104
Table 32 Age adjustment utilities in the company's model 104
Table 33 Cost effectiveness: deterministic company base case 110
Table 34: Results from the company's scenario analyses, deterministic analysis 112
Table 35 Comparison of survival predictions from the model and CASSIOPEIA 115
Table 36 Comparison of survival predictions from the model and PHE cohort 117
Table 37 Comparison of modelled outcomes for the BTd arm 118
Table 38 Additional ERG scenario analyses 119
Table 39 Results from the ERG scenario analyses, deterministic analysis 123
Table 40 ERG's preferred model assumptions, deterministic analysis 124
Table 41 Scenario analysis around ERG base case, deterministic analysis 124

Table 42 Cost effectiveness issues identified by the ERG 126
Table 43 Company's base case results, deterministic analysis 143
Table 44 Company's base case results, probabilistic analysis 143
Table 45 Deterministic sensitivity analysis, company base case (top 10 ICER impact) 143
Table 46 ERG's preferred analysis results, deterministic analysis 145
Table 47 ERG's preferred analysis results, probabilistic analysis 145
Table 48 Deterministic sensitivity analysis, ERG base case (top 10 ICER impact) 145

LIST OF FIGURES

Figure 1 Current treatment pathway for newly-diagnosed treatment eligible MM patients and the proposed position of daratumumab combination therapy 28
Figure 2 Overview of the CASSIOPEIA trial design 38
Figure 3 Landmark analysis of OS and PFS by treatment arm and MRD status at post-consolidation assessment (post hoc data cut) 60
Figure 4 Partitioned survival model structure 87
Figure 5 Response-based landmark survival analysis 87
Figure 6 KM data and parametric extrapolations for PFS (CASSIOPEIA ITT population, median follow up 29.2 months) 94
Figure 7 KM data and parametric extrapolations for OS (CASSIOPEIA ITT population, median follow up 29.2 months) 95
Figure 8: Company's extrapolation of OS and PFS BTd MRD-positive patients 99
Figure 9 Comparison of OS and PFS predicted by the model with the CASSIOPEIA (MRD-positive and MRD-negative combined) 116
Figure 1 Cost Effectiveness Acceptability Curve, company base case 144
Figure 2 Cost effectiveness scatterplot, company base case 144
Figure 3 Cost Effectiveness Acceptability Curve: ERG base case 146
Figure 4 Cost effectiveness scatterplot: ERG base case 146

LIST OF ABBREVIATIONS

AE	Adverse event
AESI	Adverse events of special interest
AIC	Academic in confidence
BCd	Bortezomib, cyclophosphamide and dexamethasone
Bd	Bortezomib and dexamethasone
BMI	Body mass index
BNF	British National Formulary
BTd	Bortezomib, dexamethasone and thalidomide
Cd	Carfilzomib and dexamethasone
CIC	Commercial in confidence
CRD	Centre for Reviews and Dissemination
CSR	Clinical study report
CTd	Cyclophosphamide, thalidomide and dexamethasone
DBd	Daratumumab, bortezomib and dexamethasone
DBTd	Daratumumab, bortezomib, thalidomide and dexamethasone
DLd	Daratumumab, lenalidomide and dexamethasone
DSA	Deterministic sensitivity analysis
DSU	Decision Support Unit
EMA	European Medicines Agency
EORTC-CLQ-C30	European Organisation for Research and Treatment of Cancer-Quality of Life Questionnaire-C30
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
ERG	Evidence Review Group
FISH	Fluorescence in situ hybridization
HCRU	Healthcare resource use
HES	Hospital Episode Statistics
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
ICU	Intensive care unit
IDMC	Independent data monitoring committee
IMWG	International Myeloma Working Group
ILd	Ixazomib, lenalidomide and dexamethasone
ITC	Indirect treatment comparison
ITT	Intent to treat
IV	Intravenous

IWRS	Interactive web response system
KM	Kaplan-Meier
LCD	Light chain disease
Ld	Lenalidomide plus dexamethasone
LDH	Lactate dehydrogenase
LOCF	Last observation carried forward
LS	Least squares
LY	Life year
MAIC	Matching-adjusted indirect comparisons
MCID	Minimal clinically important difference
MFC	Multiparametric flow cytometry
MM	Multiple myeloma
MRD	Minimal residual disease
MRD-	MRD-negative
MRD+	MRD-positive
NA	Not applicable
NGS	Next generation sequencing
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NR	Not reported
OR	Odds ratio
PAS	Patient access scheme
PBd	Panobinostat, bortezomib and dexamethasone
Pd	Pomalidomide and dexamethasone
PFS	Progression-free survival
PFS2	Progression-free survival on next line of therapy
PH	Proportional hazards
PPSRU	Personal Social Services Research Unit
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	Performance score
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality-adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
RR	Relative risk/risk ratio
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation

SE	Standard error
SEM	Standard error of the mean
SIGN	Scottish Intercollegiate Guidelines Network
SmPC	Summary of product characteristics
SoC	Standard of care
STC	Simulated treatment comparison
TA	Technology appraisal
TEAE	Treatment-emergent adverse event
UK	United Kingdom
ULN	Upper limit of normal
US	United States
VAS	Visual analogue scale

1 Executive Summary

This summary provides a brief overview of the key issues identified by the evidence review group (ERG) as being potentially important for decision making. It also includes the ERG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of key model outcomes from the company's base case analysis and the modelling assumptions that have the greatest effect on the ICER. Section 1.2 provides an overview of key issues raised by the ERG. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, health technology, evidence and information on the issues are in the main ERG report following this summary.

All issues identified represent the ERG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length of life (overall survival) and quality of life, as reflected in a quality-adjusted life year (QALY). An incremental cost-effectiveness ratio (ICER) reflects the extra cost for every QALY gained. ICERs from the company's base case analysis are shown in Table 1.

Table 1 Company's base case results, deterministic

Intervention	Total			Incremental			ICER
	Costs	LYs	QALYs	Costs	LYs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	[redacted]	-	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: from economic model by ERG, PAS price for daratumumab other drugs at list price							

The model results were most sensitive to the following scenario analysis parameters: treatment effects, hazard ratios (DBTd versus BTd) for progression free survival and overall survival in the MRD-negative and MRD-positive subgroups; and assumptions about loss of treatment effects over time.

1.2 Overview of the ERG's key issues

The ERG's key issues are summarised in Table 2. Additional minor clinical effectiveness issues which the ERG believe have low priority (e.g. because they have been resolved

within this report) are listed at the end of the clinical effectiveness section (see Table 26 below).

Table 2 Summary of key issues

Issue number	Summary of issue	Report sections
1	Uncertainty in hazard ratios (HRs) from the company's meta-analysis of the effects of minimal residual disease (MRD) status on survival outcomes	3.6.6 (ERG critique of the SLR and meta-analysis)
2	Inconsistency in the company's approach for defining and analysing MRD-negative patients	3.2.3.3 (Relationship between sCR and MRD outcomes)
3	Uncertainty in the company's adjustment of PFS to capture the effect of a second randomisation to maintenance therapy	3.2.4.6 (Adjustment for effects of second randomisation to maintenance therapy)
4	Plausibility of long-term survival with standard care (ASCT with BTd induction and consolidation).	4.2.6.2 and 5.3
5	Uncertainty over daratumumab treatment effects on PFS and OS (landmark analysis)	3.2.4.7, 3.2.7.5, 4.2.6.2.4 and 4.2.6.2
6	Waning of treatment effects for daratumumab	4.2.6.2

Of the key issues in Table 2, there are differences between the company's preferred and the ERG's preferred assumptions about loss of treatment effects over time (Key Issue 6). Parameters relating to the remaining key issues were not changed in the ERG's preferred base case. There were differences between the company's and ERG's preferred assumptions for some other parameters that were not included as key issues because they did not have a large effect on cost-effectiveness:

- Mean age of the cohort at the start of induction;
- The cost of daratumumab (based on the fixed-dose subcutaneous or weight-based intravenous formulations);
- The proportions of patients who progress to second-, third- and fourth-line treatment;
- Inclusion of panobinostat, bortezomib and dexamethasone (PBd) at third-line treatment.

1.3 The decision problem: summary of the ERG's key issues

The ERG has not identified any key issues relating to the decision problem. However, there is a discrepancy in the mean age of patients between the pivotal trial and clinical practice, which is discussed in section 1.7 below.

1.4 The clinical effectiveness evidence: summary of the ERG's key issues

The ERG identified the following three key issues regarding the clinical effectiveness evidence.

Issue 1 Uncertainty in hazard ratios (HRs) from the company's meta-analysis of the effects of minimal residual disease (MRD) status on survival outcomes. Report section 3.6.6	
Description of issue and why the ERG has identified it as important	Survival according to MRD status (positive or negative) informs the cost-effectiveness model. Hazard ratios for overall survival (OS) and progression free survival (PFS) are obtained from a meta-analysis conducted by the company. The meta-analysis and its supporting systematic literature review (SLR) have the following limitations: • Searches are 15 months out of date so it is unclear whether any relevant recent studies are missing; • Validity of included studies was not assessed, hence unclear; • Heterogeneity of included studies was not assessed adequately, hence unclear; • Uncertainty around the impact of different MRD definitions on HRs (see Issue 2).
What alternative approach has the ERG suggested?	The Evidence Review Group (ERG) requested that the company update the SLR and meta-analysis to clarify the validity and heterogeneity of the studies and reduce the risk of publication bias. The company were unable to provide an update that could be critiqued by the ERG within the timescale for preparation of this report.
What is the expected effect on the cost-effectiveness estimates?	Results of the economic model are not very sensitive to the confidence intervals for the MRD HRs from the current meta-analysis. The impact on cost-effectiveness should be checked when results from the updated of the SLR and meta-analysis are available.
What additional evidence or analyses might help to resolve this key issue?	The company's updated SLR and meta-analysis may be critiqued by the ERG in an addendum to this report, or at Technical Engagement (depending on the quantity and clarity of the information provided by the company and hence the resources required).

Issue 2 Inconsistency in the company's approach for defining and analysing MRD-negative response. Report section 3.2.3.3	
Description of issue and why the ERG has identified it as important	The company submission (CS) states that MRD-negativity was determined regardless of response. This is inconsistent with the International Myeloma Working Group (IMWG) definition of MRD negativity which requires a complete response. Consequently, the reported rates of MRD negative response from the trial exceed the reported rates of complete response and stringent complete response (≥CR). The CS notes that this is due to a lag in the decay of serum paraprotein (required for complete response) compared to clearance of malignant cells in the bone marrow (required for MRD) . However, the company do not provide a rationale for not reporting MRD negativity according to the IMWG definition, or discuss the implications for interpreting the response outcome analyses. There is also inconsistency in definitions of MRD in the studies included in the meta-analysis (Table 8 in clarification response B9) and there is some evidence from a sensitivity analysis that the different definitions of MRD status influence HRs (Table 9 in clarification response B12). However, the sensitivity analysis does not include MRD when assessed regardless of response and would need updating to include any new studies that are identified in the updated SLR and meta-analysis (see Issue 1). The economic model relies on three sets of inputs to capture the impact of MRD status and treatment on survival outcomes: • The proportion of patients who are MRD negative by treatment arm (DBTd and BTd) at post-consolidation assessment. • Results from the meta-analysis of the effect of MRD status on survival outcomes with standard care (OS and PFS HRs for MRD-negative versus MRD-positive patients treated with BTd). • Results from the CASSIOPEIA trial landmark analysis of survival outcomes by MRD status (OS and PFS HRs for DBTd versus BTd in MRD positive and negative patient groups). It appears that this was conducted with MRD negativity regardless of response, although this is not stated explicitly. The model is most sensitive to changes in the treatment effects estimated from the landmark analysis.
What alternative approach has the ERG suggested?	The company should clarify the impact of different MRD definitions in their updated meta-analysis of MRD impact on survival (Issue 1); and clarify whether MRD definitions used in the model are consistent for the meta-analysis and the landmark

	analysis. MRD definitions should be consistent across input parameters to the model.
What is the expected effect on the cost-effectiveness estimates?	The impact of differing definitions of MRD negativity for input parameters on the direction and magnitude of cost-effectiveness results is unclear. The model is not very sensitive to changes in the MRD response rates for DBTd and BTd from the trial or to results from the MRD meta-analysis. However, the ICER is very sensitive to the treatment effects from the trial landmark analysis. Estimated HRs for OS and PFS are lower, with narrower confidence intervals for patients assessed as MRD negative. It is not clear whether or how these results would change with the IMWG definition of MRD negativity.
What additional evidence or analyses might help to resolve this key issue?	Exploration of the impact of different definitions of MRD negativity (including MRD regardless of response; and MRD with complete response). This exploration should include: • Sensitivity analyses of the MRD meta-analysis when the company have updated their SLR. • Sensitivity analysis of CASSIOPEIA trial results (post-consolidation MRD negative rate and landmark analysis). • Update of the economic model to compare cost-effective results with the different definitions of MRD negativity, consistently applied across all input parameters.

Issue 3 Uncertainty in the company's adjustment of PFS to capture the effect of second randomisation to maintenance therapy. Report section 3.2.4.6	
Description of issue and why the ERG has identified it as important	Randomisation of patients to maintenance therapy (daratumumab monotherapy or observation) in Part 2 of the CASSIOPEIA pivotal trial may influence PFS estimates obtained from Part 1 (induction and consolidation therapy). The company adjusted for the effect of the maintenance re-randomisation on PFS using an inverse probability weighting (IPW) method. They explain that this analysis was pre-specified and that they cannot investigate alternative approaches because they do not have access to data from Part 2 of the trial as it is still blinded (the IPW analysis was conducted by a sequestered group). Uncertainties arise because: • The proportional hazards assumption may have been violated • The adjustment is based on immature data from the maintenance therapy period • The ERG are unable to validate the analysis as Part 2 of the trial remains blinded • There is no comparison of alternative methods of adjustment

	• OS could not be adjusted due to data immaturity • The landmark analysis of relative treatment effects on post-consolidation PFS and OS stratified by MRD status is not adjusted.
What alternative approach has the ERG suggested?	None at the moment, since Part 2 is blinded and the company cannot currently access the data
What is the expected effect on the cost-effectiveness estimates?	The model does not currently include any adjustment of PFS or OS for maintenance therapy, because survival is modelled using results from the landmark analysis. The reported IPW analysis gives similar adjusted and unadjusted PFS HRs, which suggests that no adjustment is needed in the economic model. However, this is uncertain as the impact of maintenance treatment may change as the data mature.
What additional evidence or analyses might help to resolve this key issue?	When Part 2 of the trial is unblinded the company intend to provide an updated analysis including additional data. This is expected to be at Technical Engagement. Depending on the results, further analysis of the effects of maintenance treatment on cost-effectiveness results may be appropriate.

1.5 The cost effectiveness evidence: summary of the ERG's key issues

The ERG identified the following three key issues regarding the cost effectiveness evidence.

Issue 4 Plausibility of long-term survival with standard care (ASCT with BTd induction and consolidation). Report section 4.2.6.2	
Description of issue and why the ERG has identified it as important	Overall survival data from the CASSIOPEIA trial are very immature and yield a wide range of extrapolations. Predictions from the company base case model for the standard care arm (BTd) are consistent with CASSIOPEIA and with data from a Public Health England (PHE) cohort over a period of three years from diagnosis. However, there is high uncertainty over the plausibility of survival estimates over the modelled time horizon of over 40 years because of a lack of long-term data for the population of interest. The company use a response-based approach to modelling survival. The patient cohort is split into MRD-negative and MRD-positive subgroups after consolidation treatment. For the MRD-positive BTd subgroup, survival is modelled with an exponential (constant mortality risk) extrapolation fitted to Kaplan-Meier data from the trial. This is then adjusted to model survival for the MRD-negative BTd subgroup using a constant HR (MRD-negative versus MRD-positive) estimated from the meta-analysis. Survival in the DBTd treatment arm is then modelled using constant HRs

	for DBTd versus BTd, estimated separately for the MRD subgroups in the landmark analysis (see Issue 5 below). We agree with the company's decision to use the most conservative (exponential) distribution based on the trial data. And proportional hazards assumptions do appear to hold for comparisons between MRD and treatment groups in the landmark analysis of CASISOPEIA. However, survival extrapolations for MRD-positive BTd patients exceed clinical expectations, with constant or decreasing hazards that reach general population rates within 30 years. We also note that estimated life years and QALYs from the BTd arm in the current model are considerably higher than estimates for BTd in the NICE appraisal TA311.
What alternative approach has the ERG suggested?	Alternative modelling scenarios based on current trial data cannot resolve uncertainty over the plausibility of the survival extrapolations.
What is the expected effect on the cost-effectiveness estimates?	This is uncertain, although we note that increasing the hazard for the exponential OS for BTd MRD-positive patients reduces the ICER.
What additional evidence or analyses might help to resolve this key issue?	Further clinical opinion on the plausibility of the modelled survival estimates illustrated in CS Figure 35. Updated survival analysis of CASSIOPEIA trial data when available. Comparison with other sources of survival estimates (including KM data for other studies in the MRD meta-analysis and real-world data). If other high quality, relevant data with longer follow up is available, this could be used to calibrate modelled survival estimates.

Issue 5 Uncertainty over daratumumab treatment effects on PFS and OS (landmark analysis). Report section 4.2.6.2	
Description of issue and why the ERG has identified it as important	In addition to the impact of DBTd on post-consolidation MRD response rates, the company base case includes ongoing OS and PFS benefits for both MRD-positive and MRD-negative patients treated with DBTd rather than BTd. These additional treatment effects are estimated from the CASSIOPEIA landmark analysis. The estimated risk reductions in the MRD-negative group are higher than in the MRD-positive group, and we note that there is uncertainty over these effects, as the confidence intervals include 1 for PFS and OS in the MRD-positive group, and for OS in the MRD-negative group.
What alternative approach has the ERG suggested?	We consider it appropriate to include the landmark HR estimates in our preferred analysis, but we note that they are subject to uncertainty. We tested the impact of this uncertainty on the ICER in the following scenarios:

	• Landmark effects only for MRD-negative (HR=1 for MRD-positive PFS and OS) • Landmark effects only for MRD-negative PFS (HR=1 for MRD-negative OS and for MRD-positive PFS and OS) • No landmark effects (HR=1 for OS & PFS MRD-positive and MRD-negative)
What is the expected effect on the cost-effectiveness estimates?	Incorporating the above ERG scenarios has a significant impact on the cost effectiveness results, increasing the ICERs substantially compared to the base case results: [redacted] (scenario: effects only for MRD-negative), [redacted] (scenario: effects only for MRD-negative PFS) and [redacted] (scenario: no effects in either MRD group).
What additional evidence or analyses might help to resolve this key issue?	Clinical opinion on the plausibility of the estimated risk reductions in MRD-positive and MRD-negative subgroups. Updated analysis of CASSIOPEIA, with adjustment for maintenance treatment when additional follow-up is available (as suggested in Issue 3). Sensitivity analysis based on alternative definitions of MRD negativity (as suggested in Issue 2 above).

Issue 6 Waning of treatment effects for daratumumab. Report section 4.2.6.2	
Description of issue and why the ERG has identified it as important	The company assume that the relative treatment effects of DBTd in MRD-positive and MRD-negative subgroups persist throughout the model time horizon. We view this is an optimistic assumption as the model relies on survival data from the CASSIOPEIA trial with a median follow up of 29.2 months after the start of induction. In their scenario analyses, the company show the impact of loss of treatment effects after 5 and 10 years in both MRD subgroups, and in the MRD-negative group alone.
What alternative approach has the ERG suggested?	For our base case analysis, we assumed waning of treatment effects (HR=1) five years after consolidation therapy. We also tested a scenario with loss of effects after 3 years.
What is the expected effect on the cost-effectiveness estimates?	The ERG preferred assumption of loss of treatment effects after 5 years increases the company's base case ICER to [redacted] per QALY gained. The ERG scenario of loss of effects after 3 years increases the base case ICER to [redacted], which is significantly above NICE's willingness-to-pay threshold of £30,000 per QALY.
What additional evidence or analyses might help to resolve this key issue?	Clinical opinion on the plausibility of persistence of treatment effects. Updated analysis of trial data when longer follow up is available.

1.6 Other key issues: summary of the ERG's view

The ERG have not identified any other key issues. Note that a full list of clinical effectiveness issues, including minor issues that the ERG believe have been resolved within this report, is provided in Table 26 below.

1.7 Summary of ERG's preferred assumptions and resulting ICERs

Based on the ERG critique of the company's model (discussed in section 6.1), we have identified five key aspects of the company base case with which we disagree. Our preferred model assumptions are the following:

1. **Waning of treatment effects**: The company assume that after DBTd consolidation, both MRD-positive and MRD-negative groups continue to benefit from improved survival outcomes (PFS and OS) over the modelled time horizon of about 40 years, compared with those treated with BTd (no waning of treatment effects). We think that there is insufficient evidence to support this assumption. In the ERG preferred analysis, we assume loss of treatment effects (HR=1 for DBTd versus BTd for PFS and OS in MRD-positive and MRD-negative subgroups) from 5 years after consolidation treatment.
2. **Patients' mean age:** The mean age of patients in the company's analysis (56.6 years based on patients in the CASSIOPEIA trial), does not reflect the target population. The trial excludes patients over the age of 65 years, although transplant eligibility in England is assessed on the basis of fitness, not age. Real world evidence data from a Public Heath England (PHE) dataset is that [redacted] of newly diagnosed MM patients who are transplant eligible are aged 65 years or older. We use the mean age at diagnosis from the PHE cohort ([redacted] years) in the ERG preferred analysis.
3. **Daratumumab formulation for costing**: The company estimate the cost of daratumumab administered as a fixed-dose subcutaneous (SC) injection, rather than as a weight-based intravenous (IV) infusion as used in the CASSIOPEIA trial. We note that the European Medicines Agency have accepted evidence of non-inferiority for SC versus IV, based on the COLUMBA trial. However, we prefer to use the cost for IV daratumumab for consistency with the effectiveness evidence, with the cost of SC daratumumab as a scenario.
4. **Subsequent lines of treatment:** The company include costs for second, third and fourth lines of treatment for all patients who progress after first line treatment. We understand that this will be true for some patients, indeed some may complete up to 7 lines of treatment. However, it is likely that some patients become unfit for further

treatment or die before completing the 4th line of treatment. Data on the number of patients starting subsequent lines of treatment is available in the PHE dataset, although this has limited follow up (during median follow up of [redacted] months, [redacted] started first line treatment, [redacted] second line, [redacted] third line and [redacted] fourth line). In our analysis we assume that of the patients who progress after first line treatment, 100% start second line, 75% third line and 50% fourth line treatment.

5. **Panobinostat, bortezomib and dexamethasone (PBd) at third line:** The NHS England clinical pathway for transplant eligible patients includes PBd as a treatment option at third and fourth line (CS Figure 3). In their base case, the company assume that 45%/46% of patients treated at third line after first line DBTd/BTd would receive PBd (and none at fourth line). Advice to the ERG was that PBd is not currently used at third or fourth line in practice. We therefore exclude PBd from cost calculations in the ERG preferred analysis.
6. **Age adjustment for utility:** We use EQ-5D data from the Health Survey for England in 5-year age bands up to the age of 85 (rather than 10-year estimates to age 75 in the company base case). This was included to check the impact the change of age.

The ICERs obtained using the ERG's preferred assumptions are shown in Table 3. The assumption on waning of treatment effects had the largest impact on the ICER. The change in baseline age, using the cost of IV daratumumab, assumptions about the proportions of patients with progressed disease who started subsequent lines of treatment, and the exclusion of PBd as a third line treatment led to further, small, increases of the ICER.

Table 3 Cumulative cost-effectiveness results for ERG's preferred assumptions, deterministic analysis

Individual scenarios on the base case	Treatment	Total costs (£)	Total QALYs	ICER (£/QALY)
Company base case	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
1. Loss of effects after 5 years (PFS and OS HR=1 MRD+ &MRD-)	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
2. Mean age [redacted] years (PHE dataset)	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
3. Daratumumab IV formulation	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
4. Subsequent treatment 2L (100%), 3L (75%) & 4L (50%)	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
5. No PBd at 3L	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
6. Age adjustment of utility: 5-year bands	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]
ERG preferred model	BTd	[redacted]	[redacted]	-
	DBTd	[redacted]	[redacted]	[redacted]

Source: produced from model by ERG, PAS discount for daratumumab, list prices for other drugs

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Janssen-Cilag on daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable. It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the evidence review group (ERG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on 21st September 2020. A response from the company via NICE was received by the ERG on 12th October 2020 (excluding questions B3 and B7) and 19th October 2020 (question B7) and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on multiple myeloma

An accurate overview of multiple myeloma (MM) is provided in CS section B.1.3.1, supported by a range of appropriate references.[1-17] Notable features of MM are:

- MM is characterised by proliferation of malignant plasma cells within bone marrow and overproduction of abnormal immunoglobulin proteins, referred to as M-protein or paraprotein. Disease progression and response to therapy are indicated by the concentration of M-protein in plasma, and the number of plasma cells/myeloma cells in bone marrow. To be classed as a complete response (CR) both assessments have to be normal.
- The disease follows a relapsing-remitting course in which all newly-diagnosed patients eventually become refractory, with prognosis worsening with each successive relapse.
- According to the latest data from Cancer Research UK there were 5,034 new cases of MM in England in 2017.[18]

Prognostic factors

Clinical outcomes including OS vary according to a range of prognostic factors. Key risk factors are captured within the International Staging System (ISS)[10] which classifies patients into three groups according to their prognosis (Table 4). The most recent, revised, version of the ISS, known as R-ISS, includes additional prognostic factors. A large study on 3,060 patients[10] found that at a median follow-up of 46 months, 5-year overall survival (OS) was

82% in the R-ISS stage I group, 62% in the R-ISS stage II group, and 40% in the R-ISS stage III group, with respective 5-year progression-free survival (PFS) rates of 55%, 36%, and 24%.

Table 4 Prognostic factors included in the original (ISS) and revised (R-ISS) versions of the International Staging System for multiple myeloma

Stage	ISS criteria	R-ISS criteria
Stage I (best prognosis)	People with serum β2-microglobulin level < 3.5 mg/L and serum albumin level ≥ 3.5 g/dL	People with serum β2-microglobulin level < 3.5 mg/L and serum albumin level ≥ 3.5 g/dL, no high-risk chromosomal abnormalities [del(17p) and/or t(4;14) and/or t(14;16)], and normal LDH level (less than the upper limit of normal range)
Stage II (intermediate prognosis)	People who are not ISS stage I or III	People who are not R-ISS stage I or III
Stage III (worst prognosis)	People with serum β2-microglobulin level > 5.5 mg/L	People with serum β2-microglobulin level > 5.5 mg/L and high-risk chromosomal abnormalities or high LDH level
Source: International Myeloma Working Group (IMWG)[10] LDH: serum lactate dehydrogenase		

Chromosomal abnormalities detected by fluorescent in situ hybridization (FISH) are key features of MM and are included in the revised ISS (the revised ISS expands the original ISS by adding chromosomal abnormalities and serum lactate dehydrogenase). According to the International Myeloma Working Group,[10] standard-risk newly diagnosed MM is characterized by the absence of del(17p), translocation t(4;14)(p16;q32), or translocation t(14;16)(q32;q23) and is associated with a median OS of 50.5 months, whereas high-risk disease is characterized by the presence of at least one of the previously mentioned abnormalities and is associated with a median OS of 24.5 months.

The ERG's clinical experts commented that the R-ISS captures most of the currently recognised key prognostic factors. Additionally:

- All experts agreed that renal function is also an important prognostic factor.
- One expert added that general fitness, presence of comorbidities, and the extent of extramedullary disease (that is, manifestations of MM outside the bone marrow, including plasma cells and soft tissue MM lesions) are also likely to be prognostically important.

- Another expert commented that in the UK a wider range of FISH abnormalities than the IMWG criteria is assessed, including t(14;20), gain 1q and del 1p. Two or more of the high risk abnormalities is used to define ultra high risk myeloma and has been confirmed from the Myeloma IX and Myeloma XI trials.[19-21]

The CS reports that the R-ISS was published after the company's pivotal trial (CASSIOPEIA) had commenced and therefore the baseline population characteristics in the pivotal trial (as well as in comparator trials) report the ISS rather than the R-ISS. The company made a post-hoc adjustment to derive R-ISS from the available baseline characteristics in the pivotal trial (CS section B.2.3.2). The implications of this adjustment for interpreting the prognosis of patients in the company's pivotal trial are discussed in section 3.2.1.2 below. In clarification response A7 the company explained that the original version of the ISS was the version used in all analyses reported in the CS, since this was the only version available in studies and data sets referred to in the CS.

We note that age, which is often considered to be a prognostic factor for cancer survival, is not included in the R-ISS. As explained below (section 2.2.3), one of the key goals of the treatment pathway for NDTE MM patients is to achieve an autologous stem cell transplant (ASCT). According to NICE Guidance[22] and the ERG's clinical experts, a patient's eligibility for ASCT is determined primarily by their overall fitness (i.e. degree of frailty, risk factors, and comorbidities) rather than by their age. Therefore, when assessing the comparability of trials on NDTE MM, imbalances in patient age may not necessarily introduce bias (since differences in age would not necessarily imply differences in fitness).

2.2.2 Background information on daratumumab

The mode of action of daratumumab (Darzalex®), as understood from pre-clinical in vitro data, is summarised accurately in CS Table 2 and in CS Figure 26, consistent with the mode of action described in the Summary of Product Characteristics (SmPC)[2] and the European Public Assessment Report (EPAR).[23]

Daratumumab is a human monoclonal antibody that binds to CD38, a cell surface glycoprotein that is expressed by various immune cells including white blood cells and expressed at a high level by myeloma tumour cells. CD38 performs several roles including receptor mediated adhesion, signalling, and enzymatic activity. Daratumumab directly affects CD38-expressing MM tumour cells through complement-dependent cytotoxicity, antibody-

dependent cell-mediated cytotoxicity, antibody-dependent phagocytosis, and induction of MM cell apoptosis, as well as having several immunomodulatory effects.

Daratumumab is the first human monoclonal antibody to target CD38. In the current indication daratumumab is to be used in combination with bortezomib (B), thalidomide (T) and dexamethasone (d), i.e. DBTd, for treating multiple myeloma patients who are newly-diagnosed and transplant-eligible (NDTE). Marketing authorisation for the DBTd combination therapy for NDTE patients was granted in January 2020.

Daratumumab is available as a 20mg/ml solution for intravenous (IV) infusion with a weight-based dose of 16 mg/kg and as an 1800mg solution for subcutaneous injection. The clinical effectiveness evidence for DBTd presented in the CS is taken from the company's pivotal CASSIOPEIA trial (for details of CASSIOPEIA see section 3.2.1 below) in which daratumumab was administered by IV infusion. A licence extension authorising the subcutaneous (SC) injection of daratumumab was granted in June 2020 and the company expect the SC injection will be preferred in clinical practice, due to convenience, reduced risk of infusion-related reactions, and lower cost (CS section B.3.8.3). The company use the cost of the daratumumab SC injection in their economic model base case and the cost of the daratumumab IV infusion in a scenario analysis (CS Table 81; CS section B.3.6.2).

The dosing schedule for the combination therapy of daratumumab (IV), bortezomib (SC), thalidomide (oral), and dexamethasone (oral) in the CASSIOPEIA trial (CS Figure 3) is consistent with how these drugs would be used in NHS practice in each therapy cycle, based on a cycle length of 28 days. An exception, according to one of the ERG's clinical experts, is that bortezomib would normally be given once weekly over a longer period to the same total dose overall in order to be better tolerated and reduce neuropathy. Another expert commented that most patients would be able to follow the CASSIOPEIA schedule.

In the CASSIOPEIA trial and the company's proposed indication, four pre-transplant induction cycles of therapy and two post-transplant consolidation cycles are administered (i.e. 6 cycles in total). This differs from NHS practice which does not currently include post-transplant consolidation therapy. Clinical experts advising the ERG commented that:

- In current NHS practice NDTE MM patients receive 4 to 6 cycles of induction therapy (i.e. all pre-transplant), with most patients receiving 6 cycles
- Approval of the DBTd regimen, including the consolidation therapy cycles, would be welcomed by physicians

- The proposed 4 induction and 2 consolidation cycles of DBTd could be easily integrated into existing NHS practice without requiring additional resources (other than daratumumab itself).

The CS reports the dosing interval for daratumumab when administered by SC injection to be weekly in the first two induction cycles; every 2 weeks in the third and fourth induction cycles; and weekly in each of the two consolidation cycles (CS Table 2).

2.2.3 The current treatment pathway for multiple myeloma and the proposed place of daratumumab

Treatment guidelines for the management of MM are available from NICE (Guidance NG35),[22] as well as the European Society of Medical Oncology (ESMO),[24] European Myeloma Network (EMN),[25] and National Comprehensive Cancer Network (NCCN).[26]

The current NHS treatment pathway for NDTE MM patients is summarised in CS section B.1.3.4 and in CS Figure 3, reproduced in Figure 1 below. The ERG's clinical experts agreed that the CS provides a broadly accurate representation the current care pathway and an accurate indication of the intended position of the company's proposed technology DBTd.

The CS does not explicitly discuss the diagnosis and monitoring of MM. According to NICE NG35,[22] diagnosis of MM is based initially on serum protein electrophoresis to determine the presence of M-protein, followed by bone marrow aspiration and trephine biopsy to confirm that the plasma cell morphology and frequency corresponds to MM. A FISH assay on selected plasma cells obtained at this stage is used to identify the presence of the cytogenetic risk factors listed above in section 2.2.1 which are indicative of patient prognosis. Patients with newly diagnosed MM would also receive whole-body imaging to assess for MM-related bone disease and extra-medullary lesions (i.e. those outside of the bone marrow). According to NICE NG35 people who have completed MM therapy and recovered should be monitored every 3 months for symptoms and serum disease markers (immunoglobulins and protein electrophoresis), although one of the ERG's clinical experts suggested that monitoring would be every two months for the first year post-treatment. One of the ERG's clinical experts commented that monitoring is sometimes done more frequently than 3-monthly, e.g. if MM is genetically defined as high risk.

For patients with newly-diagnosed MM the aims of treatment are to induce remission, delay progression, prolong survival and maximise quality of life. For those with newly-diagnosed MM who are fit enough, a key objective of therapy is to achieve autologous stem cell transplant (ASCT) which, although not currently curative, improves depth of response and PFS. The company estimate that approximately one third of newly-diagnosed MM patients would be expected to be eligible for ASCT. The ERG's clinical experts agree, although one expert expected nearer 40% to be eligible at the time of diagnosis, with around one third eventually achieving ASCT. The experts also commented that eligibility for ASCT is based on the patient's overall fitness and ideally achieving at least a partial response to induction treatment. The current standard of care for NDTE MM patients has three main components: bortezomib-based induction therapy to stabilise the disease; high-dose chemotherapy to kill MM cells; and ASCT.

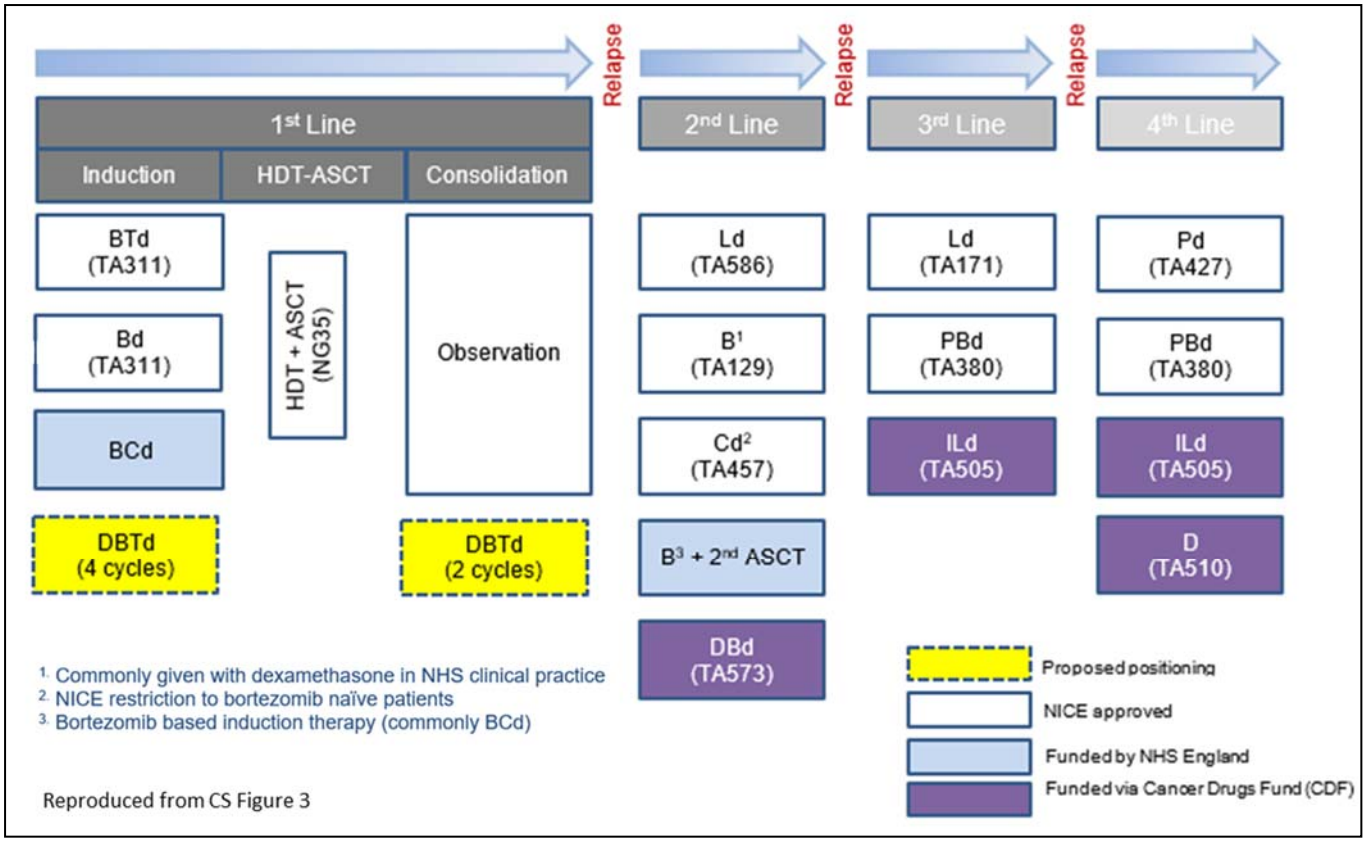


Figure 1 Current treatment pathway for newly-diagnosed treatment eligible MM patients and the proposed position of daratumumab combination therapy

As discussed above (section 2.2.2), the company's proposed DBTd therapy would be administered as four cycles of induction therapy prior to ASCT and a further two cycles of therapy post-ASCT, referred to as consolidation therapy.

In the current NHS care pathway NDTE MM patients who have received a stem cell transplant would be monitored, instead of receiving a consolidation therapy (one of the ERG's clinical experts commented that in their practice consolidation therapy would be offered if the patient had only received 4 cycles of induction therapy, although this is uncommon). On disease progression patients would then proceed to receive second-line therapy or, in a minority of patients who are suitable, a second ASCT. CS Figure 3 indicates that patients who relapse on second-line therapy may receive third-line and fourth-line therapies (Figure 1). Clinical experts advising the ERG suggested that some patients may also experience up to seven lines of therapy.

CS Figure 3 suggests that three induction therapies are currently used prior to ASCT. These are bortezomib + thalidomide + dexamethasone (BTd) and bortezomib + dexamethasone (Bd), as approved in NICE TA311; and bortezomib + cyclophosphamide + dexamethasone (BCd), not currently approved by NICE but funded by NHS England (Figure 1). The company assert that BTd is the standard of care and BCd and that Bd are seldom used in the NDTE MM population (CS section B.1.3.4). The ERG's clinical experts gave slightly different estimates of the extent to which these therapies are used, with their estimates ranging from 40% to 80% for BTd, 10% to 40% for BCd, and up to 10% for Bd, with some other therapies (e.g. cyclophosphamide + dexamethasone; Cd) also being used for a minority (<10%) of patients.

One of the ERG's clinical experts noted minor exceptions to CS Figure 3 relating to later lines of therapy in the care pathway (post-consolidation):

- CS Figure 3 suggests panobinostat + dexamethasone therapy (PBd) is used third-line and fourth-line but the expert felt that due to toxicity concerns PBd would mostly be used as a late (e.g. fifth) line of therapy.
- CS Figure 3 suggests ixazomib + lenalidomide + dexamethasone (ILd) is used at third-line and fourth-line. As a result of the COVID-19 pandemic, ILd is currently (temporarily) approved by NHS England also for second-line use.

As indicated in CS Figure 3 the care pathway is supported by several previous NICE technology appraisals, most of which refer to later lines of therapy post-ASCT:

- first-line: Bd, BTd (TA311)[27]
- second-line: bortezomib (TA129),[28] carfilzomib (TA457),[29] DBd (TA573),[30] lenalidomide (TA586)[31]

- third-line: lenalidomide (TA171),[32] PBd (TA380),[33] ILd (TA505)[34]
- fourth-line: PBd (TA380),[33] pomalidomide + dexamethasone (Pd) (TA427),[35] ILd (TA505),[34] daratumumab (TA510)[36]

The company acknowledge that most of the clinical management of MM is provided in an outpatient setting, with the bulk of care being provided by caregivers (CS section 1.3.3). The CS does not explicitly discuss current service provision (i.e. the staff and infrastructure resources required to deliver MM therapy) in their description of the care pathway in CS section B.1.3.4).

2.3 Critique of the company's definition of the decision problem

The NICE decision problem template is shown in Table 5 below. The CS only reports the population and comparator(s) components of the template (CS Table 1). The missing components were provided by the company in clarification response A1 and have been included in Table 5 below.

Table 5 Summary of the decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Population	People with previously untreated MM who are eligible for ASCT	Adult patients with newly diagnosed MM who are eligible for ASCT	This population is considered to be in line with the full marketing authorisation for this indication	The NICE scope and company decision problem are consistent except for patient age. The decision problem restriction to adult patients aged 18 to 65 years is consistent with the pivotal trial population (although the Summary of Product Characteristics [SmPC][2] does not specify an upper age limit). The ERG's clinical experts commented that in NHS practice some patients aged up to 75 years would be considered for ASCT if fit enough (and so eligible for DBTd if approved by NICE) and hence this age difference between the decision problem and NHS practice would not be expected to limit generalisability of clinical effectiveness (see section 3.2.1.2 below). However, the age difference does have an influence on cost-effectiveness of DBTd (see ERG scenario 1 in Table 39 below).
Intervention	DBTd	As per the final scope	Not applicable	The intervention is consistent with the NICE scope

Comparator(s)	• Bortezomib with dexamethasone (Bd) or with dexamethasone and thalidomide (BTd) • Bortezomib with cyclophosphamide and dexamethasone (BCd) (off-label) • Cyclophosphamide with thalidomide and dexamethasone (CTd) (off-label)	• Bd • BTd • BCd (off-label)	Janssen does not consider CTd a relevant comparator to DBTd in this indication following clinical expert feedback that CTd is rarely used as an induction therapy for NDTE MM patients in England. Real-world evidence supports limited CTd usage, with steady decline in prescribing and less than 2% of NDTE MM patients in England treated with CTd since 2018. Furthermore, CTd is not recommended by NICE, or recognised by international or European clinical practice guidelines.	The NICE scope and company decision problem are consistent, except for the exclusion of CTd by the company. The ERG's clinical expert advisors agreed that exclusion of CTd is appropriate, since CTd use in clinical practice is negligible (experts estimated <5%).We therefore agree that the company's choice of comparator therapies is appropriate and reflective of NHS practice.
Outcomes	The outcome measures to be considered include: • Overall survival • Progression-free survival • Response rate • Minimal residual disease-negative status • Proportion of people undergoing high dose chemotherapy and autologous stem cell transplantation • Adverse effects of treatment • Health-related quality of life	As per the final scope	Not applicable	The company's outcomes are consistent with the NICE scope. NB the proportion of people undergoing high dose chemotherapy and autologous stem cell transplantation is not specified as a discrete outcome in the CS; however, this information is included in CS Figure 7 and informs the economic model as an input parameter (CS Table 80).

Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies will be taken into account. The availability and cost of generic products should be taken into account.	No response provided	No response provided	The company's economic analysis is consistent with the NICE scope
Subgroups	No subgroups specified	No response provided	No response provided	Pre-specified subgroups in the pivotal CASSIOPEIA trial are listed in CS Table 4. Subgroup analysis results are reported in CS section 2.7 and CS Appendix E but do not inform the economic model.
Special considerations including issues related to equity or equality	No equity or equality issues specified	No equality issues identified (CS section B.1.4)	No response provided	The ERG are not aware of any equality or equity issues relating to the anticipated use of DBTd in clinical practice.

ASCT = autologous stem cell transplant; Bd = bortezomib and dexamethasone; BCd = bortezomib, cyclophosphamide and dexamethasone; BTd = bortezomib, dexamethasone and thalidomide; CTd = cyclophosphamide, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; MM = multiple myeloma; NDTE = newly diagnosed transplant-eligible

We note that the lower age limit for NDTE MM patients in the CS is 18 years. This is consistent with the company's pivotal CASSIOPEIA trial population and the licensed indication, which refer to the adult population. Although not explicitly stated in the CS, no data are available on the safety and efficacy of daratumumab in people aged below 18 years[2] and the number of young people who experience MM is very low (a large study found only 5% of MM patients were aged below 30 years[37] and a clinical expert advising the ERG considered the number of young people with MM to be exceptionally low). We therefore agree that the lower age limit of 18 years is appropriate.

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of the company's systematic literature review

The systematic literature review (SLR) for clinical effectiveness studies is reported in CS Appendix D (referred from CS section B.2.1). This same search informed the identification of studies for the indirect treatment comparison (CS section B.2.9.1). A summary of the ERG's critique is provided in Table 6 below.

The company searched for randomised controlled trials (RCTs), and clinical trials generally, in the NDTE MM disease population in the relevant core bibliographic databases and in five relevant conferences. Further details of the searches are in Appendix 10.1 of this report. Searching for the term 'clinical trial' in the title, abstract and publication type, means that the SLR is versatile enough to support identifying relevant studies for the matched-adjusted indirect comparison (MAIC) where non-RCT studies were eligible. The ERG believe that this, and the response to clarification question A2, explains the identification of two single-arm studies not meeting the original RCT study design criteria (CS Table 29).

Table 6 ERG appraisal of the company's systematic review methods

Systematic review components and processes	ERG response	ERG comments
Was the review question clearly defined using the PICOD framework or an alternative?	Yes	Used a PICOS-T framework, where 'T' for timepoint specifies stages in treatment and disease course that are relevant to the outcomes: post-induction, post-ASCT, and long-term survival outcomes up to 60 months.
Searches: was the literature review carried out appropriately (sources, date range, in line with PICOD, correct search terms/syntax, etc.)?	Yes	Designed to identify both RCTs and non-RCTs. Full details are in Appendix 9.1 of this report.
Searches: were any relevant studies missed?	No	Searches were four months out of date, but the ERG identified no further relevant studies published since the company's latest update in April 2020.
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes	Specified in CS Appendix Table 9. Population matches scope, broad choice of interventions/comparators informed by various treatment guidelines, and outcomes are appropriate.

Were study selection criteria applied by two or more reviewers independently?	Yes	Two independent investigators, discrepancies addressed via discussion with a third investigator resolving disagreements.
Was data extraction performed to a reasonable standard (e.g. use of two reviewers)?	Yes	Two independent investigators and a third investigator to resolve disagreements (as above).
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes	Reported in CS section B.2.5 and CS Appendix D. Risk of bias assessments were made for: • The pivotal CASSIOPEIA trial, using NICE criteria (CS section B.2.5 and CS Appendix D.3); • CASSIOPEIA, and the comparator studies IFM 2005-01 and GMMG-MM5, as included in the MAIC, using the updated version of the Cochrane Risk of Bias assessment tool (CS Appendix D.1.6).
Was risk of bias assessment (or other study assessment) conducted by two or more reviewers independently?	Unclear	Not reported. Full ERG discussion on the risk of bias assessment is in section 3.2.2 of this report.
Is sufficient detail on the individual studies presented?	Yes	ERG discussion of the pivotal study, CASSIOPEIA is in section 3.2.1 of this report.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes, albeit with limitations	Indirect treatment comparisons are reported in CS section B.2.9 and CS Appendix D. ERG discussion is in sections 3.4 and 3.5 of this report.
MAIC: Matched-adjusted indirect comparison; RCTs, randomised controlled trials		

In addition to the SLR of clinical effectiveness studies, the company conducted a SLR and meta-analysis of the impact of patients' minimal residual disease (MRD) status (positive or negative) on survival outcomes, to inform the economic model. This SLR and meta-analysis is discussed and critiqued separately, in section 3.6 of this report below.

ERG conclusions:

Overall, the company's SLR of clinical effectiveness studies was well conducted and adequately reported. All relevant studies for this submission appear to have been identified.

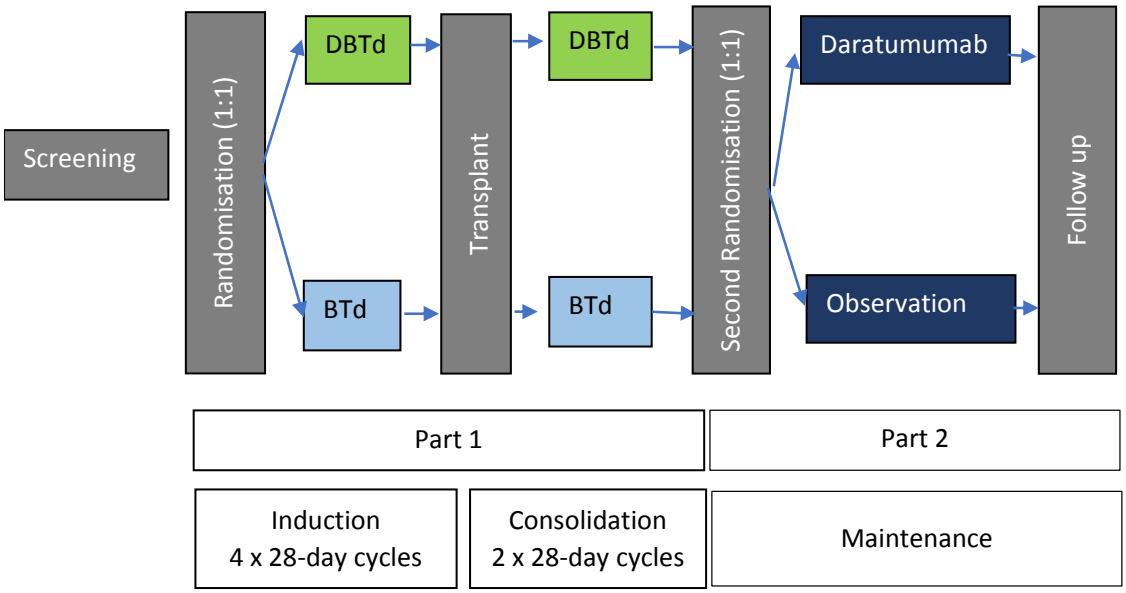
3.2 ERG critique of the included clinical effectiveness studies

3.2.1 Included and ongoing studies

The company's systematic literature review identified one study relevant to the decision problem. This study, the CASSIOPEIA trial, is a company-sponsored, phase III randomised, open-label, active-controlled, European, multi-centre trial.

3.2.1.1 Study design and conduct

CASSIOPEIA is a two-part study comparing the safety and efficacy of DBTd with BTd in patients who are newly diagnosed with multiple myeloma, who have not received previous treatment and who are eligible for ASCT. The study design is shown in Figure 2 below. The primary efficacy outcome was the proportion of patients achieving stringent complete response (sCR) post-consolidation at (or within 30 days) of day 100 post-ASCT. Outcome definitions, including those that contribute to the economic model, are described in section 3.2.3 of this ERG report.



Source: CS Figure 4 redrawn by the ERG

Figure 2 Overview of the CASSIOPEIA trial design

Part 1 comprised an induction phase (4 x 28-day cycles) followed by ASCT and a consolidation phase (2 x 28-day cycles) starting approximately 30 days after ASCT (CS section B.2.3.1). In Part 1, eligible patients were randomised to one of the following two treatment regimens:

- DBTd
    - o Daratumumab (16mg/kg) intravenous infusion once weekly for the first two 28-day induction cycles followed by once every two weeks for two remaining

induction 28-day cycles and the two post-transplant consolidation 28-day cycles, and
    - BTd as described below
- BTd
    - Bortezomib (1.3mg/m^2) subcutaneous injection twice a week for the first two weeks of all four induction cycles and the two consolidation cycles,
    - Thalidomide 100mg daily orally throughout all induction and consolidation cycles and
    - Dexamethasone orally or intravenous infusion starting at a 40mg dose on the first two days per week, reducing to a 20mg dose during later cycles (CS Table 3 and CS Figure 5 provide further details of dexamethasone dosing).

The ERG's clinical experts advised that in clinical practice the thalidomide dose would initially be titrated upwards from 50mg daily to check tolerability. The bortezomib dose may be given once weekly over a longer period in practice in some patients to reduce side effects. The ERG note that the total dose of dexamethasone in CASSIOPEIA (1200mg) is similar to that of a 4-cycle induction with BTd (1280mg) but the timing of doses is more regular in CASSEOPIA (dosing on two days per week with fewer steroid-free weeks) than that of the current BTd regime recommended in the label for bortezomib (dosing more concentrated on four days per week but for fewer weeks).[38] One of the ERG's clinical experts commented that the BTd regimen commonly used in the UK is a 4-week cycle with once weekly bortezomib but no steroid-free weeks. Another expert noted that in clinical practice patients on weekly BTd would receive dexamethasone 20mg for two days per week; response is unlikely to be affected by the timing of the therapy if the total dose is comparable (in practice steroid doses are frequently adjusted to manage toxicities).

In Part 2, patients who had achieved at least a partial response after consolidation were re-randomised to either:

- maintenance monotherapy with daratumumab 16mg/kg intravenous infusion every eight weeks until disease progression (for up to two years) followed by observation only, or
- observation only until disease progression (up to two years)

Patients who did not achieve at least a partial post-consolidation response or who did not wish to proceed to Part 2 were followed up until disease progression or death, even if they received subsequent treatment (CS section B.2.3.1 and CS Table 9).

As Part 2 is ongoing and remains blinded to the company, only results from Part 1 inform the current submission. This includes data from a pre-specified primary data cut (June 2018) and an additional unplanned post-hoc data cut (May 2019) requested by the European Medicines Agency (EMA) (for further details of the data cuts see section 3.2.4 below). The company expect Part 2 results to be available for their response to Technical Engagement. Results for longer-term efficacy outcomes in Part 1 such as progression-free survival (PFS) and overall survival (OS) may be affected by the second randomisation in Part 2. To investigate this effect, the company conducted an exploratory post hoc adjustment of PFS, as discussed in section 3.2.4.6 of this ERG report.

Key trial features for Part 1 are summarised below in Table 7.

Table 7 Summary of the trial characteristics for CASSIOPEIA

Trial characteristic	CASSIOPEIA
Study design	Phase III, two-part, randomised, open-label, active-controlled, parallel group trial
Number and location of centres	111 European sites including: France (70), Belgium (13) and the Netherlands (28). No UK sites were included.
Study population	Adults (aged 18 to 65 years) with previously untreated multiple myeloma who are eligible for ASCT
Intervention	DBTd (N=543 randomised)
Comparator	BTd (N=542 randomised)
Primary outcome	Proportion of patients achieving stringent complete response (sCR) post-consolidation at (or within 30 days) of day 100 post-ASCT
Randomisation stratification factors	Study site affiliation, ISS disease stage and cytogenic risk status. The original version of the ISS was used (clarification response A7)
Number of patients completing consolidation therapy	DBTd: 461 (84.9% of ITT population) BTd: 437 (80.6% of ITT population)
Number of patients with outcome data for primary outcome (sCR)	DBTd: 533 (98.2% of ITT population) BTd: 527 (97.2% of ITT population)
Status	Ongoing
Latest available data	1st May 2019 (unplanned post-hoc interim analysis requested by EMA) (for data cuts and follow-up see section 3.2.4 below)
Pre-specified sub-groups	Sex, age, study site affiliation, ISS staging, cytogenetic risk, baseline renal and hepatic function, type of multiple myeloma and ECOG performance status score (see section 3.2.4 of this report for further information)
Source: CS section B.2.3.1; CS Table 4; CS Figure 7; CS Appendix L.3	

The ERG note that although daratumumab is licensed for use as induction and consolidation therapy in this indication, current UK clinical practice would typically include 4-6 cycles of induction therapy without consolidation post-ASCT. The ERG's clinical experts agreed that using less intensive induction therapy, compensated by having consolidation therapy, could be an appealing option. The ERG also note that maintenance therapy with daratumumab (Part 2 of CASSIOPEIA) is not currently licensed for use in this indication.

3.2.1.2 Patients' baseline characteristics

Key inclusion and exclusion criteria are described in CS Table 4. Of note, patients older than 65 years were excluded from the CASSIOPEIA trial whereas the ERG's clinical experts confirmed that in practice patients up to around 75 years old who are otherwise fit may be considered for ASCT. According to the experts, in the NHS around one third of patients aged 65-70 years and around 15% of those aged 70-75 years would be considered eligible for ASCT based on fitness. The experts did not consider the exclusion of older patients to be prognostically important since those older patients (>65 years) who are eligible for ASCT would be expected to be similar in terms of prognosis factors (i.e. general fitness) to those aged under 65 years.

The baseline characteristics of patients in CASSIOPEIA are summarised in CS Table 6 and in Table 8 below. In general, patient characteristics were well balanced between trial arms. No information is provided in the CS to assess representativeness of the trial patients in terms of socioeconomic status or ethnicity. The ERG's clinical experts advised that the reported patient characteristics are generally representative of patients seen in UK clinical practice.

An exception was noted with respect to the ISS which uses measures of albumin and β-2-microglobulin to classify multiple myeloma patients according to prognostic risk (see Table 4 above). Following the start of the trial, revised ISS criteria, referred to as R-ISS, were introduced. The R-ISS criteria are considered a better prognostic measure as these include additional risk factors: serum LDH and presence of selected chromosomal abnormalities indicative of cytogenetic risk (for details see section 2.2.1 above). Clinical expert advice suggests that in CASSIOPEIA the proportion of patients with the worst prognostic classification (revised ISS stage III) may be lower than that seen in practice (around 20-25% in practice are classified as ISS stage III). The trial may therefore under-represent patients with more advanced disease. Furthermore, an imbalance was observed between treatment

arms when the company applied a retrospective adjustment to convert the original ISS values to the R-ISS classes: 19.3% in the DBTd arm versus 27.0% in the BTd arm were in R-ISS class I (lowest prognostic risk), suggesting that patients in the DBTd arm have a poorer prognosis. We note, however, that any bias introduced would be conservative (i.e. treatment effect would be biased in favour of BTd) (clarification response A7).

The ERG's clinical experts advised that renal function is an important prognostic factor in patients with MM. Mean creatinine levels were similar between trial arms at baseline (Table 8) and in line with the normal reference ranges (typical reference range for men: 60-110 μmol/L and for women: 45-90 μmol/L).[39] The proportion of patients with moderate or greater renal insufficiency was low in both trial arms (Table 8). [40]However, the CS does not report the proportions of patients with different degrees of severity of renal impairment in CASSIOPEIA and so it is unclear how well this reflects clinical practice.

Table 8 Patients' baseline characteristics in the CASSIOPEIA trial

Characteristic	DBTd (n=543)	BTd (n=542)	ERG Comments
Sex (Female), n (%)	227 (41.8%)	223 (41.1%)	This is consistent with the disease population as MM is more frequent in men
Age, years, n (%)			
<50	83 (15.3%)	90 (16.6%)	This is consistent with the target population
≥50-65	460 (84.7%)	452 (83.4%)	
Mean (SD)	56.8 (6.93)	56.7 (7.03)	
Median (range)	59.0 (22 to 65)	58.0 (26 to 65)	
Baseline ECOG score, n (%)			
0	265 (48.8%)	257 (47.4%)	Overall, 90% of patients had an ECOG performance score of 0 or 1 indicating good functional status
1	225 (41.4%)	230 (42.4%)	
2	53 (9.8%)	55 (10.1%)	
Type of measurable disease[a], n (%)			
IgG	331 (61.0%)	314 (57.9%)	Percentages calculated by ERG. A clinical expert reported that 60-65% of patients have IgG paraprotein; the majority of the rest of this population will have IgA paraprotein or light chain only disease.
IgA	80 (14.7%)	99 (18.3%)	
Other[b]	13 (2.4%)	22 (4.1%)	
Urine only	70 (12.9%)	67 (12.4%)	
Serum FLC only	48 (8.8%)	40 (7.4%)	
Unknown	1 (0.2%)	0 (0.0%)	

Characteristic	DBTd (n=543)	BTd (n=542)	ERG Comments
Time since initial diagnosis to randomisation (months)[c]			
Mean (SD)	1.33 (2.984)	1.37 (2.184)	Clinical experts agreed that this is consistent with UK clinical practice.
Median (range)	0.92 (0.2 to 66.6)	0.95 (0.2 to 31.0)	
ISS staging, n (%)			
I	204 (37.6%)	228 (42.1%)	The ISS staging criteria were revised after the start of the trial
II	255 (47.0%)	233 (43.0%)	
III	84 (15.5%)	81 (14.9%)	
Revised ISS staging, n (%)			
N[c]	535	540	The ERG's clinical experts advised that the proportion with revised ISS stage III was lower than expected (around 20-25% in practice)
I	103 (19.3%)	146 (27.0%)	
II	383 (71.6%)	344 (63.7%)	
III	49 (9.2%)	50 (9.3%)	
Cytogenetic risk result			
N[e]	542	540	Most patients (84.5%) were classified as having a standard risk profile with a relative minority classified as high risk (15.5%)
High risk	82 (15.1%)	86 (15.9%)	
Standard risk	460 (84.9%)	454 (84.1%)	
Creatinine (µmol/L)			
Mean	76.2	78.9	The proportion of patients with renal insufficiency was small in both treatment arms but rates of other degrees of renal impairment are not reported
Median (range)	73.0 (5 to 213)	75.0 (6.5 to 235.1)	
Percentage with renal insufficiency (creatinine>177 µmol/L)	0.2%	0.4%	

Source: Excerpts from CS Tables 6, 7, 33 and 34.
FLC = free light chains; ECOG = Eastern Cooperative Oncology Group; ISS = International Staging System; SD = standard deviation.
[a] Includes patients without measurable disease in serum and urine.
[b] Includes IgD, IgM, IgE and biclonal
[c] Incorrect "time to initial diagnosis" data were entered into the database for 4 patients. These data errors did not affect the median reported in this analysis.
[d] Excludes patients who did not have baseline measurements of chromosomal abnormalities and/or lactate dehydrogenase (clarification response A8).
[e] The analysis population is patients with risk results available.

3.2.1.3 Ongoing studies

The CS provides a summary of three ongoing phase III trials evaluating daratumumab in newly diagnosed and untreated multiple myeloma patients and one single-arm phase II trial

which also included relapsed patients with multiple myeloma. Only one of these trials, Part 2 of the CASSEOPIA trial, compares DBTd with BTd (interim data are expected to be available at Technical Engagement). The other three trials (GRIFFIN, PERSEUS and LYRA) are not relevant to the NICE scope for this STA as they use lenalidomide or cyclophosphamide as an alternative to thalidomide.

ERG conclusions:

The CASSIOPEIA trial is an appropriate study design to assess the clinical effectiveness of DBTd compared to BTd. The ERG considers the study population to be representative of the target population with the possible exception that patients classified as having the poorest prognostic stage (revised ISS stage III) may be under-represented. It is also unclear how well the trial population represents the target population with respect to renal function.

3.2.2 Risk of bias assessment

The company report validity assessments for CASSIOPEIA, using the NICE recommended criteria, in CS Table 11. The ERG's assessment of the trials, following the same criteria, is shown in Appendix 10.2 of this report. The criteria were applied by one ERG reviewer and checked by a second reviewer with differences in judgement resolved through discussion.

In addition, the company have also assessed the risk of bias in CASSEOPIA using the revised version of the Cochrane Risk of Bias Tool (CS Appendix D.1.6).

The ERG note the following minor issues/uncertainties in relation to the risk of bias in CASSIOPEIA:

Baseline imbalance between treatment arms

As noted in Table 8 above, a lower proportion of patients in the intervention arm (DBTd) were classified as having the lowest prognostic risk (revised ISS stage I) indicating poorer prognosis in this group at baseline. Any arising bias is likely to be in favour of BTd.

Impact of lack of blinding

The ERG assume that the open-label design refers to lack of blinding post-treatment allocation and that treatment allocation itself would have been concealed using the centralised IWRS. The lack of blinding post-treatment allocation would be unlikely to bias results for the key efficacy and safety outcomes as these are measured objectively through

laboratory assessments but knowledge of treatment allocation could lead to bias for more subjective such as patient-reported measures of HRQoL The direction of potential bias is unclear but could favour the intervention arm if, for example, patients allocated to the DBTd are more optimistic and report higher levels of HRQoL. Alternatively, patients in the DBTd arm may have reduced perceived HRQoL due to extra, longer visits for IV infusion with this regimen. This is not an issue for the economic model however, which applies the same HRQoL values from CASSIOPEIA to both treatment arms (section 4.2.7 below).

Influence of missing data

The ERG agree that efficacy outcomes have been analysed using the ITT population, except that the company's risk of bias assessment does not adequately describe the handling of missing outcome data for the post-consolidation MRD status outcome. The company explained in clarification responses A3 and A4 that some of the missing data (which represented 15-20% of the ITT population) were carried forward from the post-induction assessment whilst missing MRD status in patients without a post-induction assessment were assumed to be MRD-positive (for details see section 3.2.4.5 below). The implication of this for risk of bias is unclear: approximately 10% of the ITT population had last observations carried forward (LOCF). However, the ERG's clinical experts expect that MRD negativity would increase over time (from induction to post-consolidation) so LOCF imputations are more likely to classify missing MRD status conservatively as MRD positive. Furthermore, carrying forward missing MRD status as MRD-positive is more likely to impact the more effective treatment arm (as more MRD-negative cases may develop later in this arm) and any bias would thus be in favour of BTd.

ERG conclusions:

Overall, the ERG considers the CASSEOPIA trial to be well-designed with generally low risk of bias, with only minor issues noted.

3.2.3 Outcomes assessment

All planned outcomes for the CASSIOPEIA trial are clearly defined and discussed in CS sections B.2.2, B.2.3.1 and CS Appendix L. An overview of all the outcomes, with their definitions and ERG comments, is provided in Appendix 10.3 of this report.

The only outcome in the NICE scope that is not a specified clinical outcome of the CASSIOPEIA trial is the "proportion of people undergoing high dose chemotherapy and autologous stem cell transplantation". This is not explicitly discussed in the CS; however, the

CS does report the proportion of patients in each arm who had completed all four cycles of induction treatment and both cycles of consolidation treatment, and the proportion of patients who had undergone ASCT (CS Figure 7) and these are inputs to the economic model (CS Table 80).

3.2.3.1 Primary outcome

The primary outcome for the CASSIOPEIA trial is stringent complete response (sCR). This outcome is useful to measure deeper level of response than the previous definition of complete response because it is more stringent, it correlates with PFS and OS,[41] and this is fully justified in CS section B.2.3.1. The ERG's clinical experts commented that sCR is not used in clinical practice to make decisions to alter treatment (it has not been validated robustly in many studies), but it can be useful to give an idea of prognostic status.

3.2.3.2 Secondary outcomes

The secondary outcomes from CASSIOPEIA that inform the economic model are overall survival (OS), progression free survival (PFS) and minimal residual disease (MRD) status (the primary CASSIOPEIA outcome, sCR, does not inform the model).

OS and PFS outcomes that inform the economic model were obtained from a landmark analysis of the impact of MRD status on survival, described in section 3.2.4.7 below. The designated time point of follow-up was the time of the post-consolidation response assessment, i.e. 100 days post-ASCT, which is the typical assessment timepoint in clinical practice, according to the ERG's clinical experts, for disease assessment following ASCT.

MRD status is assessed by identifying residual tumour cells in the bone marrow. MRD negative status is defined as undetectable clonal or sub-clonal cancerous cells for a given analysis threshold (10^{-4}, 10^{-5}, or 10^{-6}). This outcome is recommended for clinical trials by the IMWG, with 10^{-5} being the currently recommended detection threshold in their consensus paper.[42] MRD status is not currently used in clinical practice; although the ERG's clinical experts commented that it is an appealing method that is likely to be integrated into clinical practice in future.

Two recent review articles support the use of MRD status as an outcome in clinical trials.[43,44] However, it has not yet been identified which timepoint is best to evaluate MRD status, and from a one-time assessment it is not possible to tell how long MRD-negative status may be sustained.[45] The UK Myeloma Forum state in their evidence submission for this appraisal

that “depth of response (MRD, sCR and CR) are considered clinically meaningful outcomes as they correlate with long term survival.[41]

A recent meta-analysis, identifying six randomised studies of newly-diagnosed MM patients, found the odds ratio for MRD-negative versus MRD-positive response correlated with the hazard ratio for PFS and suggested that MRD status can be a surrogate for PFS.[46] In addition, the recent UK trials Myeloma IX and Myeloma XI have shown a correlation between MRD status and OS.[19-21] The ERG are satisfied that MRD status is an appropriate outcome for informing the company’s economic analysis.

Further secondary outcomes in CASSIOPEIA are: time to progression (TTP), post-consolidation complete response (CR) rate, post-consolidation MRD-negative rate, post-induction sCR rate, PFS2 (i.e. PFS on the subsequent line of therapy), post-induction overall response rate (ORR), the rate of very good partial response (VGPR) or better, duration of CR and sCR, time to response, and HRQoL. The use of these outcomes is supported by the IMWG guidelines[47] and the criteria are reproduced in CS Appendix Table 62 (see also Appendix 10.3 of this report). The outcomes most meaningful to clinical practice according to the ERG’s clinical experts are TTP, post-consolidation MRD-negative rate, the CR and VGPR or better response rates, and the duration of CR.

Response and disease progression outcomes were assessed centrally and by a validated computer algorithm. A sensitivity analysis was performed to compare centrally assessed outcomes with investigator assessed outcomes because a more stringent evaluation method can result in higher rates of VGPR compared to CR. Level of response relating to ongoing treatment of patients was determined by the investigator (CS Table 9).

3.2.3.3 Measurement and interpretation of MRD status

MRD was assessed at post-induction and post-consolidation using EuroFlow-based multiparametric flow cytometry (MFC) of bone marrow aspirates, according to criteria described in the IMWG consensus paper and outlined in CS Table 5.[42] Assessment using next-generation sequencing (NGS) was additionally performed for a subset of patients where there was enough bone marrow sample left to do so. The clinical study report (CSR) indicates that 371 patients from the DBTd arm and 364 patients from the BTd arm were assessed using NGS. The flow cytometry assessment method is most relevant to NHS practice; a clinical expert advising the ERG commented that this method is routinely used in the UK for chronic lymphocytic leukaemia which requires the same laboratory infrastructure.

The CS states that MRD was assessed according to IMWG criteria, which require that patients achieve CR.[42] However, assessments of MRD were conducted regardless of response (CS section B.2.6.1) with an assessment of MRD status in patients who achieved CR performed as a post-hoc analysis (CS Appendix L.4). We note that CS Table 5 appears contradictory: the caption reports that the definition of MRD was according to IMWG criteria but the MRD definition omits the IMWG requirement for CR. The company acknowledge that their approach, assessing MRD status regardless of response, differs from previous trials, but they do not provide a rationale for this nor explain why it is different to the IMWG recommended approach. As noted in section 3.6.3.1 below, analyses conducted by the company suggest that the way MRD is measured (i.e. whether the assessment is restricted to patients achieved CR) can influence the HRs that result from a comparison of the impact of MRD status (negative versus positive) on PFS.

3.2.3.4 Comparison of MRD with sCR

As shown in the clinical effectiveness results section of this report (see Table 9), comparisons between sCR and MRD negative status can produce counterintuitive results, with the proportion of MRD-negative patients considerably exceeding those with sCR. A determination of sCR requires clearance of serum paraprotein (among other criteria)[42] whereas the company's definition of MRD regardless of response requires clearance of malignant cells in the bone marrow, without dependence on serum paraprotein (CS Table 5). The company explain the discrepancy between the proportion of patients with sCR and those who are MRD-negative as being an effect of the lag in decay of paraprotein, which may remain circulating in serum after the malignant MM clone in bone marrow is eliminated (CS section B.2.6.1).

Due to the lag in clearance of serum paraprotein, at the post-consolidation assessment 289 patients were MRD-negative and not in CR or sCR at the first data cut, but this had decreased to 184 patients (64%) at the second data cut (CS section B.2.6.1). At the second data cut the percentage of patients who were MRD-negative post-consolidation and had a best response of CR or sCR was 66.7% in the DBTd arm and 59.8% in the BTd arm. The remaining MRD-negative patients in each group (33.3% and 40.2% respectively) had a best response of VGPR or worse (CS Table 16). The CS does not specify the percentage of MRD-negative patients whose best response was sCR.

Four of five experts asked by the ERG considered that MRD negativity is likely to be a better predictor of outcome than sCR whilst one expert was unsure. The experts noted that due to the lag in decay of paraprotein the comparability of these measures is not fixed over time,

and one expert commented that MRD negativity becomes increasingly important as a predictor as the post-transplant time interval increases.

3.2.3.5 HRQoL outcomes

HRQoL for patients in the CASSIOPEIA trial is assessed using the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 and the EQ-5D-5L. The EQ-5D-5L is generic utility instrument whereas EORTC QLQ-C30 is specific to oncological diseases. The EORTC QLQ-C30 has a myeloma-specific module, MY-20, but it is unclear whether it is used in the CASSIOPEIA trial because the company have not mentioned it in the submission or protocol. A 2018 report by Myeloma UK assessed patient reported outcome measures that are currently in use for myeloma patients.[48] Its literature review mentions other disease-specific and symptom-specific tools, for example several different Functional Assessment of Cancer Therapy (FACT) tools that each handle bone marrow transplant, fatigue, anaemia and myeloma specifically, and the QLQ-CIPN20 for chemotherapy-induced peripheral neuropathy. The same report also demonstrates patient preferences for MyPOS, a recent myeloma-specific tool validated in the UK in 2015[49] after the initiation of the CASSIOPEIA trial, and FACT-G (general). The tools used by the company are widely used in trials and clinical practice, but the CS does not report using any known tools specific to myeloma disease or its symptoms.

The patient reported outcomes were measured at screening, post-induction and post-consolidation (CS section B.2.6.4), that is at baseline and after each stage of treatment with chemotherapy, which is standard practice.

3.2.3.6 Adverse events

CASSIOPEIA included adverse events as an outcome, and all adverse events in the CS are taken from this trial. The adverse effects of treatment are discussed in section 3.3 below and were classified according to MedDRA 2.0 criteria.

ERG conclusions:

The response outcomes in CASSIOPEIA (VGPR, CR, sCR and MRD-negative status), are clinically meaningful in relation to depth of response which is a prognostic factor in length of remission. MRD status (a secondary outcome) is used in the company's economic model whereas the trial primary outcome (sCR) is not. The ERG agree that use of MRD status in the model is appropriate as it is a more sensitive response outcome than sCR. However, there is some inconsistency and

lack of clarity in how the company have reported their assessment of MRD which might influence hazard ratios that inform the economic model (see ERG Issue 2 in Table 26 below).

3.2.4 Approach to study statistics

The CASSIOPEIA trial is designed to test the hypothesis that there is no difference in the post-consolidation sCR rate between DBTd and BTd in patients with newly diagnosed untreated multiple myeloma who are eligible for ASCT.

The CS presents analyses for two data cuts: a prespecified primary analysis and a post-hoc unplanned analysis at the request of the EMA which provides updated data for a longer period of follow up (an additional 10.4 months) to support the marketing authorisation. The European Public Assessment Report clarifies that it was to obtain further information on the PFS data as the treatment effect for PFS, in patients with PR or better, for Part 1 of CASSIOPEIA is affected by the second randomisation for consolidation treatment in Part 2.[23] We refer to these as the primary data cut and the post hoc data cut (PA1 and PHA respectively in the CS):

- Primary data cut: Median follow-up 18.8 months
- Post hoc data cut: Median follow-up 29.2 months

CS Table 10 provides a summary of the outcomes assessed at each data cut.

The company clarified that the statistical analysis for the primary outcome (sCR) was stratified by the permuted block randomisation stratification factors: original version of ISS stage, study site affiliation, and cytogenetic risk (clarification response A7). The statistical analyses for survival outcomes (PFS and OS) were not stratified by ISS stage or any other stratification factor due to the expected low numbers of events. Results are not available using R-ISS as a stratification factor.

3.2.4.1 Analysis populations

The analysis populations in CASSIOPEIA (CS Table 8) are as follows:

- Analyses of primary efficacy outcomes are based on the intention to treat (ITT) population which includes all patients randomised in Part 1 of the trial.
- Analyses of adverse events are based on the safety population which includes all patients randomised in Part 1 who received at least one dose of study medication and contributed any safety data after the start of study treatment.

- Analyses of HRQoL are based on the subset of the randomised population who had a baseline assessment and at least one post-baseline assessment

3.2.4.2 Sample size and power calculation

The sample size and power calculation criteria (CS Table 9) were as follows:

- The planned sample size of 1080 patients (540 in each trial arm) randomised in Part 1 was reached. This sample size was considered sufficient to provide at least 85% power to detect an improvement in the primary outcome (sCR) from 25% to 35% with a 2-sided alpha of 0.05.
- This sample size was also sufficient to test the hypothesis in Part 2 (that daratumumab maintenance will decrease the risk of progression or death by 25%)[50] on the assumption that 75% of patients in Part 1 would be eligible for the second randomisation.

3.2.4.3 Statistical analysis approach for each outcome

The statistical approaches for comparing outcomes between the DBTd and DBd groups (CS Table 9) were as follows:

- For response outcomes, including the primary endpoint sCR and MRD status, the proportions of patients achieving response were compared between trial arms using a Cochran-Mantel Haenszel chi-squared test. Treatment effects were estimated using a Mantel-Haenszel odds ratio with corresponding 95% confidence intervals.
- Survival outcomes (PFS, time to disease progression, PFS2 and OS) were compared between treatment arms using Kaplan Meier survival curves and Cox proportional hazards models to calculate hazard ratios for median PFS and median OS. Hazard ratios for PFS only were adjusted for the second randomisation in Part 2 using inverse probability weighting (see section 3.2.4.6 below for further details).
- Adverse event rates were summarised descriptively.
- HRQoL outcomes (EQ-5D and EORTC QLQ-C30 measures) were compared between trial arms using a mixed model for repeated measures (MMRM) to account for missing data.

Controls for Type I error are reported in CS Table 9, which involved:

- Splitting the alpha level of 0.05 between analyses within each outcome (this is stated for OS and PFS but unclear whether also applied to response outcomes)
- For key secondary outcomes a hierarchical testing procedure was used

3.2.4.4 Subgroup analyses

The following pre-specified subgroups are reported (CS Appendix E) however, the study is not powered to detect differences between sub-groups:

- Sex, age, study site affiliation, ISS staging, cytogenetic risk, baseline renal and hepatic function, type of multiple myeloma and ECOC performance score

According to forest plots reported in CS Appendix E the effects of DBTd and BTd did not differ across subgroups. Each of these subgroups either favoured DBTd over BTd or showed no statistically significant difference between them, for sCR (CS Appendix E.1), MRD negativity rate (CS Appendix E.2), and PFS (CS Appendix E3). (NB sCR and PFS subgroup analyses are reported for both data cuts; MRD negative status subgroup analyses are reported only for the primary data cut.)

3.2.4.5 Sensitivity analyses and methods for handling missing data

Sensitivity analyses

The following sensitivity analyses are reported in CS sections B.2.6.1, B.2.6.2, and CS Appendix L:

- PFS was adjusted for the second randomisation using inverse probability weighting (this is discussed in detail in section 3.2.4.6 below)
- MRD negativity response was calculated using methods with different test sensitivity thresholds (MFC and NGS)
- Additional sensitivity analyses for PFS were performed:
    - Using investigator assessments (instead of those derived by the validated computer algorithm)
    - Censored for subsequent therapy
    - Multivariate analysis (including transplant in the model)
    - Multivariate analysis including interaction of transplant and treatment

Results for MRD-negativity and PFS were robust to these sensitivity analyses (CS section B.2.6.1 and CS Table 18).

Approaches for handling missing data

If the patient did not have any post baseline disease assessments for the primary outcome (sCR) they were classified as ‘not evaluable’ (CS Appendix L.3). This applied to 10 patients (1.8%) and 15 patients (2.8%) in the in the DBTd and BTd arms respectively. The ERG

considers this unlikely to introduce significant bias. From CS Figure 7 and the company response to clarification question A3, the ERG assume (as the company's explanation is not fully explicit) that missing data from post-consolidation assessments were handled as follows for the proportion MRD negative outcome (see section 3.2.3 for definition):

- DBTd arm: 84 patients (15% of ITT population) had missing data including 46 patients with data carried forward from cycle 4, day 28 assessments and [redacted] patients imputed as MRD-positive (i.e. non-responders)
- BTd arm :106 (20% of ITT population) had missing data including 63 patients with data carried forward from cycle 4, day 28 assessments and [redacted] patients imputed as MRD-positive

The company did not perform any sensitivity analyses to explore alternative imputation approaches. Imputation using a worst-case scenario (assuming missing data as MRD positive) is a conservative approach but the use of LOCF approaches can introduce bias. The ERG would therefore have preferred a multiple imputation approach. However, as noted above (section 3.2.2) we would expect the impact of LOCF to be conservative (classifying MRD-negative patients as MRD-positive where post-consolidation data are imputed).

For HRQoL outcomes, the company used a mixed model for repeated measures (MMRM) approach. This method includes observed data only but is considered an appropriate method to handling missing data when the data are 'missing-at-random'. In response to clarification question A4, the company have provided further data on the pattern of missing data for HRQoL outcomes and results of sensitivity analyses accounting for missing data based on a pattern mixture model. The analysis accounting for missing data had no impact on EQ-5D results (section 3.2.8.1 below) and very little impact on the EORTC QLQ-C30 results (section 3.2.8.2 below).

3.2.4.6 Adjustment for effects of second randomisation to maintenance therapy

The company have used inverse probability weighting (IPW) to adjust for the effect of the second randomisation (CASSIOPEIA Part 2) on PFS. The methods are described in CS Appendix L and CS section B.2.6.2. The CS does not provide a justification for the choice of method. In response to clarification question A6 the company report that the IPW analysis for PFS was conducted by a sequestered group independent from the Janssen study team (as Part 2 of CASSIOPEIA is currently blinded). Due to the anticipated low number of events and immaturity of survival data, a similar statistical analysis was not performed for OS.

The ERG note that alternative weighting methods are suggested in the reference quoted in the CS.[51] The ERG therefore requested a sensitivity analysis for the IPW adjustment using the weighted Kaplan-Meier method with time-dependent weights as opposed to fixed weights. The company clarified that, as Part 2 of the study remains blinded, Janssen does not have access to the patient-level data necessary to perform this sensitivity analysis. Importantly, the ERG also note that the IPW methods used to calculate the adjusted hazard ratio[52] may not be appropriate when the assumption of proportional hazards is violated and suggests that alternative methods should be considered.[53] The company presented a log-log plot for PFS (Figure 3 in clarification response A11) which suggests that the assumption of proportional hazards did not hold, with evidence of varying treatment effects at different treatment stages and the strongest effects observed post-induction therapy.

As shown in the results section below (section 3.2.7.1), the IPW-adjusted and unadjusted hazard ratios for PFS are identical. The CS states that the similarity of IPW-adjusted and unadjusted results is expected because a high number of patients from CASSIOPEIA Part 1 were re-randomised in both treatment arms (more than 80% of patients had undergone re-randomisation) and duration of the maintenance therapy to date is relatively short. The median follow-up period for the Part 1 analysis was 18.8 months. However, the company have not reported the follow-up time in Part 2, which would help to more fully evaluate the face validity of the results. The ERG's clinical experts considered that short-term maintenance therapy (e.g. 6-9 months) would be unlikely to have a significant impact as most patients in both treatment arms would be expected remain in remission over this time frame since median PFS is at least 36 months with standard treatment and ASCT. One expert suggested that maintenance therapy could prevent a small proportion of partial responders progressing to frank early relapse.

The ERG are unable to validate the IPW analysis and the company are unable to conduct sensitivity analyses because Part 2 of CASSIOPEIA is currently blinded and the company do not have access to the IPD (clarification response A6).

3.2.4.7 Landmark analysis of survival by MRD status

As reported in CS section B.2.6.3, the company conducted an exploratory landmark analysis to explore the effect of MRD status (positive or negative) on PFS and OS outcomes. Survival was assessed from the time of the post-consolidation MRD response assessment (the landmark point) which differed between patients with respect to the time from randomisation.

Kaplan Meier plots and Cox proportional hazards models were used to explore whether the impact of MRD status on survival was the same for both trial arms. The analysis was not adjusted for the impact of the second randomization to maintenance therapy.

Log-cumulative hazard plots presented CS Appendix N and in Figures 4 & 5 of clarification response A11 suggest that the assumption of proportional hazards is supported for the comparison of PFS and OS in MRD positive and MRD negative patients.

Although this landmark analysis is described in the CS as exploratory, it is important because it informs the response-based modelling approach in the company's economic analysis (CS section B.3.3.1) (see section 4.2.2 below).

ERG conclusions:

Overall the approach to statistics in CASSIOPEIA is appropriate, although there are uncertainties around the IPW adjustment for the effect of maintenance therapy on PFS due to possible violation of the proportional hazards assumption, lack of exploration of alternative adjustment methods, and immaturity and confidentiality of the maintenance therapy data (ERG Issue 3 – see Table 26 below). The company's landmark analysis of the effect of MRD status on survival outcomes appears to have been conducted appropriately, although it was not adjusted for the (currently immature) effect of maintenance therapy. Some outcomes have missing data but these were generally handled appropriately and appear unlikely to impact on overall conclusions.

3.2.5 Clinical effectiveness results

3.2.6 Response outcomes

Response outcomes are shown in Table 9 for the primary data cut and Table 10 for the post hoc data cut, with time to response shown in Table 11. Except where stated otherwise the denominator for % response rates is the number randomised. Note that p-values are reported in the CS but have not been reproduced here as 95% confidence intervals are available.

At the first data cut the proportion of patients who achieved CR or better increased at each assessment timepoint (post-induction therapy, post-ASCT and post-consolidation therapy). Response outcomes favoured DBTd over BTd, with the difference being most pronounced

post-consolidation. As noted above (section 3.2.3.3), the proportion MRD-negative post-consolidation was determined regardless of response and exceeds the proportion with CR or better, favouring DBTd over BTd (Table 9). Less detailed results are reported in the CS for the second data cut, and do not include MRD status (Table 10). Median time to response was similar for the DBTd and BTd groups (Table 11).

Table 9 Response outcomes at primary data cut (median follow up 18.8 months)

Outcome & assessment time		**DBTd N=543**	**BTd N=542**	**Difference, OR (95% CI)**
sCR	Post-induction	7.4%	6.5%	0.9% OR=1.16 (0.72 to 1.86)
	Post-ASCT	13.4%	9.4%	4.0% OR=1.5 (1.02 to 2.19)
	Post-consolidation	28.9%	20.3%	8.6% OR=1.60 (1.21 to 2.12)
CR	Post-induction	7.0%	2.4%	4.6%
	Post-ASCT	9.2%	5.2%	4.0%
	Post-consolidation	9.9%	5.7%	4.2%
CR or better	Post-induction	14.4%	8.9%	5.5% OR=1.73 (1.18 to 2.53)
	Post-ASCT	22.7%	14.6%	8.1% OR=1.72 (1.26 to 2.35)
	Post-consolidation	38.9%	26.0%	12.9% OR=1.82 (1.40 to 2.36)
VGPR	Post-induction	50.5%	47.2%	3.3%
	Post-ASCT	54.1%	52.8%	1.3%
	Post-consolidation	44.6%	52.0%	-7.4%
VGPR or better	Post-induction	64.8%	56.1%	8.7% OR=1.44 (1.13 to 1.84)
	Post-ASCT	76.8%	67.3%	9.5% OR=1.6 (1.23 to 2.09)
	Post-consolidation	83.4%	78.0%	5.4% OR=1.41 (1.04 to 1.92)
PR	Post-induction	27.8%	33.8%	-5.2%
	Post-ASCT	15.5%	23.1%	-7.6%
	Post-consolidation	9.2%	11.8%	-2.6%
ORR	Post-induction	92.6%	89.9%	2.7% OR=1.41 (0.92 to 2.17)
	Post-ASCT	92.3%	90.4%	1.9% OR=1.26 (0.82 to 1.93)
	Post-consolidation	92.6%	89.9%	2.7% OR=1.41 (0.92 to 2.16)
MRD negative (10^{-5}) [a]	Post-consolidation	63.7%	43.5%	20.2% OR=2.27 (1.78 to 2.90)
MRD negative (10^{-5}) [b]	Post-consolidation	56.6% (n=371) [d]	36.8% (n=364) [d]	19.8% OR=2.26 (1.68 to 3.05)
MRD negative (10^{-6}) [c]	Post-consolidation	39.1% (n=371) [d]	22.8% (n=364) [d]	16.3% OR=2.18 (1.58 to 3.01)

Sources: CS Tables 12-13; CS Figures 8-10; CS section B.2.6.1
[a] 10^{-5} threshold, standard Euroflow assay, MRD-negative regardless of response

[b] Exploratory analysis, 10^{-5} threshold, next-generation sequencing (NGS), MRD-negative regardless of response
[c] Exploratory analysis, 10^{-6} threshold, NGS, MRD-negative regardless of response
[d] Subset of patients with sufficient bone marrow left for testing (CSR pages 137-138)

Table 10 Response outcomes at post-hoc data cut (median follow up 29.2 months)

Outcome		DBTd N=543	BTd N=542	Difference, OR (95% CI)
Best response over time	sCR	54.3%	42.1%	12.2% OR=1.64 (1.29 to 2.09)
	sCR or better	62.1%	47.6%	14.5% OR=1.80 (1.41 to 2.30)
	VGPR or better	85.5%	84.9%	0.6% OR=1.05 (0.75 to 1.47)
Source: CS Table 13				

Table 11 Time to response at primary data cut (median follow up 18.8 months)

Median (range) time to response (months) among response-evaluable patients [a]			
Outcome	**DBTd (n=513)**	**BTd (n=510)**	**Difference**
Time to first response [b]	1.02 (0.7 to 10.0) n=513	1.05 (0.8 to 10.1) n=510	-0.03
Time to VGPR or better	2.14 (0.9 to 10.6) n=454	2.83 (0.9 to 10.3) n=434	-0.69
Time to CR or better	7.23 (1.9 to 10.6) n=211	7.38 (1.9 to 11.4) n=144	-0.15
Time to sCR	7.98 (3.5 to 11.2) n=157	7.98 (3.6 to 10.8) n=113	0

Source: CS Table 14
[a] response-evaluable patients were those with PR or better by 100 days post-ASCT (or imputed date if missing) + 30 days
[b] PR or better

At a median follow up of 18.8 months the median duration of response had not been reached for either of the CASSIOPEIA study arms (CS Table 15).

3.2.7 Survival outcomes

3.2.7.1 Progression-free survival

Median PFS was not reached in either arm of CASSIOPEIA. PFS favoured the DBTd group over the BTd group, with a 53.0% reduction in the risk of disease progression or death at the primary data cut (median follow up 18.8 months) and a 50.5% reduction at the post hoc data cut (median follow up 29.2 months) (Table 12). The corresponding Kaplan-Meier plots are shown in CS Figures 11 and 12 (not reproduced here). IPW adjustment for the subsequent maintenance therapy had no impact (primary data cut) or a very small impact (post hoc data cut) on the hazard ratio (Table 12). The ERG's critique of this analysis is discussed in section 3.2.4.6 above.

PFS rates at 6, 12, 18 and 24 months are reported in CS Table 19. According to the 95% confidence intervals, at both data cuts the PFS rate was not significantly different between DBTd and DBd at 6 months but was significantly higher in the DBTd arm at 12, 18, and 24 months. PFS at 24 months at the latest (post hoc) data cut was 88.4% (95% CI 85.3 to 90.9) in the DBTd group and 77.4% (73.4 to 80.8) in the BTd group (CS Table 19).

Table 12 PFS and PFS2 at the primary and post-hoc analysis data cuts

Outcome	IPW-adjusted for maintenance therapy	DBTd N=543	BTd N=542	HR (95% CI)
PFS events, n (%), primary data cut	No	45 (8.3%)	91 (16.8%)	0.47 (0.33 to 0.67)
	Yes	Ditto	Ditto	0.47 (0.33 to 0.67)
PFS, events, n (%), post-hoc data cut	No	83 (15.3%)	151 (27.9%)	0.495 (0.38 to 0.65)
	Yes	Ditto	Ditto	0.50 (0.34 to 0.75)
PFS2 events, n (%), primary data cut	No	18 (3.3%)	37 (6.8%)	0.46 (0.26 to 0.82)
PFS2 events, n (%), post hoc data cut	No	33 (6.1%)	60 (11.1%)	0.51 (0.33 to 0.78)
Source: CS section B.2.6.2; CS Tables 17, 19, 20				

The company conducted sensitivity analyses for investigator (rather than central) assessment of PFS; PFS censored for subsequent therapy; and including the effect of transplant in the multivariate analysis (CS Table 18). These gave very similar HRs (ranging from 0.48 to 0.49).

3.2.7.2 Progression-free survival on next line of therapy

PFS2 data are immature, with fewer than 12% of patients experiencing an event at the post hoc data cut follow up of 29.2 months (Table 13). The corresponding Kaplan-Meier plot for PFS2 is shown in CS Figure 14. The reduction in the risk of disease progression or death on the next line of therapy in the DBTd arm was 54% at the primary data cut and 49% at the post hoc data cut.

3.2.7.3 Time to progression

Time to progression data are immature, with median TTP not reached at either data cut (CS Table 22). Rates of progression were lower in the DBTd arm, with an overall reduction of the

risk of progression (including death due to progression) of 48.0% at both data cuts (Table 13).

Table 13 Progression rates at the primary and post-hoc analysis data cuts

Outcome [a]	DBTd N=543	BTd N=542	HR (95% CI)
Events, n (%), primary data cut	42 (7.7%)	76 (14.0%)	0.52 (0.36 to 0.76)
Events, n (%), post hoc data cut	79 (14.5%)	136 (25.1%)	0.52 (0.39 to 0.68)
Source: CS Table 22 [a] includes progression and death due to progression			

3.2.7.4 Overall survival

OS data are immature, with median OS not reached at the latest data cut (CS Table 24). A larger proportion of deaths occurred in the BTd arm at both data cuts, with the reduction in risk of death on DBTd therapy being 57.0% at the primary data cut and 48.0% at the post hoc data cut (Table 14). The corresponding Kaplan-Meier plots are shown in CS Figures 15 and 16.

Table 14 Mortality rates at the primary and post-hoc analysis data cuts

Outcome	DBTd N=543	BTd N=542	HR (95% CI)
Events, n (%), primary data cut	14 (2.6%)	32 (5.9%)	0.43 (0.23 to 0.80)
Events, n (%), post hoc data cut	26 (4.8%)	48 (8.9%)	0.52 (0.33 to 0.85)
Source: CS Tables 23 & 24			

According to the 95% confidence intervals, the OS rate was not significantly different between the DBTd and BTd groups at 6, 12, 18 or 24 months at either data cut (CS Tables 23 and 24). OS at 24 months at the latest (post hoc) data cut was 96.6% (95% CI 94.7 to 97.9) in the DBTd group and 93.2% (90.6 to 95.0) in the BTd group (CS Table 24).

3.2.7.5 Impact of MRD status on survival (landmark analysis)

Results of the Cox proportional hazards models show that DBTd appears to improve OS and PFS in both the MRD-positive and MRD- groups, but only the hazard ratio for PFS in the MRD-negative group is statistically significant (the 95% confidence interval excludes 1.0)

(Table 15). These results should be considered illustrative, since the analysis was exploratory and not powered statistically for this comparison.

Table 15 Landmark analysis hazard ratios for OS and PFS by MRD status

Subgroup	DBTd versus BTd	
	HR for OS (95% CI)	HR for PFS (95% CI)
MRD-negative	[redacted]	[redacted]
MRD-positive	[redacted]	[redacted]
Source: Reproduction of CS Table 25		

The Kaplan-Meier OS and PFS data that inform the company's economic model are shown in CS Figures 17 and 18, reproduced in Figure 3 below. Patients achieving post-consolidation MRD negativity (the dashed lines in Figure 3) showed improved OS and PFS compared to those who were MRD-positive. The company claim that patients treated with DBTd who are MRD-negative have OS which resembles that of the general population when matched on age and gender, suggestive of long-term disease control and a possible functional cure for some patients. The ERG's clinical experts did not agree that the company have sufficient data to make such a claim.

Figure 3 Landmark analysis of OS and PFS by treatment arm and MRD status at post-consolidation assessment (post hoc data cut)

CS Figure 19 presents additional evidence for the impact of MRD status on PFS for newly-diagnosed treatment-ineligible MM patients [redacted] who were treated with daratumumab and a comparator therapy. This is described as an exploratory analysis and these populations are outside the scope of the current technology appraisal but median follow-up duration was longer than the CASSIOPEIA trial. The Kaplan-Meier curves in CS Figure 19 show that [redacted]. The CS does not explain why a similar analysis is not reported for OS.

3.2.8 Health-related quality of life

The CS reports results of the two HRQoL measures that were employed in the CASSIOPEIA trial: the EORTC QLQ-C30 and the EQ-5D-5L. As noted in CS section B.3.4.1 only the EQ-5D data inform the economic model.

3.2.8.1 EQ-5D scores

EQ-5D-5L index and visual analogue scale (VAS) data were collected from patients in CASSIOPEIA at 3 timepoints: at baseline, post-induction (cycle 4 day 28), and post-consolidation. The mean EQ-5D index values were similar for the DBTd and BTd arms at each timepoint (CS Appendix Table 68). The company pooled the mean EQ-5D-5L values across the arms for each of these timepoints for use in the economic model, as explained in CS section B.3.4.1 and in section 4.2.7 below.

The CS also reports the least squares mean change in EQ-5D index and VAS scores from baseline to post-induction and from baseline to post-consolidation, assessed using a mixed effects model with repeated measures, stratified by study site, ISS score and cytogenetic risk (these analyses do not inform the economic model). Results are presented in CS Table 28 (and CS Appendix Table 69), reproduced in Table 16 below.

Table 16 Change from baseline in EQ-5D-5L index and VAS scores

EQ-5D score and timepoint		LS means of change from baseline (95% CI) [a]		Difference Mean (95% CI)
		DBTd	BTd	
	Post-induction	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)	0.0 (-0.02, 0.02)

Index score	**Post-consolidation**	0.17 (0.14, 0.19)	0.16 (0.13, 0.18)	0.01 (-0.01, 0.04)
VAS score	**Post-induction**	2.7 (0.5, 4.8)	2.2 (0.1, 4.4)	0.4 (-1.8, 2.7)
	Post-consolidation	8.6 (6.5, 10.8)	7.7 (5.5, 9.9)	0.9 (-1.4, 3.2)
Source: reproduction of CS Table 28 LS: least squares [a] Mixed effects model with repeated measures, with independent variables baseline, visit, treatment, visit by treatment interaction, and randomisation stratification factors (site affiliation, ISS stage and cytogenetic risk) as fixed effects and individual subject as a random effect				

The EQ-5D scores improved from baseline to post-induction and from baseline to post-consolidation similarly in the DBTd and BTd arms, with no statistically significant difference between the treatment arms at either timepoint. These results suggest that the addition of daratumumab to BTd does not influence the improvement in patients' HRQoL that results from induction and post-consolidation therapy with BTd.

The company explained in clarification response A4 why some EQ-5D data were missing, and they ran a pattern mixture model to iteratively impute the missing data. Results of the pattern mixture modelling approach (Table 1.7.2 in clarification response A4) were very similar to those reported above in Table 16, indicating that the results are robust to the missing data.

3.2.8.2 EORTC QLQ-C30

The CS reports EORTC QLQ-C30 scores for the Global Health Scale (GHS), three symptom scales (fatigue, nausea and vomiting, and pain), and five functional scales (cognitive, emotional, physical, role, and social functioning). Patients' mean baseline EORTC QLQ-C30 scores in the CASSIOPEIA trial were generally similar between the DBTd and BTD arms for each of these scales (CS table 27).

A summary of the EORTC QLQ-C30 results is provided in Table 17. Overall, patients' HRQoL improved following induction and consolidation therapy relative to baseline, except for decreases in cognitive functioning and social functioning. For all scales except emotional functioning HRQoL did not differ significantly between the DBTd and BTd groups. Emotional functioning scores differed significantly at the post-consolidation timepoint and favoured DBTd over BTd.

Table 17 Changes from baseline in EORTC QLQ-C30 scores

EORTC QLQ-C30 scale	Summary	Where reported

Global Health Score	Improvement from baseline to post-induction & post- consolidation statistically significant at both time points and clinically significant (>8 points) at post-consolidation, but differences between DBTd and BTd not statistically or clinically significant	CS section B.2.6.4; CS Figure 20; Table 1.7.4 in clarification response A4
Physical functioning	Improvement from baseline to post-induction & post-consolidation but only the latter statistically significant. Differences between DBTd and BTd groups not statistically significant.	CS Appendix Table 70
Role functioning	Improvement from baseline to post-induction (statistically significant for BTd) & baseline to post-consolidation (statistically significant for both groups), but differences between DBTd and BTd groups not statistically significant.	CS Appendix Table 71
Emotional functioning	Improvement from baseline to post-induction & post- consolidation, statistically favouring DBTd at post-consolidation (groups not statistically significantly different at post-induction)	CS section B.2.6.4; CS Figure 22; CS Appendix Table 72
Cognitive functioning	Clinically non-significant (<10 point) decrease from baseline to post-induction & post-consolidation, statistically favouring DBTd (smaller decrease) at post-consolidation, but groups not statistically significantly different at post-induction	CS section B.2.6.4; CS Figure 23; CS Appendix Table 73
Social functioning	Decrease from baseline to post-induction but improvement from baseline to post-consolidation; larger increase in DBTd group but differences between groups not statistically significant	CS Appendix Table 74
Fatigue	Increase from baseline to post-induction but decrease from baseline to post-consolidation; differences between DBTd and BTd not statistically significant	CS Appendix Table 75
Pain	Decrease from baseline which was clinically significant (>15.7 points) for DBTd group post-induction and for both groups post-consolidation. Difference between groups statistically significant at post-consolidation, favouring DBTd (larger decrease), but not statistically significant post-induction.	CS section B.2.6.4; CS Figure 21; CS Appendix Table 76
Nausea & vomiting	No difference from baseline except for slight decrease in DBTd group at post-consolidation;	CS Appendix Table 77

	differences between groups not statistically significant	

The company's pattern mixture model to investigate the impact of missing data (see section 3.2.8.1) gave almost identical results to those reported above in Table 17, except that the only statistically significant difference between DBTd and BTd, for emotional functioning, became non-significant.

3.3 Safety outcomes

All safety outcomes reported in the CS are taken from the CASSIOPEIA trial primary data cut (median follow up 18.8 months) (CS section B.2.10 and CS Appendix F). Adverse events reported in the CS are classified according to the Medical Dictionary for Regulatory Agencies (MedDRA) version 20.0.

3.3.1 Treatment exposure

Median duration of treatment exposure was 8.9 months for the DBTd group and 8.7 months for the BTd group, and both groups received a median of 6 (range 1 to 6) cycles of treatment. For bortezomib, thalidomide and dexamethasone the median dose intensities were similar between the DBTd and BTd groups (CS Table 38).

3.3.2 Treatment-emergent adverse events

Nearly all patients in both treatment groups had at least one TEAE (Table 18). Overall TEAE frequencies were similar in the two treatment groups except for a slightly higher rate of Grade3/4 TEAEs in the DBTd group (mainly driven by haematological events including neutropenia and lymphopenia) and a higher number of TEAE-emergent deaths in the BTd group.

Table 18: Summary of treatment-emergent adverse events in the CASSIOPEIA trial

Event [a]	DBTd (n=536)	BTd (n=538)
Any TEAE, n (%)	535 (99.8%)	536 (99.6%)
Grade 3/4 TEAE, n (%)	432 (80.6%)	408 (75.8%)
Serious TEAE, n (%)	251 (46.8%)	255 (47.4%)
TEAE leading to discontinuation, n (%)	40 (7.5%)	45 (8.4%)
TEAEs leading to death, n (%)	1 (0.2%)	9 (1.7%)
Source: Reproduction of CS Table 39		

[a] TEAEs during induction, ASCT, or consolidation Treatment Phase; incidence reflects the number of patients experiencing at least one TEAE associated with at least one of the study treatments. Note: Adverse events emerging during ASCT phase related to the planned procedures were not reported.

The most common specific TEAEs, of any grade, with a frequency of ≥20% in either group (DBTd versus BTd) were peripheral sensory neuropathy (58.6% vs 63.2%), constipation (50.7% vs 48.7%), nausea (30.2% vs 24.2%), asthenia (31.9% vs 28.8%), peripheral oedema (30.2% vs 27.5%), neutropenia (29.3% vs 16.5%), pyrexia (26.1% vs 21.2%), paraesthesia (22.0% vs 20.1%), and thrombocytopenia (20.3% vs 13.6%) (CS Table 40).

Frequent TEAEs (≥10% in either group) that differed in frequency by at least 5 percentage points between the DBTd and BTd groups are listed in Table 19. Haematological TEAEs, bronchitis, nausea, vomiting and cough were more frequent in the DBTd group than the BTd group.

Table 19: Most frequent TEAEs of any grade (≥10% in either group) that differ by ≥5% between study groups

Event	DBTd (n=536)	BTd (n=538)
Neutropenia	157 (29.3%)	89 (16.5%)
Thrombocytopenia	109 (20.3%)	73 (13.6%)
Lymphopenia	99 (18.5%)	67 (12.5%)
Bronchitis	102 (19.0%)	66 (12.3%)
Nausea	162 (30.2%)	130 (24.2%)
Vomiting	87 (16.2%)	52 (9.7%)
Cough [a]	90 (16.8%)	49 (9.1%)
Source: Excerpt from CS Table 40 Text in CS section B.2.10 gives different % values for cough: DBTd 17.2%; BTd 10.4%		

The most frequent Grade 3 or 4 TEAEs are listed in Table 20. Neutropenia, thrombocytopenia and lymphopenia were more frequent in the DBTd group than the BTd group. The remaining Grade 3 or 4 TEAEs differed in frequency by <5 percentage points between the groups.

Table 20: Most frequent Grade 3 or 4 TEAEs (≥5% in either group)

Event	DBTd (n=536)	BTd (n=538)
Neutropenia	148 (27.6%)	79 (14.7%)
Lymphopenia	91 (17.0%)	52 (9.7%)
Stomatitis	68 (12.7%)	88 (16.4%)
Thrombocytopenia	59 (11.0%)	40 (7.4%)

Peripheral sensory neuropathy	47 (8.8%)	46 (8.6%)
Febrile neutropenia	36 (6.7%)	28 (5.2%)
Source: Excerpt from CS Table 40		

The increased rate of neutropenia in patients receiving daratumumab was not associated with any increased risk of neutropenic fever, as patients in the both treatment groups reported comparable levels of febrile neutropenia.

The CS reports that approximately half the patients in each group had serious TEAEs (DBTd 46.8%, BTd 47.4%). The most commonly reported specific serious TEAEs (≥2% in either group) (DBTd vs BTd) were pyrexia (2.8% vs 4.3%), neutropenia (3.9% vs 1.5%), pulmonary embolism (1.5 vs 3.7%), pneumonia (3.5%, 1.7%), febrile neutropenia (2.2% vs 2.8%), peripheral sensory neuropathy (2.1% vs 2.8%), thrombocytopenia (2.2% vs 0.7%), lung disorder (2.1% vs 1.1%), sepsis (1.3% vs 2.0%), and febrile bone marrow aplasia (1.3% vs 2.0%) (CS Table 41).

3.3.3 Infusion-related reactions

The CS reports that at a median follow-up of 18.8 months 35.4% of patients in the DBTd arm (N=536) experienced an infusion-related reaction (IRR) of any grade. IRRs were mostly Grade 1 or Grade 2 events (together 31.8%), with Grade 3 events experienced by 3.2% of patients, Grade 4 events experienced by 0.4% of patients, and no IRR-related fatalities (CS Table 42).

IRRs were more frequent at the first infusion (26.9%) and at the first infusion post-ASCT (i.e. the first infusion of the 5th DBTd cycle) (10.7%) with the IRR rate at the second infusion of the first DBTd cycle being only 1.9. Only 4.7% of patients experienced IRRs in more than one infusion (CS Table 42). The CS states that, overall, IRRs were manageable, with only 0.6% leading to discontinuation, but details of how IRRs were managed are not reported in the CS or Appendices.

The most common IRRs were general disorders and administration site conditions which included chills (5.6%) and pyrexia (3.7%) (CS table 42).

As reported in CS section B.2.10, a licence extension for a subcutaneous (SC) formulation of daratumumab was received in June 2020. Results from the non-inferiority phase III study COLUMBA[54,55] demonstrated that the rate of IRRs was significantly reduced with SC versus

IV daratumumab (12.7% vs 34.5%; odds ratio, 0.28; 95% CI, 0.18-0.44). The company therefore argue that IRRs associated with administering DBTd will be substantially reduced following the availability of daratumumab as a SC injection. We note that the population in COLUMBA was patients with relapsed or refractory MM which is outside the scope of the current appraisal, and included older patients than in CASSIOPEIA (23% in the IV group and 18% in the SC group were aged ≥75 years). However, the ERG's clinical experts agreed that these population differences between COLUMBA and CASSIOPEIA would be unlikely to influence the rate of IRRs.

3.4 Critique of studies included in the indirect treatment comparisons

3.4.1 Rationale for indirect comparisons

The comparisons of interest in untreated MM when subjects are suitable for ASCT are DBTd, BTd, BCd, and Bd. The company state that BTd is standard of care whilst other comparators (BCd, Bd) are used only in a minority of patients but have been included in indirect comparisons for completeness. The ERG's clinical experts mostly agreed that BCd and BTd are used infrequently (see section 2.2.3 above). The CASSOPEIA trial directly compares DBTd vs BTd, hence an indirect treatment comparison was necessary for the remaining comparisons of DBTd against BCd and Bd. Hazard ratios from the indirect comparisons do not directly inform the economic model but are used to support an assumption that BTd is equivalent BCd and that both BTd and BCd are superior to Bd.

3.4.2 Identification, selection and feasibility assessment of studies for ITC

The company's systematic review of clinical effectiveness studies (section 3.1 above) identified two studies which could form a network with CASSIOPEIA. Study IFM 2013-04 compared BTd versus BCd[56] and study IFM 2007-02compared BTd versus Bd.[57] However, both were excluded from a network meta-analysis (NMA) due to insufficient outcome reporting. IFM 2007-02 reported only PFS whilst IFM 2013-04 reported neither OS nor PFS. Neither trial reported MRD negativity.

Hence, the company broadened their inclusion criteria (based on the SLR reported in section 3.1 above) to identify comparator studies which could potentially contribute to an unanchored MAIC. Two studies were selected, both of which reported OS and PFS: study IFM 2005-01 compared Bd against VAd (vincristine + doxorubicin + dexamethasone) (N=240)[58] whilst study GMMG-MM5 compared BCd against PAd (doxorubicin + dexamethasone) (N=251).[59] A further 11 studies were excluded due to inappropriate

comparators, lack of reporting OS and/or PFS, reporting outcomes only by cytogenetic markers (presence/absence of del 17p and t[4; 14]), or being single-arm phase II studies (CS Table 29).

The ERG disagree that excluding single-arm studies is appropriate for a MAIC analysis. However, of the 11 studies excluded from MAIC analysis listed in CS Table 29, only one, by Kumar et al.[60] appears potentially relevant for the MAIC. The Kumar study[60] reports KM survival curves for PFS and OS but has a small sample size (N=33) compared to N=251 in the GMMG-MM5 study, so the latter is preferable for inclusion in the MAIC analysis.

3.4.3 Clinical heterogeneity assessment

In an unanchored MAIC (i.e. one with no common comparator), Technical Support Document (TSD) 18 recommends that all potential prognostic factors should be included in the analysis.[61]

Baseline characteristics of the three studies are compared in CS Table 32. Heterogeneity is evident in terms of:

- CASSIOPEIA had a higher proportion of ISS stage II patients and a lower proportion of ISS stage III patients than IFM 2005-01 and GMMG-MM5
- CASSIOPEIA patients had a lower median creatinine concentration than those in IFM 2005-01 and GMMG-MM5
- CASSIOPEIA had a lower proportion of patients with serum LDH (lactate dehydrogenase) below the upper limit of normal (ULN) and a higher proportion above the ULN compared to GMMG-MM5
- CASSIOPEIA had a lower proportion of patients with renal insufficiency than GMMG-MM5.
- Some patients in IFM 2005-01 (20.8%) and GMMG-MM5 (27.8%) received a second ASCT unlike patients in CASSIOPEIA (0%).
- IFM 2005-01 reported that 42% of patients had the chromosomal abnormality del(13) as detected by FISH whilst GMMG-MM5 reported that 37% of patients had Gain 1q21 (>2 copies) positive. Neither of these abnormalities are reported for CASSIOPEIA.
- There were differences in inclusion/exclusion criteria between the studies. IFM 2005-01 excluded subjects with end-stage renal failure.
- Only GMMG-MM5 included patients with ECOG performance score 3 (although only 4 patients; 1.6%).

Key prognostic factors for MM are captured in the ISS (see Table 4 in section 2.2.1 above). These include serum β2 microglobulin and serum albumin in the original version of the ISS, and, additionally, LDH above ULN, and the presence of chromosomal abnormalities del(17p) and/or t(4;14) and/or t(14;16) in the revised ISS. The ERG's clinical experts agreed that the revised ISS captures most of the key prognostic factors for MM, but commented that renal impairment and the extent of extramedullary disease are additional prognostic factors that should be considered, and the chromosomal abnormalities t(14;20), gain 1q and del 1p are also assessed in UK practice (section 2.2.1 above). Although the studies report the original ISS rather than the revised ISS, all the constituent factors in the revised ISS are represented among the baseline characteristics of the studies, together with renal function, and hence could be adjusted for in the MAIC. The company noted in clarification question A12 that extramedullary disease was not reported by the comparator trials but as this affects less than 10% of subjects it would not be expected to influence the MAIC results. The ERG's clinical experts agreed.

In summary, the ERG note that there is heterogeneity between studies, meaning that an adjusted comparison is necessary, and we agree that most, but not all, potential prognostic factors were reported in the studies.

3.4.4 Similarity of treatment effects

Two outcomes were included in the MAIC analyses: OS and PFS. Each of these outcomes appears comparable across the studies (CS Appendix Table 14).

3.4.5 Risk of bias assessment for studies included in the ITC

The CS assessed the risk of bias of the CASSIOPEIA, IFM 2005-01 and GMMG-MM5 studies as reported in CS Appendix Table 17. Overall the studies were rated favourably, but many of the risk of bias criteria are not relevant to a MAIC since randomization is broken and homogeneity of populations between trial arms becomes irrelevant (instead, heterogeneity assessment noted above in section 3.6.3 is important). The company noted that outcomes data were available for 91.4% of the randomised participants in IFM 2005-01 (i.e. a small proportion of data were missing but do not specify for which outcomes. All studies were open-label and the company suggest that outcome assessors were probably aware of participant allocations in all studies. However, this is unlikely to be a source of bias since the key outcome assessments were objective measures of disease response and survival.

3.5 Critique of the ITC methods

3.5.1 Data inputs to the MAICs

The summary data inputs to the MAICs are reported in CS Table 33 for the matching to Bd and CS Table 34 for BCd. The impact of inclusion/exclusion of specific prognostic factors on the effective sample size (ESS) is described in section 3.5.2 below.

3.5.2 Statistical methods for the MAICs

The company elected to conduct a population matching exercise using unanchored MAIC. Estimates of OS and PFS from CASSIOPEIA, the "reference trial", where individual patient level data (IPD) are available were adjusted to be comparable to a "target" trial where only aggregate data are available. The data in the "reference" trial are reweighted so that the mean baseline characteristics match the "target" trial.[61] All prognostic factors and treatment effect modifiers should be included in the analysis. The company argue that sufficient overlap of baseline characteristics precluded the need for a simulated treatment comparison (STC), an alternative population matching technique.

The ERG disagree there is a clear rationale for the use of the MAIC over the STC methodology. There is a lack of overlap in certain characteristics such as renal insufficiency (BCd) and serum LDH (Bd), which could not be matched on and creatinine (Bd) which led to a large reduction in ESS and an associated large number of subjects with zero weights. Hence the ERG requested the company to provide an STC as a scenario analysis (clarification question A16). The company responded that STC was not a suitable alternative method. The ERG generally agree with the company's argument that there are limitations to the use of STC, including the scale conflict between the outcome and predictor variables. However, there are work-arounds which the company note, and the ERG would have preferred STC to be employed as a scenario analysis.

No statistical code or data for the MAIC were provided with the submission. The ERG requested the statistical code for the MAIC and associated data in clarification question A17. The company provided the code, but the IPD for CASSIOPEIA were considered confidential and not provided to the ERG. Twenty-four SAS files were provided, however, these were opaque and referred to a number of macros. We therefore could not verify that the MAIC had been correctly implemented nor check the results.

In terms of generalisability the ERG's clinical experts agreed that the BCd and Bd populations in the studies were generally similar to those observed in UK practice. However,

they noted that the 21-day cycles of Bd and BCd in which bortezomib is administered twice-weekly are rarely used in the UK which favours better tolerated 28-day cycles and weekly bortezomib. This is likely to be conservative (i.e. favour the comparators rather than DBTd) and in any case the total dose is likely to be similar.

The company performed two separate MAIC analyses, using the individual patient data (IPD) from CASSIOPEIA for the DBTd and BTd groups to form comparisons against Bd (using aggregate data from IFM 2005-01) and BCd (using aggregate data from GMMG-MM5). As noted above, only OS and PFS were compared. Baseline characteristics pre- and post-matching are reported in CS Table 33 (comparison versus Bd) and CS Table 34 (comparison versus BCd).

3.5.3 Comparison of DBTd and BTd against Bd

Prognostic factors included in the MAIC analyses versus Bd are shown in Table 21. LDH was not reported in IFM 2005-01 so could not be matched. Creatinine (a marker of renal function) was excluded from the primary (base case) MAIC due to a lack of overlap of values between CASSIOPEIA and IFM 2005-01.

Table 21 Factors included in the MAICs of DBTd and BTd versus Bd

MAIC base case	Sensitivity analysis
Age, years (median) Gender, % male ISS class, % in each class β2-microglobulin, %: >3mg/L Abnormal cytogenetics: % with del17p and/or t(4;14) Haemoglobin, g/dL (mean, median) Calcium, mmol/L (mean, median)	All base case factors plus creatinine, µmol/L (mean, median)
Source: Summarised from CS section B.2.9.3	

Exclusion of creatinine from the MAIC base case resulted in 24% and 27% reductions in the effective sample size (ESS), for DBTd and BTd, respectively. However, inspection of the rescaled weights (CS Appendix Figures 5 and 6) shows that no very large weights were attributed which could skew the analysis. The company conducted a MAIC sensitivity analysis that included creatinine, and this led to a greater reduction in ESS (90% and 77%) and revealed a skewed distribution with some very large weights (CS Appendix Figures 7 & 8).

The baseline characteristics post-matching reported in CS Table 33 show that matching was effective at balancing variables included in the analysis for the base case and sensitivity analysis. However, the large reduction in ESS and a skewed distribution of weights when creatinine was included in the MAIC suggests there was limited overlap between studies. Furthermore, LDH, another key prognostic factor, was not reported in IFM 2005-01 and thus could not be included in the matching. The impact of excluding these prognostic factors is uncertain.

The Company also conducted a naïve ITC which did not adjust for any prognostic factors. Results were consistent across the naïve and matched analyses and showed a benefit for DBTd and BTd over Bd (CS Table 35).

3.5.4 Comparison of DBTd and BTd against BCd

Prognostic factors included in the MAIC analyses versus BCd are shown in Table 22. Anaemia was excluded from the MAIC base case due to lack of overlap between CASSIOPEIA and GMMG-MM5. Renal insufficiency was also excluded from the MAIC; there were only two patients in CASSIOPEIA with renal insufficiency (compared to 15.5% in GMMG-MM5) which precluding matching. However, the creatinine levels were included in the matching and the ERG's clinical experts agreed that this is an adequate measure of renal function for comparison purposes provided that the age and gender compositions of the populations are similar.

Table 22 Factors included in MAICs of DBTd and BTd versus BCd

MAIC base case	Sensitivity analysis 1	Sensitivity analysis 2
Age, years (median) Gender, % male ECOG/WHO performance status (PS), % in each PS class Heavy-chain isotope, %: IgG, IgA, LCD, other Calcium elevation, %: >2.65mmol/L Bone disease, %: lytic lesions ISS class, % in each class LDH , %: >ULN	All base case factors plus anaemia, %: Hb <10 g/dL or 2 g/dL < normal	All base case factors except LDH, %: > ULN

Abnormal cytogenetics, %: del17p and/or t(4:14) Calcium, mmol/L: median, % above 2.4 Creatinine, mg/dL: median, % above 1 Haemoglobin, g/dL: median, % above 10.7 Platelets, per nL: median, % above 240		
Source: Summarised from CS section B.2.9.3 LCD: light chain disease NB β2-microglobulin was not reported in GMMG-MM5 (CS Table 32)		

In the MAIC base case there were large reductions in ESS (62% for DBTd and 61% for BTd) accompanied by a large proportion of zero weights (CS Appendix Figures 9 & 10). Two further sensitivity analyses were conducted: the first added anaemia but led to similar reductions in ESS (64% & 62%, respectively). The second sensitivity analysis excluded LDH which led to slightly lower reductions in ESS (50% & 50%, respectively) (CS Appendix Figures 11-14). Baseline characteristics post-matching are reported in CS Table 34. Anaemia was only matched in the first sensitivity analysis but was not identified as a key prognostic factor by the ERGs experts. Matching appeared successful for the remainder of the characteristics.

The company also conducted a naïve ITC which did not adjust for any prognostic factors. Results were consistent with the MAIC approach and showed a benefit for DBTd over BCd and similarity of BTd and BCd (Table 36).

3.5.5 External validation of MAIC results against real-world data

The company conducted an external validation of the MAIC results using a Public Health England (PHE) dataset. This consisted of a large dataset of newly diagnosed MM patients in England between 2015 and 2018. Unadjusted comparisons of OS and PFS from subjects receiving BTd (n=1,218), BCd (n=588), and Bd (n=248) showed that BTd and BCd were approximately equivalent and superior to Bd (CS Table 37). However, the distribution of prognostic factors between arms is neither reported nor adjusted for in this naïve analysis, and it is not clear why 2019 data were not included.

3.5.6 Results of the indirect comparisons

The company's ITC results (naïve comparison and MAICs) versus Bd and versus BCd are presented below in Table 23 and

Table 24, respectively.

Table 23 Results of the naive comparison and MAIC (DBTd and BTd versus Bd)

	Naïve comparison		MAIC (Base case)		MAIC (Sensitivity analysis)	
	PFS	**OS**	**PFS**	**OS**	**PFS**	**OS**
DBTd vs Bd						
ESS (DBTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BTd vs Bd						
ESS (BTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Reproduction of CS Table 35. ESS added by ERG						

Table 24 Results of the naive comparison and MAIC (DBTd and BTd versus BCd)

	Naïve comparison		MAIC (Base case)		MAIC (Sensitivity analysis 1)		MAIC (Sensitivity analysis 2)	
	PFS	OS	PFS	OS	PFS	OS	PFS	OS
DBTd vs BCd								
ESS (DBTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BTd vs BCd								
ESS (BTd)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P-value	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Reproduction of CS Table 36. ESS added by ERG								

3.5.7 Summary of the ERG's critique of the indirect comparisons

- The ERG agree that the MAIC approach was the most appropriate method for the ITC. We would have liked to have seen a scenario analysis using STC; however, the results of the indirect comparisons are not used directly in the economic model.
- Although not all prognostic factors could be included, the ERG are satisfied that all available prognostic factors that could be matched were included in the analyses.
- The MAICs and naïve comparisons are supportive of the company's assumption that BTd is equivalent BCd and that both BTd and BCd are superior to Bd.
- These methods are subject to limitations, including that the ERG could not validate the methods as IPD were unavailable. However, the ERG's clinical experts agreed that the company's assumption about relative treatment effectiveness is appropriate.

3.6 Systematic literature review and meta-analysis of the impact of MRD status on survival outcomes

The company's approach to response-based modelling requires an estimate of the impact

of MRD status (MRD-negative versus MRD-positive) on PFS and OS (CS section B.3.3.2). The company therefore conducted a systematic literature review (SLR) and meta-analysis of the impact of MRD status on PFS and OS (reported in CS section B.3.3.2 and CS Appendix M). The company state that their meta-analysis was an "expanded" meta-analysis (reported in an abstract by Munshi et al. 2019[62]) that was based on a previous meta-analysis conducted by Munshi et al. 2017.[63] The CS does not clearly explain the relationship between the two meta-analyses, the study selection criteria that were used, the characteristics, heterogeneity or validity of the included studies, or the statistical methods used for data synthesis. The ERG therefore sought further explanation of the company's methods (clarification questions B1 to B14).

3.6.1 Identification and selection of studies

3.6.1.1 Searches

Three brief search strategies are reported in CS Appendix M.1: for MEDLINE articles, EMBASE articles, and EMBASE conference abstracts. Additional hand searches were carried out on relevant conference websites and the bibliographies of systematic reviews on MM identified through the database searches. The company's searches were conducted in June 2019 and were 15 months out of date when the CS was received by the ERG. In clarification response B3 the company report that they checked whether any RCTs identified in their clinical effectiveness SLR searches (which were conducted more recently) reported survival (OS or PFS) by MRD status. No relevant trials were identified.

Given that evaluations of the relationship between MRD status and survival outcomes are often single-arm studies (as acknowledged by the company in clarification response B7) non-randomised studies should be included. The company state in clarification response B3 that they would provide an updated SLR and meta-analysis that includes non-randomised studies published after May 2019. The results of this update were not available to the ERG at the time of submission of this report.

3.6.1.2 Study selection

The company explain in clarification response B1 that the original SLR reported by Munshi et al. 2017[63] included both transplant-eligible and ineligible patients. The company's expanded SLR[62] (CS Appendix M) was limited to studies where: transplant was performed; MRD was measured at 100 days post-ASCT; and studies represented standard of care. The full eligibility criteria are provided in clarification response B4, reproduced in Table 25 below. CS Appendix M states that the study selection process was conducted by two independent

reviewers, with disagreements resolved by a third investigator, but does not mention whether any of the reviewers were independent of the company.

Table 25 Eligibility criteria for the SLR of MRD status on survival outcomes

	Inclusion Criteria	Exclusion Criteria
Population	Newly diagnosed transplant-eligible patients with MM	Patients without a primary diagnosis of MM; Relapsed/refractory patients, transplant-ineligible patients.
Intervention/ comparator	Any treatment representing standard of care, with ASCT transplant	• Allogenic stem cell transplant • ASCT not performed • Studies with DBTd (CASSIOPEIA DBTd arm)
Outcomes	• OS and/or PFS stratified by MRD status (using any MRD definition) • Any PRO, TTP or PFS2 reported by MRD status • MRD measured at 100 days post-ASCT	• Survival data that cannot be extracted or are not available • MRD measured in peripheral blood • MRD assessed by PET-CT • MRD not measured at 100 days post-ASCT
Study Design	RCTs and non-RCT study design	Economic models, case reports, comments and editorials, animal/in-vitro studies
Date Limit	• No date limit applied on indexed databases search • Conference abstract and other materials (grey literature): 3 years (2016-2019): EHA, ASH, ISPOR, ASCO. • SLRs: 5 years (2014-2019)	• Conference abstract or other materials (grey literature) published before 2016 • SLRs published before 2014
Language	English language	Non-English language
Source: Reproduction of Table 5 in clarification response B4 adapted by ERG		

In clarification response B5 the company provided a list of the 820 studies that had been included in full-text screening in their original systematic review up to May 2019 and the 677 studies excluded at full-text screening, with reasons for exclusion (clarification response B5). These studies are consistent with the PRISMA flow chart reported in CS Appendix Figure 61 for the company's original SLR.

The CS states that “the SLR/meta-analysis included a number of older trials which do not capture the shift in outcomes for MM patients due to the introduction of novel agents as well as trials with a range of MRD sensitivity thresholds” (CS section B.3.3.2). The ERG queried why the company had included older trials in their original SLR/meta-analysis given this statement; and we requested a sensitivity analysis to explore the impact of the older trials on the meta-analysis. In clarification response B2 the company provided sensitivity analyses which show that the PFS and OS hazard ratios are not sensitive to study publication date. We therefore agree that inclusion of all studies in the SLR, irrespective of publication date, is appropriate.

3.6.2 Assessment of study validity

An assessment of validity is important to ensure that any studies at risk of bias can be identified and their impact on the meta-analysis investigated. However, validity assessment for the studies included in the SLR is not reported in the CS. The company state in clarification response B11 that they had employed the “modified STROBE Statement” (no reference provided) to assess the risk of bias and they considered the studies “all sufficiently robust and of a decent quality for inclusion”. However, no results of the STROBE assessment have been provided. The ERG note that the STROBE Statement is a reporting guideline, not a risk of bias assessment tool, and likely would not have identified key risks of bias. The ERG’s concern was communicated to the company via NICE prior to the company’s clarification response being received. The company took this into consideration and, as stated in clarification response B7, conducted a risk of bias assessment using the Cochrane ROBINS-I tool (Risk of Bias in Non-Randomised Studies of Interventions). Partial results of this assessment, for the RCTs only, were made available to the ERG close to the deadline for submission of this report and have not yet been critiqued by the ERG. The company intended to provide the full risk of bias assessment, including non-randomised studies identified in their updated SLR, but this had not been received by the ERG at the time of submission of this report.

3.6.3 Assessment of study heterogeneity

The CS does not report baseline characteristics of the included studies, nor a statistical assessment of heterogeneity. The appropriateness of statistically combining the studies in a meta-analysis is therefore unclear.

In clarification response B9 the company provide a tabulation of the baseline characteristics of the studies included in their original SLR, together with a discussion of their comparability

to the CASSIOPEIA trial (noting that most of the studies included were non-randomised). The table of study characteristics (Table 7 in clarification response B9) is limited to only five population variables: age, sex, ISS category (I, II, III), type of measurable disease (IgG, IgA, other), and cytogenetic risk profile (% high risk thereby missing serum LDH and renal function which are key prognostic factors. The company note some differences between the studies in the proportions of patients with ISS category I, and with high cytogenic risk. The company note some differences between the studies in the proportions of patients with ISS category I, and with high cytogenic risk.

The company state in clarification response B12 that they conducted several subgroup analyses to identify potential effect modifiers (i.e. those variables that significantly impact the association of MRD and PFS or OS). However, these subgroup analyses are limited to varying the MRD detection threshold (10^{-4}, 10^{-5}, 10^{-6}) and limiting to only patients who achieved CR, or those who achieved at least VGPR (Tables 9 and 10 in clarification response B12). They do not include any prognostic factors or any other patient demographic variables, and a rationale for this (e.g. that the factors were not reported in the studies) is not provided.

3.6.3.1 Heterogeneity in the MRD assessment approach

In clarification response B9 the company report that around half (N=8) of the trials included in the meta-analysis assessed MRD status regardless of response. The remaining trials assessed MRD in patients who achieved CR only (3 trials) or in patients who achieved VGPR or better (3 trials) (in one trial data were not available).

In clarification response B12 the company present a sensitivity analysis comparing HRs for the impact of MRD status on PFS for MRD measured only in patients who achieved CR (3 trials) versus MRD measured in patients who achieved at least VGPR (3 trials). The analysis does not include a comparison with MRD assessed for any response (8 trials). The HRs for PFS comparing MRD-negative versus MRD-positive are [redacted] when only patients achieving CR were assessed and [redacted] when patients achieving at least VGPR were assessed. A similar comparison was not possible for OS as there were no observations in the CR only group.

These results suggest that the way MRD is assessed influences HRs, although the HR confidence intervals are wide, indicating uncertainty. The ERG would have preferred the sensitivity analysis to have included MRD measured regardless of response, as this is

consistent with the MRD approach used in the CASSIOPEIA trial. Further, an update of the sensitivity analysis to include any additional trials identified by the company in their updated search would be helpful.

3.6.3.2 Statistical heterogeneity

The company explain in clarification response B10 that statistical heterogeneity was assessed using the maximum likelihood estimator to derive the I-squared statistic and a chi-squared statistic (Q) to test the significance of I^2. For the PFS analysis [redacted] and for the OS analysis [redacted]. The company conclude that the studies were relatively homogeneous. The ERG note that there is some evidence of heterogeneity (I^2 >0) and uncertainty in the I^2 estimates (wide confidence intervals) but there is not strong statistical evidence of substantial heterogeneity (I^2 <50 and Q non-significant).

In summary, the comparability of the studies included in the company's original meta-analysis remains unclear to the ERG, due to the limited characteristics reported, limited consideration of prognostic factors when comparing the studies, and lack of any prognostic factors in the sensitivity analyses conducted.

3.6.4 Method of data synthesis

The statistical method of data synthesis is not reported in the CS. The company explain in clarification responses B8 and B14 that a frequentist meta-analysis was used to synthesise mean hazard ratios, fitting a random effects model to obtain the pooled estimate; generating the pooled KM curves based on the simulated IPD; and conducting subgroup analysis to address the potential bias caused by the differences in disease setting, eligibility for MRD assessment by conventional response, MRD assay and sensitivity and time of MRD assessment. According to clarification response B14 the analysis was performed using the "metafor" R package (v2) for frequentist meta-analyses, within JASP open-source software. The company could not extract the source code from the JASP interface but tabulated the data input inputs for PFS and OS from each of the studies included in their original SLR (tables in clarification response B14). (NB "TDE" appears to be a typographic error which should read "PFS" in Table 11 of clarification response B14). The ERG agree that the PFS and OS input data for the meta-analysis appear appropriate and the use of a random effects model (justified in clarification response B8) is also appropriate. The ERG were able to validate the method and reproduce the company's forest plots which are provided in clarification response B2.

3.6.5 Results of the meta-analysis

The results of the meta-analysis are used to model PFS and OS for the MRD-negative group in the BTd arm of the economic model (the general approach employed is summarised in Table 29 below and in CS Figure 31). The hazard ratios for MRD- versus MRD-positive were applied to BTd MRD-positive curves as shown in CS Table 51 for PFS and CS Table 54 for OS. Note that these hazard ratios are taken from the meta-analysis based on the company's original SLR, and do not include any studies identified in the company's updated searches.

3.6.6 ERG critique of the SLR and meta-analysis

The general statistical approach to the meta-analysis appears appropriate and is reproducible. However, as noted in the sections above, key limitations are:

- The meta-analysis is based on the company's original SLR of studies published up to June 2019 and does not include non-randomised studies published during the past 15 months. It is unclear whether any relevant studies are missing.
- Except for CASSIOPEIA, the validity of the studies included in the meta-analysis had not been assessed in time for the ERG to provide a critique within this report. It is unclear whether any of the comparator studies are at risk of bias.
- The comparability of the studies included in the meta-analysis is unclear, as the company's assessments of clinical heterogeneity explored a limited set of patient characteristics which did not include key prognostic factors such as serum LDH and renal function.
- To reduce uncertainty in the results of the company's SLR and meta-analysis, the company's updated SLR and meta-analysis should address the shortcomings noted above, including clarification of the impact on HRs of different approaches for measuring MRD (ERG Issue 1 – see Table 26 below).

3.7 Conclusions of the clinical effectiveness section

The clinical effectiveness issues identified by the ERG are summarised in Table 26 below. As noted in the Executive Summary (section 1.3) only the first three issues are considered to be key issues. We believe that the remaining issues are relatively minor, because they are not expected to influence cost effectiveness and/or they have been resolved (e.g. through clinical experts' opinion) within this report.

Table 26 Clinical effectiveness issues identified by the ERG

Key issues (1-3)	Summary	Priority and action

1	**SLR and meta-analysis of the effects of minimal residual disease (MRD) status on survival outcomes (section 3.6.6)** Issue: Uncertainty in hazard ratios due to limitations in the methods	SLR and meta-analysis limitations: • Searches are 15 months out of date so unclear whether any relevant recent studies are missing; • The validity of included studies is unclear; • The heterogeneity of included studies is unclear; • The impact of different methods for measuring MRD status is unclear; a sensitivity analysis excluded MRD when assessed regardless of response, which was the approach employed in around half the included studies (see also Issue 2).	**High priority**. Informs the economic model. The company agreed to provide an update of the SLR and meta-analysis but not in time for the ERG to critique within this report. The ERG will provide a critique either in a report addendum or at Technical Engagement, depending on the quantity and clarity of information provided by the company.
2	**Company's approach for defining and analysing MRD-negative patients (section 3.2.3.3)** Issue: Inconsistency in the method of assessing MRD	• The CS states that MRD-negativity was determined regardless of response. This is inconsistent with the IMWG definition of MRD negativity, which requires a complete response. • The definition of MRD used in the company's landmark analysis may therefore differ from the definition used in the studies included in the company's meta-analysis of the impact of MRD status on survival outcomes.	**High priority**. Both the landmark analysis and the meta-analysis of MRD impact on survival outcomes inform the economic model and the impact of different definitions of MRD status should be taken into account when the company update their analyses. A sensitivity analysis reported in clarification response B12 suggests the MRD definition may affect HRs. Implications for the comparability of HRs from the landmark analysis and the meta-analysis should be clarified.
3	**Adjustment of PFS to capture the effect of second randomisation to maintenance therapy (section 3.2.4.6)** Issue: Limitations in the methods	Analysis limitations: • The proportional hazards assumption appears to have been violated; • Limited follow up in the maintenance period; • The ERG are unable to validate the analysis as the maintenance phase is currently blinded.	**Medium priority.** Adjustment for maintenance therapy is not currently factored into the landmark analysis which informs the economic model. The company may address these limitations at the Technical Engagement step when they provide an updated analysis after unblinding of the trial maintenance phase.
Minor issues (4-6)			

4	**MAIC analyses (section 3.5.4)** Issue: Uncertainty in the reliability of HRs due to limitations of the methods	MAIC limitations: • Unclear whether MAIC is the most appropriate method (lack of overlap of population characteristics); • The ERG are unable to validate the MAIC analyses as IPD are confidential; • As an unanchored approach to ITC, the method has inherent uncertainty.	**Low priority**. HRs from the MAICs do not directly inform the economic model, but they inform a key assumption guiding the modelling approach (effectiveness of BTd = BCd > Bd). The ERG's clinical experts agreed with the company's assumption.
5	**Applicability of DBTd to the care pathway (section 2.2.2)** Issue: The technology, DBTd, includes post-ASCT consolidation therapy which is not used in current clinical practice	The DBTd intervention involves four induction and two post-ASCT (consolidation) cycles of therapy, compared to 4-6 cycles of induction-only BTd therapy in clinical practice.	**Low priority**. The ERG's clinical experts agreed that consolidation therapy could be readily integrated into NHS practice, and would be welcomed, in clinical practice.
6	**Decision problem (section 2.3)** Issue: CTd comparator therapy excluded from the decision problem	CTd therapy is specified as a comparator in the NICE scope but excluded by the company from the decision problem.	**Low priority**. The ERG's clinical experts agreed with the company that CTd is rarely used and that is appropriate to exclude this from the comparison.

4 COST EFFECTIVENESS

The company submission includes:

- A systematic review of published economic evaluations of interventions for transplant-eligible patients with newly diagnosed MM (CS section B.3.1 and CS Appendix G);
- A description of the company's de novo model developed to assess the cost-effectiveness of DBTd versus BTd for patients with newly diagnosed MM who are eligible for ASCT (CS sections B.3.2 to B.3.11).

4.1 ERG critique of the company's review of cost-effectiveness evidence

The company conducted a combined systematic literature search to identify published economic evaluations and cost/resource use studies of interventions for the treatment of patients with newly diagnosed MM. The search was conducted in April 2018 and updated in May 2020, see CS Appendix G for details. Although the search did not use a published economics filter, the choice of economic and cost search terms has reasonable coverage. The combined search required combined inclusion/ exclusion criteria and PRISMA flow charts, but the reporting remained clear. Results are presented in CS section B.3.1 and CS Appendix G for economic evaluations; and CS Appendix I for the review of costs and healthcare use. The company state that no UK studies of costs or resource use were identified (CS section B.3.5).

Nine relevant economic evaluation studies were identified (CS Appendix Table 32, CS Appendix G.3.1). Two of these studies were UK based, both HTA submissions for bortezomib (NICE TA311 and SMC ID927/13) (CS Table 44).[27,64] The company stated that these HTAs were used to inform parameters and assumptions for the current appraisal. No studies evaluating the cost-effectiveness of daratumumab in the population of interest were identified.

ERG conclusions:

The company's search strategy and eligibility criteria for their review of cost-effectiveness studies are appropriate. The search did not identify any economic evaluations of daratumumab in the population of interest. The company used TA311 to inform inputs and assumptions for their model.

4.2 ERG summary and critique of the company's submitted economic

4.2.1 NICE reference case checklist

The ERG assessed the company's economic evaluation against NICE Reference Case requirements as shown in Table 27. We consider that all criteria are met.

Table 27 NICE reference case

Issue	Reference case	ERG comment
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and PSS	Yes
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes, a lifetime time horizon in the base case. The economic model has the flexibility to run the analysis for shorter time horizons (10, 20 years).
Synthesis of evidence on health effects	Based on systematic review	Yes. The company model compared DBTd with BTd (other comparators were excluded), and the company SLR (CS section B.2.1 and CS Appendix D) confirmed CASSIOPEIA is the only relevant source of effectiveness evidence for this comparison. The model also uses external evidence on the relationship between MRD status and long-term survival outcomes from a SLR and meta-analysis (CS section B.3.3.2).
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes. The model outputs include QALYs. Pre-progression utilities are derived from CASSIOPEIA EQ-5D-5L data and post-progression utility from EQ-5D-3L data reported by van Agthoven et al. (CS section B.3.4.1).[65]
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes

Issue	Reference case	ERG comment
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes. CASSIOPEIA EQ-5D-5L data are valued using the van Hout crosswalk algorithm and UK value set and the van Agthoven value is based on the EQ-5D-3L UK social tariff (CS section B.3.4.1).[65,66]
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
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	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes
PSS, personal social services; QALYs, quality-adjusted life years; EQ-5D, standardised instrument for use as a measure of health outcome.		

4.2.2 Model structure

The company describe the structure and key features of their model in CS section B.3.2.2. Assumptions are summarised in CS Tables 46 and 81; and parameters in CS sections B.3.3 to 3.5 and CS Table 80. The model uses a response-based, partitioned survival structure, with a cycle length of 4 weeks, a lifetime horizon and half-cycle correction. Costs and QALYs are discounted at 3.5% per year.

The model consists of three 'partitioned survival' health states, as illustrated in Figure 4:

- *Progression-free (PF):* calculated as the proportion of patients alive and progression-free (PFS).
- *Progressed disease (PD):* calculated as the proportion of patients alive (OS) minus the proportion of patients alive and progression-free (PFS).
- *Death:* calculated as one minus the proportion of patients alive (OS)

The Survival extrapolations (PFS and OS) are estimated separately for patients assessed as MRD negative or MRD positive after consolidation therapy (as illustrated in Figure 5). The

company argue that this response-based approach reflects patient heterogeneity and enables the use of external data to link "deep" MRD response to better long-term survival outcomes, reducing uncertainty caused by immaturity of survival data from the CASSIOPEIA trial. They also note that the model in the NICE appraisal of bortezomib for this population (TA311) used a response-based approach, although it differed from the current model in several respects, see CS Table 46 for the company's justification of the differences.[27]

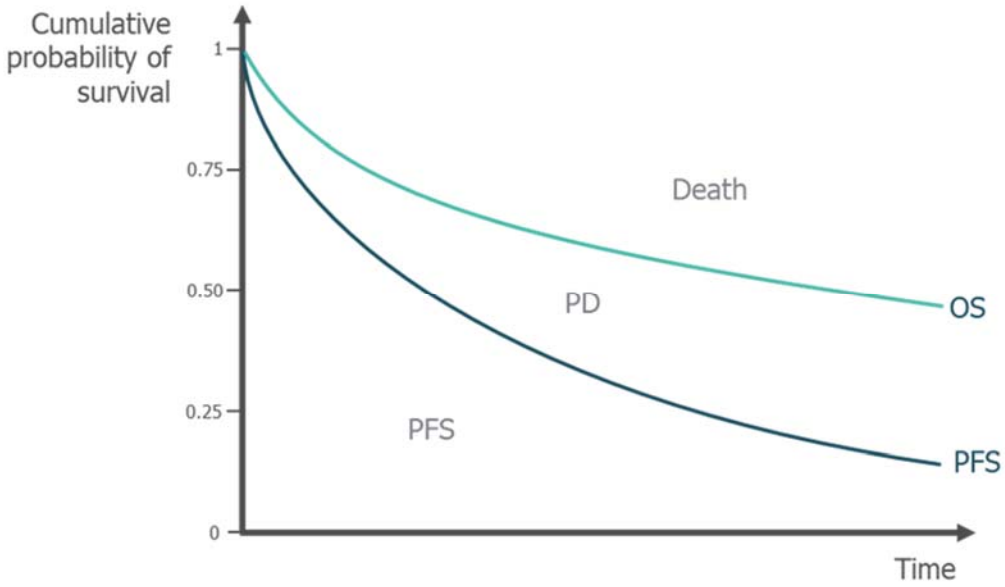


Figure 4 Partitioned survival model structure
Source: reproduced from CS Figure 28

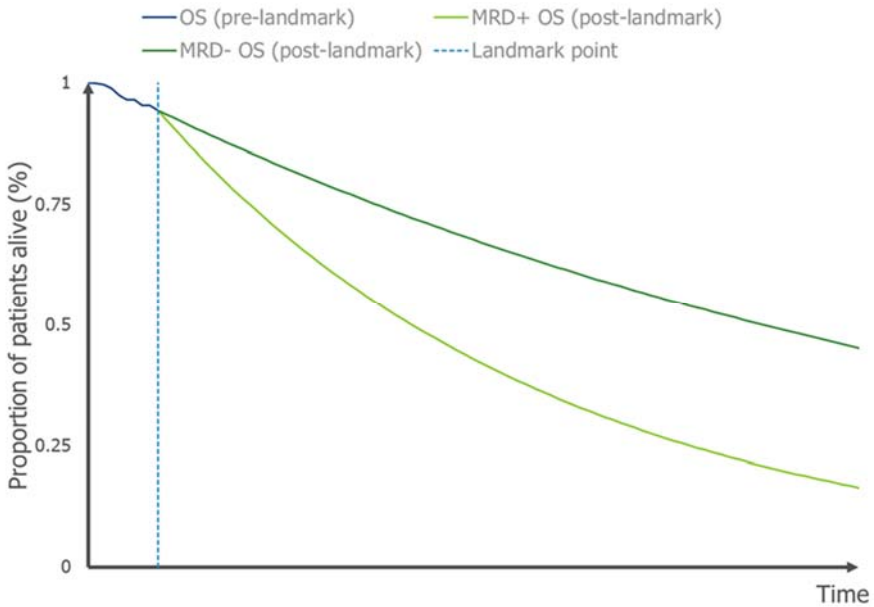


Figure 5 Response-based landmark survival analysis
Source: reproduced from CS Figure 28

In CASSIOPEIA, MRD status was assessed after consolidation therapy: 100 days post ASCT, with a mean time from the start of induction of 37 weeks, or approximately 9 model cycles. Before this 'landmark' timepoint, the survival curves follow KM data from the trial. After the landmark, separate survival curves are estimated for MRD negative and positive subgroups using results from the CASSIOPEIA landmark survival analysis (section 3.2.4.7 and 3.2.7.5 above and CS section B.2.6.3) and a hazard ratio for the relationship between MRD status and survival from the meta-analysis of published data (section 3.6 above and CS section B.3.3.2). We describe and critique the company's approach to fitting response based PFS and OS curves in section 4.2.6.2 below.

In their base case, the company apply constant hazard ratios for DBTd versus BTd to post-consolidation survival throughout the time horizon in both MRD negative and positive subgroups. This assumes a lasting treatment effect for daratumuab, on top its direct effect on MRD response. The company cite TA311, TA510 and TA573 as precedents for the assumption that daratumumab effects do not wane over time.[27,30,36] However, we note that the TA311 model did not assume any treatment-specific survival benefit after the initial induction response, so waning was not relevant to this model. In both TA510 and TA573, simple (not response-based) parametric survival extrapolations were used. Furthermore, in TA510 the committee took account of a company scenario assuming no further survival benefit after trial follow-up (paragraph 4.19 in the NICE TA510 guidance). And in TA573, the committee preferred the Weibull survival curve for the daratumumab arm, which had an increasing risk of death over time (paragraph 3.13 in the NICE TA573 guidance).

Following the usual partitioned survival approach, the company do not model progression through subsequent lines of treatment.[67] The PD health state includes cost estimates for second, third and fourth lines of treatment, but survival linked to specific treatment regimens is not modelled. The company argue that this is a reasonable simplification given the similarity of the subsequent treatment pathway after first line ASCT with DBTd or BTd.

We discuss input parameters and assumptions relating to adverse events, utilities and costs in sections 4.2.6.4, 4.2.7 and 4.2.8 respectively below.

ERG conclusions:

- The model structure is appropriate and accurately implemented. We agree with the partitioned survival approach. This is consistent with reported survival outcomes from CASSIOPEIA and facilitates comparison with other data sources,

such as the PHE dataset and studies in the MRD systematic review and meta-analysis for the comparator arm.

- There is also a good rationale for taking a response-based approach to survival modelling, as this can reflect heterogeneous patient response to cancer immunotherapies. Ultimately though, long-term prediction from immature trial data can only be improved with external information.[68,69] The company adopt a landmark approach to survival extrapolation and use external data on the relationship between MRD status and survival from published studies. This is appropriate, but reliant on the robustness of the MRD review and meta-analysis (see section 3.6 above for ERG critique). We also note that the model only makes use of pooled estimates of relative survival from the review (HRs for MRD-negative versus MRD-positive). Evidence on absolute survival in other studies with longer-follow up is not used to inform the model (section 4.2.6.2.3 below).
- In their base case model, the company apply treatment effects for the DBTd arm to survival outcomes in MRD negative and MRD positive subgroups from the landmark analysis. We note that there is uncertainty over these effects, with only the HR for PFS in the MRD negative subgroup reaching statistical significance. The company base case also assumes that additional daratumumab treatment effects on OS and PFS persist throughout the time horizon in both MRD groups. The ERG considers that there is insufficient evidence to support this assumption given current trial follow up. See section 4.2.6.2.4 below for further discussion.
- The model does not include separate health states for subsequent lines of treatment. Costs for second, third and fourth lines of treatments are included in the 'progressed disease' health state, but it is not possible to adjust survival estimates to reflect the mix of subsequent treatment regimens in clinical practice (see section 4.2.8.4). We consider this a reasonable simplification, which is common for partitioned survival models.[67]

4.2.3 Population

The company model a population of adults with newly diagnosed MM who are eligible for ASCT (CS section B.3.2.1). Baseline characteristics of the modelled cohort are based on those of patients in CASSIOPEIA: mean age 56.6 years, 41.5% female (see CS Tables 6 and 45).

The CASSIOPEIA trial excluded patients over 65 years of age. The company provide real world evidence from a Public Health England (PHE) dataset for patients with NDTE MM,

diagnosed between 1st January 2015 and 31st December 2018 (see CS Table 56).[70] The PHE cohort had a mean age of [redacted] years at diagnosis, with [redacted] aged 65 years or older.

Experts consulted by the ERG advised that (with the exception of age and general fitness), the CASSIOPEIA and PHE cohorts are broadly representative of patients seen in clinical practice.

> **ERG conclusions:** The modelled population is consistent with the licensed indication for daratumumab and the clinical trial evidence. However, the CASSIOPEIA trial only enrolled patients aged 18-65 years which does not reflect current NHS practice. For the ERG base case, we assume a mean age of [redacted] years (as in the PHE dataset), rather than 56.6 years (as in the CASSIOPEIA trial and company base case).

4.2.4 Interventions and comparators

The company describe the intervention and comparators included in their model in CS section B.3.2.3.

Intervention

The modelled intervention is DBTd administered as in the licensed indication and trial:

- 4 cycles as an induction therapy prior to ASCT and
- 2 cycles as a consolidation therapy post-ASCT

Consolidation therapy after transplant is not current practice in the UK. Expert advice to the ERG is that if DBTd were to be recommended, implementation of consolidation would not be a problem and that it would be welcomed by clinicians.

In the CASSIOPEIA trial daratumumab was administered as a weight based IV formulation (16 mg/kg). This is inconsistent with the company's base case analysis, in which daratumumab is costed as a fixed dose SC formulation (1800 mg). The company cite the COLUMBA trial as demonstrating non-inferiority between the IV and SC formulations.[54,71] This was the primary source of evidence for the EMA licence extension granted in June 2020. Our experts agreed that the fixed dose SC injection is likely to have equivalent safety and efficacy as the weight based IV infusion. The company conducted a scenario analysis with costs for the IV infusion.

Comparators

The only comparator considered in the economic model is BTd, which the company describes as standard of care. They argue that CTd is very rarely used and that other comparators in the NICE scope (BCd and Bd) are also less commonly used than BTd. In the PHE dataset, [redacted] of NDTE MM patients received BTd at first line, [redacted] of patients received BCd and [redacted] received Bd (CS section B.3.3.2 page 125).[70] The ERG's clinical experts estimated similar use of these therapies in their practice, but with some variation: BCd (10% to 40%) and Bd (negligible to 10%). We conclude that BTd and BCd are relevant comparators, but that Bd is less relevant as it is not commonly used in this population and triplet therapy is preferred.

The company argue that their MAIC analysis supports the conclusion that [redacted] (CS section B.2.9.5). Based on this and the similar cost of the generic regimens, they conclude that if DBTd is cost-effective versus BTd, then it will also be cost-effective versus BCd and Bd. The company do not present comparative costs for BTd and BCd at first line, but we confirm that costs are similar if the same number of treatment cycles is used for induction (6 cycles).

The modelled treatment protocol for BTd aligns with that in the CASSIOPEIA trial: 4 cycles of BTd as induction therapy and 2 cycles of BTd consolidation after ASCT. Four cycles of induction therapy are consistent with the bortezomib SmPC; but there is no precedent in UK clinical practice for consolidation therapy. Our experts state that between 4 and 6 cycles of BTd are currently used pre-ASCT in UK clinical practice; the number of cycles depending on the rapidity of response. In the model, treatment costs are based on the proportions of patients who received the induction cycles, ASCT and consolidation cycles in the ITT population of the CASSIOPEIA trial (CS Table 47). It is not clear whether costs for the BTd arm reflect the 4-6 induction cycles currently used in practice.

There are some inconsistencies in thalidomide and dexamethasone dosing in the CASSIOPEIA protocol and the bortezomib SmPC (see CS B.3.2.3). The company report a scenario analysis with costs based on dosing recommendations for thalidomide and dexamethasone as in the bortezomib SmPC; this has a minimal impact on the cost-effectiveness results.

ERG conclusions:

- The company use the cost of fixed-dose SC daratumumab in their base case, rather than the weight-based IV formulation as in the trial. There is evidence that the SC formulation is non-inferior, and it is also likely to be preferred by patients and clinicians because of convenience. However, inconsistency between daratumumab costs and effects may bias the ICER. We therefore prefer to use costs for IV daratumuab in the ERG base case, with an SC scenario.
- The company include BTd as the only comparator in the economic model, omitting other comparators from the NICE scope (BCd, CTd and Bd). We agree with the omission of CTd and Bd, as they are not commonly used in this population. BCd is used by around [redacted] of patients (PHE data). However, based on the company's MAIC analysis and clinical opinion, we agree that overall treatment effects and costs are likely to be similar for BTd and BCd. It is therefore reasonable to omit BCd as a comparator.
- The use of BTd for consolidation after ASCT in CASSIOPEIA does not align with current practice, although the total number of treatment cycles in the trial (up to 4 induction plus 2 consolidation) is within the range of 4-6 induction cycles used in practice and clinical advice to the ERG is that outcomes are unlikely to differ.
- Discrepancies between the CASSIOPEIA protocol and bortezomib SmPC in the dosing of thalidomide and dexamethasone are minor, make little difference to costs, and were not a source of concern for clinical experts consulted by the ERG.

4.2.5 Perspective, time horizon and discounting

The company use a lifetime horizon and take the perspective of the NHS and PSS in England. Both costs and outcomes (life years and QALYs) are discounted at 3.5%, in line with the NICE guidance.

4.2.6 Treatment effectiveness and extrapolation

The company report two approaches to modelling progression free and overall survival:

- **Direct extrapolation**; parametric OS and PFS survival models for BTd and DBTd fitted using individual data from the CASSIOPEIA trial. This approach was not used to derive ICERs.
- **Response based modelling**; OS and PFS survival dependent on post-consolidation MRD status (MRD-positive and MRD-negative) and treatment arm (DBTd and BTd). This was used in the company's base case and scenario analyses.

4.2.6.1 Direct extrapolation of survival outcome

In this approach, OS and PFS data for DBTd and BTd were extrapolated using parametric survival functions fitted to ITT CASSIOPIEA data from the most recent, post-hoc analysis (median follow-up of 29.2 months). Kaplan-Meier plots for PFS and OS used in these analyses are shown in CS Figures 12 and 16, respectively. We note that these data are immature:

- PFS 27.9% (151/542) BTd and 15.3% (83/543) DBTd events (CS Table 19)
- OS 8.9% (48/542) BTd and 4.8% (26/543) DBTd events (CS Table 24)

We also note that the KM data used for survival extrapolations are not adjusted for the second randomisation to maintenance treatment and that proportional hazards assumptions were not met for PFS (clarification response A11). Piecewise sensitivity analysis for PFS showed lower HR after ASCT (consolidation and maintenance phases) than during the induction phase (CS Table 18, median follow-up 18.8 months).

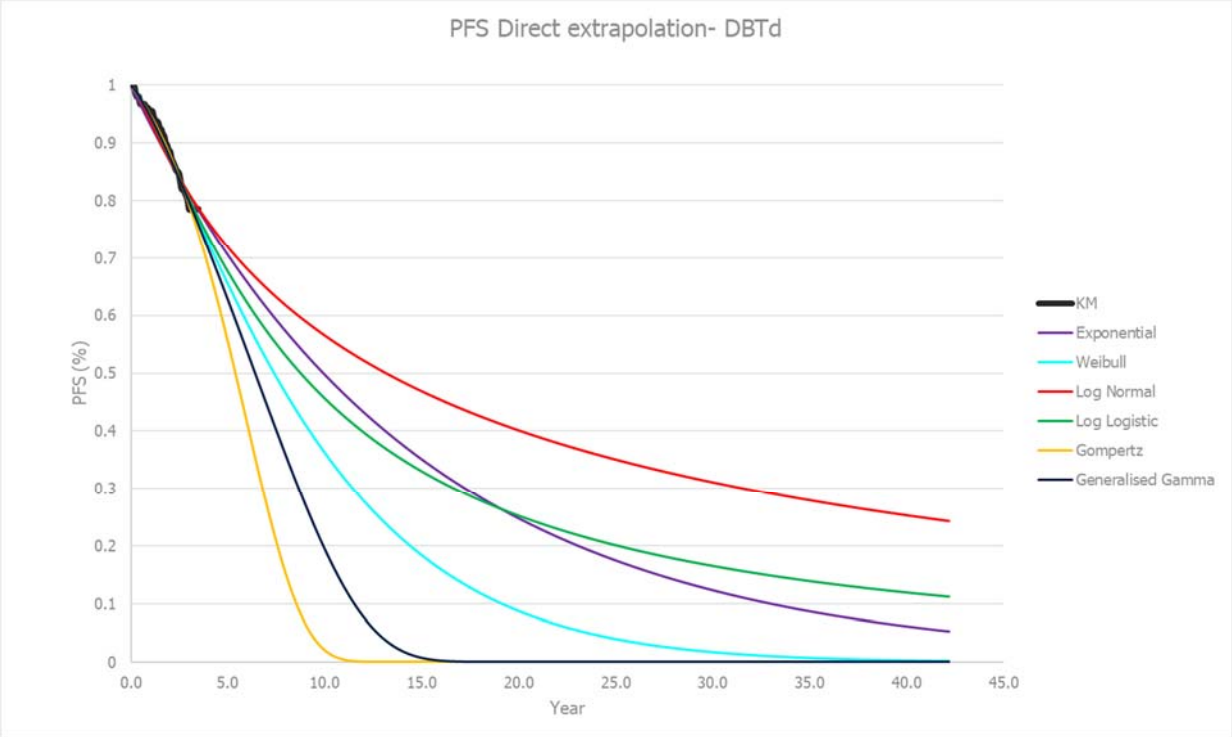
CS Figures 29 and 30 show the six parametric OS extrapolations. The company do not include equivalent graphs for PFS, but parameters for the fitted models and extrapolations are included in the economic model. The ERG have prepared graphs showing KM data and parametric extrapolations for PFS and OS (Figure 6 and Figure 7 below, respectively).

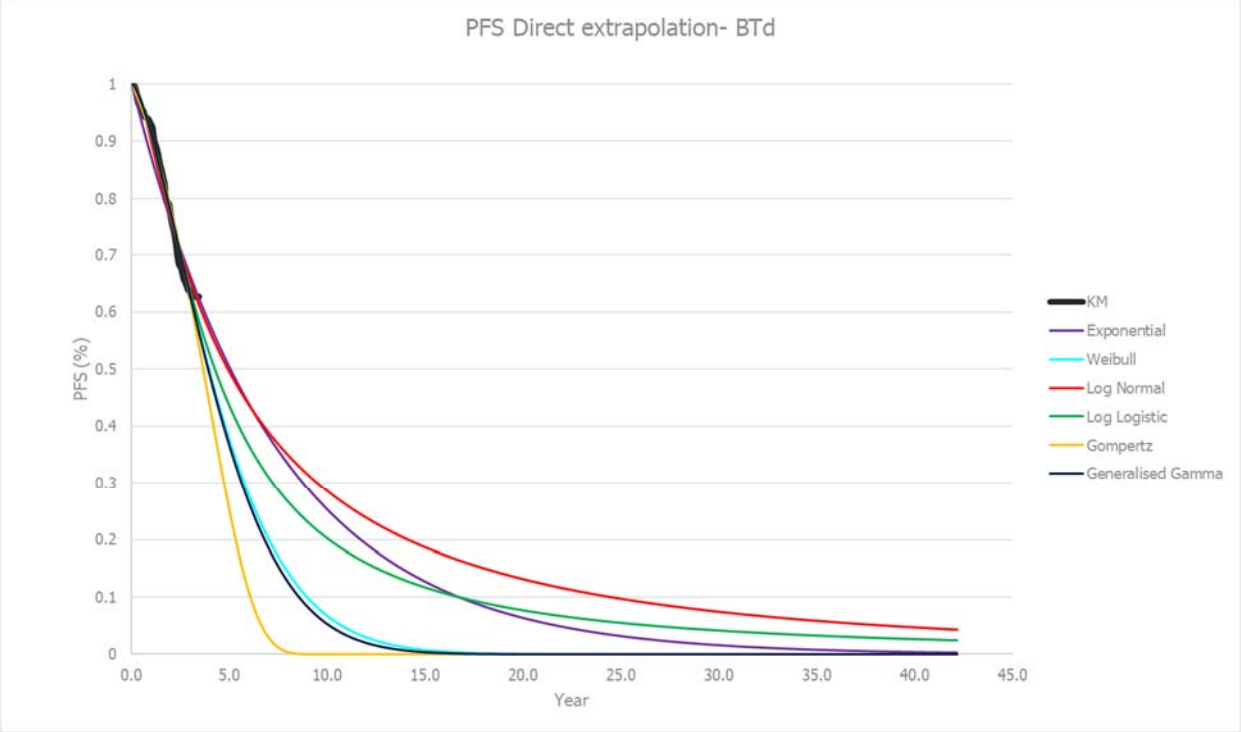
The company conclude that the wide variation in the OS predictions would translate to high uncertainty in the cost-effectiveness results (CS section B.3.3.2). They did not present cost-effectiveness estimates based on direct parametric extrapolation of OS and PFS.

ERG conclusions:

- The ERG consider that CASSIOPEIA OS data are too immature for simple extrapolation with parametric survival functions to be robust.
- Although less immature, there is also high uncertainty over parametric extrapolations of PFS.
- We consider that the use of more flexible parametric survival curves is unlikely to improve the reliability of predictions based on direct extrapolation of OS and PFS data, due to the immaturity of the data and heterogeneity of patient responses.

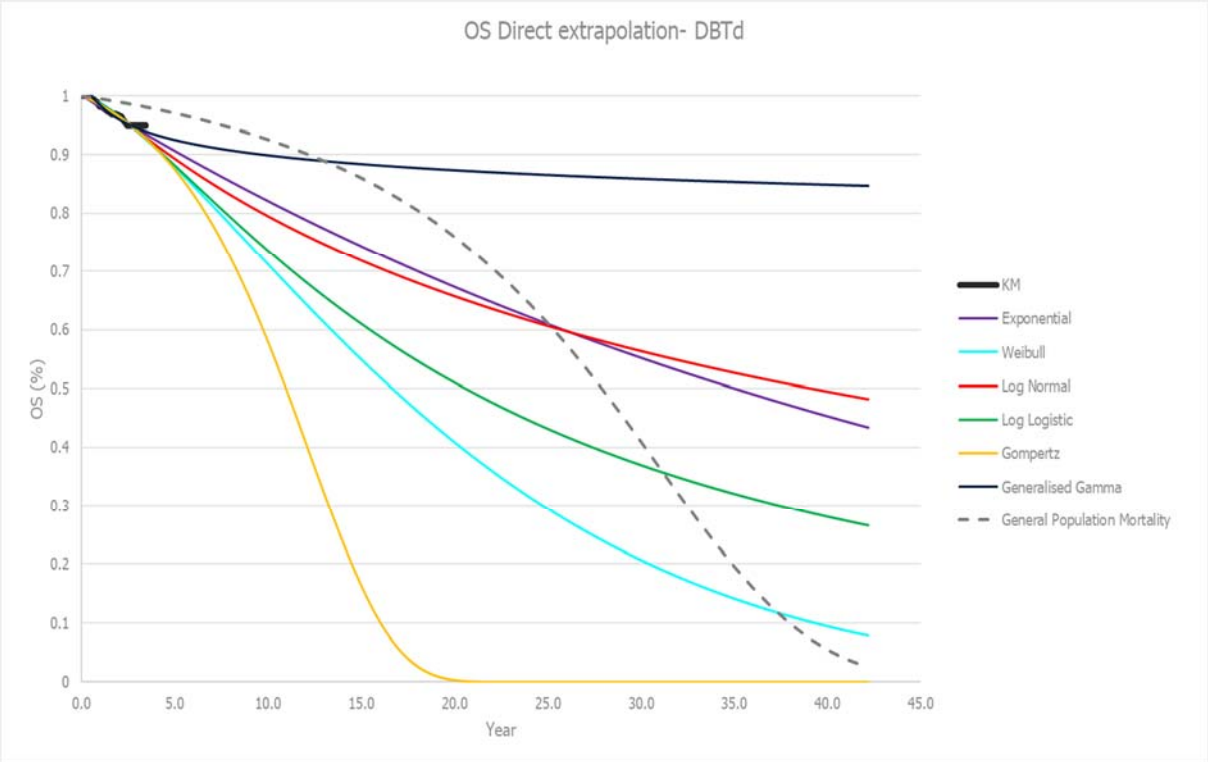
Figure 6 KM data and parametric extrapolations for PFS (CASSIOPEIA ITT population, median follow up 29.2 months)

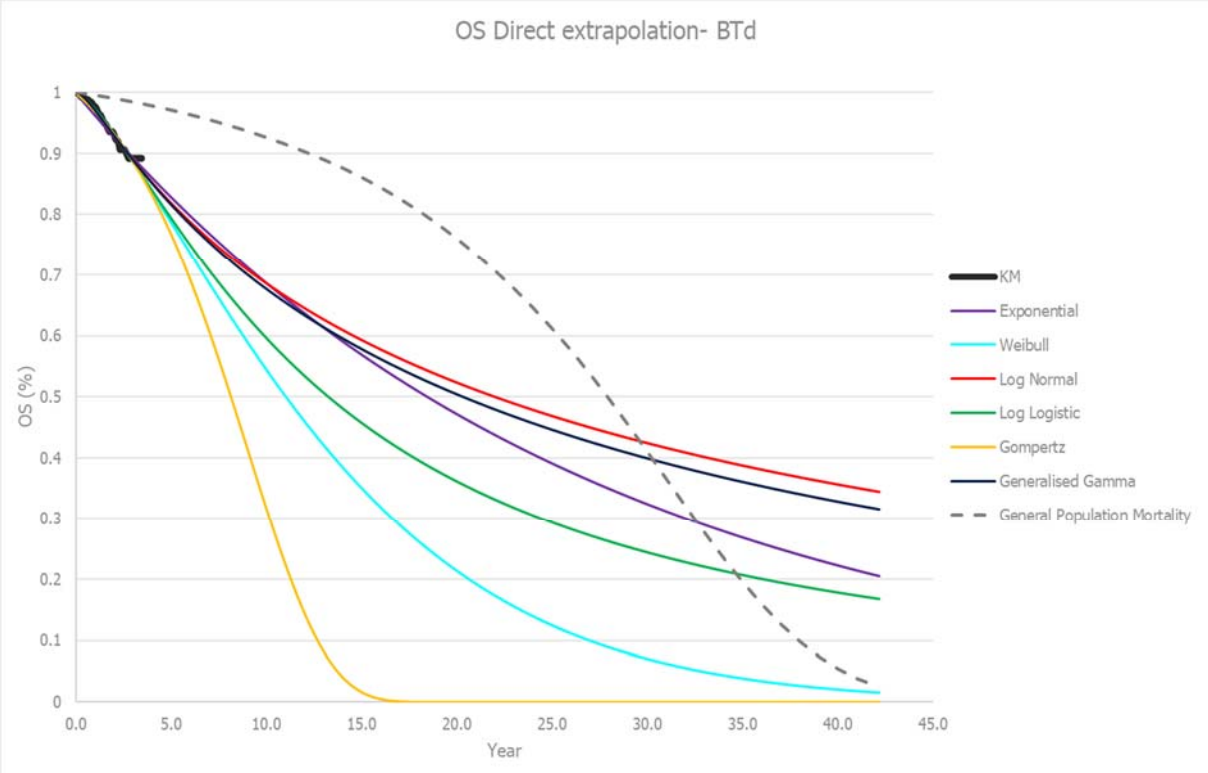




Source: Produced by the ERG from the company model

Figure 7 KM data and parametric extrapolations for OS (CASSIOPEIA ITT population, median follow up 29.2 months)





Source: Produced by the ERG from the company model

4.2.6.2 Response-based modelling of survival outcomes

4.2.6.2.1 Overview

As an alternative to simple direct extrapolation, the company adopt a response-based approach to extrapolating PFS and OS. The cohort is split into MRD-negative and MRD-positive subgroups at the 'landmark' post-consolidation assessment timepoint at the end of the eighth 28-day model cycle (approximately 100 days post ASCT).

PFS and OS in the pre-landmark period (model cycles 0-8)

Survival estimates follow PFS and OS KM data for DBTd and BTd in the pre-landmark period (up to approximately month 9 in CS Figures 12 and 16).

Proportion of MRD negative patients at the landmark timepoint

The model splits the cohort according to the proportion of the CASSIOPEIA ITT population achieving MRD negativity (as assessed with multiparametric flow cytometry at a sensitivity threshold of 10^{-5}) at the post-consolidation assessment, as shown in Table 28. We note that this MRD response assessment does not follow IMWG criteria, which require a conventional response of at least CR (CR or sCR) for MRD negativity. A lower proportion of patients in both arms were MRD-negative at this timepoint according to IMWG criteria (see sections 3.2.3.3 and 3.2.6 above).

Table 28 Proportion MRD negative at post consolidation assessment

MRD status	DBTd	BTd
MRD-negative	63.7% (95% CI: 59.5, 67.8)	43.5% (95% CI: 39.3, 47.8)
MRD-positive	36.3%	56.5%
Source: CS Table 48		

PFS and OS in the post-landmark period (model cycle 9 onwards)

From cycle 9, PFS and OS are modelled separately for MRD-negative and MRD-positive subgroups in the two treatment groups. This analysis is informed by two sources of data:

- the landmark analysis of CASSIOPEIA data (CS section B.2.6.3 and section 3.2.4.7 above); and
- the company's meta-analysis of the evidence on the relationship between MRD status and PFS/OS (CS section B.3.3.2 and CS Appendix M; for ERG critique see section 3.6 above).

The company's base case approach to modelling post-landmark PFS and OS entails three steps:

Step 1: Fitting parametric curves to post-landmark CASSIOPEIA data for MRD-positive patients in the BTd arm

Step 2: Applying hazard ratios for MRD-negative versus MRD-positive from the company's meta-analysis to obtain curves for MRD-negative patients in the BTd arm.

Step 3: Applying hazard ratios for DBTd versus BTd from the landmark analyses to obtain curves for MRD-positive and MRD-negative patients in the DBTd arm.

This base case approach is summarised in Table 29 below and we discuss the three steps in more detail in the following sections. Note that steps 2 and 3 require the assumption of proportional hazards for comparisons between MRD-positive and MRD-negative subgroups, and between DBTd and BTd arms in both MRD subgroups.

Table 29: Overview of company's approach to OS and PFS extrapolation (base case)

	BTd	DBTd
MRD positive	**Step 1**: Parametric curves fitted to post-landmark trial data (PFS Weibull & OS exponential)	**Step 3**: HRs for DBTd versus BTd from landmark analysis applied to BTd MRD-positive curves
MRD negative	**Step 2**: HRs for MRD-negative versus MRD-positive (from meta-analysis) applied to BTd MRD-positive curves	**Step 3**: HRs for DBTd versus BTd from landmark analysis applied to BTd MRD-negative curves

In addition to the base case, the company present seven scenario analyses to assess the impact of alternative assumptions for modelling clinical effectiveness data (CS Table 84):

- **CS scenario 1A** – parametric curves fitted to post-landmark trial data for BTd MRD-negative patients (Weibull for PFS and OS), see CS Appendix O. In this scenario, external data from the MRD meta-analysis is not used.
- **CS scenario 1B** – DBTd MRD-negative curves estimated by applying HRs (MRD-negative versus MRD-positive) from the meta-analysis to the DBTd MRD-positive curves. This effectively assumes that the DBTd treatment effect in the MRD-negative group is equal to that in the MRD-positive group.
- **CS scenario 2** – exponential PFS for the BTd MRD-positive reference group, rather than Weibull as in the base case.
- **CS scenario 3A/3B** – no additional DBTd effects after 5/10 years. HRs for DBTd versus BTd = 1 after 5/10 years (MRD-positive and MRD-negative)
- **CS scenario 3C/3D** – no additional DBTd effect after 5/10 years for MRD-negative group. HRs for DBTd versus BTd = 1 after 5/10 years (MRD-negative only)

4.2.6.2.2 Survival extrapolations for the BTd MRD-positive subgroup

Survival outcomes (PFS and OS) for MRD-positive patients in the BTd arm are extrapolated using post-landmark individual patient data from the CASSIOPEIA trial. The company explain that they chose BTd MRD-positive as the reference group, as most patients in the BTd arm were MRD-positive (56.5%) and this group experienced the highest number of events.

In line with NICE TSD14, the company fitted six parametric distributions (CS Figures 32 and 33, reproduced in Figure 8 below).[72] The company state that their choice of distributions was based on assessment of goodness-of-fit using Akaike information criteria (AIC) and Bayesian information criteria (BIC) and visual inspection, and clinical plausibility elicited from a panel of 3 UK clinicians (for further details, see CS section 3.3.2 and CS Tables 49, 50, 52 and 53).

For the base case, the company chose a Weibull model for PFS (exponential scenario) and exponential for OS (no scenarios). These choices are largely driven by estimates of clinical plausibility.

- Choice of PFS function: The ERG's clinical experts agreed with the company's estimates of plausible PFS (20-30% at 5 years, <10% at 10 years and <1% at 20 years). Based on these estimates, the lognormal and loglogistic predictions appear overly optimistic (8% and 13% alive after 20 years) but it is difficult to distinguish between the other functions.

- Choice of OS function: The company's clinical experts estimated that OS would not exceed 70% at 5 years or 44% at 10 years. The ERG's clinical experts gave a wide range of estimates: 5 years (ranging between 40-50% and between 70-80%) and 10 years (ranging between 35-40%, and approximately 15%), but agreed that few in this group would survive to 20 years. All of the fitted curves (except for the constant hazard exponential distribution) showed declining hazards over time. The company point out that all six parametric OS predictions exceed clinical expectations of survival for this patient population. They also note that, with the exception of exponential and Weibull, all parametric predictions exceed age-sex matched general population survival within 30 years (see Figure 8).

We note that the model sets a minimum mortality rate based on the age and sex matched general population mean, so modelled OS cannot exceed general population survival (as shown in CS Figures 34 and 35, reproduced in Figure 9 below).

Extrapolation of PFS for BTd MRD-positive (landmark analysis)

Extrapolation of OS for BTd MRD-positive (landmark analysis)

Figure 8: Company's extrapolation of OS and PFS BTd MRD-positive patients
(reproduced from CS Document B Figures 32 and 33)

ERG conclusions:

- We agree with the use of the BTd MRD-positive subgroup as the reference group for survival extrapolation, as it has more mature data than the MRD-negative subgroup. The survival extrapolations are accurately implemented in the model.
- The company's choice of PFS extrapolation for the BTd MRD-positive group is reasonable (Weibull base case and exponential scenario). We also run a Gompertz scenario, as it has similar AIC/BIC statistics to the Weibull and survival predictions within the range of clinical expectations.
- Company and ERG clinical experts agree that the OS extrapolations for the BTd MRD-positive group are unrealistically high. We agree with the company's rationale for choosing the exponential distribution for their base case, as this gives the lowest survival predictions, which are closest to clinical expectations. The company do not report any scenario analysis. For comparison, we report a scenario with Weibull OS for BTd MRD-positive. The Weibull is the next most conservative OS extrapolation (after the exponential), and we note that the company select a Weibull for the BTd MRD-negative subgroup in scenario 1A (CS Appendix O.1).
- The model limits mortality rates to be no lower than age and sex matched general population means. This is appropriate as it prevents predictions of better survival outcomes for people with MM than for the general population.

4.2.6.2.3 Survival extrapolations for the BTd MRD-negative subgroup

In the company base case, PFS and OS curves for the BTd MRD-negative subgroup are estimated from the parametric curves for the BTd MRD-positive subgroup, adjusted with hazard ratios from the MRD meta-analysis (Table 30). See section 3.6 above for the ERG critique of the MRD review and meta-analysis.

Table 30: Hazard ratios used for modelling PFS and OS

Comparison	PFS HR (95% CI)	OS HR (95% CI)
BTd MRD- versus MRD-positive	[redacted]	[redacted]
DBTd versus BTd (MRD-positive)	[redacted]	[redacted]
DBTd versus BTd (MRD-negative)	[redacted]	[redacted]
Source: CS Tables 51 and 54		

The use of constant HRs for this adjustment requires the assumption of proportional hazards. The company provide log-cumulative hazard plots (based on the landmark

analyses of the CASIOPEIA ITT data) to support this assumption (see CS Appendix N and clarification response B15).

ERG conclusions:

- The assumption of proportional hazards for MRD status appears to hold for PFS and OS, based on log-cumulative hazard plots of CASSIOPEIA data. It is not clear if proportional hazards hold for other data included in the MRD meta-analysis.

4.2.6.2.4 Survival extrapolations for DBTd MRD-positive and MRD-negative subgroups

The company applied constant hazard ratios for DBTd versus BTd to the survival outcomes (PFS and OS) throughout the entire model time horizon in both the MRD-positive and MRD-negative subgroups. The hazard ratios (Table 30) are obtained from the CASSIOPEIA landmark analysis (section 3.2.4.7 above).

We note that the confidence intervals for the DBTd versus BTd HR estimates cross 1 for PFS and OS in the MRD-positive subgroup, and also for OS in the MRD-negative subgroup. This indicates uncertainty over the additional treatment effects for DBTd. Company sensitivity analysis indicates that the model is very sensitive to the HRs for additional treatment effects on OS (see tornado plot in CS Figure 38). Company scenario 1B also shows that the model is sensitive to assuming that the additional treatment effects in the MRD-negative subgroup are the same as in the MRD-positive subgroup.

The assumption of proportional hazards for treatment comparisons in the landmark analysis is supported by log-cumulative hazard plots in CS Appendix N. The base case assumption that these HRs are maintained throughout the modelled time horizon is more difficult to justify. Company scenarios 3A to 3D with treatment waning (HR=1) after 5 and 10 years increased the ICER considerably.

- The company use hazard ratios estimated from the landmark analysis of CASSIOPEIA to adjust OS and PFS extrapolations for DBTd relative to the fitted extrapolations for BTd in both MRD subgroups. Proportional hazards assumptions do appear to hold for these comparisons, based on the log-cumulative hazard plots reported in the CS and their clarification response.
- We agree with the inclusion of these additional treatment effects regardless of MRD status in the base case analysis, as they reflect best available data. However, we

note that the landmark analysis is exploratory and that the confidence intervals for the DBTd versus BTd HRs span 1, except for PFS in the MRD-negative subgroup. This indicates uncertainty over these effects and the model is highly sensitive to the magnitude of these additional treatment effects on overall survival. We therefore conduct additional scenario analysis to demonstrate the impact of alternative assumptions.

- The company base case includes constant hazard ratios for DBTd versus BTd in both MRD subgroups, applied throughout the time horizon. This is a strong assumption in the absence of longer follow up. We therefore apply an assumption of waning in the ERG base case and scenarios.

4.2.6.3 Mortality

General population mortality, based on life tables for England 2016-2018 adjusted for age and gender, were used to cap OS in the model. This adjustment reflects an expectation that the risk of mortality is higher in people with MM than for people of the same age and gender in the general population.

4.2.6.4 Adverse events

The economic model includes Grade 3 and 4 adverse events that were reported in at least 5% of the patients in the CASSIOPEIA trial, as summarised in CS Table 57. The associated costs and disutilities are applied in the first model cycle.

The company report that 3.5% of patients experienced a Grade 3 or 4 infusion related reaction (IRR) associated with daratumumab in CASSIOPEIA, with only 0.6% discontinuing treatment (CS Table 42). They argue that the incidence of IRRs in the COLUMBA trial was substantially reduced with SC injection compared with IV infusion[55,71] (section 3.3.3 above).

ERG conclusions:

The AEs were correctly implemented in the model. Apart from those included by the company, the ERG's clinical experts suggested that other events, including diarrhoea, upper respiratory tract infections and infusion related reactions may occur in MM patients. We note that these AEs do not meet the company's threshold for inclusion in the economic model (5% or greater incidence of grade 3 or 4), and that they are unlikely to affect the cost-effectiveness results.

4.2.7 Health related quality of life

The health state utilities used in the economic model are based on EQ-5D-5L data collected in the CASSIOPEIA trial and a published study that has been used in previous TA311 (CS B.3.4). Values used in the base case and scenario analyses are reported in CS Tables 60 and 61 respectively.

We note a small discrepancy in reporting of the duration of induction therapy (CS Table 59), which they address in their response to ERG clarification question B20. This did not influence cost effectiveness results.

4.2.7.1 Health related quality of life studies

The company conducted a systematic literature review to identify HRQoL, patient-reported outcomes and utility for patients with newly diagnosed MM (further details in CS Appendix H). Of the identified studies, only one reported utility values in the patient population of interest. This US based study by Abonour et al. reported EQ-5D in patients who received ASCT and did or did not receive maintenance. For further details, see CS section B.3.4.3. This study reported the following utilities: 0.75 at baseline; 0.79 at pre-ASCT; 0.83 during follow-up from 100 days post ASCT; and 0.79 at disease progression. We note that these values are consistently higher than those reported in the CASSIOPEIA trial.

4.2.7.2 Utility for progression-free survival

Patients randomised in CASSIOPEIA were asked to complete the EQ-5D-5L questionnaire at three timepoints: baseline, Cycle 4 Day 28 (end of induction) and Day 100 post-ASCT (end of consolidation). The EQ-5D-5L dimension scores were mapped to the EQ-5D-3L UK value set, using the van Hout et al. 'crosswalk' methods, as recommended by NICE.[66] The reported utility values for the two treatment arms at these time points are similar (CS Appendix L Tables 68 and 69). Because of this similarity in values, the company pooled EQ-5D-5L data across the treatment arms.

The utility values for the PF health state in the economic model are shown in Table 31. Baseline utility was applied for model cycles 0-3; utility assessed at day 28 of induction cycle 4 was applied for post-induction to post-consolidation response (model cycles 4-8); and utility assessed 100 days after ASCT was applied post consolidation for the remainder of the time horizon for patients who remained progression-free (model cycles 9 onwards).

4.2.7.3 Utility for progressed disease

Utility for the PD health state was obtained from published literature. In TA311, utility values were derived from van Agthoven et al. which reported values of 0.69 for second and third lines of treatment and 0.64 for further treatment lines. The value of 0.69 used in the current appraisal is derived from averaging utility values from second and third lines, as in TA311.

The company conducted a range of scenario analyses with respect to alternate utility values for PFS and PD health states (presented in CS Table 61). Changing the utility values did not influence the overall cost effectiveness results, as shown in CS Table 84.

Table 31 Utilities used in the base case model

Health state		Model cycle	Utility Mean (SD)	Source
Progression free	Induction therapy	0-3	0.57 (0.31)	CASSIOPEIA Baseline
	Post-induction to post-consolidation response	4-8	0.68 (0.22)	End of induction
	Post-consolidation	9+	0.73 (0.17)	Response assessment
Progressed disease			0.69 (-)	Van Agthoven et al. (2004), TA311[27,65]
Source: CS Table 60				

4.2.7.4 Adjustment of utilities for age

The company adjusted for age using EQ-5D population norms from the Health Survey for England 2008.[73] These adjustments are reproduced below in Table 32. In their response to ERG clarification question B16, the company state that these utility adjustments are applied by dividing the utility for current age by the age at entry into the model (to align with the mean age of the sample informing health state utility values), and multiplying this adjustment factor by the health state utility value. This applies a reduction in the health state utility with increasing age.

Table 32 Age adjustment utilities in the company's model

Age	Value
18-24	0.929
25-34	0.919
35-44	0.893
45-54	0.855
55-64	0.810
65-74	0.773
75+	0.703
Source: Company's economic model	

4.2.7.5 Adverse event related disutility

The economic model incorporated utility decrements for patients who experienced adverse events associated with induction therapy, ASCT and consolidation therapy. The utility decrements used in the economic model are derived from sources used in previous TAs (see CS Table 62). Duration of disutility was assumed to last one cycle of induction therapy, which is consistent with assumption in TA510.

ERG conclusions:

- Overall, the company's approach to estimating utilities is reasonable. We agree with use of EQ-5D data from the CASSIOPEIA trial for pre-progression utilities, and the van Agthoven et al. (2004) utility for progressed disease as in TA311.[27,65] The company conducted scenario analyses with pre-progression utilities from TA311 and alternative estimates for progressed disease. For completeness, we conducted a scenario analysis with the utilities reported by Abonour et al.[74]
- It is appropriate to adjust utilities for age, although the age categories used are broad (55-64, 65-74, 75+). This is potentially important for our analysis in which we change the initial age of the cohort to reflect PHE data. It is notable that whilst changing utility values for PF and PD health states has little effect on the ICER, the utility age adjustments are among the top 10 parameters in the tornado plot (CS Figure 38). We think this is misleading, as it is based on a wide variation of 20%. Nevertheless, we tested the impact of using finer-grained utility age adjustments with our analysis in which we changed the initial age of the cohort in line with PHE data.

4.2.8 Resources and costs

The economic model includes estimates of costs (discussed in CS Document section B.3.5) for: drug acquisition and administration for induction/consolidation and subsequent therapies; concomitant medication for induction/consolidation therapies; ASCT; monitoring; and management of adverse events. The CS reports a systematic literature review conducted to identify resource use and costs (Appendix G and Appendix I).

4.2.8.1 Drug acquisition costs

The dosing regimens for induction and consolidation therapies are summarised in CS Table 64; the costs of administration of the individual therapies are presented in CS Tables 65-66; and the total cost of induction and consolidation therapy applied in the economic model in CS Table 67. The dosing regimens are based on treatment protocols in the CASSIOPEIA trial, except for daratumumab. In their base case the company apply fixed-dose SC

daratumumab rather than the weight-based IV formulation. They argue that the EMA have accepted non-inferiority for SC versus IV, based on the COLUMBA trial.[55,71] The company conducted a scenario analysis using the weight-based formulation of daratumumab which increased the base case ICER.

In their submission, the company assume no vial sharing, rounding up the number of vials required per administration to the nearest integer based on the mean weight of the population. In response to a clarification question, the company revise cost estimates for weight based IV daratumumab using the distribution of weight in the trial population (CR B17). This causes a small increase in the ICER with IV daratumumab (scenario 4). The company note that vial usage for bortezomib has not been revised, as the distribution of BSA in the CASSIOPEIA population is not reported in the CSR.

We also note discrepancies between the CASSIOPEIA protocol and bortezomib SmPC in the dosing of thalidomide and dexamethasone, although these are minor and make little difference to costs (explored in the company's scenario 6, CS Table 84). It is not clear if these differences would have impacted on effectiveness results from the trial, but this was not a source of concern for clinical experts consulted by the ERG.

4.2.8.2 Concomitant medication costs

The resource use and monitoring costs of concomitant medications included in the economic model are summarised in CS Table 68. These are applied as a single cost in the first cycle of the model, along with the main treatment costs.

4.2.8.3 ASCT

The costs of ASCT included in the economic model are presented in CS Table 69. These are applied to those patients who receive ASCT following induction therapy (shown in CS Table 47). The included costs are consistent with those used in previous TA311, except that in CASSIOPEIA, a proportion of patients in both the treatment (20.3%) and comparator arms (7.2%) also received plerixafor. The associated cost of plerixafor are included in the economic model. All the remaining costs between DBTd and BTd are same.

4.2.8.4 Subsequent therapies

In the economic model, the cost of subsequent treatments is implemented as a single, per-cycle cost that is applied across all the cycles for patients in the PD health state. The total cost of each treatment was based on:

- Assumptions about the proportion of patients receiving different regimens at second, third and fourth line (shown in CS Table 70)
- Median Time to Progression (TTP) (or PFS) obtained from pivotal clinical trials of each regimen (shown in CS Table 71)

The per cycle cost of subsequent treatments is then estimated as the sum of total costs of treatments at second, third and fourth lines divided by the mean time spent in the model in the PD state. This means that in the company's model, all surviving patients are assumed to receive subsequent therapy at second-, third- and fourth lines following progression.

The company exclude therapies recommended via the Cancer Drugs Fund (CDF) from their base case, which is consistent with the NICE position statement. Results including CDF treatment regimens are reported in CS Scenario 5.

To reflect clinical practice, the company assume that some patients (10% in DBTd versus 8% in the BTd arm respectively) undergo a second ASCT following progression. Clinical experts consulted by the ERG agreed that a higher proportion in the DBTd arm is likely, given the deeper response and more prolonged remission achieved with DBTd than with BTd. The company assume that patients undergoing a second ASCT would receive BCd as induction therapy for 3 cycles of 21 days. Our experts agreed that this is a likely reinduction schedule.

The treatment cost of second ASCT is estimated as the total cost of BCd induction therapy and ASCT divided by the median PFS for second ASCT. The company report use of a similar approach to estimate costs for Bd at second line and PBd at third line. Further description of the company's assumptions is given in CS section B.3.5.1. Dosing regimens of subsequent therapies used in the economic model are shown in CS Table 72. The unit costs and total costs per administration associated with the individual therapies included within the subsequent treatment regimens are described in CS Table 73 and CS Table 74 and the average costs per model cycle are presented in CS Table 75 and CS Table 76.

4.2.8.5 Administration costs

Administration costs for first line and subsequent treatments in the economic model are presented in CS Table 77. The company made the following assumptions:

- A one-off cost on treatment initiation for oral chemotherapy regimens (i.e. thalidomide and dexamethasone);

- Costs for each administration for therapies administered via SC injection (i.e. daratumumab and bortezomib);
- Cost of only one administration on days where both bortezomib and daratumumab are administered by the same nurse;
- Cost of a blood test prior to the first administration of daratumumab was included in the cost of administration for DBTd.

In their scenario analysis for the IV formulation of daratumumab, the administration cost is based on the cost of delivering complex chemotherapy. Furthermore, on days when both bortezomib and daratumumab are administered together, the economic model includes only the higher cost of IV infusion.

4.2.8.6 Monitoring costs

Estimates of the type and frequency of monitoring visits and tests are based on TA573. CS Table 78 reports the estimated proportions of NDTE MM patients who have monitoring tests per 4-week cycle while:

- 'on treatment' (i.e. from the start of induction to 100 days post ASCT [end of consolidation], and for subsequent treatments) and
- 'off treatment' (i.e. progression free after the first 100 days post ASCT).

The ERG's clinical experts identified some inconsistencies. They noted that UK clinical practice is not to measure urinary light chain excretion (as assumed by the company), which is an expensive procedure. Secondly, in clinical practice, patients 'on treatment' (i.e. from the start of induction to 100 days) are likely to have blood tests at each cycle, followed by a break during the stem cell harvest and transplant, and then would have 1 or 2 full sets of tests at day 100. For patients who are off treatment, the blood tests are likely to be every 3 months.

4.2.8.7 Adverse event costs

Costs associated with adverse events are summarised in CS Table 79. These are applied to patients who experience the events (shown in CS Table 57) and to only the first model cycle.

ERG conclusions:

- The company's approach to estimation of drug acquisition costs is reasonable. The ERG spotted some inconsistencies in values reported in the CS and those in the model. The company confirmed that values in the model are correct, so this does not impact on cost-effectiveness results (clarification responses B18 and

B19)

- The ERG also spotted an error in the calculation of average cost per model cycle for the CDF treatments DBd and daratumumab monotherapy (CS Table 76), which the company correct in their clarification response. This has a small effect on the results for Scenario 5 but does not impact the base case results or other scenarios (which exclude CDF drugs).
- Clinical experts advised the ERG that the PBd regimen is not currently used at third or fourth line. For the ERG analysis, we therefore assume that all patients receive the other available regimens: Ld at third line and Pd at fourth line.
- Estimates of administration, monitoring and adverse event costs are consistent with previous MM appraisals. Experts consulted by the ERG suggested that some of the company's assumptions about monitoring do not reflect current clinical practice. These differences would have minimal impact on cost estimates in the model.

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company report their deterministic base case results in CS Table 82. This analysis is conducted with list prices for all drugs. A Patient Access Scheme (PAS) with a simple price discount for daratumumab has been agreed. We report cost-effectiveness results including this discount in Table 33 below, and in subsequent results tables in this report. All comparator, concomitant and subsequent treatments are costed at list price in this report. We present results with all available PAS/CMU price discounts in a confidential addendum to this report. The incremental cost effectiveness ratio (ICER) for DBTd versus BTd is [redacted] per QALY gained, with the PAS discount for daratumumab and all other drugs costed at list price.

Table 33 Cost effectiveness: deterministic company base case

Intervention	Total			Incremental			ICER
	Costs	LYs	QALYs	Costs	LYs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	[redacted]	-	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: estimated from model by ERG, PAS discount for daratumumab, list price for other drugs							

Disaggregated results are reported in CS Appendix J. The base case model predicts that patients treated with DBTd spend a mean 6.4 years longer before progression than patients treated with BTd (means of 147 and 70 months respectively, CS Appendix Table 56). Mean time with progressed disease is similar for DBTd and BTd (124 versus 127 months). Overall survival (undiscounted) is 6.1 years longer for DBTd with 60% of the cohort predicted to be alive after 10 years, compared with 36% in the control group (CS Appendix Table 57). Total costs are [redacted] higher for DBTd than BTd. This is caused by higher costs accrued before progression [redacted], which are offset to some extent by lower costs after progression (due delayed progression and discounting). Excess pre-progression costs for DBTd are largely attributable to the price of daratumumab (CS Appendix Tables 58 and 59).

5.2 Company's sensitivity analyses

5.2.1 Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA), with input parameter distributions as reported in CS Appendix P, Table 89. The company's probabilistic base case results are reported in CS Table 83. The cost-effectiveness scatter plot and acceptability curve are shown in CS Figures 36 and 37.

We consider that some of the distributions in the company's PSA are inappropriate: i.e. a normal for median time to progression for subsequent treatments and for the hazard for BTd MRD-positive OS. In addition, a simple percentage variation (e.g. 10% or 20%) is assumed for many parameters when empirical data to estimate variance is available: for example, for the MRD response rates and proportions of patients who undergo ASCT, for which confidence intervals are available.

5.2.2 Deterministic sensitivity analysis

The company report results from their one-way, deterministic sensitivity analysis in the tornado plot in CS Figure 38. This has similar limitations as the PSA, with variation for most input parameters based on simple assumed percentages rather than empirical evidence. This applies to the response rates and utility age adjustment parameters in the tornado plot, as well as resource use and cost parameters. However, the ranges tested for the OS extrapolation parameters for the BTd MRD-positive group, hazard ratios for MRD, and treatment effects and utilities estimated from the trial are based on 95% confidence intervals, so sensitivity analysis around these parameters is more meaningful. This indicates that the ICER is very sensitive to the OS treatment effects and extrapolation.

ERG conclusions:

- The company's deterministic and probabilistic sensitivity analyses do not provide an accurate reflection of parametric uncertainty because the variance assumed for many of the input parameters is not based on the available evidence.
- The ERG has revised these analyses to include the PAS for daratumumab and to provide a better reflection of uncertainty around the input parameters (see Appendix 10.4 below).

5.2.3 Scenario analysis

The company report 14 scenario analyses (CS Table 84). We present results including the PAS for daratumumab and all other drugs at list price in Table 34 below.

Table 34: Results from the company's scenario analyses, deterministic analysis

Scenario		Intervention	Costs	QALYs	ICER (£/QALY)
Company base case		BTd DBTd	[redacted]	[redacted]	- [redacted]
Extrapolation of baseline survival curves					
CS2	BTd MRD+ PFS extrapolation (Exponential)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Relative survival MRD-negative versus MRD-positive					
CS1A	BTd MRD- extrapolations (Weibull for PFS & OS)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Treatment effects (DBTd versus BTd) in MRD-negative and MRD-positive groups					
CS1B	DBTd MRD- extrapolations from MRD meta-analysis HR	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS3A	No additional DBTd effect after 5 years (MRD+ and MRD-)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS3B	No additional DBTd effect after 10 years (MRD+ and MRD-)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS3C	No additional DBTd effect after 5 years (MRD- only)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS3D	No additional DBTd effect after 10 years (MRD- only)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Utilities					
CS7	With vial sharing	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS8A	PD utility 0.644 from van Agthoven et al. (2004) (TA311)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS8B	PD utility 0.57 from Palumbo et al. (2013) (TA510)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS8C	PF and PD utilities from van Agthoven et al. (2004)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Resource use and costs					
CS4	Daratumumab IV formulation	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS5	With CDF subsequent therapies	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS6	Dosing for BTd based on bortezomib SmPC	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS7	With vial sharing	BTd DBTd	[redacted]	[redacted]	- [redacted]

Source: estimated from model by ERG, PAS discount for daratumumab, list price for other drugs
MRD- : MRD-negative; MRD+ : MRD-positive

With the PAS (and list prices for other drugs) [redacted] The ICER is most sensitive to the additional treatment

effects for DBTd estimated from the landmark analysis. Scenario 1B (which implicitly assumes the same treatment effect for DBTd versus BTd in the MRD-negative group as in the MRD-positive) and scenarios 3A to 3D (waning of treatment effects) give higher ICERs than the base case. The higher cost of IV daratumumab (scenario 4) causes a moderate increase of the ICER and the inclusion of costs for CDF treatments at second, third and fourth lines of treatment (scenario 5) DBTd dominates BTd.

We report additional ERG scenario analyses in section 6 below.

5.3 Model validation and face validity check

5.3.1 Processes for model checking

The company approach to validation is described in CS section B.3.10. This included:

- Quality control by an analyst not involved in model development (checking of formulae and extreme value 'stress testing');
- Assessment of clinical plausibility of PFS and OS extrapolations by UK clinicians;
- Internal validity: comparison of model results with outputs from the CASSIOPEIA trial, which was used for parameter estimation.
- External validity: comparison of model results with external data from the PHE cohort.

The ERG have also conducted a series of quality checks of the company model. This included: checking that the input parameters in the model matched the values cited in the CS and in the original sources; and validating the results of the scenario and sensitivity analyses as reported by the company. We also conducted a series of 'white box' and 'black box' checks to validate the model. We spotted a few inconsistencies between parameters in the model and values reported in the CS; these have been described in our critique above. We corrected the cost calculations for CDF daratumumab and DBd subsequent treatments, which affected the company's scenario 5. Other inconsistencies did not affect any results.

As noted above (section 5.2.1), we consider that that the deterministic (tornado) and probabilistic sensitivity analyses do not provide an accurate reflection of parameter uncertainty, as variation for many parameters was based on a simple assumed percentage (10% or 20% around the mean), rather than empirical evidence sources (see CS Appendix P, Table 89). We have revised these analysis to reflect available evidence of parameter uncertainty (see Appendix 10.4 below).

5.3.2 Face validity

To check plausibility of the survival inputs used in the model, the company obtained feedback from four UK based clinicians. Furthermore, three of these clinicians are reported to have validated the MM treatment pathway in the UK and the generalisability of the CASSIOPEIA trial. The CS states that long-term survival extrapolations were based on-statistical goodness of fit, visual inspection, real-world evidence of outcomes for UK patients, and clinical experts' opinion. The clinical experts consulted by the ERG agree with the MM treatment pathway and how this is encapsulated in the company model structure. However, overall survival predictions from extrapolated trial data for MRD-positive patients treated with BTd were higher than expected. There remains high uncertainty over the long-term survival extrapolations, as discussed in the previous sections.

5.3.3 Internal validation: comparison of survival predictions with CASSIOPEIA data

For internal validation, the company explain that model programming was quality assessed by an analyst who was not involved in the original model development. They also state that the efficacy and cost of BTd consolidation therapy in the economic model reflect the CASSIOPEIA trial protocol. This is true, except that the cost of daratumumab in the company base case assumed use of the fixed dose SC formulation, rather than the weight based IV infusion as in the trial (see section 4.2.4 above).

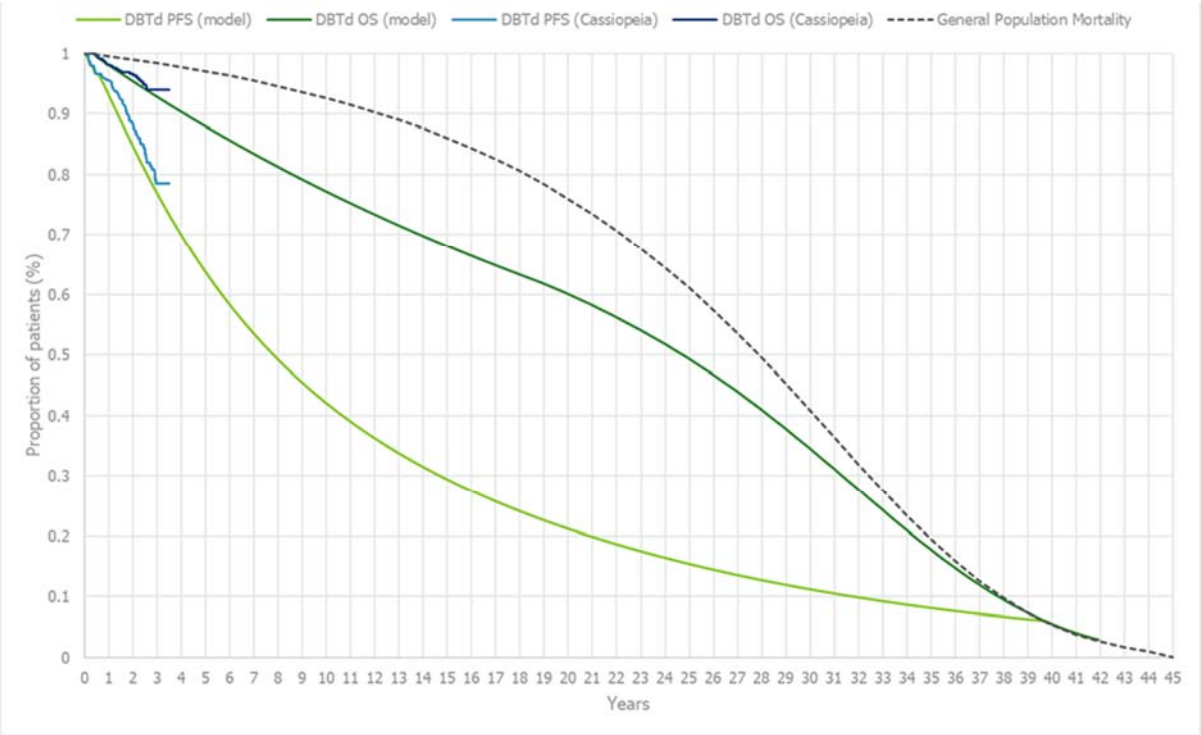
The company provide a comparison of the modelled survival estimates with observed data from the CASSIOPEIA trial (CS Figures 34 and 35 and CS Table 55, reproduced below in Figure 9 and Table 35). We note that these results are for the MRD-positive and MRD-negative combined cohort, weighted by the proportion of patients achieving post-consolidation MRD negativity in the trial. The modelled OS estimates also restrict the mortality rate to the age-sex matched rate for the general population. Whilst the model OS estimates are comparable with those in the CASSIOPEIA trial, the model PFS estimates are consistently lower than in the trial. The company argue that this underestimation of PFS was consistent across both arms and is unlikely to influence the relative treatment effect or model results.

Table 35 Comparison of survival predictions from the model and CASSIOPEIA

Timepoint	DBTd		BTd	
	Model	CASSIOPEIA	Model	CASSIOPEIA
Overall Survival				
1-year	98%	98%	97%	98%
2-years	95%	97%	92%	93%
3-years	93%	-	87%	-
Progression Free Survival				
1-year	93%	95%	89%	93%
2-years	85%	88%	74%	77%
3-years	77%	-	61%	-

Source: Reproduced from CS Table 55

DBTd



BTd

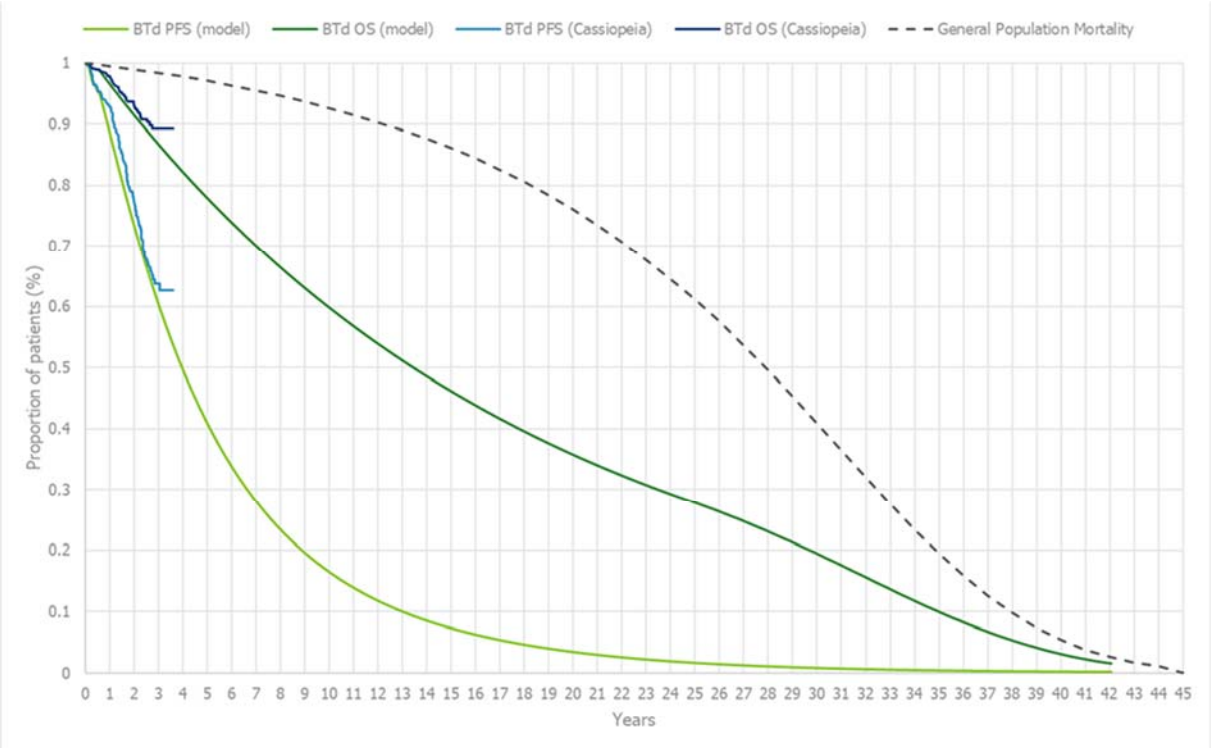


Figure 9 Comparison of OS and PFS predicted by the model with the CASSIOPEIA (MRD-positive and MRD-negative combined)
Source: Reproduced from CS Figures 34 and 35

5.3.4 External validation: comparison of survival predictions with PHE data

The company also validated the model findings against an external dataset from Public Health England (PHE). This is described in CS section B.3.3.2. The cohort included patients diagnosed with MM in England from January 2015 to December 2018, including [redacted] patients who underwent ASCT. Median follow up in the ASCT group was [redacted] months. Most of the patients with ASCT received BTd ([redacted]) as first line treatment, followed by BCd ([redacted]) and Bd ([redacted]).

Comparison of patient characteristics from the CASSIOPEIA BTd arm and PHE cohort

Patient characteristics from the PHE dataset are compared with those for the BTd arm of the CASSIOPEIA trial (CS Table 56). The mean age of patients in the CASSIOPEIA BTd arm (56.7 years; N=542) was slightly lower than that in the PHE dataset ([redacted] years; N=[redacted]). In particular, we note that [redacted] of the PHE cohort patients were aged 65 years and above; patients in this age category were not included in the CASSIOPEIA trial. Furthermore, there are differences in patient distributions across the two datasets across the different ECOG scores and ISS stages. We present a scenario analysis with age based on the mean for the PHE cohort (see section 6.1 below).

Comparison of modelled survival estimates with PHE data

A comparison of the survival rates from the PHE dataset with the model predictions from extrapolation of trial data are presented in CS Table 55, reproduced below in Table 36.

Table 36 Comparison of survival predictions from the model and PHE cohort

Timepoint	OS			PFS		
	Model BTd	PHE all 1L [redacted]	PHE BTd 1L [redacted]	Model BTd	PHE all 1L [redacted]	PHE BTd 1L [redacted]
1-year	97%	[redacted]	[redacted]	89%	[redacted]	[redacted]
2-years	92%	[redacted]	[redacted]	74%	[redacted]	[redacted]
3-years	87%	[redacted]	[redacted]	61%	[redacted]	[redacted]
Source: Reproduced from CS Table 55						

The model OS estimates are comparable with the PHE cohort (for BTd and all first-line therapy), but modelled PFS estimates are higher than in the PHE cohort, despite the model underestimating PFS in the trial. The company argue this was likely due to the impact of consolidation and maintenance treatment in the trial, which are not funded by NHS England. If so, this does raise a question of whether the model results are transferable to the clinical population of interest.

5.3.5 Cross validation: comparison with TA311 BTd extrapolation

We compare the modelled QALY and life year estimates for BTd from the current appraisal with those from the previous NICE appraisal of bortezomib for MM TA311 (Table 37). The estimates in the current appraisal are consistently higher than those reported for TA311 (from the published pre-meeting briefing), despite the similar patient population in both the appraisals. The large difference in the estimates may be due to different modelling approaches in the two appraisals (i.e. a Markov state transition cohort model in TA311 versus a response-based partitioned survival model in the current appraisal) and/or different data sources.

Table 37 Comparison of modelled outcomes for the BTd arm

Source (time horizon)	QALY (discounted)	LYs (discounted)
Current appraisal (lifetime)	[redacted]	[redacted]
TA311 (lifetime)	4.02	6.03
LY: life years		

ERG conclusions:

- Overall survival predictions from the company base case model are consistent with the CASSIOPEIA trial data and with PHE data over a period of 1-3 years from diagnosis. However, modelled PFS predictions over this period exceed the observed rates in the PHE cohort, despite underestimating PFS from the trial. The company suggest that the higher rates of PFS in CASSIOPEIA may be due to the use of consolidation and maintenance therapy in the trial that is not currently funded by NHS England.
- Clinical opinion suggests that longer-term survival extrapolations from trial data are unrealistically high for the BTd MRD-positive group. Overall estimates from the model (for MRD-negative and MRD-positive patients) suggest a median survival with current treatment (BTd) of 13 years. This increases to nearly 25 years in the modelled DBTd arm. Estimated life years and QALYs for BTd from the current model are much higher than from the model used in the previous appraisal of bortezomib in this indication TA311. The reason for this difference is not clear.
- These validation results indicate high uncertainty over the modelled survival estimates. This is not surprising given the immaturity of the trial data. Although the model includes external data on relative survival for MRD-negative versus MRD-positive, it is not informed by other sources of data on long-term absolute survival rates.

6 ERG ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the ERG

Table 38 below summarises ERG conclusions from our critique of the company's cost-effectiveness analysis and explains our reasons for conducting additional analyses.

Table 38 Additional ERG scenario analyses

Issue	Summary	ERG analyses
The mean age of patients in the company's economic analysis does not reflect the target population	The company assume a mean age of 56.6 years, as in the CASSIOPEIA trial. This does not reflect NHS practice as the trial excluded patients over the age of 65. PHE data indicate that [redacted] of patients with newly diagnosed transplant eligible MM are aged 65 years or older.	**ERG Scenario 1**: mean age from PHE dataset ([redacted] years)
OS data from the trial are very immature and yield a wide range of extrapolations for standard care (BTd) which do not reflect clinical expectations	Parametric survival extrapolations for the BTd MRD-positive group exceed clinical expectations, with constant or declining hazards that reach general population rates within 30 years. We agree with the company's decision to use the most conservative (exponential) extrapolation for the BTd MRD-positive group. It is also reasonable to estimate the MRD-negative curve relative to MRD-positive with a constant hazard ratio, as the proportional hazards (PH) assumption appears to hold between MRD subgroups in CASSIOPEIA. For comparison, we ran a scenario with a Weibull OS extrapolation for the BTd MRD-positive group. This is the next most conservative extrapolation and is consistent with the distribution for BTd MRD-negative patients in company scenario 1A. Weibull has a decreasing hazard for MRD-positive and a slowly increasing hazard for MRD-negative.	**ERG Scenario 2**: Weibull OS for BTd MRD+ (fixed HR for MRD- versus MRD+)
There is some uncertainty over PFS extrapolations for BTd, although differences between alternative functions are not large	The company chose a Weibull distribution for BTd MRD-positive PFS in their base case (with constant HRs for MRD and treatment comparisons). **Company Scenario 2** uses an exponential PFS, which is slightly more favourable than the Weibull. The PH assumption does appear to hold between MRD subgroups for PFS. Generalised gamma gives similar projections to the Weibull. Gompertz projections are less	**ERG Scenario 3**: Gompertz PFS for BTd MRD+ (fixed HR for MRD- versus MRD+)

Issue	Summary	ERG analyses
	favourable than the exponential, with decreasing hazards for both MRD-positive and MRD-negative groups. We agree that lognormal and loglogistic extrapolations are unrealistically optimistic.	
Uncertainty over the reliability of the company's MRD meta-analysis adds to uncertainty over cost-effectiveness.	In the base case, the company models BTd OS and PFS for the MRD-negative group using HRs from the MRD meta-analysis. The PH assumption appears to hold for the comparison of MRD subgroups based on log-cumulative hazard plots from CASSIOPEIA. It not clear if the PH assumption holds for other data in the meta-analysis. **CS Scenario 1A** suggests that the ICER is not very sensitive to using survival curves for the MRD-negative group fitted directly to trial data (instead of using the meta-analysis HRs). This scenario uses a Weibull distribution for MRD-negative OS (rather than exponential as in the base case). The MRD meta-analysis supplements immature trial data with external evidence but the ERG have some concerns about the reliability of the SLR (see Table 26 above). We illustrate the impact of uncertainty with additional scenario analyses.	**ERG Scenario 4**: MRD HRs at lower 95% limits: PFS [redacted] OS [redacted] **ERG Scenario 5**: MRD HRs at upper 95% limits: PFS [redacted] OS [redacted] **ERG Scenario 6**: MRD HRs from CASSIOPEIA: PFS [redacted] OS [redacted] **ERG Scenario 7**: BTd MRD- curves fitted to CASSIOPEIA data (PFS Weibull and OS exponential, as in base case)
Uncertainty over additional effects of daratumumab from the landmark analysis	The company base case uses HRs from the CASSIOPEIA landmark analysis to model OS and PFS extrapolations for DBTd relative to BTd in each MRD group. This assumes ongoing treatment benefits, in addition to the effect on MRD response. **CS Scenario 1B**, relies on MRD meta-analysis results to model OS and PFS for DBTd MRD-negative relative to DBTd MRD-positive. This implicitly assumes that the additional effects of DBTd versus BTd in the MRD-negative group are the same as in the MRD-positive group. We note that the landmark HRs for DBTd versus BTd do not reach statistical significance, except for PFS in the MRD-negative group. Given the magnitude of the landmark HRs, we consider it appropriate to include them in the model, but we report additional ERG scenario analyses to illustrate their impact.	**ERG Scenario 8**: Landmark effects only for MRD- (HR=1 for MRD+ PFS and OS) **ERG Scenario 9**: Landmark effects only for MRD- PFS (HR=1 for MRD- OS and for MRD+ PFS and OS) **ERG Scenario 10**: No landmark effects (HR=1 for OS/PFS MRD+ and MRD-)

Issue	Summary	ERG analyses
Waning of treatment effects	The model relies on survival data from CASSIOPEIA with median follow up of 29.2 months (maximum about 40 months) after the start of induction. In the company's base case, the relative treatment effects of DBTd in MRD-positive and MRD-negative subgroups are assumed to persist throughout the lifetime model horizon. **CS Scenarios 3A-3D** show the impact of loss of treatment effects after 5/10 years in both MRD groups or in the MRD-negative group alone. We report an additional scenario with loss of effects after 3 years.	**ERG Scenario 11**: Loss of effects at 3 years (HR=1 for both MRD groups, PFS and OS)
Survival outcomes used in the economic model are not adjusted for the second randomisation to maintenance treatment in the CASSIOPEIA trial	The company present PFS results adjusted for maintenance in a prespecified IPW analysis conducted by a sequestered group. Landmark PFS and OS analyses (on which the economic model is based) are not adjusted for maintenance. We understand that it is not currently possible for the company to conduct further analysis as maintenance phase data is still blinded.	*Check impact of maintenance on landmark PFS and OS estimates after unblinding of the maintenance phase*
Health state utility estimates used in the model are appropriate but we conduct additional scenario analyses for completeness	We agree with use of EQ-5D data from the CASSIOPEIA trial for pre-progression utilities, and the van Agthoven et al. (2004) utility for progressed disease as in TA311. **CS Scenarios 8A to 8C** test alternative estimates used in TA311 and TA510. For completeness, we conducted a scenario analysis with the utilities from a US registry, reported by Abonour et al. (2018)[74] The company adjusts utilities for age using general population utilities for 10-year age bands up to 75 years.[73] This raised uncertainty over the ERG analysis with an older cohort, so we report results with a general population utility data for 5-year age bands up to age 85 years.[75]	**ERG Scenario 12**: Utilities from Abonour et al. (2018)[74] **ERG Scenario 13**: utility age adjustment from Ara and Brazier (2011)[75]
Daratumumab is costed as a fixed dose SC injection, although trial data relates to the weight based IV formulation	The company note that the EMA have accepted the case for non-inferiority of SC versus IV based on the COLUMBA trial and that patients and clinicians will prefer the convenience of SC injections. **CS Scenario 4** includes costs for the IV formulation. We prefer this scenario for consistency with clinical evidence.	
Uncertainty over the use of subsequent treatments	The model includes costs for three further lines of treatment for all patients after first progression, based on the NHS England	**ERG Scenario 13**: 100%, 75% and 50% of patients who

Issue	Summary	ERG analyses
has a moderate impact on cost estimates	pathway. Data on the number of patients starting subsequent lines of treatment is available in the PHE dataset, although this has limited follow up (during median follow up of [redacted] months, [redacted] started first line treatment, [redacted] second line, [redacted] third line and [redacted] fourth line). The base case excludes CDF treatments, which is appropriate as they are not routine practice. **CS Scenario 5** includes costs for CDF regimens. Expert advice to the ERG is that PBd is not currently used at third line (as assumed in the company base case).	progress start 2L, 3L and 4L treatment, respectively **ERG Scenario 14**: No PBd at 3L
Deterministic and probabilistic sensitivity analyses do not reflect available evidence on parameter uncertainty	The company's sensitivity analyses do not provide an accurate reflection of parametric uncertainty because the variance assumed for many of the input parameters is not based on the available evidence. Distributions assumed for some input parameters are also inappropriate.	We present revised sensitivity analysis results for the company's base case and for the ERG preferred assumptions in Appendix 10.4 below.

6.2 Impact on the ICER of additional ERG analyses

Results from the ERG scenarios applied to the company base analysis are shown in Table 39 below. The ICER is most sensitive to changes to the additional treatment effects for DBTd compared with BTd in the MRD subgroups. Scenario ERG9, in which only statistically significant landmark HRs are used (PFS in the MRD-negative group) gives the highest estimated ICER ([redacted] per QALY gained). Assuming that the treatment effects from the landmark analysis do not persist beyond 3 years post-consolidation is also associated with a large increase in the ICER ([redacted] per QALY gained).

The ICER is not sensitive to variation of the PFS/OS HRs for MRD-negative versus MRD-positive between confidence limits from the meta-analysis (ERG4 and 5) or to reliance only on data on this relationship from the CASSIOPEIA trial (ERG 6 and 7).

Table 39 Results from the ERG scenario analyses, deterministic analysis

Scenario		Intervention	Costs	QALYs	ICER (£/QALY)
Company base case		BTd DBTd	[redacted]	[redacted]	- [redacted]
Mean age of population at start of induction					
ERG1	PHE dataset: [redacted] years	BTd DBTd	[redacted]	[redacted]	- [redacted]
Extrapolation of baseline survival curves					
ERG2	Weibull OS for BTd MRD+ (fixed HR for MRD- vs MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG3	Gompertz PFS for BTd MRD+ (fixed HR for MRD- vs MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Relative survival MRD-negative versus MRD-positive					
ERG4	HRs from MA lower 95% limits: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG5	HRs from MA upper 95% limits: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG6	HRs from CASSIOPEIA only: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG7	Independently fitted BTd MRD- PFS Weibull; OS exponential	BTd DBTd	[redacted]	[redacted]	- [redacted]
Treatment effects (DBTd versus BTd) in MRD-negative and MRD-positive groups					
ERG8	Effects only for MRD- (HR=1 for MRD+ PFS and OS)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG9	Effects only for MRD- PFS (HR=1 for OS and PFS MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG10	No effects in either MRD group	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG11	Loss of effects MRD- & MRD+ 3 years post consolidation	BTd DBTd	[redacted]	[redacted]	- [redacted]
Utilities					
ERG12	EQ-5D from US registry (Abonour et al 2018)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG13	Age adjustments 5-year bands to > 85 years (Ara and Brazier 2011)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Resource use and costs					
ERG14	PD patients starting 2L, 3L and 4L treatment (100%, 75%, 50%)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG15	No PBd at 3L	BTd DBTd	[redacted]	[redacted]	- [redacted]

Source: produced from model by ERG, PAS discount for daratumumab, list prices for other drugs

6.3 ERG's preferred assumptions

We summarise the cumulative effect of applying the ERG preferred scenarios to the company base case results in Table 40**.** Collectively, these assumptions increase the ICER to [redacted] per QALY gained. The most influential assumption is that the treatment effects for DBTd compared with BTd (estimated from the CASSIOPEIA landmark analysis) do not persist beyond 5 years after consolidation. We consider this a reasonable assumption, given the immaturity of the CASSIOPEIA survival data and the lack of other evidence for long-term effects on survival.

Table 40 ERG's preferred model assumptions, deterministic analysis

Preferred assumption		Section in ERG report	Cumulative ICER £/QALY
Company base-case			[redacted]
CS3A	No additional DBTd effect after 5 years	4.2.6.2.4	[redacted]
ERG1	Mean age PHE dataset: [redacted] years	4.2.3	[redacted]
CS4	Daratumumab IV formulation	4.2.4	[redacted]
ERG14	PD 2L (100%), 3L (75%) & 4L (50%)	4.2.8.4	[redacted]
ERG15	No PBd at 3L	4.2.8.4	[redacted]
ERG13	Utility adjustment: 5-year age bands	4.2.7.4	[redacted]
Source: produced from model by ERG, PAS discount for daratumumab, list prices for other drugs			

We explore the effect of other uncertainties on the ERG preferred analysis results in Table 41. Again, this highlights the sensitivity of results to assumptions about the extent and persistence of additional treatment effects of daratumumab on survival outcomes in the MRD subgroups.

Table 41 Scenario analysis around ERG base case, deterministic analysis

Scenario		Intervention	Costs	QALYs	ICER (£/QALY)
ERG preferred analysis		BTd DBTd	[redacted] [redacted]	[redacted] [redacted]	- [redacted]
Mean age of population at start of induction					
ERG1	PHE dataset: 56.6 years	BTd DBTd	[redacted] [redacted]	[redacted] [redacted]	- [redacted]
Extrapolation of baseline survival curves					
ERG2	Weibull OS for BTd MRD+ (fixed HR for MRD- vs MRD+)	BTd DBTd	[redacted] [redacted]	[redacted] [redacted]	- [redacted]
CS2	Exponential PFS for BTd MRD+ (fixed HR for MRD- vs MRD+)	BTd DBTd	[redacted] [redacted]	[redacted] [redacted]	- [redacted]

Scenario		Intervention	Costs	QALYs	ICER (£/QALY)
ERG3	Gompertz PFS for BTd MRD+ (fixed HR for MRD- vs MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
Relative survival MRD-negative versus MRD-positive					
ERG4	HRs from MA lower 95% limits: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG5	HRs from MA upper 95% limits: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG6	HRs from CASSIOPEIA only: PFS [redacted] and OS [redacted]	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS1A	Independently fitted BTd MRD- PFS Weibull; OS Weibull	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG7	Independently fitted BTd MRD- PFS Weibull; OS exponential	BTd DBTd	[redacted]	[redacted]	- [redacted]
Treatment effects (DBTd versus BTd) in MRD-negative and MRD-positive groups					
ERG8	Effects only for MRD- (HR=1 for MRD+ PFS and OS)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG9	Effects only for MRD- PFS (HR=1 for OS and PFS MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG10	No effects in either MRD group	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS1B	DBTd MRD- estimated from MRD HRs (same as DBTd MRD+)	BTd DBTd	[redacted]	[redacted]	- [redacted]
ERG11	Loss of effects MRD+ and MRD- 3 years post consolidation	BTd DBTd	[redacted]	[redacted]	- [redacted]
CS3B	Loss of effects MRD+ and MRD- 10 years post consolidation	BTd DBTd	[redacted]	[redacted]	- [redacted]
CSBC	No loss of effects	BTd DBTd	[redacted]	[redacted]	- [redacted]
Resource use and costs					
CSBC	SC formulation	BTd DBTd	[redacted]	[redacted]	- [redacted]
CSBC	PD patients starting 2L, 3L and 4L (100%, 100%, 100%)	BTd DBTd	[redacted]	[redacted]	- [redacted]
CSBC	Include PBd at 3L	BTd DBTd	[redacted]	[redacted]	- [redacted]
Source: produced from model by ERG, PAS discount for daratumumab, list prices for other drugs					

6.4 Conclusions of the cost effectiveness section

The main conclusions from the ERG critique of the company's cost-effectiveness analysis and additional ERG analysis are summarised in Table 42 below.

Table 42 Cost effectiveness issues identified by the ERG

	Issue	Summary	Priority and action
1	OS data from the trial are very immature and yield a wide range of extrapolations for standard care (BTd) which do not reflect clinical expectations. (section 4.2.6.2.2)	In the short term (up to 3 years after diagnosis), modelled survival is a good fit to CASSIOPEIA trial data and real-world data from the PHE cohort. However, extrapolations for the BTd MRD-positive reference group exceed clinical expectations, with constant or decreasing hazards that reach general population rates within 30 years. We agree with use of the most conservative (exponential) distribution for OS as the base case, but this still yields high survival estimates. Estimated life years and QALYs for BTd from the current model are also much higher than estimates from the model used in TA311. The reason for this difference is not clear.	High priority – the model is not very sensitive to OS extrapolations from the trial data. However, long-term predictions from these data appear overly optimistic. Further opinion and alternative data sources to validate the survival extrapolations would be helpful.
2	Uncertainty over additional effects of daratumumab from the landmark analysis (section 4.2.6.2.4)	The company base case uses HRs from the CASSIOPEIA landmark analysis to model OS and PFS for DBTd relative to BTd in each MRD group. The estimated risk reductions in the MRD-negative group are higher than in the MRD-positive group. And the confidence intervals include 1 for PFS and OS in the MRD-positive group, and for OS in the MRD-negative group. We consider it appropriate to include the landmark analysis estimates in our preferred analysis, but we note that they are subject to uncertainty and have a large impact on the ICER.	High priority – the model is sensitive to the landmark analysis estimates of treatment effects Further opinion on the plausibility of these effects would help to resolve uncertainty
2	Assumptions about persistence of treatment effects are highly uncertain (section 4.2.6.2.4)	In the company's base case, the relative treatment effects of DBTd in MRD-positive and MRD-negative subgroups are assumed to persist throughout the lifetime model horizon. This is a strong assumption, as the model relies on immature survival data from the CASSIOPEIA trial with median follow up of 29.2 months (maximum about 40 months) after the start of induction.	High priority – the model is very sensitive to assumptions about the persistence of treatment affects. Further clinical opinion on this issue would be useful

	Issue	Summary	Priority and action
4	Uncertainty over the reliability of the company's MRD meta-analysis adds to uncertainty over cost-effectiveness. (section 4.2.6.2.3)	In the base case, the company model BTd OS and PFS for the MRD-negative group using HRs from the MRD meta-analysis. The MRD meta-analysis supplements immature trial data with external evidence but the ERG have some concerns about the reliability of the meta-analysis (see Table 26 above).	Medium priority – scenario analysis indicates that the ICER is not very sensitive to MRD HR estimates. Further clarification on the completeness and methods of the meta-analysis would help to resolve uncertainty.
5	Survival outcomes used in the economic model are not adjusted for the re-randomisation to maintenance treatment in the CASSIOPEIA trial (section 3.2.4.6)	The company present PFS results adjusted for maintenance in a prespecified IPW analysis conducted by a sequestered group. Landmark PFS and OS analyses (on which the economic model is based) are not adjusted for maintenance. We understand that it is not currently possible for the company to conduct further analysis as maintenance phase data are still blinded.	Medium priority – Current indications from the IPW analysis are that maintenance has a negligible impact on PFS. However, this could change with longer follow up. Check impact of maintenance on landmark PFS and OS estimates after unblinding of the maintenance phase
5	The mean age of patients in the company's economic analysis does not reflect the target population (section 4.2.3)	The company assume a mean age of 56.6 years, as in the CASSIOPEIA trial. This does not reflect NHS practice as the trial excluded patients over the age of 65. PHE data indicate that [redacted] of patients with newly diagnosed transplant eligible MM are aged 65 or older.	Medium priority – the ICER is moderately sensitive to the higher mean age of patients in the PHE cohort data
10	Daratumumab is costed as a fixed dose SC injection, although trial data relates to the weight based IV formulation (section 4.2.4)	The company note that the EMA have accepted the case for non-inferiority of SC versus IV based on the COLUMBA trial and that patients and clinicians will prefer the convenience of SC injections.	Medium priority – We prefer the scenario with costs for IV daratumumab for consistency with clinical evidence. This has a moderate impact on the ICER.
11	Uncertainty over the use of subsequent treatments has a moderate impact on cost estimates (section 4.2.8.4)	The model includes costs for three subsequent lines of treatment based on the NHS England pathway. The model assumes that patients who progress all start second, third and fourth lines of treatment. This seems unrealistic as some patients will die or become unfit for further treatment before completing fourth line treatment.	Low priority – exploratory scenario analysis suggests that subsequent treatment costs have a small impact on the ICER

7 END OF LIFE

The company do not make a claim for the applicability of end of life criteria for this appraisal. Without discounting, mean survival in the comparator arm (BTd) is estimated at over 16.4 years in the company's base case analysis. With ERG preferred assumptions expected life expectancy is only slightly shorter (15.9 years for BTd).

8 INNOVATION, EQUALITY AND DIVERSITY

8.1 Innovation

The company argue that daratumumab is innovative in this indication (combination therapy for NDTE MM patients) based on the following key points (CS section B.2.12):

- Daratumumab is a first-in-class therapy, targeting the CD38 protein which is expressed on MM cells;
- Daratumumab targets the tumour both directly and indirectly, as well as modulating the immune system through several mechanisms (CS Figure 26) not typically seen in monoclonal antibodies;
- The multi-modal action of daratumumab means that daratumumab is effective irrespective of MM clonal heterogeneity and increases the depth and durability of response.

The ERG note that one of the appraisal consultees, the UK Myeloma Forum, concur that daratumumab is an innovative therapy for these reasons. The other consultee organisation (Myeloma UK) did not comment on innovation in this technology appraisal, although they had agreed in the NICE appraisal of daratumumab monotherapy for previously-treated MM patients (NICE TA 510) that daratumumab is an innovative therapy.

We agree that, in terms of its mode of action, daratumumab is an innovative therapy. We also note that daratumumab can be administered either as an IV or SC therapy, although not orally. The safety profile of daratumumab is manageable, which makes it suitable for the DBTd combination therapy. However, whilst we agree that daratumumab is innovative, we are unaware of any evidence that daratumumab has demonstrable and distinctive benefits of a substantial nature other than those already captured in the QALY measure.

8.2 Equality and diversity

Neither the company, nor the consultee organisations (UK Myeloma Forum and Myeloma UK) raised any concerns regarding equality or diversity if DBTd is recommended for use in NHS practice. One of the ERG's clinical experts stressed that DBTd availability should not be limited to patients aged under 65 (i.e. the pivotal trial population age cut-off) because eligibility for therapy in NHS practice is based on fitness rather than patient age, and in practice patients may be treated into their mid-70s. The NICE Committee in NICE TA510 when considering equality and diversity noted that MM is more common in men than women and the incidence is also reported to be higher in people of African American family origin. The ERG are not aware of any other issues pertinent to equality and diversity.

9 References

1. European Medicines Agency (EMA). *Public Summary of Opinion on Orphan Designation: Daratumumab for the treatment of plasma cell myeloma. 2013.* https://www.ema.europa.eu/en/documents/orphan-designation/eu/3/13/1153-public-summary-opinion-orphan-designation-daratumumab-treatment-plasma-cell-myeloma_en.pdf.
2. European Medicines Agency (EMA). *Summary of Product Characteristics. Darzalex 20 mg/mL concentrate for solution for infusion*. https://www.medicines.org.uk/emc/product/7250/smpc#CLINICAL_PARTS.
3. European Medicines Agency (EMA). *Summary of Product Characteristics. Darzalex 1,800 mg solution for injection* https://www.medicines.org.uk/emc/product/11488/smpc.
4. Food and Drug Administration (FDA). *Orphan Drug Designations and Approvals. Daratumumab*. https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=397113.
5. Kumar SK, Vij R, Noga SJ, et al. Treating Multiple Myeloma Patients With Oral Therapies. *Clinical Lymphoma, Myeloma and Leukemia* 2017;17(5):243-51.
6. Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. *Leukemia* 2009;23(1):3-9.
7. Röllig C, Knop S, Bornhäuser M. Multiple myeloma. *The Lancet* 2015;385(9983):2197-208.
8. Bianchi G, Anderson KC. Understanding biology to tackle the disease: multiple myeloma from bench to bedside, and back. *CA: a cancer journal for clinicians* 2014;23(10):422-44.
9. Palumbo A, Sezer O, Kyle R, et al. International Myeloma Working Group guidelines for the management of multiple myeloma patients ineligible for standard high-dose chemotherapy with autologous stem cell transplantation. *Leukemia* 2009;23(10):1716-30.
10. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *Journal of Clinical Oncology* 2015;33(26):2863-9.
11. Richardson P, Mitsiades C, Schlossman R, et al. The treatment of relapsed and refractory multiple myeloma. *Hematology Am Soc Hematol Educ Program* 2007:317-23.
12. Brenner H, Gondos A, Pulte D. Recent major improvement in long-term survival of younger patients with multiple myeloma. *Blood* 2008;111(5):2521-6.

13. Kumar SK, Rajkumar SV, Dispenzieri A, et al. Improved survival in multiple myeloma and the impact of novel therapies. *Blood* 2008;111(5):2516-20.
14. Pratt G, Jenner M, Owen R, et al. Updates to the guidelines for the diagnosis and management of multiple myeloma. *Br J Haematol* 2014;167(1):131-3.
15. Usmani S, Ahmadi T, Ng Y, et al. Analysis of real-world data on overall survival in multiple myeloma patients with ≥3 prior lines of therapy including a proteasome inhibitor (PI) and an immunomodulatory drug (IMiD), or double refractory to a PI and an IMiD. *Oncologist* 2016;21(11):1355-61.
16. Yong K, Delforge M, Driessen C, et al. Multiple myeloma: patient outcomes in real-world practice. *Br J Haematol* 2016;175(2):252-64.
17. Voorhees P, Fonseca R, Cote S, et al. Characterization of frontline treatment patterns and the proportion of patients reaching subsequent lines of therapy in transplant eligible patients with newly diagnosed multiple myeloma. *HemaSphere* 2019;3:274-5.
18. Cancer Research UK (CRUK). Myeloma Statistics 2017. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma. 2020.
19. Rawstron AC, Gregory WM, de Tute RM, et al. Minimal residual disease in myeloma by flow cytometry: independent prediction of survival benefit per log reduction. *Blood* 2015;125(12):1932-35.
20. de Tute R, Rawstron AC, Cairns D, et al. Minimal residual disease in transplant ineligible myeloma patients: results from the UK NCRI Myeloma XI trial. *Clin Lymphoma Myeloma Leuk* 2015;15:e84.
21. De Tute R, Cairns D, Rawstron A, et al. Sequential minimal residual disease (MRD) monitoring: Results from the UK Myeloma XI trial. *Clinical Lymphoma, Myeloma and Leukemia* 2019;19(10):e45-e46.
22. National Institute for Health and Care Excellence (NICE). *NG35. Myeloma: diagnosis and management*. https://www.nice.org.uk/guidance/ng35.
23. European Medicines Agency (EMA). *Assessment Report. Darzalex* https://www.ema.europa.eu/en/documents/variation-report/darzalex-h-c-4077-ii-0002-epar-assessment-report-variation_en.pdf.
24. Moreau P, San Miguel J, Sonneveld P, et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncololgy* 2017;28(suppl_4):iv52-iv61.
25. Engelhardt M, Terpos E, Kleber M, et al. European Myeloma Network recommendations on the evaluation and treatment of newly diagnosed patients with multiple myeloma. *haematologica* 2014;99(2):232-42.

26. National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology. Multiple Myeloma. Version 4.2020. Available at: https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf. 2020.
27. National Institute for Health and Care Excellence (NICE). *TA311: Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation*. https://www.nice.org.uk/guidance/ta311.
28. National Institute for Health and Care Excellence (NICE). *TA129: Bortezomib monotherapy for relapsed multiple myeloma*. https://www.nice.org.uk/Guidance/TA129.
29. National institute for Health and Care Excellence (NICE). *TA457: Carfilzomib for previously treated multiple myeloma*. https://www.nice.org.uk/guidance/ta457.
30. National Institute for Health and Care Excellence (NICE). *TA573: Daratumumab with bortezomib and dexamethasone for previously treated multiple myeloma*. https://www.nice.org.uk/guidance/TA573.
31. National Institute for Health and Care Excellence (NICE). *TA586: Lenalidomide plus dexamethasone for multiple myeloma after 1 treatment with bortezomib*. https://www.nice.org.uk/guidance/TA586.
32. National institute for Health and Care Excellence (NICE). *TA171: Lenalidomide for the treatment of multiple myeloma in people who have received at least 2 prior therapies*. https://www.nice.org.uk/guidance/ta171.
33. National Institute for Health and Care Excellence (NICE). *TA380: Panobinostat for treating multiple myeloma after at least 2 previous treatments*. https://www.nice.org.uk/guidance/ta380.
34. National Institute for Health and Care Excellence (NICE). *TA505: Ixazomib with lenalidomide and dexamethasone for treating relapsed or refractory multiple myeloma*. https://www.nice.org.uk/guidance/TA505.
35. National Institute for Health and Care Excellence (NICE). *TA427: Pomalidomide for multiple myeloma previously treated with lenalidomide and bortezomib*. https://www.nice.org.uk/guidance/TA427.
36. National Institute for Health and Care Excellence (NICE). *TA510: Daratumumab monotherapy for treating relapsed and refractory multiple myeloma* https://www.nice.org.uk/guidance/ta510.
37. Ludwig H, Durie BGM, Bolejack V, et al. Myeloma in patients younger than age 50 years presents with more favorable features and shows better survival: an analysis of 10 549 patients from the International Myeloma Working Group. *Blood Cancer J* 2008;111(8):4039-47.

38. European Medicines Agency (EMA). *Velcade. EPAR - Product Information. Summary of Product Characteristics*. https://www.ema.europa.eu/en/medicines/human/EPAR/velcade.
39. BMJ Best Practice. *Assessment of serum creatinine*. https://bestpractice.bmj.com/topics/en-gb/935 (accessed 29/10/2020).
40. Myeloma UK. *Myeloma and the kidney*. https://www.myeloma.org.uk/documents/myeloma-and-the-kidney-infoguide/ (accessed 29/10/2020).
41. UK Myeloma Forum. Professional organisation submission: daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]: UK Myeloma Forum, [2020].
42. Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncology* 2016;17(8):e328-e46.
43. Kostopoulos IV, Ntanasis-Stathopoulos I, Gavriatopoulou M, et al. Minimal residual disease in multiple myeloma: current landscape and future applications with immunotherapeutic approaches. *Frontiers in Oncology* 2020;10(860).
44. Mina R, Oliva S, Boccadoro M. Minimal residual disease in multiple myeloma: state of the art and future perspectives. *Journal of Clinical Medicine* 2020;9(7):2142.
45. Moreau P, Zamagni E. MRD in multiple myeloma: more questions than answers? *Blood Cancer J* 2017;7(12):639:1-4.
46. Avet-Loiseau H, Ludwig H, Landgren O, et al. Minimal residual disease status as a surrogate endpoint for progression-free survival in newly diagnosed multiple myeloma studies: a meta-analysis. *Clinical Lymphoma Myeloma & Leukemia* 2020;20(1):e30-e37.
47. Rajkumar SV, Harousseau JL, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. *Blood* 2011;117(18):4691-5.
48. Myeloma UK. Patient reported outcome measures (PROMS) in myeloma: are they fit for the future? Edinburgh: Myeloma UK, 2018:46.
49. Osborne TR, Ramsenthaler C, Schey SA, et al. Improving the assessment of quality of life in the clinical care of myeloma patients: the development and validation of the Myeloma Patient Outcome Scale (MyPOS). *BMC Cancer* 2015;15:280.
50. Janssen. [Data on File] MMY3006 (CASSIOPEIA). Clinical Protocol. 2015.
51. Miyahara S, Wahed AS. Weighted Kaplan-Meier estimators for two-stage treatment regimes. *Stat Med* 2010;29(25):2581-91.

52. Lokhnygina Y, Helterbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics* 2007;63(2):422-8.
53. Tang X, Wahed AS. Cumulative hazard ratio estimation for treatment regimes in sequentially randomized clinical trials. *Statistics in Biosciences* 2015;7(1):1-18.
54. Mateos M-V, Nahi H, Legiec W, et al. Efficacy and safety of the randomized, open-label, non-inferiority, phase 3 study of subcutaneous (SC) versus intravenous (IV) daratumumab (DARA) administration in patients (pts) with relapsed or refractory multiple myeloma (RRMM): COLUMBA. *Journal of Clinical Oncology* 2019;37(15 Suppl).
55. Mateos M-V, Nahi H, Legiec W, et al. Subcutaneous versus intravenous daratumumab in patients with relapsed or refractory multiple myeloma (COLUMBA): a multicentre, open-label, non-inferiority, randomised, phase 3 trial. *The Lancet Haematology* 2020;7(5):e370-e80.
56. Moreau P, Hulin C, Macro M, et al. VTD is superior to VCD prior to intensive therapy in multiple myeloma: results of the prospective IFM2013-04 trial. *Blood* 2016;127(21):2569-74.
57. Moreau P, Avet-Loiseau H, Facon T, et al. Bortezomib plus dexamethasone versus reduced-dose bortezomib, thalidomide plus dexamethasone as induction treatment prior to autologous stem cell transplantation in newly diagnosed multiple myeloma. *Blood* 2011:blood-2011-05-355081.
58. Harousseau JL, Attal M, Avet-Loiseau H, et al. Bortezomib plus dexamethasone is superior to vincristine plus doxorubicin plus dexamethasone as induction treatment prior to autologous stem-cell transplantation in newly diagnosed multiple myeloma: results of the IFM 2005-01 phase III trial. *Journal of Clinical Oncology* 2010;28(30):4621-9.
59. Mai EK, Bertsch U, Dürig J, et al. Phase III trial of bortezomib, cyclophosphamide and dexamethasone (VCD) versus bortezomib, doxorubicin and dexamethasone (PAd) in newly diagnosed myeloma. *Leukemia* 2015;29(8):1721-9.
60. Kumar S, Flinn I, Richardson PG, et al. Randomized, multicenter, phase 2 study (EVOLUTION) of combinations of bortezomib, dexamethasone, cyclophosphamide, and lenalidomide in previously untreated multiple myeloma. *Blood, The Journal of the American Society of Hematology* 2012;119(19):4375-82.
61. Phillippo DM, Ades AE, Dias S, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. 2016. Available at http://www.nicedsu.org.uk. 2016.

62. Munshi N, Avet-Loiseau H, Anderson KC, et al. Expanded meta-analysis confirms the association between MRD and long-term survival outcomes in multiple myeloma (MM). Poster presented at American Society of Hematology (ASH). 2019 2019.
63. Munshi NC, Avet-Loiseau H, Rawstron AC, et al. Association of minimal residual disease with superior survival outcomes in patients with multiple myeloma: a meta-analysis. *JAMA Oncol* 2017;3(1):28-35.
64. Scottish Medicines Consortium (SMC). *Bortezomib 3.5mg powder for solution for injection (Velcade®) SMC No. (927/13). December 2013*. https://www.scottishmedicines.org.uk/files/advice/bortezomib_Velcade_FINAL_December_2013_for_website.pdf.
65. van Agthoven M, Segeren CM, Buijt I, et al. A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study. *European Journal of Cancer* 2004;40(8):1159-69.
66. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708-15.
67. Woods B, Sideris E, Palmer S, et al. Partitioned survival analysis for decision modelling in health care: a critical review. NICE DSU technical support document 19, 2017.
68. Bullement A, Latimer NR, Gorrod HB. Survival extrapolation in cancer immunotherapy: a validation-based case study. *Value in Health* 2019;22(3):276-83.
69. Ouwens MJ, Mukhopadhyay P, Zhang Y, et al. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. *Pharmacoeconomics* 2019;37(9):1129-38.
70. Janssen. [Data on File] Public Health England datasets. Standing cohort study of newly diagnosed multiple myeloma cancer patients in England (January 2015 to December 2018). (July 2020 Report). 2020.
71. European Medicines Agency (EMA). *Darzalex Assessment Report Extension* https://www.ema.europa.eu/en/documents/variation-report/darzalex-h-c-4077-x-0032-epar-assessment-report-extension_en.pdf.
72. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011.
73. Janssen B, Szende A. Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J (eds.) *Self-Reported Population Health: An International Perspective based on EQ-5D*. Dordrecht: Springer Netherlands; 2014 p19-30.

74. Abonour R, Wagner L, Durie BG, et al. Impact of post-transplantation maintenance therapy on health-related quality of life in patients with multiple myeloma: data from the Connect® MM Registry. *Ann Hematol* 2018;97(12):2425-36.
75. Ara R, Brazier JE. Using Health State Utility Values from the General Population to Approximate Baselines in Decision Analytic Models when Condition-Specific Data are Not Available. *Value in Health* 2011;14(4):539-45.
76. Wong R, Paisley S, Carroll C. Assessing searches in NICE single technology appraisals: practice and checklist. *International Journal of Technology Assessment in Health Care* 2013;29(3):315-22.
77. Moreau P, Attal M, Hulin C, et al. Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label, phase 3 study. *Lancet* 2019;394(10192):29-38.

10 Appendices

10.1 ERG critique of clinical effectiveness searches

Clinical effectiveness searches (CS sections B.2.1 and D.1.1)		ERG comments
Dates covered	Original search: database inception to 31/05/2018 Update search performed: 01/05/2020 NB – only 1995 publications onwards in PICO inclusion criteria. Using the mid-1990s as a start date limit means that all three commonly used interventions used as single, doublet and triplet therapies (thalidomide, lenalidomide and bortezomib) are included as comparators.	Searches were four months old at time of submission. A targeted ERG search identified no further relevant studies.
Reporting	Concise: original and update search strategies are the same, the numbers of results are reported in separate columns of one search strategy table for each database. PRISMA flow diagram included. List of publications excluded at full text screening reported with exclusion reason.	No concerns
Strategy overall	Searched for randomised controlled trials (RCTs) and 'clinical trials' generally. Broad.	No concerns
Strategy PICO and terms	Population: no search terms for 'transplant eligible' which would have narrowed results too much. No intervention, comparator or outcomes concept – keeps search broad. Date limit for inclusion criteria is since 1995, and for conferences it is from 2015-2020.	No concerns
Strategy subject headings	MeSH used in Medline and Cochrane. Embase search table footnote refers to MeSH instead of Emtree but on checking the thesaurus the headings used in the search string are from Emtree. RCT concept does not use subject headings. Appropriate terms are available in MeSH and Emtree (as per published RCT filters).	RCT concept lacks relevant subject headings. RCT search string is not systematic. Same SLR used to identify non-RCTs for the MAIC.
Strategy free-text terms	Yes, uses relevant synonyms for naïve/newly diagnosed to combine with MM terms to achieve results for the relevant population.	No concerns
Strategy syntax	Correct	No concerns

Strategy structure	Condensed: one concept per search line rather than one search term per line.	No concerns
Sources	Medline and Medline In-Process, Embase, Cochrane Library. Abstracts from five conferences were searched for separately in Embase: ASCO, ASH, EHA, ISPOR and IMWG. Abbreviations not spelled out in CS. American Society of Clinical Oncology American Society of Hematology European Haematological Association International Society for Pharmacoeconomics International Myeloma Working Group A relevant UK conference would be the British Society for Haematology (BSH) Annual Scientific Meeting. The ERG searched conference abstracts published in the British Journal of Haematology supplements since 2015 using Embase and found one relevant study (UK Myeloma XI) already identified and included by the company.	Cochrane Library's CENTRAL is the only trial-specific source searched
Limits	All searches are limited to results with abstracts. No line in the searches for this so we cannot see how many results were without abstracts. Only likely to miss short communications, letters, newsy magazine type journals. Excluded animal studies, letters, editorials, non-systematic reviews.	No concerns
Filters	No search filters used. RCT search string checked against published filters: Ovid database filters, Scottish Intercollegiate Guidelines Network (SIGN), Canadian Agency for Drugs and Technologies in Health (CADTH), BMJ (British medical Journal) Best Practice	Using a published filter would be better than creating a RCT search string without relevant subject headings for RCTs.
Translation	Searching was consistent across the databases.	No concerns
Missing studies	A targeted update by the ERG in Medline identified no further studies.	No
The criteria are adapted from a published checklist.[76] RCTs: randomised controlled trials		

10.2 Company and ERG risk of bias assessments

ASSESSMENT CRITERIA	COMPANY JUDGEMENT		ERG JUDGEMENT
	CASSIOPEIA	**Risk of bias**	
Was randomisation carried out appropriately?	Yes, randomisation was carried out as per the pre-specified randomisation method; patients were randomised using a central IWRS	Low	Agree: low risk of bias
Was the concealment of treatment allocation adequate?	CASSIOPEIA was open label	Low, as patients were randomised using a central IWRS	Agree: low risk of bias The ERG assume that the open-label design refers to lack of blinding post-treatment allocation and that treatment allocation itself would have been concealed using a centralised IWRS
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes, demographic and baseline characteristics were well balanced between the two treatment groups with no categories having a difference of ≥10%	Low	Agree: low risk of bias We note imbalance with respect to the revised ISS classification which suggested a higher proportion of patients in the DBTd group had a stage II or III disease indicating poorer prognosis. Any arising bias is likely to be in favour of BTd..
Were the care providers, participants and outcome assessors blind to treatment allocation?	CASSIOPEIA was open label and only Janssen was blinded to the results	Low, as an IDMC reviewed the data	<u>Efficacy outcomes measures</u> • Agree: low risk of bias (outcomes are objective and standardised.) <u>HRQoL outcomes</u> • Unclear risk of bias as these outcomes are more subjective in nature and may be influenced by knowledge of treatment allocation <u>Adverse event outcomes</u>

			Agree: low risk of bias (most adverse events considered in the economic model are measured objectively e.g. using blood tests)
Were there any unexpected imbalances in drop-outs between groups?	No, of the 1,085 patients randomised (543 in the DBTd group and 542 in the BTd group), 1,074 received study treatment: 536 patients received DBTd and 538 patients received BTd	Low	Agree: low risk of bias • A slightly higher proportion of patients in the DBTd arm (84.9%) completed consolidation therapy compared to the BTd arm (80.6%), • Rates of discontinuations were similar in both groups for adverse events (approximately 10%) and progressive disease (3.7%). • The number of deaths differed however with no deaths in the DBTd arm and 7 deaths in the BTd arm.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	None	Low	Agree: low risk of bias
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, the ITT population was used for analysis of the primary outcome and other time-to-event efficacy outcomes, which included all randomised patients	Low	Disagree: unclear risk of bias In particular, the use of last observation carried forward imputations to account for missing post-ASCT outcome data for MRD status may introduce a potential bias. The ERG estimates that around 10% of the ITT population had MRD status estimated at the earlier cycle 4, day 28 assessment. The magnitude of any bias is unknown but is likely to be conservative in direction (favouring BTd).
Sources: CS section B.2.5; CS Tables 4, 6, 7; CS Figure 7; Moreau et al,[77] and clarification response A3 IDMC: Independent Data Monitoring Committee; IWRS: interactive web response system			

10.3 Overview of response rate outcomes assessed in the CASSIOPEIA trial

All responses were measured at post-induction, post-transplant and post-consolidation, except for MRD status which was measured at post-induction and post-consolidation.

Response outcome	Outcome definition	IMWG criteria	How measured in CASSIOPEIA
sCR (stringent complete response)	Percentage of patients achieving complete response in addition to having a normal serum free light chain ratio and an absence of clonal cells in bone marrow by immunohistochemistry, immunofluorescence or 2- to 4-colour flow cytometry.	• CR as defined below, plus • Normal free light chain ratio, and • Absence of clonal plasma cells by immunohistochemistry, immuno-fluorescence or 2- to 4-colour flow cytometry.	• sCR was confirmed centrally by a minimum of 4 colour flow cytometry, for which a fresh bone marrow aspirate was required, as per the study protocol. (In clinical practice confirmation with a fresh bone marrow biopsy is not required.) • As per IMWG criteria
MRD-negative rate	Proportion of patients who achieved MRD-negative status	Requires a complete response, plus One of: Sustained MRD-negative, Flow MRD-negative, Sequencing MRD-negative, or Imaging plus MRD-negative • Flow MRD-negative: Absence of phenotypically aberrant clonal plasma cells by next-generation flow on bone marrow aspirates using EuroFlow (or validated equivalent method) with a minimum sensitivity of 1 in 10^5 nucleated cells or higher. • Sequencing MRD-negative: Absence of clonal plasma cells by NGS on bone marrow aspirate. Presence of a clone is defined as <2 identical sequencing	• MRD-negative rate measured regardless of response, so differs from IWMG criteria for MRD (rationale not provided) • Multiparametric flow cytometry and 10^{-5} was used for the ITT population; • Next generation sequencing and 10^{-6} for a selection of patients was used for an exploratory analysis. • Both measured from bone marrow aspirates as per IMWG criteria.

Response outcome	Outcome definition	IMWG criteria	How measured in CASSIOPEIA
		reads from bone marrow aspirates using the LymphoSIGHT platform (or validated equivalent method) with a minimum sensitivity of 1 in 10^5 nucleated cells or higher.	
CR (in CASSIOPEIA referred to as ≥CR, i.e. CR or better)	Proportion of patients who achieved complete response (or better)	• Negative immunofixation on the serum and urine, and • Disappearance of any soft tissue plasmacytomas, and • <5% plasma cells in bone marrow.	• Absence of clonal plasma cells in the bone marrow (<5%) is detected by measuring paraprotein levels in the blood and urine. • Confirmation with a fresh bone marrow biopsy is not required. • As per IMWG criteria.
≥VGPR (very good partial response or better)	Proportion of patients who have achieved VGPR or better	• Serum and urine M-component detectable by immunofixation but not on electrophoresis, or • ≥90% reduction in serum M-protein <100 mg/24 hours.	As per IMWG criteria
ORR (overall response rate)	Proportion of patients who have achieved partial response or better	Criteria for sCR, CR, VGPR, PR, SD and progressive disease are reproduced from the IMWG guidelines in CS Appendix Table 62.	As per IMWG criteria
Sources: CS Tables 3, 4, 5 and 10; CS Appendix L			

10.4 Revised sensitivity analysis produced by the ERG

10.4.1 Analysis conducted on the company's base case

Table 43 Company's base case results, deterministic analysis

Intervention	Total		Incremental		ICER
	Costs	QALYs	Costs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Produced from model by ERG, PAS discount for daratumumab, list prices for other drugs					

Table 44 Company's base case results, probabilistic analysis

Intervention	Total		Incremental		ICER
	Costs	QALYs	Costs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Produced from model by ERG, PAS discount for daratumumab, list prices for other drugs					

Table 45 Deterministic sensitivity analysis, company base case (top 10 ICER impact)

Parameter	Base case	Lower limit		Upper limit	
		Value	ICER	Value	ICER
OS, DBTd vs BTd HR (MRD-)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, DBTd vs BTd HR (MRD+)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, MRD+ Extrapolation, Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PFS, DBTd vs BTd HR (MRD+)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response Rates, DBTd, MRD-	0.637	0.595	[redacted]	0.678	[redacted]
PFS, DBTd vs BTd HR (MRD-)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response Rates, BTd, MRD-	0.435	0.393	[redacted]	0.478	[redacted]
PFS, MRD- vs MRD+ HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, MRD- vs MRD+ HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Resource Use, Pre-Progression	£64.21	£51.37	[redacted]	£77.05	[redacted]
Source: Produced from model by ERG, PAS discount for daratumumab, list price for other drugs					

Figure 10 Cost Effectiveness Acceptability Curve, company base case

Figure 11 Cost effectiveness scatterplot, company base case

10.4.2 Analysis conducted on the ERG's preferred base case

Table 46 ERG's preferred analysis results, deterministic analysis

Intervention	Total		Incremental		ICER
	Costs	QALYs	Costs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Produced from model by ERG, PAS discount for daratumumab, list prices for other drugs					

Table 47 ERG's preferred analysis results, probabilistic analysis

Intervention	Total		Incremental		ICER
	Costs	QALYs	Costs	QALYs	(£/ QALY)
BTd	[redacted]	[redacted]	-	-	-
DBTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Source: Produced from model by ERG, PAS discount for daratumumab, list prices for other drugs					

Table 48 Deterministic sensitivity analysis, ERG base case (top 10 ICER impact)

Parameter	Base case	Lower limit		Upper limit	
		Value	ICER	Value	ICER
OS, DBTd vs BTd HR (MRD-)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, DBTd vs BTd HR (MRD+)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, MRD+ Extrapolation, Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response Rates, DBTd, MRD-	0.637	0.595	[redacted]	0.678	[redacted]
Response Rates, BTd, MRD-	0.435	0.393	[redacted]	0.478	[redacted]
PFS, DBTd vs BTd HR (MRD-)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PFS, DBTd vs BTd HR (MRD+)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS, MRD- vs MRD+ HR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean weight of patients (kg)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Health state utilities, Pre-Progression	0.731	0.720	[redacted]	0.740	[redacted]
Source: estimated from model by ERG, PAS discount for daratumumab, list price for other drugs					

Figure 12 Cost Effectiveness Acceptability Curve: ERG base case

Figure 13 Cost effectiveness scatterplot: ERG base case

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

ERG report – factual accuracy check and confidential information check

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

'*Data owners will be asked to check that confidential information is correctly marked in documents created by others in the technology appraisal process before release; for example, the technical report and ERG report.*' (Section 3.1.29, Guide to the processes of technology appraisals).

You are asked to check the ERG 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 19 November 2020** 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.

Major issues

Issue 1

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 15: 'The CASSIOPEIA trial protocol states that MRD would be assessed for complete responders, which is consistent with the International Myeloma Working Group (IMWG) definition of MRD negativity. However, the company submission (CS) states that MRD-negativity was determined regardless of response.'	Please amend to: '~~The CASSIOPEIA trial protocol states that MRD would be assessed for complete responders, which is consistent with the International Myeloma Working Group (IMWG) definition of MRD negativity. However, t~~The company submission (CS) states that MRD-negativity was determined regardless of response which is inconsistent with the International Myeloma Working Group (IMWG) definition of MRD negativity which is an assessment for complete responders.'	The CASSIOPEIA trial protocol submitted as a reference to the CS included an amendment which revised the eligibility criteria for MRD assessment for Part 1 to include all patients in the induction/consolidation phases regardless of response.	Not a factual inaccuracy. Protocol page 85 states MRD is defined per IMWG criteria. There is no clear protocol amendment stated within the protocol regarding the assessment of MRD. However, in the interests of accuracy we have reworded the text on page 15 as suggested.

Issue 2

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 82: 'The trial protocol states that MRD would only be assessed in patients achieving CR, but the company determined MRD status regardless of response. A rationale for this is not provided.	Please amend to: 'The CS states that MRD-negativity was determined regardless of response which is inconsistent with the IMWG definition of MRD negativity, which is an assessment	The CASSIOPEIA trial protocol submitted as a reference to the CS included an amendment which revised the eligibility criteria for MRD assessment for Part 1 to include all patients in the	Not a factual inaccuracy. Protocol page 85 states MRD is defined per IMWG criteria. There is no clear protocol amendment stated within the protocol regarding the assessment of MRD.

The definition of MRD used in the company's landmark analysis may therefore differ from the definition used in the studies included in the company's meta-analysis of the impact of MRD status on survival outcomes.'	for complete responders. The definition of MRD used in the company's landmark analysis may therefore differ from the definition used in the studies included in the company's meta-analysis of the impact of MRD status on survival outcomes.'	induction/consolidation phases regardless of response.	However, in the interests of accuracy we have reworded the text on page 82 as suggested.

Minor issues

Issue 3

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 60: 'The company claim that patients treated with DBTd who are MRD-negative have OS which resembles that of the general population when matched on age and gender, suggestive of long-term disease control and a possible cure for some patients.'	Please amend to: 'The company claim that patients treated with DBTd who are MRD-negative have OS which resembles that of the general population when matched on age and gender, suggestive of long-term disease control and a possible functional cure for some patients.'	Janssen recognise multiple myeloma as an incurable disease however claim a possible functional cure for patients that achieve the deepest levels of response, where mortality resembles that of the general population (when matched for age and gender).	Thank you for alerting us to this typographic error. We have made the suggested amendment on page 60.

Issue 4

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 106: 'The company conducted	Please amend to:	Presenting cost-effectiveness	Thank you for raising this point. It

a scenario analysis using the weight-based formulation of daratumumab which increased the cost-effectiveness ICER significantly to over £30,000 per QALY.'	'The company conducted a scenario analysis using the weight-based formulation of daratumumab which increased the cost-effectiveness ICER ~~significantly~~ to over £30,000 per QALY [redacted]	results without a PAS discount for daratumumab is not consistent with the results presented in section 5 of the report.	is not a factual inaccuracy, but it is potentially confusing. We have amended the text on page 106 to say simply that the change 'increased the base case ICER.' Details of the ICERs are better left to the results section, where the PAS is introduced.

Issue 5

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 110: 'The incremental cost effectiveness ratio (ICER) for DBTd versus BTd is [redacted] per QALY gained, with all drugs costed at list price'	Please amend to: 'The incremental cost effectiveness ratio (ICER) for DBTd versus BTd is [redacted] per QALY gained, [redacted]	Cost-effectiveness results presented in the report include the PAS discount for daratumumab; all other drugs are at list price.	Thank you, we have made the suggested amendment on page 110.

Misreporting from the CS and typographical errors

Issue 6

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 12: 'Section 1.1 provides an overview of the key issues. Section 1.1 provides an overview of key model outcomes and the modelling	Please amend to: 'Section 1.~~1~~2 provides an overview of the key issues. Section 1.1 provides an	Typographical error.	Thank you for alerting us to this typographic error on page 12. We have corrected this. We have also made a minor change to the text

assumptions that have the greatest effect on the ICER.'	overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. '		in the second paragraph on page 12 to improve clarity of signposting within the report.

Issue 7

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 27: 'The CS does not report the dosing interval for daratumumab when administered by SC injection, but according to the SmPC[3] daratumumab SC is administered weekly in the first two induction cycles; every 2 weeks in the third and fourth induction cycles; and weekly in each of the two consolidation cycles.'	Please amend to: 'The CS ~~does not~~ reports the dosing interval for daratumumab when administered by SC injection to be weekly in the first two induction cycles; every 2 weeks in the third and fourth induction cycles; and every 2 weeks in each of the two consolidation cycles (CS page 16, Table 2).'	Misreporting of information from the CS.	Thank you for alerting us to this error. We have made the correction on page 27 as suggested.

Issue 8

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 48: 'At the second data cut the percentage of patients who were MRD-negative post-consolidation and had a best response of CR or sCR was 66% in the DBTd arm and 59.8% in the BTd arm.'	Please amend to: 'At the second data cut the percentage of patients who were MRD-negative post-consolidation and had a best response of CR or sCR was ~~66%~~ 66.7% in the DBTd arm and 59.8% in the BTd arm.'	Misreporting of data from the CS.	Thank you for alerting us to this typographic error. We have made the correction on page 48.

Issue 9

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 48: 'The CS states that MRD was assessed according to IMWG criteria, which require that patients achieve CR and the CASSIOPEIA CSR states that "complete responders will be utilized to investigate the prognostic effect of MRD on PFS" (CSR page 89). However, assessments of MRD were conducted regardless of response (CS section B.2.6.1) and an assessment of MRD status in patients who achieved CR is not reported for CASSIOPEIA.'	The CSR does not state on page 89 that "complete responders will be utilized to investigate the prognostic effect of MRD on PFS". Please amend to: 'The CS states that MRD was assessed according to IMWG criteria, which require that patients achieve CR~~ and the CASSIOPEIA CSR states that "complete responders will be utilized to investigate the prognostic effect of MRD on PFS" (CSR page 89)~~. However, assessments of MRD were conducted regardless of response (CS section B.2.6.1) with an assessment of MRD status in patients who achieved CR performed as a post-hoc analysis (CS Appendix L.4).'	Typographical error from the CSR and misreporting from the CS.	Thank you for alerting us to this error: "CSR" should read "protocol". However, we have made the suggested alternative change on page 48 in the interests of accuracy and completeness.

Issue 10

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 48: 'The median follow-up period for the Part 1 analysis was 18.8 months. However, the company have not reported the follow-up time in Part 2, which would help to more	Please amend to: 'The median follow-up period for the Part 1 and Part 2 analysis was 18.8 months and 29.2 months, respectively (CS Tables 17 and 20). ~~However, the company have~~	Misreporting of information from the CS.	NB this issue refers to page 54, not page 48. Not a factual inaccuracy. The

fully evaluate the face validity of the results'	~~not reported the follow-up time in Part 2, which would help to more fully evaluate the face validity of the results.~~'		company's issue here mixes up the post-hoc analysis (PHA) with Part 2 of CASSIOPEIA which refers to the maintenance phase (as clearly stated in CS section B.2.3.1 and CS Table 9). The ERG report is correct in stating that the duration of follow-up for Part 2 is not reported in the CS. No change made.

Issue 11

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 56: 'As noted above (section 3.2.3.3), the proportion MRD-negative, which was assessed only post-consolidation , was determined regardless of response and exceeds the proportion with CR or better, favouring DBTd over BTd (Table 9).'	Please amend to: 'As noted above (section 3.2.3.3), the proportion MRD-negative, which was assessed post-induction and ~~only~~ post-consolidation, was determined regardless of response and exceeds the proportion with CR or better, favouring DBTd over BTd (Table 9).'	Misreporting of information from the CS.	Thank you for alerting us to this discrepancy. Given that post-induction MRD is not reported in ERG Table 9 we have made the following change on page 56: *As noted above (section **Error! Reference source not found.**), the proportion MRD-negative post-consolidation was determined regardless of response and exceeds the proportion with CR or better, favouring DBTd over BTd (**Error! Reference source not found.**).*

Issue 12

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 56, Table 9: Outcome & assessment time, MRD neg-ative (10^{-6}) [b]	Please amend to: Outcome & assessment time, MRD neg-ative (10^{-5}) [b]	Typographical error from the CS.	Thank you for alerting us to this typographic error in Table 9. We have corrected this.

Issue 13

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 79: 'In clarification response B10 the company present a sensitivity analysis comparing HRs for the impact of MRD status on PFS for MRD measured only in patients who achieved CR (3 trials) versus MRD measured in patients who achieved at least VGPR (3 trials).'	Please amend as follows: 'In clarification response B~~10~~12 the company present a sensitivity analysis comparing HRs for the impact of MRD status on PFS for MRD measured only in patients who achieved CR (3 trials) versus MRD measured in patients who achieved at least VGPR (3 trials).'	Typographical error from the clarification response.	Thank you for alerting us to this typographic error on page 79. We have corrected this.

Issue 14

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 89: 'Baseline characteristics of the modelled cohort are based on those of patients in CASSIOPEIA: mean age 56.6 years, 41.1% female	Please amend as follows: 'Baseline characteristics of the modelled cohort are based on those of patients in CASSIOPEIA: mean age 56.6 years,	Misreporting of data from the CS.	Thank you for alerting us to this typographic error on page 89. We have corrected this.

(see CS Tables 6 and 45).'	~~41.1%~~ 41.5% female (see CS Tables 6 and 45).'		

Issue 15

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 93, 'We also note that the KM data used for survival extrapolations are not adjusted for the second randomisation to maintenance treatment and that proportional hazards assumptions were not met for OS or PFS (clarification response A11).'	Please amend to: 'We also note that the KM data used for survival extrapolations are not adjusted for the second randomisation to maintenance treatment and that proportional hazards assumptions were not met for ~~OS or~~ PFS (clarification response A11).'	Misreporting of information from the CS.	Thank you for alerting us to this typographic error on page 93. We have corrected this.

Issue 16

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 109: 'The ERG also spotted an error in the calculation of average cost per model cycle for the CDF DBTd regimen and daratumumab monotherapy in CS Table 84 Scenario 5, which the company correct in their clarification response.'	Please amend to: 'The ERG also spotted an error in the calculation of average cost per model cycle for the CDF ~~D~~BTd regimen and daratumumab monotherapy in CS Table 84 Scenario 5, which the company correct in their clarification response	Typographical error from the clarification response.	Thank you for noticing this typographical error. The CDF treatments affected are DBd and daratumumab monotherapy (as in CS Table 76). We have corrected this on page 109.

Issue 17

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 115, Table 35: 'CASSIOPEIA, BTd, progression free survival, 1-year: 94%'	Please amend to: 'CASSIOPEIA, BTd, progression free survival, 1-year: ~~94%~~ 93%'	Misreporting of data from the CS.	Thank you for alerting us to this typographic error in Table 35. We have corrected this.

CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR Systematic Reviews Programme on behalf of NICE

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

ADDENDUM

ERG critique of the company's updated systematic literature review and meta-analysis of the impact of MRD status on survival outcomes (response to clarification question B3)

Produced by	Southampton Health Technology Assessments Centre (SHTAC)
Authors	As per main ERG report
Correspondence to	Dr Geoff Frampton Southampton Health Technology Assessments Centre (SHTAC) Wessex Institute Alpha House Enterprise Road, University of Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	26th November 2020

Markup: [redacted]

[redacted]

1 INTRODUCTION

The company submission (CS) reports a systematic literature review (SLR) and meta-analysis of the impact of minimal residual disease (MRD) status on overall survival (OS) and progression-free survival (PFS). Hazard ratios (HRs) (MRD-negative versus MRD-positive) for OS and PFS obtained from the meta-analysis inform the company's economic model (as explained in section 4.2.6.2 and summarised in Table 29 of the ERG report).

The SLR presented in the CS had some limitations which were addressed by the company in clarification responses B1 to B14. However, the company were unable to fully update their SLR and meta-analysis (clarification response B3) in time for the ERG to critique this within the ERG report. We have therefore provided the ERG's critique of the updated SLR and meta-analysis in this addendum to the ERG report.

As stated in the ERG report, the ERG consider uncertainty around the HRs obtained from the meta-analysis to be a key issue, since these HRs inform the economic model. The HRs are uncertain because:

- The SLR and meta-analysis as reported in the CS was 15 months out of date and might therefore have missed recent studies;
- The heterogeneity of studies, and hence their appropriateness for pooling in a meta-analysis, was considered for a limited range of population baseline characteristics (clarification response B9), omitting some key prognostic factors such as serum lactate dehydrogenase (LDH) and renal function.
- The company did not assess the validity (i.e. risk of bias) of studies included in the meta-analysis (NB the company subsequently provided validity assessment, for some of the studies, using the ROBINS-I tool, in clarification response B7).

2 ERG CRITIQUE METHODS

To enable the critique to be provided within the limited available timescale for the technology appraisal we have specifically focused our critique on the following aspects of the company's updated SLR and meta-analysis:

- A check that the search strategy is fit for purpose and that relevant studies are unlikely to have been missed.
- An evaluation of the company's overall approach to the statistical analysis and a check that the HRs obtained by the company from the included studies are consistent with the information reported in the study publications.

- An assessment of the heterogeneity of key prognostic factors (i.e. those factors, where reported, that could influence the appropriateness of combining studies in meta-analysis).
- An abbreviated assessment of the risks of bias for the included studies.

2.1 Critique of the searches and study selection

An experienced information specialist checked the company's search strategy and study selection process.

2.2 Critique of the statistical approach

An experienced statistician assessed the appropriateness of the company's overall meta-analysis approach and two reviewers independently checked the accuracy of the HRs reported by the company against the information reported in each study.

2.3 Assessment of clinical heterogeneity

The company's SLR Update Report presents limited baseline characteristics for the included studies, which excludes some prognostic factors such as serum LDH and renal function (Tables 5 and 6 in the SLR Update Report). Two reviewers checked whether further information on study characteristics is available in the study publications and whether the available information on participant characteristics is sufficiently homogeneous across the studies to justify statistical pooling of the studies in meta-analysis.

2.4 Assessment of risk of bias

The company assessed risks of bias in 14 of the 21 included studies using the ROBINS-I tool (clarification response B7). Given the limited timescale available to the ERG we have not provided a detailed critique of the company's ROBINS-I judgements but have instead focused on assessing the following key risks of bias in the studies:

- Risk of bias in the selection of the MRD status subgroups
- Risk of bias due to missing or unaccounted for data
- Risk of bias due to confounding (presence of variables that may affect both MRD status and survival outcomes) if this is not accounted for in analyses

Two ERG reviewers independently assessed each study for any indication that these types of bias risk were not adequately controlled for, recording answers, where possible, as "low" or "high" risk of bias. Where insufficient information was available to make a clear judgement we recorded the risk of bias as "unclear". A summary risk of bias judgement from both

reviewers was then tabulated alongside the company's risk of bias judgements for comparison. Note that the ERG's risk of bias classification (low/high/unclear risk) differs from the ROBINS-I tool classification which also permits a judgement of "moderate" risk of bias.

In addition, we comment on statistical conclusion validity (i.e. the reliability of the HRs reported in the original study publications) which is an important consideration for meta-analyses that synthesise existing effect estimates.

2.5 ERG sensitivity analyses

Based on the information obtained from steps 2.1 to 2.4 above we conducted sensitivity analyses to explore the impact of the different aspects of study heterogeneity and validity on the pooled HR estimates from the company's updated SLR. Frequentist analyses were conducted in STATA (version 16.1) on log-transformed HR estimates and their standard errors using a random-effects model. Statistical heterogeneity was assessed using the I^2 statistic, where I^2>50% is considered indicative of substantial heterogeneity.

Results of the ERG's sensitivity analyses were taken into consideration in the economic analysis (see section 4 below).

3 ERG CRITIQUE RESULTS

3.1 Searches and study selection

Searches were conducted in Medline, Embase and selected conferences which the ERG agree is appropriate. The search strategy, reported for Medline only, is consistent with that of the original SLR reported in CS Appendix M, but includes additional terms to focus on newly-diagnosed or untreated MM. Overall the search strategy and sources searched are broadly appropriate but not exhaustive. The majority of relevant studies are likely to have been identified.

The reported eligibility criteria (Table 2 in the SLR Update Report) are consistent with those provided in clarification response B4 which the ERG previously considered to be appropriate (ERG report section 3.6.1.2). A PRISMA chart is provided in Figure 1 of the SLR Update Report which shows that 149 references were excluded at full-text screening. However, a list of these excluded references and their reasons for exclusion is not provided, so the ERG are unable to check which studies were considered and whether the exclusion reasons were appropriate. The PRISMA chart further shows that following full-text screening 19 references

were eligible for inclusion in the meta-analysis, of which 11 were excluded, mainly due to lack of extractable hazard ratios. The reasons for excluding these 11 references are provided (Table 3 in the SLR Update Report) and appear appropriate.

In conclusion, although the search and selection process is broadly appropriate, the absence of information on which references were excluded at full text screening and why they were excluded is a key limitation. Overall, we believe that a substantial proportion of the relevant evidence has been identified and included, but the possibility that studies have been missed or selectively excluded cannot be ruled out. Sensitivity analyses conducted by the ERG (section 3.5 below) suggest that the company's meta-analysis results are relatively insensitive to the omission of individual studies.

3.2 Statistical approach

The company's SLR Update Report states only that "The meta-analysis was repeated using the same methodology as the meta-analysis previously reported" using a random effects model conducted on log-transformed values. We assume the company are referring to the method mentioned in clarification response B8, which states that the analysis approach repeated a method that is reported in an unpublished manuscript by Munshi et al. 2020 which has been accepted for publication. No PDF copy or any further information on the Munshi et al. 2020 manuscript was provided to the ERG. According to clarification response B8, the company's approach involved simulating individual patient data (IPD) to generate Kaplan-Meier curves and conducting subgroup analysis to address the potential bias caused by the differences in disease setting, eligibility for MRD assessment by conventional response, MRD assay and sensitivity and time of MRD assessment.

There are several computational steps required for the company's analyses, none of which are described. For instance:

- The studies varied in how much information they reported and whether data were missing. It is unclear how missing information was handled.
- The technique used to simulate IPD is not specified.
- Studies differed in their definitions of OS and PFS, which related to different timepoints (time since diagnosis, time since start of therapy, time since ASCT, time since a specified landmark). It is unclear how these differences were standardised.

The company's meta-analysis approach using simulated IPD is therefore a "black box" which cannot be checked; as we have only the original data reported in the studies, and the

company's estimated HRs (reported in Tables 9 and 10 of the SLR Update Report) without any of the intermediate calculations, assumptions, or data provided. However, inspection of the studies shows that where relevant HRs and their confidence intervals are directly reported in the study publications these have been used directly by the company, i.e. not requiring the IPD simulation approach.

Twelve of the 20 studies reporting PFS and six of the studies reporting OS provided relevant HRs in their study publications and we have verified that these agree with the company's HR estimates used in their meta-analysis (Appendix 1 below). In the remaining studies it is not possible to determine how the company's HR estimates and confidence intervals were obtained from the data reported in the study publications. In the case of the CASSIOPEIA trial, which provided HRs for both PFS and OS, the company derived their HRs from the landmark analysis reported in CS Figures 17 and 18 but employed confidential IPD so the ERG are unable to check the accuracy of the resulting HRs.

In conclusion, we were able to reproduce the company's frequentist random-effects meta-analysis results, indicating that the overall meta-analysis approach is appropriate. However, we could not verify the HRs used as input data for the analysis for 8/20 studies reporting MRD effect on PFS and 5/11 studies reporting MRD effect on OS, meaning that the pooled HR estimates and their confidence intervals are uncertain. We explored this uncertainty further in sensitivity analyses (see section 3.5 below).

3.3 Clinical heterogeneity

The company have tabulated nine participant baseline characteristics for each of the included studies (see Table 1 below). We checked the study publications to determine whether any further patient baseline characteristics could be extracted, particularly for further prognostic factors such as serum LDH and renal function (serum creatinine).

Of the total 21 studies that report PFS and/or OS, only 3 report serum LDH, 5 report serum creatinine, and 6 report β2 microglobulin. However, the studies differ in the measures reported for these prognostic factors meaning that no consistent comparisons can be made across studies. We therefore agree that the company's tabulation of baseline characteristics represents the best available data on which to assess study clinical heterogeneity. An overview of how these nine baseline characteristics vary across the studies is given in Table 1 below.

Table 1 Summary of patient baseline characteristics across studies

Baseline characteristic	ERG comments
Median age (years)	Median age (not reported in Hahn, 2019) ranged from 54 to 63 years with a modal age of 58 years and no discernible outliers.
Sex (% male)	The % male (not reported in 3 studies) ranged from 52% to 68% in most studies apart from two outliers (37% in Sololev, 2016 and 46% in Parrondo, 2019).
ISS class (I, II, III)	Six studies did not report the ISS class, and in a further 6 studies some ISS data were missing. Among the majority of studies the proportions of patients were generally spread across classes I (lowest risk) to III (highest risk). Notable exceptions are Cohen, 2016 and Clark, 2018 which had ≤5% in class III (although in Cohen 43% of data overall were missing). Rossi, 2018 and CASSIOPEIA also had a relatively small proportion in class III (10% and 15% respectively). Notably, in Rossi, 2018 most patients (83%) were in class I.
Type of measurable disease (IgG, IgA, other)	Around half the studies (n=11) reported the type of measurable disease which was generally consistent across the studies, ranging from 55% to 68% IgG, 12% to 35% IgA, and 4% to 30% other type.
High cytogenetic risk (%)	Thirteen studies reported the proportion of patients with high cytogenetic risk, which ranged from 14% to 100%. Two studies with 100% of patients in the high cytogenetic risk group (Chakraborty, 2019 and Parrondo, 2019) are outliers, with all remaining studies having 14% to 47% patients in the high risk group.
Eligibility for MRD assessment by response (any response, CR only, at least CR, VGPR, at least VGPR)	Nine studies assessed MRD status regardless of response. The remaining studies required a very good partial response (VGPR) (n=1) or at least VGPR (n=4), or a complete response (CR) (n=4) whilst three studies did not report the eligibility for MRD assessment by response. We conducted an exploratory subgroup analysis comparing the impact or response eligibility on MRD HRs for PFS (see Appendix 2). Studies requiring at least VGPR have a lower HR than those permitting any

	response or requiring CR, but with overlapping confidence intervals.
MRD assay type	Most (n=18) studies assessed MRD status using multiparametric flow cytometry (MFC). One study used next-generation sequencing (NGS), one used a polymerase chain reaction (PCR) assay and one study did not report the method.
MRD assay sensitivity (10^{-4}, 10^{-5})	Six MFC studies used a 10^{-5} sensitivity, 11 MFC studies used a 10^{-4} sensitivity and one MFC study did not report the sensitivity. The remaining three studies used NGS with a 10^{-5} sensitivity, PCR with a 10^{-4} sensitivity, or did not report any details of the assay. We conducted an exploratory subgroup analysis comparing the impact of MFC assay sensitivity on HRs (see Appendix 2). Whilst studies with the higher (10^{-5}) sensitivity have a slightly lower HR, confidence intervals are relatively wide, indicating uncertainty of these HRs.
Treatment	The studies varied considerably in the therapy given to patients, as reported in Table 6 in the SLR Update Report. The extent to which the therapy was reported for induction, consolidation and maintenance also varied across the studies. It is not possible to determine the effect of therapy on the MRD status HRs since nearly all studies had different therapies.

Overall, despite the company's claim in clarification response B2 that the studies are generally homogeneous, there is evidence for substantial heterogeneity across studies, e.g. in ISS scores and the therapies received. This heterogeneity is not easy to capture in sensitivity or subgroup analyses since individual studies are subject to multiple possible limitations (e.g. see section 3.5 below). The limited exploratory subgroup analyses that were feasible (Appendix 2) did not detect any clear systematic effects of study heterogeneity on HRs.

The company do not discuss heterogeneity in the number of ASCT transplants conducted. We note that tandem ASCT transplants were referred to in five studies:

- Rossi, 2019: all participants received tandem transplants
- Sololev, 2018: 13/70 patients received a tandem transplant
- Sololev, 2016: 16/52 patients received a tandem transplant

- Hahn, 2019: The study was based on a trial in which patients in one arm received a single transplant whilst patients in another arm received a double transplant, but it is not reported which arm(s) the MRD HR was calculated from.
- Bakkus, 2004: Patients were eligible for double transplant but it is not reported whether any double transplants took place.

Given the differences in these studies it is difficult to get a clear picture of whether having a second transplant would affect the pooled meta-analysis MRD HRs. The one study in which all patients received a second transplant (Rossi, 2019) is small (16 patients achieved VGPR, of which only 3 achieved MRD negativity) and is not influential in the PFS meta-analysis.

In conclusion, we believe that the company have made the best quantitative use of the available data in their meta-analysis, but an inevitable consequence of including relatively heterogeneous studies is that the pooled HR estimates are subject to additional uncertainty which is not reflected in the confidence intervals.

3.4 Risk of bias

The company's and ERG's risk of bias assessments are summarised in Appendix 3 below.

3.4.1 Company assessments

The company concluded that of the 14 studies they assessed, 9 have moderate risk of bias and 5 have low risk of bias, although they do not explain how these classifications should be interpreted for meta-analysis. Of the five studies that the company judged to have low risk of bias, the ERG's opinion is that each of these had an unclear risk of selection bias, attrition bias and/or confounding (Appendix 3).

In addition to assessing the risks of bias that are inherent in the original studies, the company have acknowledged in their judgements that their approach for estimating HRs from the studies using simulated IPD is also a source of bias, since their method of IPD simulation has limitations. As noted above (section 2.2), no details of the company's IPD simulation method have been provided. The company's "black box" IPD simulation approach for estimating HRs is applicable to those studies which do not directly report relevant HRs as well as their confidence intervals and/or standard errors (approximately half of the studies) (Appendix 1) (excluding CASSIOPEIA, for which the company had access to real IPD).

3.4.2 ERG assessments

For most of the 21 studies we were unable to make a definitive risk of bias judgement due to inadequate reporting of methodology in the study publications. Overall, 19/21 studies were judged to have an unclear risk of selection bias, attrition bias and/or confounding, with the remaining two (Chakraborty, 2017 and Parrondo, 2019) considered to be at high risk of selection bias due to their retrospective designs (which increases the risk of preferential selection of cases) (Appendix 3). In summary:

- Selection bias: 15/21 studies were judged to have unclear risk, 4/21 low risk and 2/21 high risk.
- Attrition bias: 17/21 studies were judged to have unclear risk and 4/21 to have low risk.
- Confounding: 18/21 studies were judged to have unclear risk of confounding, with 3/21 having low risk.

Risk of bias judgements are inherently subjective, and some disagreement between the company and ERG is not surprising, especially since information in the papers was sometimes ambiguous or incomplete (10 of the 21 studies were reported only in abstracts).

3.4.3 Statistical conclusion validity

For meta-analyses that synthesise existing HRs the reliability of the original HRs reported in the study publications should be considered. We note that a landmark analysis was deemed necessary in the CASSIOPEIA trial to prevent the risk of immortal time bias (CS section B.2.6.3), but the company do not specify whether they also required studies included in their meta-analyses to have used a landmark analysis approach. Four of the included studies referred to landmark analyses, as follows:

- Rawstron, 2013: All statistical analyses were landmarked from the date of MRD assessment (no specific methodological details given).
- Paiva, 2008: Analyses were landmarked from day 100 after transplant (no specific methodological details given); however, it is unclear whether the MRD HR from this study used by the company was from this analysis.
- Schinke, 2017: The HRs used by the company from this study appear to be based on a landmark analysis at end of consolidation (reported as an 8 months assessment in the publication).
- Chakraborty, 2017: A landmark analysis was conducted 1 year post-transplant, although the MRD HRs from this study for the current meta-analyses are taken from a standard non-landmark analysis 3 months post-transplant.

The fact that most of the included studies did not state explicitly that they used a landmark analysis, and those studies that specified a landmark approach did not explicitly define it, could mean that immortal time bias has not been fully accounted for in the meta-analysis. However, when the CASSIOPEIA trial (which employed a landmark analysis) was included and excluded in sensitivity analyses this had little impact on the overall HRs (see section 3.5 below).

In conclusion, the majority of studies were rated as having a similar (i.e. unclear) risk of bias in our assessments, so there are limited opportunities for subgroup or sensitivity analyses to explore the impact of risks of bias in the meta-analysis. We conducted a sensitivity analysis to explore the impact of excluding studies for which we could not verify the hazard ratios (i.e. those studies that the company regarded as being at moderate risk of bias due to requiring simulation of IPD to estimate the HRs); and we conducted a sensitivity analysis to explore the impact of excluding the two retrospective studies which we judged to be at high risk of selection bias (see section 3.5 below).

3.5 ERG sensitivity analyses

Based on the methodological issues identified in sections 3.2 to 3.4 above we conducted the following sensitivity analyses to explore the impact of study limitations on the results of the meta-analyses for PFS and OS:

(1) Retaining only those studies in the analysis that directly reported relevant HRs in the study publication. This analysis excludes the uncertainty associated with the company's IPD simulation approach for estimating HRs. As shown in Appendix 1 there were 7 PFS studies and 4 OS studies where the HR is not directly reported and these were excluded in this sensitivity analysis.

(2) Of the studies remaining in sensitivity analysis (1), we were unable to verify the confidence intervals for the HR in 2 studies (as shown in Appendix 1). We therefore excluded these studies to eliminate possible uncertainty around the accuracy of these confidence intervals.

(3) Three of the studies (Chakraborty, 2017; Gu, 2018; Luoma, 2019) reported HRs based on both univariate and multivariate analyses, and the multivariate analyses give more conservative results (higher HRs and/or wider confidence intervals) (Appendix 1). The

company had selected the HRs based on the univariate analyses. In this sensitivity analyses we replaced the univariate HRs with the multivariate HRs from these studies (2 PFS studies and 2 OS studies).

(4) As noted above (section 3.4), two of the studies were retrospective and therefore potentially at high risk of selection bias. We excluded these two studies in this sensitivity analysis. Note that this is an illustrative analysis because it is not possible to say with certainty that these were biased studies; the risk of bias is inferred from a lack of study features that would protect against preferential patient selection.

For each of these sensitivity analyses we ran the analysis (a) with CASSIOPEIA included and (b) with CASSIOPEIA excluded, given that this is the pivotal trial, utilised a landmark analysis approach, and has a large weight in the analysis.

Results of these sensitivity analyses are presented in Tables 3 and 4 below. We conducted several other sensitivity analyses (e.g. excluding studies for which we could not verify the timing of the MRD assessment; excluding the two outlier studies which had 100% of patients with high cytogenetic risk; and excluding studies only reported in abstracts) but these gave similar results to those shown in Tables 3 and 4 and are not presented here.

Table 3 MRD HR estimates for PFS from the company's meta-analyses and ERG sensitivity analyses

MRD impact on PFS		No. of studies	Pooled HR	Lower 95% CI	Upper 95% CI	I^2
Company's original SLR		15	[redacted]			22.4%
Company's updated SLR		20				59.84%
ERG sensitivity analyses (a) including CASSIOPEIA, (b) excluding CASSIOPEIA						
(1) Excluding 7 studies for which ERG could not verify the HR	**(a)**	13	0.32	0.24	0.43	68.00%
	(b)	12	0.30	0.22	0.42	67.21%
(2) Analysis #1 repeated excluding a further 2 studies for which ERG could not verify the 95% CI or SE of the HR	**(a)**	11	0.31	0.21	0.44	69.23%
	(b)	10	0.29	0.19	0.43	67.17%
(3) Analysis #2 repeated using a more conservative adjusted HR where available in 3 studies	**(a)**	11	0.32	0.22	0.47	67.83%
	(b)	10	0.30	0.20	0.46	65.99%
	(a)	9	0.29	0.18	0.45	68.83%

(4) Analysis #3 repeated excluding 2 retrospective studies						
	(b)	8	0.26	0.16	0.43	63.50%

Table 4 MRD HR estimates for OS from the company's meta-analyses and ERG sensitivity anlayses

MRD impact on OS		**No. of studies**	**Pooled HR**	**Lower 95% CI**	**Upper 95% CI**	**I^2**
Company's original SLR		9	[redacted]			17.3%
Company's updated SLR		11				15.97%
ERG sensitivity analyses (a) including CASSIOPEIA, (b) excluding CASSIOPEIA						
(1) Excluding 4 studies for which ERG could not verify the HR	**(a)**	7	0.54	0.42	0.70	37.99%
	(b)	6	0.49	0.39	0.60	0.00%
(2) Analysis #1 repeated excluding a further 2 studies for which ERG could not verify the 95% CI or SE of the HR	**(a)**	5	0.46	0.28	0.75	57.50
	(b)	4	0.37	0.21	0.64	29.25%
(3) Analysis #2 repeated using a more conservative adjusted HR where available in 2 studies	**(a)**	5	0.63	0.49	0.82	5.70%
	(b)	4	0.48	0.29	0.79	8.94%
(4) Analysis #3 repeated excluding 1 retrospective study	**(a)**	4	0.49	0.27	0.89	37.52%
	(b)	3	0.31	0.14	0.68	0.00%

As can be seen in Tables 3 and 4, the pooled HRs from the meta-analyses for PFS and OS are mostly robust to the sensitivity analyses, but with high heterogeneity (I^2>50%) for the PFS analyses. The inclusion of CASSIOPEIA gives marginally higher HRs in all cases than when the trial is excluded, but with relatively wide confidence intervals. The exclusion of studies with non-verified HRs has relatively little impact. This suggests that the company's approach to estimating HRs using simulated IPD, although associated with uncertainty, has not generated substantive systematic error. Note, however, that the results of the meta-analyses are subject to an untested assumption that those HRs which are reported directly in the study publications are reliable.

The influence of the alternative HRs from the ERG's sensitivity analyses on the cost-effectiveness results is shown in section 4 below.

4 IMPLICATIONS FOR THE ECONOMIC ANALYSIS

The sensitivity of the company's base case ICER to changes in MRD HR estimates from the updated SLR is shown in **Table** 5 below. The ICER is around £1,000 per QALY higher with the updated PFS and OS HRs from the company's meta-analysis. The PFS HR is robust to the ERG sensitivity analysis presented above, and exclusion of selected studies does not reduce heterogeneity. The OS HRs are less stable, ranging from [redacted] in ERG analyses 3a and 4b. Table 5 shows that the company's base case ICER increases with lower OS HR estimates, but that the ICER remains below the £30,000 per QALY threshold across the wide range of HR estimates tested.

Table 5 Scenarios for updated MRD SLR: company base case deterministic analysis

Scenario		Incremental Costs	Incremental QALYs	ICER (£/QALY)
Company base case Original SLR: PFS [redacted] and OS [redacted]		[redacted]		
Relative survival MRD-negative versus MRD-positive				
ERG16	HRs from updated SLR: PFS [redacted] and OS [redacted]	[redacted]		
ERG17	OS HR scenario 3a upper limit: PFS [redacted] and OS [redacted]			
ERG18	OS HR scenario 3a mean: PFS [redacted] and OS [redacted]			
ERG19	OS HR scenario 4b mean: PFS [redacted] and OS [redacted]			
ERG20	OS HR scenario 4b lower limit: PFS [redacted] and OS [redacted]			

The impact of the company's updated MRD SLR for the ERG base case is shown in Table 6 below. In this case, the ICER is lower with the updated MRD HR estimates and the declines with increasing OS HRs. This difference in the direction of effect compared with that shown above for the company base case, results from the interaction of the MRD HRs with assumptions about the persistence of daratumumab treatment effects. However, as with the analysis for the company's base case, interpretation of the cost-effectiveness results with ERG preferred assumptions does not change at a cost effectiveness threshold of £30,000 per QALY gained: the ICER remains above £30,000 per QALY gained across the range of OS HRs tested.

Table 6 Scenarios for updated MRD SLR: ERG base case, deterministic analysis

Scenario		Incremental Costs	Incremental QALYs	ICER (£/QALY)
ERG preferred analysis Original SLR: PFS [redacted] and OS [redacted]		[redacted]		
Relative survival MRD-negative versus MRD-positive				
ERG16	HRs from updated SLR: PFS [redacted] and OS [redacted]	[redacted]		
ERG17	OS HR scenario 3a upper limit: PFS [redacted] and OS [redacted]			
ERG18	OS HR scenario 3a mean: PFS [redacted] and OS [redacted]			
ERG19	OS HR scenario 4b mean: PFS [redacted] and OS [redacted]			
ERG20	OS HR scenario 4b lower limit: PFS [redacted] and OS [redacted]			

5 REFERENCES

For references to the studies included in the MRD meta-analyses please refer to Table 5 in the company's Updated SLR Report.

6 APPENDICES

APPENDIX 1 Input data for the company's meta-analyses, showing data verified by the ERG (green cells) and data that the ERG were unable to verify (red cells)

(a) PFS meta-analysis

Study	HR PFS	HR PFS 95% LL	HR PFS 95% UL	HR PFS log	HR PFS SE log
Rawstron, 2013	0.56	0.42	0.75	-0.59	0.15
Paiva, 2008	0.28	0.16	0.40	-1.29	0.23
Popat, 2017 (arm 1)	0.38	0.13	1.15	-0.96	0.56
Cohen, 2016	0.28	0.10	0.75	-1.28	0.51
Clark, 2018	0.13	0.02	0.81	-2.02	0.92
Chakraborty, 2017	0.45	0.31	0.66	-0.80	0.19
Schinke, 2017	0.55	0.22	1.37	-0.61	0.47
Bakkus, 2004	0.27	0.14	0.55	-1.29	0.36
Rawstron, 2002	0.31	0.13	0.73	-1.18	0.44
Gu, 2018	0.29	0.13	0.65	-1.24	0.41
Rossi, 2019	0.13	0.03	0.51	-2.07	0.71
CASSIOPEIA (BTd)					
Hahn, 2019	0.48	0.31	0.74	-0.73	0.22
Solovev, 2018	0.50	0.21	2.00	-0.70	0.58
Solovev, 2016	0.59	0.29	0.83	-0.53	0.27
Luoma 2019	0.23	0.11	0.48	-1.46	0.37
Ribolla 2020	0.20	0.07	0.54	-1.61	0.52
Parrondo 2019	0.44	0.23	0.86	-0.82	0.34
Paiva 2020	0.12	0.08	0.19	-2.14	0.24
Garifullin 2019	0.33	0.83	11.39	-1.12	0.67
Alternative multivariate HR estimates used in ERG sensitivity analyses for PFS					
Gu, 2018 univariate	0.29	0.13	0.65	-1.24	0.41
Gu, 2018 multivariate	0.26	0.09	0.76	-1.35	0.54
Chakraborty, univariate	0.45	0.31	0.66	-0.80	0.19
Chakraborty, multivariate	0.49	0.31	0.77	-0.71	0.23
Luoma 2019 univariate	0.23	0.11	0.48	-1.46	0.37
Luoma 2019 multivariate	0.40	0.18	0.87	-0.92	0.40

(b) OS meta-analysis

Study	HR OS	HR OS 95% LL	HR OS 95% UL	HR OS log	HR OS SE log
Rawstron, 2013	0.63	0.43	0.90	-0.47	0.19
Paiva, 2008	0.50	0.42	0.71	-0.70	0.13
Chakraborty, 2017	0.55	0.32	0.92	-0.60	0.27
Schinke, 2017	0.84	0.30	2.39	-0.17	0.53
Bakkus, 2004	0.59	0.21	1.65	-0.53	0.53
Rawstron, 2002	0.53	0.16	1.76	-0.63	0.61
Gu, 2018	0.23	0.09	0.59	-1.47	0.48
CASSIOPEIA (BTd)					
Hahn, 2019	0.77	0.35	1.67	-0.26	0.40
Ribolla, 2020	0.20	0.04	0.99	-1.61	0.82
Chan, 2019	0.34	0.10	1.20	-1.08	0.63
Alternative multivariate HR estimates used in ERG sensitivity analyses for OS					
Gu, 2018 univariate	0.23	0.09	0.59	-1.47	0.48
Gu, 2018 multivariate	0.38	0.11	1.34	-0.97	0.64
Chakraborty, univariate	0.55	0.32	0.92	-0.60	0.27
Chakraborty, multivariate	0.63	0.36	1.10	-0.46	0.29

APPENDIX 2 Analyses of the impact of heterogeneity in patient baseline characteristics on HRs

Any systematic impacts of clinical heterogeneity on MRD HRs should be detectable in subgroup analyses. Among the baseline characteristics available (Table 1 above) it was only feasible to investigate the effects of the MRD assay sensitivity (Table A2 below) and the eligible response for MRD assessment (Table A3 below). Although there are some differences in HRS, the confidence intervals overlap in all cases.

Table A2 Subgroup analysis of MRD effect by MFC assay sensitivity

Outcome	MFC assay sensitivity	No. of studies	Pooled HR	Lower 95% CI	Upper 95% CI	I^2
MRD effect on PFS	10^{-5}	6	0.30	0.18	0.51	77.49%
	10^{-4}	10	0.34	0.26	0.46	43.77%
MRD effect on OS	10^{-5}	4	0.49	0.26	0.92	64.08%
	10^{-4}	5	0.53	0.44	0.65	0%

Table A3 Subgroup analysis of MRD effect by eligible response for MRD assessment

Outcome	Eligible response	No. of studies	Pooled HR	Lower 95% CI	Upper 95% CI	I^2
MRD effect on PFS	Any response	8	0.41	0.34	0.51	31.93%
	At least VGPR	4	0.22	0.10	0.46	68.76%
	CR	4	0.41	0.25	0.68	36.45%
NB insufficient data for a subgroup analysis on OS						

APPENDIX 3 Company and ERG risk of bias assessments

Study	Company assessment (ROBINS-I tool)	ERG comments	
		MRD group selection and attrition	**Confounding**
Rawstron, 2013 (full paper)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics of patients included in the MRD assessment were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	**Unclear risk of bias** MRD patient subgroup not defined (unclear risk of selection bias), so unclear whether any data missing (unclear risk of attrition bias).	**Unclear risk of bias** MRD status was associated with the treatment received so this could be a confounder. Patient characteristics at baseline not reported. KM plots only reported (not HRs) and do not appear to have been adjusted for covariates.
Paiva, 2008 (full paper)	**Low risk of bias** HR was reported, statistical methods were clearly described. Baseline characteristics were reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted multiple subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	**Unclear risk of bias** Rationale for patient selection reported. PFS compared in those with and without an MRD assessment and no difference (low risk of selection bias). Not reported whether any data missing during follow up (unclear risk of attrition bias)	**Low risk of bias** Multivariate analysis adjusted for wide range of relevant confounders
Popat, 2007 (abstract only)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. Potential sources of confounding were described but not addressed. However,	**Unclear risk of bias** Not all study participants assessed for MRD (reasons not stated) (unclear risk of selection	**Unclear risk of bias** Potential confounders not mentioned and do not appear to have been included in analyses.

	the population was uniform in terms of known confounders: all patients achieved at least CR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment.	bias). Not reported whether any data missing during follow up (unclear risk of attrition bias).	
Cohen, 2016 (abstract only)	**Moderate risk of bias** HR was reported; however, the statistical methods were not clearly described. The population was not uniform in terms of the treatment received. The conventional responses are unknown. The authors did not conduct subgroup or interaction analyses, or discuss the limitations of the MRD analysis.	**Unclear risk of bias** Rationale for patient selection not reported (unclear risk of selection bias). All enrolled patients accounted for (low risk of attrition bias)	**Unclear risk of bias** PFS Figure missing from paper. Potential confounders not mentioned and do not appear to have been included in analyses.
Clark, 2018 (full paper)	**Low risk of bias** HR was reported, statistical methods were clearly described. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors did not conduct subgroup and interaction analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.	**Unclear risk of bias** Rationale for patient selection not reported (unclear risk of selection bias). Not reported whether any data missing during follow up (unclear risk of attrition bias)	**Unclear risk of bias** Cox proportional hazards model appears only to have included MRD status and ISS (stage I vs I/III). Baseline characteristics reported per MRD subgroup but small sample size (n=7) for MRD-positive so difficult to assess homogeneity of the MRD status subgroups. Paper states study was underpowered to evaluate the impact of cytogenetics on MRD status.
Chakraborty, 2017 (full paper)	**Low risk of bias** HR was reported, statistical methods were clearly described. Background characteristics were clearly described. The population was uniform in terms of cytogenetic risk, but not conventional response. It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. The authors conducted subgroup analyses for MRD. Both direction and magnitude of any potential bias were discussed in the publication.	**High risk of bias** Retrospective study (high risk of selection bias). No information on missing data (unclear risk of attrition bias).	**Low risk of bias** Analysis adjusted for age, ISS, response level, post-transplant maintenance therapy. ISS and age not included in model after univariate analysis. Only maintenance treatment significant in final model.

Schinke, 2017 (full paper)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was uniform in terms of known confounders: all patients achieved at least VGPR, MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. The authors conducted several subgroup analyses.	**Unclear risk of bias** Rationale and number of patients selected for MRD assessment stated (low risk of selection bias) but not reported whether any data were missing during follow up (unclear risk of attrition bias).	**Unclear risk of bias** KM curves reported without HR; unclear whether any adjustment made for potential confounders
Bakkus, 2004 (full paper)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	**Unclear risk of bias** Rationale for selection and numbers eligible for MRD assessment stated but unclear whether any patients received a tandem ASCT (unclear risk of selection bias). Not reported whether any data were missing during follow up (unclear risk of attrition bias).	**Unclear risk of bias** Cox model used with relevant confounders but appears to be for PFS analysis only. Baseline covariates for MRD groups given in Table 1 but appear unbalanced, so difficult to assess effect of confounding on the OS HR estimate
Rawstron, 2002 (full paper)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The background characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR outcome. However, the authors conducted several subgroup and interaction analyses, discussed the limitations of the study, taking into account the sources of potential bias or imprecision. Both direction and magnitude of any potential bias were discussed in the publication.	**Unclear risk of bias** Patient selection not reported (unclear risk of selection bias). Missing data not reported (unclear risk of attrition bias).	**Unclear risk of bias** Multivariate analysis conducted adjusting for Hb, β2 microglobulin, MRD status and response (age and sex not significant in univariate), but other prognostic factors (e.g. treatment, ISS) not included
Gu, 2018 (full paper)	**Low risk of bias** HR was reported, statistical methods were clearly described. Background characteristics were extensively reported. The population was uniform in terms of most of the confounders. The authors	**Unclear risk of bias** Selection of patient subgroup not defined (unclear risk of selection	**Unclear risk of bias** Analysis adjusted for age, ISS, LDH and cytogenetic factors identified by interphase fluorescence in situ

	conducted subgroup analyses on MRD. Both direction and magnitude of any potential bias were discussed in the publication.	bias). No missing MRD data (low risk of attrition bias).	hybridization, but not adjusted for treatment or response (some variation evident in response and treatments received). Patients with high cytogenetic risk should be 19%, not 59% as stated in SLR Update Report Table 5.
Rossi, 2019 (full paper)	**Moderate risk of bias** HR was reported. Background characteristics were reported. The population was uniform in terms of most of the confounders (conventional response, assay, sensitivity, treatment). The authors conducted subgroup analyses on time of MRD assessment only, however, the subgroups were not well defined. Both direction and magnitude of other potential bias were discussed in the publication, but not addressed in the analyses.	**Unclear risk of bias** Patient selection adequately described, but small sample size (16 patients achieved VGPR, of which only 3 achieved MRD negativity) (unclear risk of selection bias). Sample sizes not reported for each analysis time point so unclear whether any missing data (unclear risk of attrition bias).	**Unclear risk of bias** Cox proportional hazards analyses included tumour reduction post-induction, post-ASCT1 and post ASCT2; MRD at same 3 timepoints; ISS (=1 or >1), presence or absence of adverse cytogenetic and phenotype profile, MMR > 0.01% (not defined, typo?) at the 3 timepoints. Description/categories for some covariates unclear.
CASSIOPEIA (full report)	Not assessed by the company using the ROBINS-I tool (NB the company did assess risk of bias in the parent CASSIOPEIA trial [CS section B.2.5], but this is not relevant to the MRD subgroup analysis)	**Low risk of bias** MRD group selection clear and no missing data; large proportion of patients had a MRD assessment (BTd arm) (low risk of selection bias). Company state in clarification response A9 that no data missing from landmark analysis (low risk of attrition bias).	**Unclear risk of bias** Distribution of prognostic factors between MRD groups not provided and unclear what the Cox model was adjusted for.
Hahn, 2019 (abstract only)	**Moderate risk of bias** HR was reported, but statistical methods were not clearly described. Baseline characteristics were not reported. The population was not uniform in terms of some confounders: patients eligible for MRD assessment had different conventional response (mix of CR, VGPR and less than VGPR). It is known that the distribution of patients by conventional response in MRD+ arm has an influence on the HR	**Unclear risk of bias** Patient selection not reported. Trial had single-ASCT and tandem-ASCT arms but not clear which arm(s) MRD HR was calculated for (unclear risk of selection bias). Missing data not	**Unclear risk of bias** Multivariate analysis adjusted for disease risk but not described further, so unclear whether other factors adjusted for. Source of ISS data reported by company in SLR Update Report Table 5 unclear.

	outcome. The authors did not conduct subgroup and interaction analyses for MRD, other than time of MRD assessment. Neither direction nor magnitude of a potential bias (other than time of assessment) were discussed in the publication.	reported (unclear risk of attrition bias).	
Solovev, 2018 (abstract only)	**Moderate risk of bias** HR was not reported. HR was estimated based on the simulated IPD and reconstructed KM curve, which can introduce some bias. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. However, the baseline characteristics were not well described.	**Unclear risk of bias** Patient selection described adequately but limited baseline characteristics reported (unclear risk of selection bias). No missing patients at 2-year assessment (low risk of attrition bias).	**Unclear risk of bias** Limited information reported, potential confounding factors do not appear to have been analysed
Solovev, 2016 (abstract only)	**Low risk of bias** The authors reported that the compared groups were overall balanced for known prognostic factors. The population was very uniform in terms of known confounders: all patients achieved at least CR (no other response), MRD was measured using the same assay, sensitivity and at the same time, all patients received the same treatment. HR was reported in the publication, however, the method of HR estimation was not specified.	**Unclear risk of bias** Patient selection described but limited baseline characteristics reported (unclear risk of selection bias), Not reported whether any data missing during follow up (unclear risk of attrition bias).	**Unclear risk of bias** Limited information reported, potential confounding factors do not appear to have been analysed
Luoma, 2019 (full paper)	Not assessed by the company	**Unclear risk of bias** Patient selection clearly described (low risk of selection bias). Intention to treat analysis conducted but unclear whether this applied to the univariate and multivariate HR analyses (unclear risk of attrition bias).	**Low risk of bias** Although not explicit in the statistics methods section, HR analyses appear to have been adjusted for age, treatment, IMWG risk group, ISS risk group (both ISS and revised ISS), high-risk cytogenetics and best serological response
Ribolla, 2020 (abstract only)	Not assessed by the company	**Unclear risk of bias** Patient selection not reported (unclear risk of selection bias). and amount of missing data not reported Not reported whether	**Unclear risk of bias** Paper states the MRD subgroups did not differ on patient characteristics, therapy, and

		any data missing during follow up (unclear risk of attrition bias).	cytogenetic risk, but no supporting data provided.
Parrondo, 2019 (abstract only)	Not assessed by the company	**High risk of bias** Retrospective study and patient selection not reported (high risk of selection bias). Not stated whether all patients in study were included in HR analyses (unclear risk of attrition bias).	**Unclear risk of bias** Analysis indicated PFS HR was affected individually by MRD negativity, induction therapy, consolidation therapy and tandem ASCT but not stated whether these or any other confounding factors were collectively adjusted for
Paiva, 2020 (abstract only)	Not assessed by the company	**Unclear risk of bias** Unclear why not all patients with available bone marrow aspirates were analysed for MRD, and unclear why number with MRD classifications is smaller than the stated number analysed (unclear risk of selection bias). Not reported whether any data missing during follow up (unclear risk of attrition bias).	**Unclear risk of bias** Limited information reported, potential confounding factors other than treatment do not appear to have been analysed
Garifullin, 2019 (abstract only)	Not assessed by the company	**Unclear risk of bias** Rationale for patient selection not reported (unclear risk of selection bias). Not reported whether any data missing during follow up (unclear risk of attrition bias	**Unclear risk of bias** Limited information reported; unclear whether any confounding factors were adjusted for
Chan, 2019 (abstract only)	Not assessed by the company	**Unclear risk of bias** Cohort not well described: HR appears to related to 40 patients but not explicit (unclear risk of selection bias). Amount of missing data not reported (unclear risk of attrition bias).	**Unclear risk of bias** KM curves reported without HR; unclear whether any adjustment made for potential confounders

NICE National Institute for Health and Care Excellence

Technical engagement response form

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments **5:00pm, 11 January 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response 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.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. 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.

- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- 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 replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, 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 during engagement 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.

About you

Your name	**Keith Stubbs**
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Janssen-Cilag Ltd**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**N/A**

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Uncertainty in hazard ratios from the company's meta-analysis of the effects of minimum residual disease (MRD) status on survival outcomes	**YES**	As part of the Janssen clarification response, an SLR update covering the period from June 1, 2019 to present (October 24, 2020) was submitted to NICE and reviewed as an Addendum to the ERG report. Consistent with the SLR update to June 2019, the latest update included a review of MEDLINE (via PubMed) and EMBASE databases in addition to a hand search of conference websites (ASH, ASCO and EHA) to ensure that all relevant material were identified. The SLR update identified eight publications that were eligible for inclusion in the meta-analysis, adding five progression-free survival (PFS) and two overall survival (OS) observations to the base case. Two further studies were included as a sensitivity analysis. The meta-analysis was repeated using the same methodology as the original meta-analysis and as reported by Munshi et al. 2020.[1] A random effects model was fitted to obtain a pooled effect estimate of the hazard ratio (HR) for MRD negativity versus MRD positivity. Heterogeneity in design and population among the studies eligible for meta-analysis was assessed with I^2 test using the maximum likelihood (ML) estimator. The statistical significance level was set at $p<0.05$. The base case analysis for PFS was based on 20 observations and resulted in a hazard ratio (HR) of [redacted]. The heterogeneity was estimated at [redacted] with a p-value of [redacted] based on a chi-square distribution. The base case analysis for OS was based on 11 observations and resulted in the HR of [redacted]. The heterogeneity was estimated at [redacted], with a p-value of [redacted] based on a chi-square distribution.

		Whilst the HR for PFS has improved [redacted] with the SLR update, the HR for OS remained broadly stable [redacted], supporting the robustness and overall reliability of the results. Following the SLR update, Janssen extended the risk of bias assessment (originally submitted 16th of October 2020) to include the newly identified studies. This assessment identified two publications as being at a serious risk of bias. As a result, Janssen performed an additional sensitivity analysis to exclude results from the related studies. Again, consistent results were reported with a PFS HR of [redacted] based on 19 observations and an OS HR of [redacted] based on 10 observations. Refer to the finalised SLR/meta-analysis update, risk of bias assessment and further clarification response included as supporting references.[2,3,4]
Key issue 2: Inconsistency in the company's approach for defining and analysing MRD-negative patients	**YES**	As per Janssen's factual accuracy response (Issue 1), the CASSIOPEIA study protocol included an amendment to the eligibility criteria for MRD assessment for Part 1 to include all patients in the induction/consolidation phases regardless of response. This reflects the fact that, when the study design for CASSIOPEIA was being finalised in 2015, the International Myeloma Working Group (IMWG) guidelines were not clear on which patients should be tested for MRD. A broader collection was therefore chosen as it had the potential to provide more information on MRD negativity regardless of response. With the changing landscape, Janssen adapted the analysis and analysed the MRD data in line with the IMWG guidelines of MRD assessment only in patients achieving a minimum complete response (CR). An analysis applying the IMWG guidance for assessing MRD negativity was performed as a post-hoc analysis and presented in Company Submission (CS) Appendix L.4. The results of this analysis are broadly consistent with the results of post consolidation assessment of MRD in the intention to treat (ITT) population. That is, the odds ratio (OR), DBTd versus BTd, of achieving both MRD negativity and complete response or better 100 days post-ASCT (OR: 2.06; 95% CI: 1.56, 2.72; p<0.0001; 33.7% vs 19.9%) is similar to the OR of achieving MRD negativity regardless of response (OR: 2.27; 95% CI: 1.78, 2.90; p<0.0001; 63.7% vs 43.5%).[5] As requested by the ERG, Janssen performed an additional subgroup analysis exploring the updated meta-analysis results when MRD was assessed in all patients, regardless of response (i.e. consistent definition of MRD assessment as per the landmark analysis). This subgroup analysis included nine observations for PFS and seven for OS and resulted in the HRs of [redacted] and [redacted]

		respectively. These results are broadly consistent with the original submitted base case and updated meta-analysis results including all studies. The company base case has been updated accordingly, applying a consistent definition for MRD. Refer to the 'Summary of changes to the company's cost-effectiveness estimate' section below. Note that a similar scenario exploring the cost-effectiveness results applying the IMWG definition for MRD negativity was not possible because no studies identified in the SLR update reported OS results based on the IMWG definition.
Key issue 3: Uncertainty in the company's adjustment of progression-free survival (PFS) to capture the effect of a second randomisation to maintenance therapy	**YES**	In response to the ERG's concern regarding the impact of Part 2 re-randomisation in CASSIOPEIA on absolute and relative survival outcomes for Part 1, Janssen has performed updated landmark analysis for PFS and OS, censoring patients re-randomised to daratumumab maintenance therapy (refer to Issue 5 below for further details). The company base-case has also been updated to reflect the revised landmark analysis results, thus ensuring that the survival outcomes modelled are not biased as a consequence of the study design and re-randomisation to maintenance therapy (refer to the 'Summary of changes to the company's cost-effectiveness estimate' section below). On the basis that the updated cost-effectiveness results are not impacted by the effects of maintenance treatment, Janssen does not consider Issue 3 per the ERG report to remain a key issue relevant for decision making. However, further context regarding the rational for selecting the IPW method and its appropriateness is provided below for clarification. Inverse Probability Weighting (IPW) is a widely used statistical method which is recognised and accepted by regulatory agencies worldwide. Indeed, for CASSIOPEIA, it was a U.S. Food and Drug Administration (FDA) requirement to use the IPW method to account for the second randomisation with pre-specified weights independent of time preferred by the regulator due to its simplicity and the relative ease to verify results. As such, Janssen did not consider other methods such as the marginal mean model-based estimator or weighted risk set estimator. In their report, the ERG noted that the IPW method may not be appropriate when the assumption of proportional hazards is violated and highlighted overlap of the intervention and comparator curves at early time points in the log-log plot for PFS (refer to Section 3.2.4.6 of the ERG report and company clarification response A11). To further investigate the HR across different treatment phases, Janssen conducted a piecewise HR by study phase (induction, ASCT/consolidation, maintenance) which was reported in Section B.2.6.2 of the CS. Results from this analysis demonstrate a consistent benefit for DBTd compared to BTd across the

different treatment phases, indeed the magnitude of the treatment effect is shown to increase over time. Given this, Janssen consider the IPW to remain an appropriate method to adjust for the second randomisation and is not invalidated by early violation of the PH assumption.

For completeness, in Table 1 and Table 2 below we present updated PFS and OS results including IPW adjustments based on the latest (August 2020) data cut from CASSIOPEIA. Consistent with earlier data cuts, results indicate minimal impact of the second randomisation on the outcomes for Part 1 with longer study follow-up. The similarity of adjusted and unadjusted analyses results was expected given the high proportion of patients re-randomised in both treatment groups. Note median follow-up for Part 2 based on the August 2020 data-cut is 35.5 months representing follow-up from the date of re-randomisation (maintenance study).

Table 1: PFS results with and without IPW adjustments (ITT population)

	Induction / ASCT / consolidation		
	PA1 (median follow-up = 18.8 months)	**PHA (median follow-up = 29.2 months)**	**PHA2 (median follow-up = 44.5 months)**
Analysis without adjustment for second randomisation; DBTd versus BTd			
HR (95% CI)	0.47 (0.33, 0.67)	0.495 (0.38, 0.65)	[redacted]
P-value	<0.0001	<0.0001	[redacted]
IPW Analysis; DBTd versus BTd			
HR (95% CI)	0.47 (0.33, 0.67)	0.50 (0.34, 0.75)	[redacted]
P-value	<0.0001	0.0005	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; IPW = inverse probability weighting; ITT = intention-to-treat; PA1 = Primary Analysis for Part 1; PHA = Post-hoc Interim Analysis (May 2019 data cut); PHA2 = Post-hoc Analysis 2 (August 2020 data cut); PFS = progression-free survival.

Table 2: OS results with and without IPW adjustments (ITT population)

	Induction / ASCT / consolidation		
	PA1 (median follow-up = 18.8	**PHA (median follow-up = 29.2**	**PHA2 (median follow-up = 44.5**

<table>
<tr><td></td><td></td><td>
<table>
<tr><th></th><th>months)</th><th>months)</th><th>months)</th></tr>
<tr><td colspan="2">Analysis without adjustment for second randomisation; DBTd versus BTd</td><td></td><td></td></tr>
<tr><td>HR (95% CI)</td><td>0.43 (0.23, 0.80)</td><td>0.52 (0.33, 0.85)</td><td>[redacted]</td></tr>
<tr><td>P-value</td><td>0.0065</td><td>0.0070</td><td>[redacted]</td></tr>
<tr><td colspan="2">IPW Analysis; DBTd versus BTd</td><td></td><td></td></tr>
<tr><td>HR (95% CI)</td><td>n/a</td><td>n/a</td><td>[redacted]</td></tr>
<tr><td>P-value</td><td>n/a</td><td>n/a</td><td>[redacted]</td></tr>
<tr><td colspan="4">Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; IPW = inverse probability weighting; ITT = intention-to-treat; OS = overall survival; PA1 = Primary Analysis for Part 1; PHA = Post-hoc Interim Analysis (May 2019 data cut); PHA2 = Post-hoc Analysis 2 (August 2020 data cut).</td></tr>
</table>
</td></tr>
<tr><td>Key issue 4: Plausibility of long-term survival with standard care (autologous stem cell transplant [ASCT] with bortezomib, thalidomide and dexamethasone [BTd] induction and consolidation)</td><td>YES</td><td>In response to the ERG's concern regarding the plausibility of long-term survival with standard of care (BTd), Janssen has updated the survival analysis based on results from a revised landmark analysis incorporating the August 2020 data-cut from CASSIOPEIA, representing an additional 15.3 months of study follow-up (refer to Issue 5 below for further details). Importantly, the revised landmark analysis censored patients re-randomised to daratumumab maintenance therapy, thus eliminating any absolute survival benefit (or risk of bias) introduced by Part 2 and improving the trial's generalisability to clinical practice in England where maintenance therapy does not represent standard of care.

Consistent with the original CS, extrapolation of PFS and OS for BTd patients with a post-consolidation MRD-positive response was performed in accordance with the guidance provided in the NICE DSU Technical Support Document (TSD) 14.[6] Refer to Appendix A for further details, including the goodness-of-fit statistics for each parametric distribution explored and the extrapolated survival curves.

Based on an assessment of statistical goodness-of-fit, visual inspection of the survival curves to the observed data from the CASSIOPEIA trial, and clinical plausibility of long-term survival predictions, the exponential distribution was selected for both PFS and OS. A scenario exploring cost-effectiveness results selecting the Weibull and exponential curves for PFS and OS respectively per the original CS is presented in Appendix D.</td></tr>
</table>

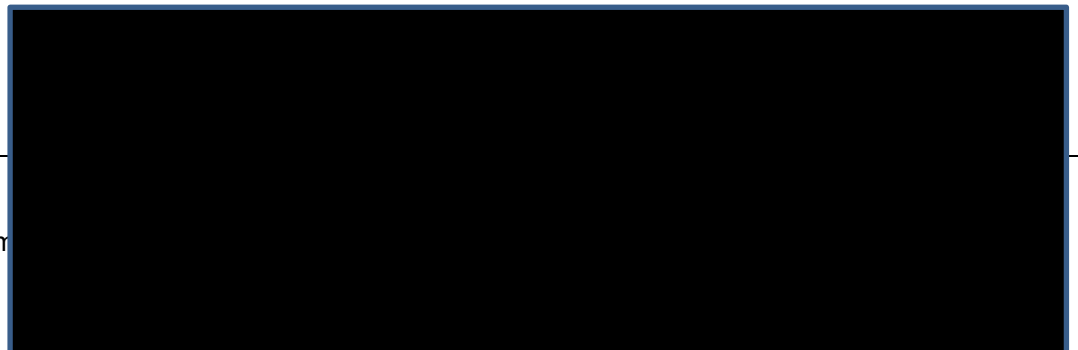
		Consistent with the CS, survival for BTd MRD-negative patients was modelled via application of a hazard ratio from the updated meta-analysis results (refer to Issue 1). However, in response to key issue 2, the hazard ratio selected was based on a subgroup of studies where MRD was assessed in all patients, regardless of response, consistent with the MRD definition prespecified per the study protocol for Part 1, and per the landmark analysis. The updated OS and PFS outcomes predicted by the model for the overall cohort (i.e. BTd MRD-negative and MRD-positive combined, weighted by the proportion of patients achieving post-consolidation MRD negativity), are presented in Figure 1 and Figure 2 respectively with a comparison of survival predictions against the original model presented in Table 3. **Figure 1: Comparison of BTd OS predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined), censoring for maintenance**  **Figure 2: Comparison of BTd PFS predicted by the model versus CASSIOPEIA (MRD+ and MRD- combined), censoring for maintenance** 

Table 3: BTd PFS and OS predictions (months) – comparison of original and revised economic model

Treatment	Median PFS	Mean PFS	Median OS	Mean OS
Original model	48	71	160	196
Revised model	37	59	146	185
CASSIOPEIA (censoring for daratumumab maintenance)	■	n/a	n/a	n/a

Key: BTd = bortezomib, thalidomide and dexamethasone; n/a = not available OS = overall survival; PFS = progression-free survival

There is a paucity of long-term survival data for BTd when administered as an induction (only) therapy. To help validate long-term survival predictions for standard of care, Janssen performed a naïve comparison of survival rates (PFS and OS) from the updated model with rates reported from a range of sources including randomised controlled trials (RCTs), observational studies and data from the Office for National Statistics (ONS). Despite inherent

challenges in terms of generalisability, the range of evidence provide a useful upper and lower bound to characterise the uncertainty with results presented in Table 4.

Table 4: Naïve comparison of survival rates predicted by the model for BTd compared to RCT and observational data sources

Data source	PFS			OS		
	3-Yr	5-Yr	10-Yr	3-Yr	5-Yr	10-Yr
Revised model (Exponential/Exponential)	52%	33%	12%	86%	76%	57%
CASSIOPEIA (censoring for daratumumab maintenance)	[redacted]	n/a	n/a	[redacted]	n/a	n/a
GIMEMA[7,9]	68%	50%	34%	86%	79%*	60%
PHE cohort[8]	[redacted]	n/a	n/a	[redacted]	n/a	n/a
US RWE (SEER/OPTUM)[10]	n/a	n/a	n/a	n/a	74%	68%
ONS (55-64)[11] **	n/a	n/a	n/a	n/a	64%	43%

Key: n/a = not available; ONS = Office for National Statistics; OS = overall survival; PFS = progression-free survival; PHE = Public Health England; RCT = randomised controlled trial; RWE = real world evidence; US = United States; Yr = year

*Janssen estimate based on visual inspection of the published Kaplan-Meier curves from Tacchetti et al. 2020

** All patient estimate for newly diagnosed MM including mixed population of transplant-eligible and ineligible patients

The revised model results in a downward shift in the PFS curve which closely aligns with the 3-year survival rate from CASSIOPEIA after censoring patients re-randomised to daratumumab maintenance (52% versus [redacted] respectively). On the basis that OS, rather than PFS, is a key driver of the economic model results, the following discussion is focussed on the plausibility of long-term survival for BTd in terms of OS.

Comparison of survival rates per the revised model versus CASSIOPEIA and PHE datasets

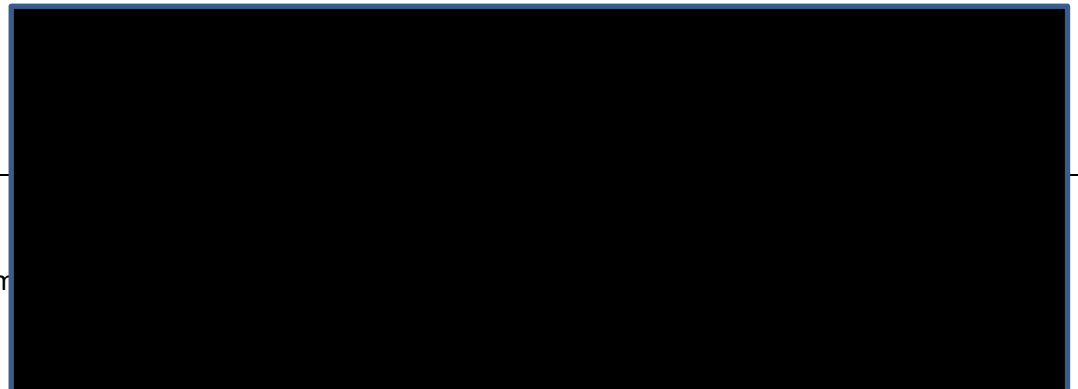
		Survival per the revised model at the 3-, 5- and 10-year timepoints remains broadly consistent with the original model, albeit marginally lower in the outer years. At 3-years, survival per the revised model (86%) is in line with both the CASSIOPEIA study and survival rates reported in a recent update to the Public Health England (PHE) cohort study which incorporates an additional 6-months follow-up to 30th June 2020.[8] **Comparison of survival rates per the revised model versus GIMEMA** The GIMEMA study was a phase III randomised controlled trial which compared BTd with thalidomide and dexamethasone (Td) as induction and consolidation therapy after double autologous haematopoietic stem-cell transplantation (ASCT) for newly diagnosed multiple myeloma (NDMM).[9] Between May 10, 2006, and April 30, 2008, the study recruited 474 patients (BTd group 236; Td group: 238) with median follow-up for surviving patients of 124.1 months. Key prognostic factors for BTd patients in GIMEMA were broadly similar to CASSIOPEIA with a mean age of 56.3 years and 56.5 years respectively, while the proportion classified as International Staging System (ISS) stage III was 16% versus 15% in CASSIOPEIA. A comparison of demographic and baseline characteristics is provided in Appendix C. In GIMEMA, induction therapy consisted of three 21-day cycles compared to four 28-day cycles in CASSIOPEIA, while consolidation was two 35-day cycles versus two 28-day cycles in CASSIOPEIA. A comparison of the BTd dosing schedule applied in GIMEMA and CASSIOPEIA is provided in Appendix C. Despite differences in study design, long-term follow-up data from GIMEMA provide a useful point of reference to compare the modelled survival predictions for BTd. The 3-year survival rate from GIMEMA is consistent with the revised model, CASSIOPEIA and PHE datasets while survival at 5- and 10-years is marginally above the revised model at ~79% and 60% respectively. It is important to highlight that GIMEMA is an older study than CASSIOPEIA and the myeloma treatment pathway has evolved considerably over the last decade. Thus, whilst patients in GIMEMA received a double as opposed to single transplant, patients will now benefit from more efficacious subsequent treatments. **Comparison of survival rates per the revised model versus US RWE**

		In this retrospective cohort study, treatment patterns and outcomes among front-line transplant eligible multiple myeloma (MM) patients were investigated using patient-level data from three US databases: – OPTUM Commercial Claims database (January 2000-March 2017); – OPTUM Electronic Medical Records (EMR) database (January 2007-March 2016); – Surveillance, Epidemiology, and End Results (SEER)-Medicare Linked database (January 2007-December 2014).[10] Among the 1,599 patients who received ASCT at any time during treatment, 1,003 received ASCT after frontline induction therapy of which 51 (5.1%) received BTd.[10] This study reports similar 5-year survival rates to the revised model (74% and 76% respectively) however 10-year survival rates are higher at 68% and 57% respectively. That said, there is evidence of high censoring after five years and results should be interpreted with caution. **Comparison of survival rates per the revised model versus ONS** The ONS report predicted estimates of 1-, 5- and 10-year net survival for myeloma patients by gender and age group. Whilst the 55-64 age group is considered the most representative of a transplant-eligible population, Janssen note that the 1-year survival estimate from ONS for this cohort is notably lower compared to the PHE dataset for BTd (89.4% and [redacted] respectively).[11] Interestingly, the PHE all-patient estimate for BTd ([redacted]) is comparable to the ONS 1-year estimate suggesting that age is not a particularly good indicator for transplant eligibility and that the ONS survival estimates based on the 55-64 subgroup are overly conservative for a transplant-eligible population. This may explain why the ONS 5- and 10-year survival estimates of 64% and 43% respectively are considerably lower than both the RCT and observational evidence as well as the revised model. To conclude, BTd survival predictions from the revised model are broadly comparable to survival rates recently reported for the GIMEMA study which provides extended 10-year median follow-up. Reassuringly, long-term follow-up from a US cohort study provides similar 5-year survival rates, further supporting the modelled survival predictions for BTd.
Key issue 5: Uncertainty over daratumumab treatment effects on PFS	YES	The treatment effect for DBTd is driven by the depth of post-consolidation response achieved with this quadruplet combination and is founded on biological plausibility. That is, deeper responses improve the prognoses of patients with extensive evidence demonstrating improved

and overall survival (OS)		survival outcomes with deeper responses.[1,12,13,14,15,16,17] Treatment with DBTd resulted in deeper responses in both MRD negative (MRD negativity at higher sensitivity thresholds) and MRD positive patients (deeper conventional response measured by IMWG criteria). In this sense, the treatment effect modelled is not a claim unique to daratumumab, but rather any treatment that achieves an equivalent depth of response and resulting shift in patient's prognosis. As requested by the ERG, to help reduce uncertainty over the benefit of the deep responses achieved with daratumumab on PFS and OS, Janssen has updated the CASSIOPEIA landmark analysis by censoring patients re-randomised in Part 2 to daratumumab maintenance therapy. The revised landmark analysis was performed using the August 2020 data-cut from CASSIOPEIA, representing a median follow-up of 44.5 months (an additional 15.3 months follow-up). Updated Kaplan-Meier plots for both PFS and OS from the time of post-consolidation response assessment, by treatment arm and MRD status, are presented in Figure 3 and Figure 4 respectively. **Figure 3: Landmark analysis: PFS by treatment arm and MRD status at the time of the post-consolidation assessment, censoring patients re-randomised to daratumumab maintenance therapy (ITT population, median follow-up = 44.5 months)** [redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention to treat; MRD = minimal residual disease; PFS = progression-free survival.
Note: comparing numbers at risk with the May 2019 data cut indicates one additional MRD-positive BTd patient and two additional MRD-positive DBTd patients at baseline. Further discussion of this discrepancy is provided in the accompanying statistical analysis file note.[18]

Figure 4: Landmark analysis: OS by treatment arm and MRD status at the time of the post-consolidation assessment, censoring patients re-randomised to daratumumab maintenance therapy (ITT population, median follow-up = 44.5 months)



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention to treat; MRD = minimal residual disease; OS = overall survival.

On visual inspection, longer follow-up from the updated landmark analysis indicate consistent results with the original landmark analysis presented in Section B.2.6.3 of the CS, with clear separation of the curves for both PFS and OS favouring the DBTd arm regardless of MRD status.

A comparison of the Cox proportional hazard model results for PFS and OS from the original and updated landmark analysis is presented in Table 5. Results from the updated tests of proportional hazards are included in Appendix B.

Table 5: Cox proportional hazard model results

	Original landmark analysis (median follow-up = 29.2	Updated landmark analysis (median follow-up = 44.5

		months) / months, censoring for maintenance)

	BTd	DBTd	BTd	DBTd
PFS				
DBTd MRD+ versus BTd MRD+ HR (95% CI)	[redacted]		[redacted]	
n/N (%)	[redacted]	[redacted]	[redacted]	[redacted]
DBTd MRD- versus BTd MRD-) HR (95% CI)	[redacted]		[redacted]	
n/N (%)	[redacted]	[redacted]	[redacted]	[redacted]
OS				
DBTd MRD+ versus BTd MRD+ HR (95% CI)	[redacted]		[redacted]	
n/N (%)	[redacted]	[redacted]	[redacted]	[redacted]
DBTd MRD- versus BTd MRD-) HR (95% CI)	[redacted]		[redacted]	
n/N (%)	[redacted]	[redacted]	[redacted]	[redacted]

Key: CI = confidence interval; BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MRD = minimal residual disease; n = number of events; OS = overall survival; PFS = progression-free survival.

Note: comparing the May 2019 and August 2020 data cuts, there is one additional MRD-positive BTd patient and two additional MRD-positive DBTd patients at baseline for progression-free survival. Further discussion of this discrepancy is provided in the accompanying statistical analysis file note.[18]

Despite longer study follow-up, the relative proportion of events increased only marginally with the updated landmark analysis due to additional censoring for maintenance (note, for the analysis of OS in MRD-negative patients, the relative proportion of events actually decreased for BTd and remained stable for DBTd). As a consequence of a smaller effective sample, the confidence interval for OS in both MRD-positive and MRD-negative subgroups has also increased.

There is therefore a clear trade-off associated with censoring for maintenance whereby the

		risk of bias is eliminated at the cost of reduced precision in the point estimate of the benefit of the deeper responses achieved with daratumumab on OS. Arguably, based on the IPW results that consistently demonstrate minimal impact of maintenance on the relative benefit of daratumumab, uncertainty in terms of OS is minimised based on results from the original landmark analysis despite shorter median follow-up. We therefore present an additional scenario analysis in Appendix D which combines HRs from the original landmark analysis for OS and updated landmark analysis for PFS for the comparison of DBTd versus BTd.
Key issue 6: Waning of treatment effects for daratumumab	**YES**	As noted for key issue 5 above, the treatment effect modelled for DBTd is driven by the depth of post-consolidation response compared with BTd. The key question for clinicians is therefore whether deeper responses are expected to result in a fundamental shift in a patient's prognosis, or whether the survival benefit associated with deeper responses is expected to wane over time. Janssen has performed updated landmark analysis based on the latest (August 2020) data cut, censoring patients re-randomised to daratumumab maintenance. With median follow-up approaching four years, the revised analysis continues to demonstrate a relative benefit in favour of DBTd driven by deeper responses for both MRD-positive and MRD-negative subgroups, with no evidence to suggest a possible waning of relative benefit over time. For PFS, with longer follow-up, the relative benefit of the deeper responses achieved with daratumumab in the MRD-positive subgroup is now statistically significant and has increased compared to the original landmark analysis with a 40% reduction in the risk of disease progression or death for patients treated with DBTd. Similarly, for the MRD-negative subgroup, the relative benefit of DBTd has become more pronounced, with a 64% reduction in the risk of disease progression or death and a narrower confidence interval indicating a more precise estimate. For OS, a low event rate and additional censoring for maintenance has reduced precision in the point estimate of effect compared with the original landmark analysis, with a wider confidence interval crossing 1 for both the MRD-positive and MRD-negative subgroups. Due to data immaturity for OS, it becomes important to consider the likelihood of waning in relative benefit over time in the context of other relevant evidence. **Improvement in survival outcomes at deeper sensitivity thresholds** In a recently published meta-analysis of front-line and relapsed/refractory MM studies, Munshi

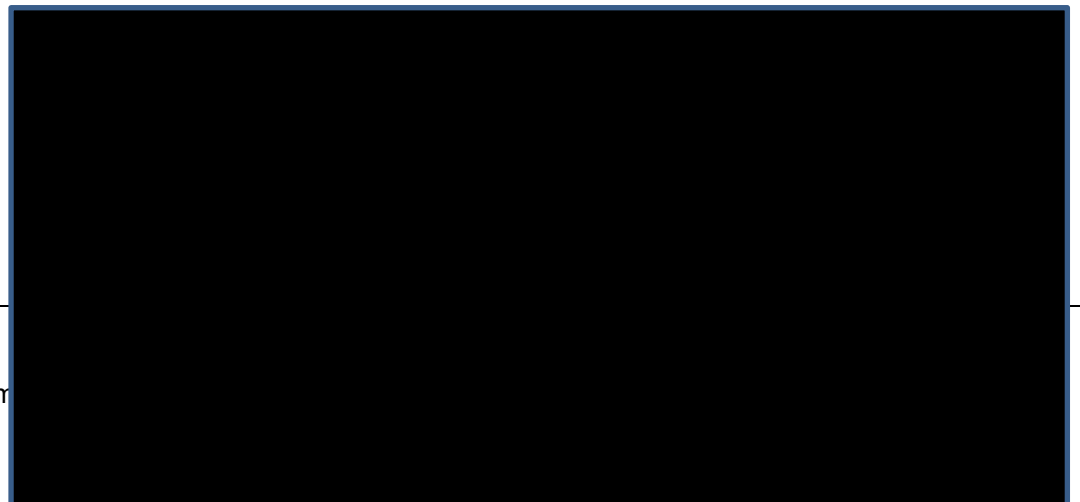
et al. (2020) report improvements in OS (and PFS) outcomes associated with increasingly stringent MRD sensitivity thresholds with OS most improved with MRD negativity at the sensitivity threshold of 10^{-6} (HR: 0.26; 95% CI: 0.13, 0.51; p<0.001).[1]

Survival extrapolation in the economic model was based on MRD negativity measured using multiparametric flow cytometry (MFC) at sensitivity threshold 10^{-5}. However, as highlighted in Section B.2.6.1 of the CS, at the higher sensitivity threshold of 10^{-6}, DBTd almost doubled the rate of MRD negativity compared to BTd (NGS 10^{-6}: 39.1% vs 22.8%; OR: 2.18; 95% CI: 1.58, 3.01; p<0.0001) indicating significantly deeper levels of response which helps to explain the relative benefit of daratumumab in the MRD-negative subgroup observed in the original and revised landmark analysis.

Deeper responses supported by sustained MRD negativity

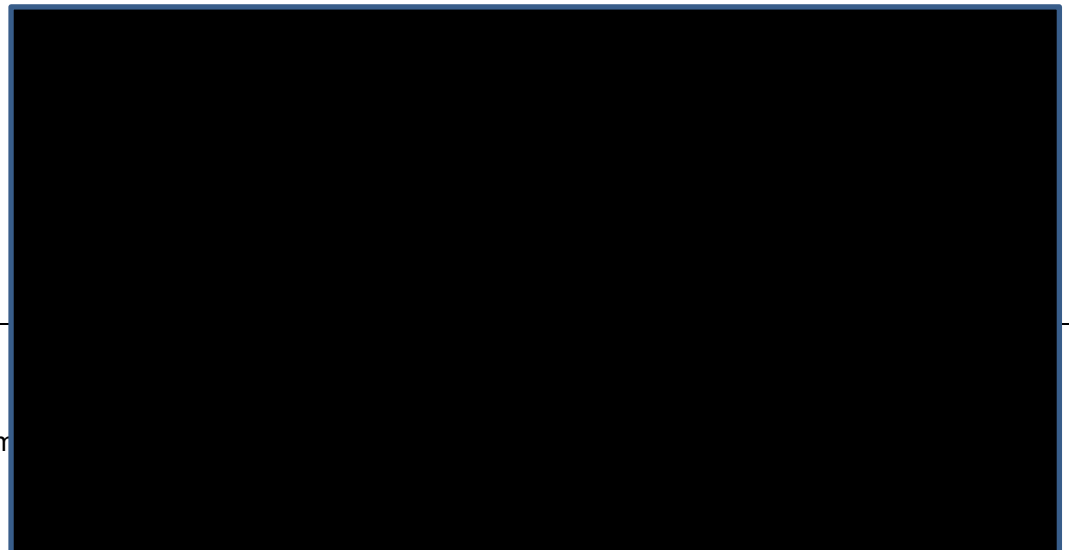
Whilst the prognostic utility of sustained MRD-negativity is not widely reported, analysis from Part 2 of CASSIOPEIA demonstrates improved outcomes for BTd and DBTd patients with 2-year sustained MRD negativity (Figure 5 and Figure 6 respectively). Note, MRD analysis reported by next generation sequencing (NGS) was available within the timeframe for technical engagement response, hence why this is presented rather than MFC.

Figure 5: Kaplan-Meier plot of PFS from 2nd randomisation by 2-year MRD sustained negative rate by NGS at 10^{-5} in bone marrow from post consolidation in BTd patients (maintenance-specific ITT population, median follow-up = 35.5 months)



Key: DARA = re-randomised to daratumumab; ITT = intention to treat; MRD = minimal residual disease; NGS = next generation sequencing; OBS = re-randomised to observation; PFS = progression-free survival
Subjects with post consolidation MRD negative, non-positive MRD at wk25, wk52 and negative MRD at week 105 are considered 2-year sustained MRD from post consolidation.
Note: Data for OBS-MRD- and DARA-MRD- is not mature and the estimation is not robust.

Figure 6: Kaplan-Meier plot of PFS from 2nd randomisation by 2-year MRD sustained negative rate by NGS at 10^{-5} in bone marrow from post consolidation in DBTd patients (maintenance-specific ITT population, median follow-up = 35.5 months)



		Key: DARA = re-randomised to daratumumab; ITT = intention to treat; MRD = minimal residual disease; NGS = next generation sequencing; OBS = re-randomised to observation; PFS = progression-free survival. Subjects with post consolidation MRD negative, non-positive MRD at wk25, wk52 and negative MRD at week 105 are considered 2-year sustained MRD from post consolidation. Note: Data for OBS-MRD- and DARA-MRD- is not mature and the estimation is not robust. On the BTd arm, [redacted] of the [redacted] patients with a post-consolidation MRD-negative assessment achieved sustained MRD negativity for 2-years (note, this number represents the summation of patients re-randomised to receive observation and daratumumab maintenance for Part 2). That compares with [redacted] of the [redacted] DBTd patients. The DBTd combination therefore achieves an increase in both the absolute number, and relative proportion, of patients achieving 2-years sustained MRD-negativity which supports deeper responses for daratumumab and maintenance of a long-term relative benefit on both PFS and OS. **Evolution of response and MRD negativity conversion** In Table 6, we present a comparison of DBTd versus BTd CR or better and MRD negativity rate post-consolidation and at 1- and 2-year timepoints for patients re-randomised to observation (Part 2). Despite patients not assigned to active treatment during the maintenance phase of the study, results indicate a deepening of response rates over time which is more pronounced for DBTd reflecting the impact of daratumumab's unique mechanism of action which is to modulate the

body's own immune system to better fight the disease.[19] This observation is consistent with the response analysis results presented in Section B.2.6.1 of the CS for Part 1 intention to treat (ITT).

Table 6: Evolution of CR or better response and MRD negativity rate (maintenance-specific ITT population)

	Induction / ASCT / consolidation	
	BTd – OBS	**DBTd – OBS**
Analysis set: maintenance specific ITT	[redacted]	[redacted]
Post-consolidation response assessment		
MRD-negative (NGS 10^{-5})	[redacted]	[redacted]
≥ CR	[redacted]	[redacted]
1-Year follow-up (Part 2)		
MRD-negative (NGS 10^{-5})	[redacted]	[redacted]
≥ CR	[redacted]	[redacted]
2-Year follow-up (Part 2)		
MRD-negative (NGS 10^{-5})	[redacted]	[redacted]
≥ CR	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention to treat; MRD = minimal residual disease; NGS = next generation sequencing; OBS = re-randomised to observation.

A comparison of MRD negativity conversion supports a more pronounced deepening of response over time on the daratumumab arm with [redacted] of DBTd patients assessed as MRD-positive post-consolidation converting to MRD-negative after 2-years follow-up compared to [redacted] on the BTd arm (Table 7).

Table 7: Summary of MRD negative conversion rate by NGS 10^{-5} (maintenance-specific ITT population, median follow-up = 35.5 months)

	Induction / ASCT / consolidation	
	BTd – OBS	**DBTd – OBS**
Analysis set: maintenance specific ITT	[redacted]	[redacted]
Patients with post-consolidation positive MRD measured on Day 100 post-ASCT	[redacted]	[redacted]
MRD-negative conversion rate during maintenance (10^{-5} in bone marrow)[a]	[redacted]	[redacted]
95% CI of MRD-negative conversion rate[b]	[redacted]	[redacted]

[a] Percentages are calculated with the number of patients with post-consolidation MRD positive status per NGS method at day100 post-ASCT in each group as denominator.
[b] Exact 95% confidence interval

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ITT = intention to treat; MRD = minimal residual disease; NGS = next generation sequencing; OBS = re-randomised to observation.

Whilst the confidence intervals for the two groups overlap, this analysis supports the evolution of response analysis presented in Table 6 and persistency of the relative benefit of daratumumab on PFS and OS.

Progression-free survival on next line of therapy (PFS2)

PFS2 was reported in Section B.2.6.2 of the CS. In Table 7 below, we present updated PFS2 results based on the most recent (August 2020) data cut from CASSIOPEIA including and excluding daratumumab maintenance therapy.

Table 7: PFS2 (ITT population, median follow-up 44.5 months)

	Induction / ASCT / consolidation			
	BTd[a]	**DBTd[a]**	**BTd - OBS[b]**	**DBTd - OBS[b]**
Analysis set: ITT	[redacted]	[redacted]	[redacted]	[redacted]
Progression-free survival (months)				

Number of events (%)	[redacted]	[redacted]	[redacted]	[redacted]		
Number censored (%)	[redacted]	[redacted]	[redacted]	[redacted]		
Median (95% CI) [c]	[redacted]	[redacted]	[redacted]	[redacted]		
Hazard ratio [d]	[redacted]		[redacted]			
P-value [e]	[redacted]		[redacted]			

[a] Including all subjects randomised in Part I regardless of second randomization.
[b] Including patients re-randomised to the observation maintenance arm and subjects not 2nd randomised.
[c] Based on Kaplan-Meier product limit estimates adjusted by inverse-probability-weight (IPW) method.
[d] Hazard ratio and 95% CI from a Cox regression analysis with inverse probability weight (IPW) and treatment as the sole explanatory variable.
[e] p-value is based on the log-rank test with risk factor Adjusted by inverse-probability-weight (IPW) method.
Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; ITT = intention to treat; MRD = minimal residual disease; OBS = observation.

Consistent with results from both the primary data cut (June 2018) and post-hoc data cut (May 2019) reported in Section B.2.6.2 of the CS, the updated PFS2 analyses demonstrate that the PFS benefit of DBTd is maintained beyond the next line of therapy received which is indicative of persistency of relative benefit beyond progression and improved long-term outcomes.

External evidence supporting maintenance of relative benefit driven by deeper response

Tacchetti et al. (2020) report the final analysis of the GIMEMA study comparing front-line induction/consolidation therapy with BTd versus Td. After median follow-up of 10-years, this study demonstrates a persistent relative benefit of BTd versus Td for PFS

		that translated to a statistically significant improvement in OS. Indeed, visual inspection of the published Kaplan-Meier analysis for OS indicate curves that continue to diverge over time (most notably beyond 4-years) with no suggestion of waning in relative benefit despite the extended follow-up. While there are differences in study design between GIMEMA and CASSIOPEIA (refer to key issue 4), the large patient population and 10-year follow-up make it an important and relevant study for comparison. To conclude, the economic model is built on the evidence-based relationship between MRD status and long-term outcomes (PFS and OS). It is important to note, however, that MRD status incorporated into the economic model is based on assessment at a single point in time (100 days post-consolidation) and at a sensitivity threshold of 10^{-5}. This means that the additional prognostic benefits of: MRD status assessed at the more stringent sensitivity threshold of 10^{-6}, deepening response rates and MRD negativity over time, sustained MRD negativity and deeper conventional responses in MRD positive patients are not taken into account by the model structure. As such, landmark analysis of CASSIOPEIA is used to account for these additional benefits which, while not explicitly captured in the model structure, are implicit within the CASSIOPEIA data. Long-term follow-up data from the GIMEMA study provides compelling evidence that deeper responses translate to significant improvements in long-term outcomes including OS, with no indication of a treatment waning effect. With median follow-up from CASSIOPEIA approaching 4-years, and evidence of deepening response rates for DBTd relative to BTd with extended study follow-up, it is reasonable to conclude that the benefit of such important prognostic factors will not wane over time.

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response

Additional issue 1: Daratumumab formulation for costing	Section 1.7 (pp20, 22) Section 4.2.4 (p92) Section 6.1 (p121) Section 6.3 (p124) Section 6.4 (p127)	YES	The ERG include the intravenous (IV) costs for daratumumab in their preferred analysis, highlighting consistency with the clinical evidence from CASSIOPEIA. However, following the licence extension for the subcutaneous (SC) formulation in June 2020,[19] this is inconsistent with how daratumumab is expected to be administered in clinical practice. For example, with DBd at second-line, there was a rapid switchover from IV to SC with [redacted] market share achieved within 3-months of licence (market share at the beginning of December had reached [redacted]). On the basis that the COLUMBA study[20] demonstrated non-inferiority of the SC versus IV formulation, and strong uptake demonstrating its acceptance in routine clinical practice in England, it is reasonable to model the SC rather than IV costs for daratumumab.
Additional issue 2: No PBd at 3L	Section 1.7 (pp21, 22) Section 6.1 (p122) Section 6.2 (p123) Section 6.3 (p124)	NO	The ERG exclude the cost of PBd from cost calculations in their preferred analysis, highlighting advice that PBd is not currently used at third or fourth line in clinical practice. In the clinical advisory board meeting held by Janssen, feedback indicated that patients without prior exposure to lenalidomide would receive lenalidomide plus dexamethasone (or ixazomib plus lenalidomide and dexamethasone on the CDF) with the remainder treated with PBd.[21] With lenalidomide now available earlier in the treatment pathway, PBd is the only NICE

			recommended treatment option available at third-line for patients who are prior-lenalidomide exposed. In practice, Janssen understand that clinicians may use regimens such as alkylators (e.g. cyclophosphamide) or steroids (e.g. dexamethasone) instead of PBd as a bridging strategy to access novel agents at fourth-line but would not re-treat with lenalidomide.
Additional issue 3: Patients' mean age	Section 1.7 (pp20, 22) Section 4.2.3 (pp89, 90) Section 5.3.4 (p117) Section 6.1 (p119) Section 6.2 (p123) Section 6.4 (p127)	NO	As is common in RCTs, the mean age of patients in CASSIOPEIA is slightly younger than expected in clinical practice. The ERG assume a mean age of 59.3 years per the Public Health England dataset rather than 56.6 years per the CASSIOPEIA study and company base case. This, however, is inconsistent with all other efficacy inputs in the model sourced from CASSIOPEIA, reflective of the younger population. Janssen therefore do not consider this adjustment appropriate.

NICE National Institute for Health and Care Excellence

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the company's base-case ICER

Key issue 1: Uncertainty in hazard ratios from the company's meta-analysis of the effects of minimum residual disease (MRD) status on survival outcomes Key issue 2: Inconsistency in the company's approach for defining and analysing MRD-negative patients	MRD SLR 15 months out of date Inconsistent definitions for MRD between the CASSIOPEIA landmark analysis which assessed MRD regardless of response and meta-analysis which included all studies with mixed definitions for MRD	MRD meta-analysis results updated to include new studies identified in recent SLR update to October 2020 Consistent definitions for MRD have been applied between the CASSIOPEIA landmark analysis and meta-analysis, with MRD assessed regardless of response. Note that a scenario exploring the cost-effectiveness results applying the IMWG definition for MRD negativity was not possible because no studies identified in the SLR update reported OS results based on the definition per the IMWG guidance.	Revised base-case ICER with PAS = £15,822 ([redacted])
Key issue 4: Plausibility of long-term survival with standard care (autologous stem cell transplant [ASCT] with bortezomib, thalidomide and dexamethasone [BTd] induction and consolidation) Key issue 5: Uncertainty over daratumumab treatment effects on PFS	The landmark analysis per the original company base case included patients re-randomised in Part 2 to daratumumab maintenance therapy (median follow-up = 29.2 months) BTd MRD+ extrapolation (PFS) = Weibull BTd MRD+ extrapolation (OS) = Exponential	The updated landmark analysis incorporates the latest available data-cut from CASSIOPEIA (August 2020) and censors patients re-randomised in Part 2 to daratumumab maintenance therapy (median follow-up = 44.5 months) BTd MRD+ extrapolation (PFS) = Exponential BTd MRD+ extrapolation (OS) = Exponential	Revised base-case ICER with PAS = £17,842 ([redacted])

and overall survival (OS)			
Company's preferred base case following technical engagement	Incremental QALYs: [redacted]	Incremental costs: [redacted]	Revised base-case ICER with PAS = £17,957 ([redacted])

References

[1] Munshi et al. 2020. A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma. Blood Adv. 2020 Dec 8;4(23):5988-5999. doi: 10.1182/bloodadvances.2020002827
[2] Janssen. [Data on File] MRD SLR/meta-analysis update
[3] Janssen. [Data on File] MRD SLR risk of bias assessment update
[4] Janssen. [Data on File] MRD SLR/meta-analysis further clarification response
[5] Janssen. [Data on File] MMY3006. Clinical Study Report: Part 1. 2019.
[6] Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available at: http://nicedsu.org.uk/.
[7] Cavo M, et al. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. The Lancet. 2010;376(9758):2075-85.
[8] Janssen. [Data on File] Public Health England datasets. Standing cohort study of newly diagnosed multiple myeloma cancer patients in England (January 2015 to December 2018). (December 2020 Report).
[9] Tacchetti et al. 2020. Bortezomib, thalidomide, and dexamethasone followed by double autologous haematopoietic stem-cell transplantation for newly diagnosed multiple myeloma (GIMEMA-MMY-3006): long-term follow-up analysis of a randomised phase 3, open-label study. Lancet Haematol. 2020 Dec;7(12):e861-e873. doi: 10.1016/S2352-3026(20)30323-9. PMID: 33242443.

[10] Mehra et al. An Adjusted Treatment Comparison of Bortezomib/Cyclophosphamide/Dexamethasone and Bortezomib/Thalidomide/Dexamethasone from Real-World Data in Patients with Newly Diagnosed Multiple Myeloma Who Are Transplant Eligible. Poster presented at American Society of Hematology (ASH). 2019
[11] Office for National Statistics (ONS). Cancer survival in England – adults diagnosed (2013-2017). Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed
[12] Lahuerta JJ, Mateos MV, Martinez-Lopez J, Rosinol L, Sureda A, de la Rubia J, et al. Influence of pre- and post-transplantation responses on outcome of patients with multiple myeloma: sequential improvement of response and achievement of complete response are associated with longer survival. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(35):5775-82.
[13] Chanan-Khan AA, Giralt S. Importance of achieving a complete response in multiple myeloma, and the impact of novel agents. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(15):2612-24.
[14] Kapoor P, Kumar SK, Dispenzieri A, Lacy MQ, Buadi F, Dingli D, et al. Importance of achieving stringent complete response after autologous stem-cell transplantation in multiple myeloma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31(36):4529-35.
[15] Harousseau JL, Avet-Loiseau H, Attal M, Charbonnel C, Garban F, Hulin C, et al. Achievement of at least very good partial response is a simple and robust prognostic factor in patients with multiple myeloma treated with high-dose therapy: long-term analysis of the IFM 99-02 and 99-04 Trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(34):5720-6.
[16] Munshi NC, Avet-Loiseau H, Rawstron AC, Owen RG, Child JA, Thakurta A, et al. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. JAMA oncology. 2017;3(1):28-35.
[17] Munshi, N. et al. Expanded Meta-Analysis Confirms the Association Between MRD and Long-term Survival Outcomes in Multiple Myeloma (MM). Poster presented at American Society of Hematology (ASH). 2019
[18] Janssen. [Data on File] Updated Landmark Analyses: Statistical note to file.
[19] Electronic Medicines Compendium (EMC). Darzalex 1,800 mg solution for injection: Summary of Product Characteristics 2020 [Available from: https://www.medicines.org.uk/emc/product/11488/smpc.
[20] Mateos M-V, Nahi H, Legiec W, Grosicki S, Vorobyev V, Spicka I, et al. Efficacy and safety of the randomized, open-label, non-inferiority, phase 3 study of subcutaneous (SC) versus intravenous (IV) daratumumab (DARA) administration in patients (pts) with relapsed or refractory multiple myeloma (RRMM): COLUMBA. American Society of Clinical Oncology; 2019.
[21] Janssen. [Data on File] Clinical Advisory Board Meeting Minutes. August 2020

NICE National Institute for Health and Care Excellence

Technical engagement response form - Appendix

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments **5:00pm, 11 January 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response 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.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. 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.

- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- 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 replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, 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 during engagement 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.

About you

Your name	**Keith Stubbs**
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Janssen-Cilag Ltd**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**N/A**

NICE National Institute for Health and Care Excellence

Appendix A: Extrapolation of BTd MRD+ OS and PFS

Table 1: Goodness-of-fit statistics for BTd MRD+ OS (updated landmark analysis) survival models

Survival model	AIC	BIC
Exponential	458.65	**462.32**
Weibull	459.41	466.75
Lognormal	**458.06**	465.40
Loglogistic	459.11	466.45
Gompertz	458.89	466.23
Generalised Gamma	460.05	471.06

Key: AIC = Akaike's information criterion; BIC = Bayesian information criterion; BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival.

Table 2: Comparison of predicted survival rates for BTd MRD+ OS (updated landmark analysis) survival models

Survival model	OS survival rates			
	5 years	10 years	20 years	30 years
Clinician estimate	≤70%[a]	44%[b]	-	-
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	[redacted]
Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Generalised Gamma	[redacted]	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; OS = overall survival.
[a] Feedback from UK clinician, not part of the clinical advisory board meeting for DBTd [1]
[b] Feedback from clinical advisory board meeting for DBTd with reference to the all patient estimate for newly diagnosed MM including mixed population of transplant-eligible and ineligible patients from the Office for National Statistics (ONS)[2]

Figure 1: Extrapolation of OS for BTd MRD+ (updated landmark analysis)

[redacted]

Table 3: Goodness-of-fit statistics for BTd MRD+ PFS (updated landmark analysis) survival models

Survival model	AIC	BIC
Exponential	893.16	**896.73**
Weibull	891.59	898.74
Lognormal	909.81	916.96
Loglogistic	898.27	905.42
Gompertz	**890.66**	897.81
Generalised Gamma	892.58	903.31

Key: AIC = Akaike's information criterion; BIC = Bayesian information criterion; BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.

Table 4: Comparison of predicted survival rates for BTd MRD+ PFS (updated landmark analysis) survival models

Survival model	PFS survival rates

	5 years	10 years	20 years
Clinician estimate	20–30%[a]	<10%[b]	<1%[b]
Exponential	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]
Loglogistic	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]
Generalised Gamma	[redacted]	[redacted]	[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; MRD = minimal residual disease; PFS = progression-free survival.
[a] Feedback from UK clinician, not part of the clinical advisory board meeting for DBTd[1]
[b] Feedback from clinical advisory board meeting for DBTd[2]

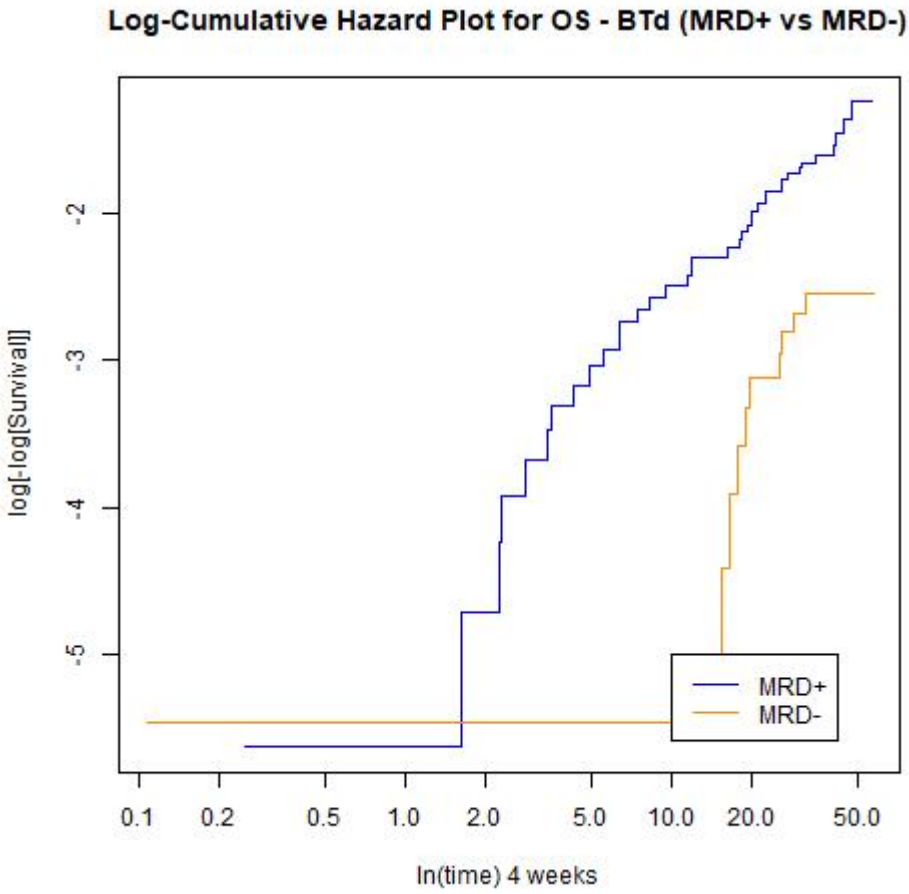
Figure 2: Extrapolation of PFS for BTd MRD+ (updated landmark analysis)

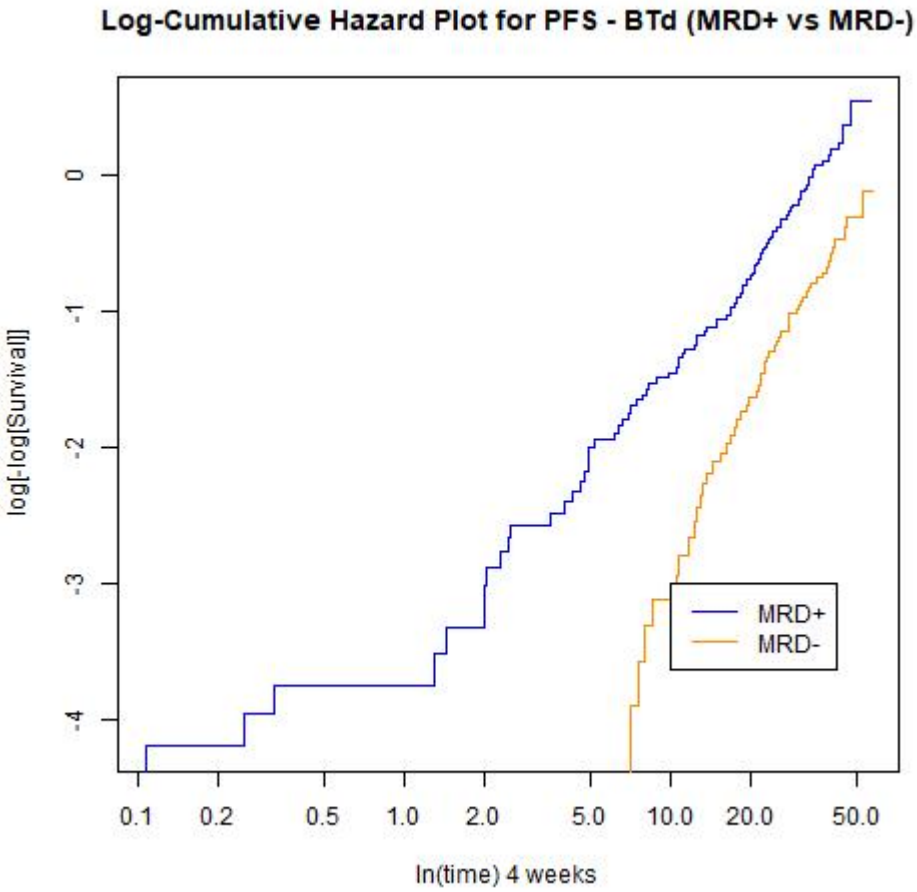
[redacted]

NICE National Institute for Health and Care Excellence

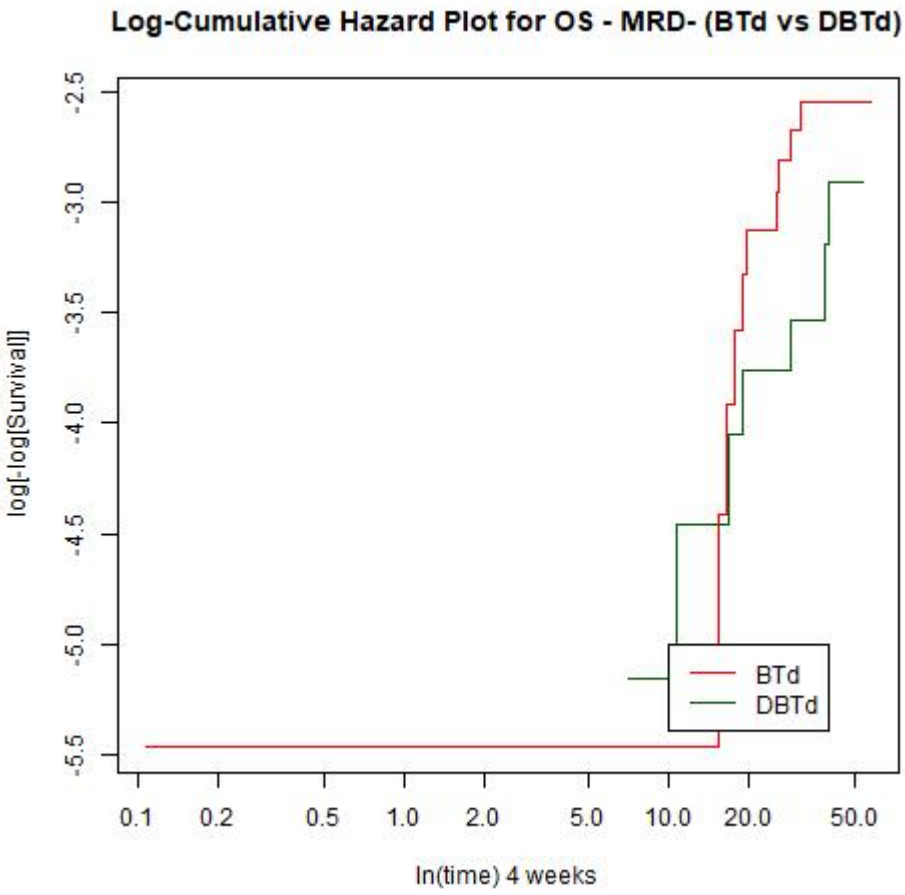
Appendix B: Updated landmark analysis - tests of proportional hazards

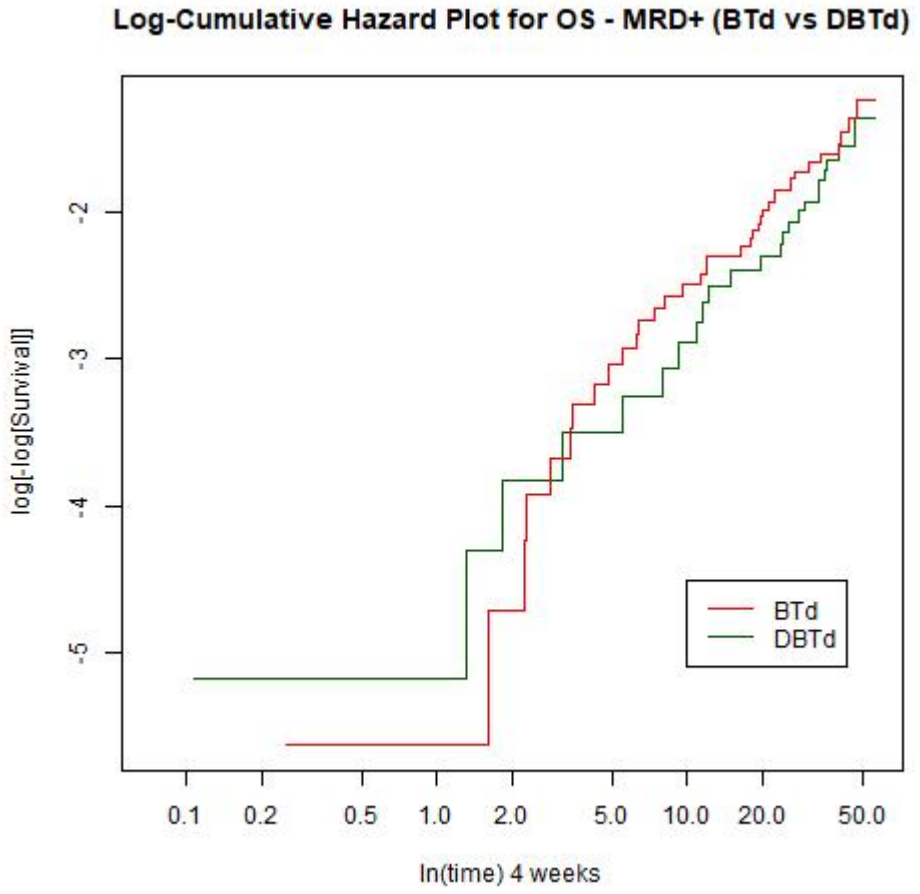
B.1. MRD status for BTd

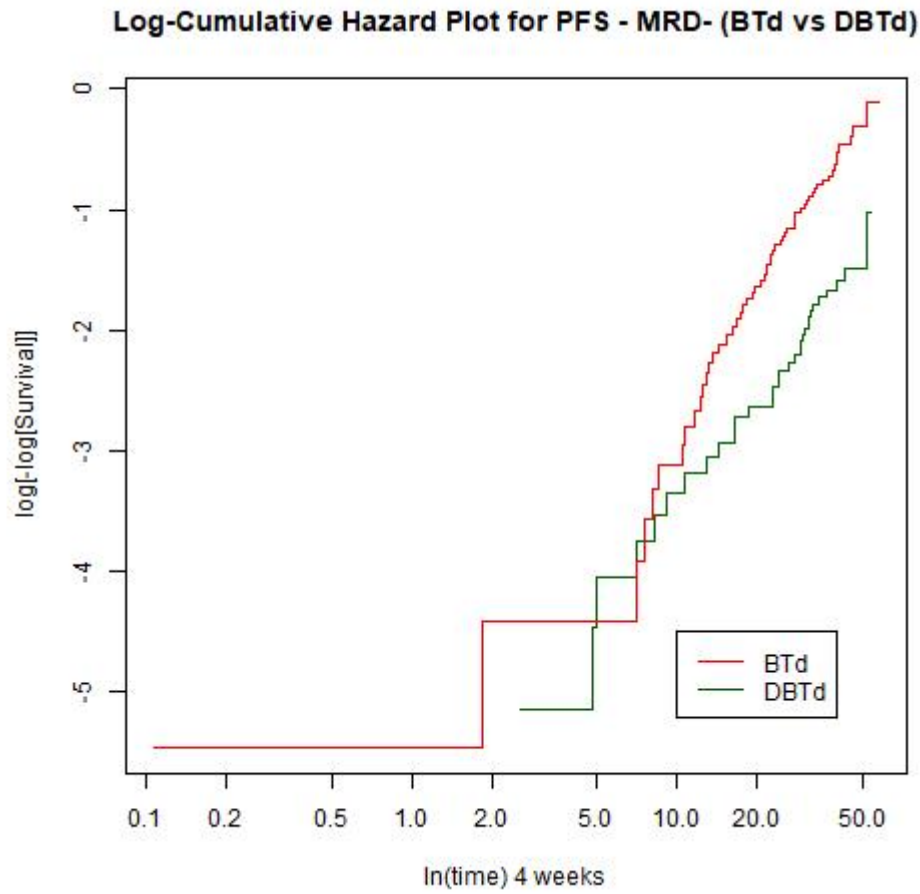


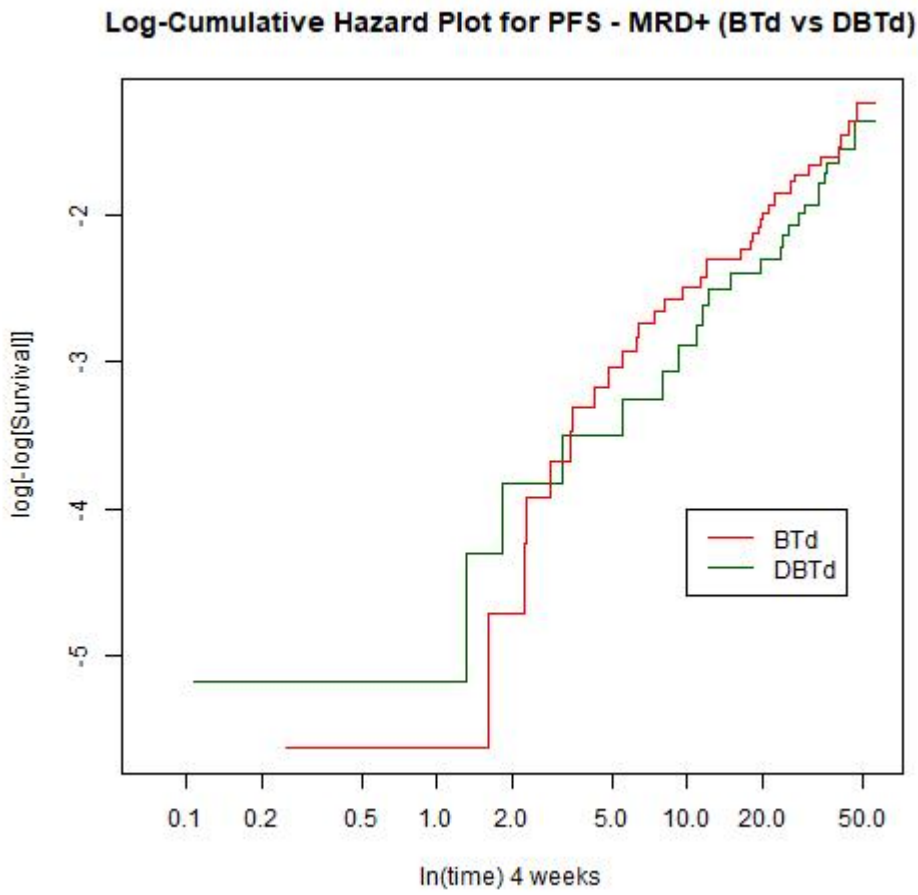


B.2. Treatment by MRD status


Log-Cumulative Hazard Plot for OS - MRD- (BTd vs DBTd)
log[-log[Survival]]
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
-5.5
0.1
0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
ln(time) 4 weeks
BTd
DBTd


Log-Cumulative Hazard Plot for OS - MRD+ (BTd vs DBTd)
log[-log[Survival]]
-2
-3
-4
-5
BTd
DBTd
0.1
0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
ln(time) 4 weeks


Log-Cumulative Hazard Plot for PFS - MRD- (BTd vs DBTd)
log[-log[Survival]]
0
-1
-2
-3
-4
-5
BTd
DBTd
0.1
0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
ln(time) 4 weeks



B.3 Schoenfeld residuals

Table 5: Schoenfeld residuals test

	PFS MRD+			PFS MRD-			OS MRD+			OS MRD-		
	chisq	df	p	chisq	df	p	chisq	df	p	chisq	df	p
Treatment	0.741	1	0.39	0.0481	1	0.83	0.279	1	0.6	0.202	1	0.65
Global	0.741	1	0.39	0.0481	1	0.83	0.279	1	0.6	0.202	1	0.65

Key: chisq: chi-square; df: degrees of freedom; MRD: minimal residual disease; OS: overall survival; PFS: progression-free survival

Appendix C: Comparison of CASSIOPEIA and GIMEMA phase III studies

Technical engagement response form
Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Table 6: Comparison of demographic and disease characteristics of BTd patients at baseline[3,4,5]

Characteristic	CASSIOPEIA (n=542)	GIMEMA (n=236)
Age (years)		
Median	58.0	58.0
Mean	56.5	56.3
Gender		
Male	319 (59%)	137 (58%)
Female	223 (41%)	99 (42%)
Myeloma subtype		
IgG	333 (61%)	154 (65%)
IgA	104 (19%)	41 (17%)
Light chain	66 (12%)	40 (17%)
Other	39 (8%)	1 (<1%)
ISS disease stage		
I	228 (42%)	107 (45%)
II	233 (43%)	91 (39%)
III	81 (15%)	38 (16%)
B2-microglobulin (mg/L)		
Median	NR	3.0
Mean	NR	3.8
Albumin (g/L)		
Median	NR	38.3
Mean	NR	38.3
Creatinine (µmol/L)		
Median	73	84.5
Mean	76.2	88.5
Haemoglobin (g/L)		
Median	115.0	111.5
Mean	114.7	111.0
Platelets (x10^9 per L)		
Median	250.0	231.5
Mean	253.5	243.7
Bone marrow plasma cells		
Median	NR	50%
Mean	NR	52.4%
FISH analysis for cytogenetic abnormalities[a]		
Absence of del(13q), t(4;14), or del(17p)	NR	100 (46%)
Presence of del(13q)	NR	103 (47%)
Presence of t(4;14) with or	NR	53 (24%)

without del(17p)		
Cytogenetic profile		
T(4; 14)		
N[b]	503	NR
Normal	450 (89.5%)	NR
Abnormal	53 (10.5%)	NR
Del17P		
N[c]	503	NR
Normal	464 (92.2%)	NR
Abnormal	39 (7.8%)	NR
Risk result		
N[d]	540	NR
High risk	86 (15.9%)	NR
Standard risk	454 (84.1%)	NR

Key: BTd = bortezomib, thalidomide, and dexamethasone; FISH=fluorescence in situ hybridisation; ISS=International Staging System; NR = not reported.

[a]218 patients in the VTD group and 223 in the TD group were available for assessment.

[b] Subjects with t(4; 14) measured (normal or abnormal).

[c] Patients with Del17p measured (normal or abnormal).

[d] Includes patients with risk results available.

Table 7: Comparison of BTd dosing schedules[3,5]

Study / Treatment	Treatment Phase	Duration	Drugs per cycle

CASSIOPEIA / BTd	Induction	4 cycles of 28 days	Bortezomib – 1.3 mg/m^2 BW for 2 weeks
			Thalidomide – 100 mg QD
			Dexamethasone – 40 mg BW for 2 cycles
			Dexamethasone – 40 mg BW week 1, 20 mg BW weeks 2, 3, for 2 cycles
	Consolidation	2 cycles of 28 days	Bortezomib – 1.3 mg/m^2 BW for 2 weeks
			Thalidomide – 100 mg QD
			Dexamethasone – 20 mg BW for weeks 1–3
	Maintenance (Part 2)	Until disease progression or a maximum of 2-years	Patients with a PR or better randomised 1:1 to observation or daratumumab 16 mg/kg Q8W
GIMEMA / BTd	Induction	3 cycles of 21 days	Bortezomib – 1.3 mg/m^2 BW for 2 weeks
			Thalidomide – 100 mg QD for 2 weeks, 200 mg QD thereafter
			Dexamethasone – 40 mg 4QW for 2 weeks
	Consolidation	2 cycles of 35 days	Bortezomib – 1.3 mg/m^2 QW for 4 weeks
			Thalidomide – 100 mg QD
			Dexamethasone – 40 mg BW for 4 weeks
	Maintenance	Until disease progression or intolerance	Dexamethasone – 40 mg on days 1-4 every 28 days

Key: ASCT = autologous stem cell transplant; BTd = bortezomib, thalidomide and dexamethasone; BW = bi-weekly; PR = partial response; QD = daily; QW = every week; Q2W = every 2 weeks; Q8W = every 8 weeks; 4QW = 4 times per week.

Appendix D: Updated cost-effectiveness analysis results (with PAS for daratumumab)

Base case results

Table 8: Deterministic base case results

Intervention	Total costs	Total LYs	Total QALYs	Inc. costs	Inc. LYs	Inc. QALYs	ICER

DBTd	[redacted]	[redacted]	[redacted]	-	-	-	-
BTd	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£17,957

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; ICER = incremental cost-effectiveness ratio; LY = life year; QALY = quality-adjusted life year.

Probabilistic sensitivity analysis

Table 9: Average probabilistic cost-effectiveness results

Comparison versus	Inc. costs	Inc. QALYs	ICER
BTd	[redacted]	[redacted]	£21,474

Key: BTd = bortezomib, thalidomide and dexamethasone; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year.

Figure 3: Cost-effectiveness plane for DBTd versus BTd

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; QALY = quality-adjusted life year.

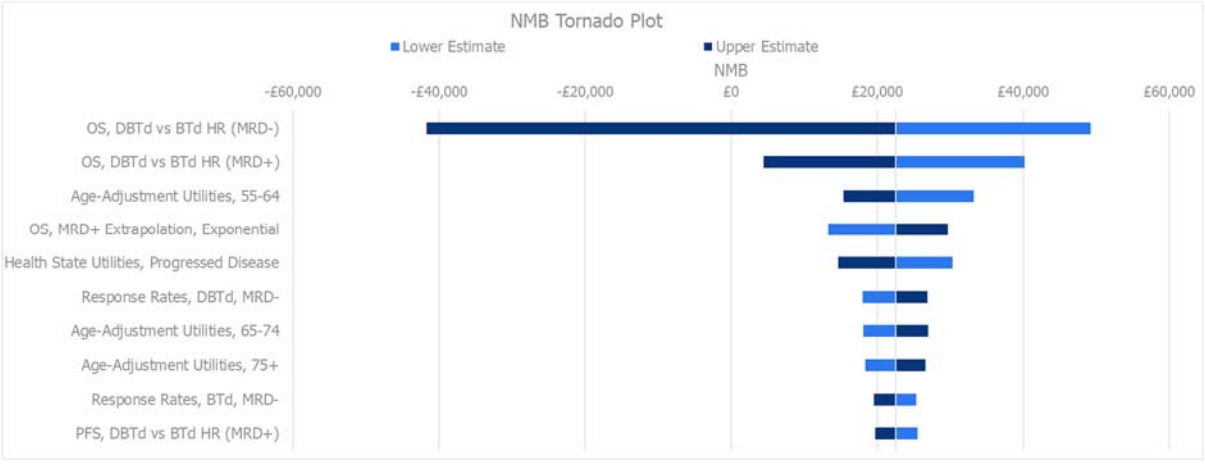
Figure 4: Cost-effectiveness plane for DBTd versus BTd

[redacted]

Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone.

Deterministic sensitivity analysis

Figure 5: NMB tornado plot from deterministic sensitivity analyses – top 10 parameters



Key: BTd = bortezomib, thalidomide and dexamethasone; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; MRD = minimal residual disease; NMB = net monetary benefit; OS = overall survival; PFS = progression-free survival.

Scenario analysis results

Table 10: Summary of results from scenario analyses

Scenario	Inc. costs	Inc. QALYs	ICER (£ per QALY)
Base case	[redacted]	[redacted]	£17,957
1A: Approach to modelling BTd MRD- (Extrapolation of BTd MRD- from CASSIOPEIA using exponential for PFS and exponential for OS)	[redacted]	[redacted]	£18,436
1B: Approach to modelling DBTd MRD- (Using HR for MRD- versus MRD+ from the updated SLR/meta-analysis, MRD assessed regardless of response)	[redacted]	[redacted]	£36,369
2: Extrapolation of BTd MRD+ PFS (Weibull)	[redacted]	[redacted]	£15,827
3A: No additional treatment effect of DBTd after 5 years (MRD+ and MRD-)	[redacted]	[redacted]	£29,787
3B: No additional treatment effect of DBTd after 10 years (MRD+ and MRD-)	[redacted]	[redacted]	£19,904
3C: No additional treatment effect of DBTd after 5 years (MRD- only)[a]	[redacted]	[redacted]	£26,882
3D: No additional treatment effect of DBTd after 10 years (MRD- only)[a]	[redacted]	[redacted]	£19,162
4: Daratumumab IV formulation	[redacted]	[redacted]	£20,562
5: Inclusion of subsequent therapies recommended via the Cancer Drugs Fund	[redacted]	[redacted]	Dominant
6: Dosing for BTd (based on bortezomib SmPC)	[redacted]	[redacted]	£16,037
7: With vial sharing	[redacted]	[redacted]	£17,884
8A: PD utility = 0.644 from van Agthoven et al. (2004) (TA311)	[redacted]	[redacted]	£17,157
8B: PD utility = 0.57 from Palumbo et al. (2013) (TA510)	[redacted]	[redacted]	£16,009
8C: Utility values from van Agthoven et al. (2004)	[redacted]	[redacted]	£17,192
9: PFS HR using updated landmark analysis; OS HR using original landmark analysis	[redacted]	[redacted]	£14,181

Key: BTd = bortezomib, thalidomide and dexamethasone; CI = confidence interval; DBTd = daratumumab, bortezomib, thalidomide and dexamethasone; HR = hazard ratio; ICER = incremental cost-effectiveness ratio; Inc. = incremental; IV = intravenous; MRD = minimal residual disease; OS = overall survival; PD = progressed disease; PFS = progression-free survival; QALY = quality-adjusted life year; SLR = systematic literature review; SmPC = Summary of Product Characteristics.

[a] In this scenario, the treatment effect is still applied across the entire model time horizon for MRD-positive patients.

References

[1] Janssen. Personal Communication with Consultant Haematologist in the UK. May 2020.
[2] Janssen. [Data on File] Clinical Advisory Board Meeting Minutes. August 2020
[3] Moreau P, Attal M, Hulin C, Arnulf B, Belhadj K, Benboubker L, et al. Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA): a randomised, open-label, phase 3 study. Lancet (London, England). 2019;394(10192):29-38.
[4] Janssen. [Data on File] MMY3006. Clinical Study Report: Part 1. 2019.
[5] Cavo M, et al. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. The Lancet. 2010;376(9758):2075-85

NICE National Institute for Health and Care Excellence

Clinical expert statement & technical engagement response form

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Thank you for agreeing to comment on the ERG report for this appraisal, and for providing your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature. The ERG report and stakeholder responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

Information on completing this form:

- In **part 1** we are asking you to complete questions where we ask for your views on this technology. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.
- In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.
- The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a clinical perspective could help either:
- resolve any uncertainty that has been identified
  OR
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

Please return this form by **5pm** on **11 January 2021**

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Important information on completing this expert statement

- 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 want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Do not include medical information about yourself or another person that could identify you or the other person.
- 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.If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

PART 1 – Treating a patient with multiple myeloma and current treatment options	
About you	
1. Your name	**Dr Karthik Ramasamy**
2. Name of organisation	**Oxford University Hospitals NHS FT**
3. Job title or position	**Consultant Haematologist**
4. Are you (please tick all that apply):	☐ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☒ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
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)	☐ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☒ other (they didn't submit one, I don't know if they submitted one etc.)

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.	**Nil**
The aim of treatment for this condition	
8. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	The aim of Daratumumab with Bortezomib Thalidomide and dexamethasone is to induce myeloma remission prior to autologous transplant.

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.)	There are internationally agreed criteria for assessing response (International Myeloma Working Group Rajkumar et al. Blood 2011;117:4691-4695 These are based on the proportional reduction of serum paraprotein / serum free light chains (serological markers of myeloma), urine monoclonal protein and the bone marrow proportion of myeloma plasma cells. Clinical response (IMWG criteria) – Very good partial response or better (VGPR) would be a significant treatment response in this patient population
10. In your view, is there an unmet need for patients and healthcare professionals in this condition?	The current induction regimen for patients when transplant is suitable, Bortezomib thalidomide and dexamethasone (TA331) typically induced VGPR or better in upto 60% of newly diagnosed myeloma patients (Rosinol et al 2012 Blood). This proportion has to be significantly increased with future treatment modalities
What is the expected place of the technology in current practice?	
11. How is the condition currently treated in the NHS?	Induction with Bortezomib thalidomide and dexamethasone
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	NICE NG 35
• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please	Care pathway is well defined

state if your experience is from outside England.)	
• What impact would the technology have on the current pathway of care?	The new technology DVTD does provide significantly better responses for patients prior to stem cell transplantation based on the Cassiopeia trial results
12. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Yes
• How does healthcare resource use differ between the technology and current care?	Healthcare resource use will not be significantly different. Currently patients attend atleast weekly for Bortezomib injection subcutaneously, if this technology is approved Daratumumab will also be administered at the same visit
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Secondary care chemotherapy day units
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	Nil. Daratumumab is currently used for relapsed myeloma patients
13. Do you expect the technology to provide clinically meaningful	Yes

benefits compared with current care?	
• Do you expect the technology to increase length of life more than current care?	There is a potential to improve overall survival for patients with longer follow up. To characterise depth of response, minimal residual disease (MRD) was assessed in the CASSIOPEIA trial: - 63.7% of DVTd patients achieved MRD negative status vs 43% of VTd patients Achieving MRD negative status after treatment for a newly diagnosed myeloma patients is associated with a longer PFS and OS (Munshi et al JAMA 3,1 (2017):28-35)
• Do you expect the technology to increase health-related quality of life more than current care?	Yes, as patients achieve deeper responses and these responses are sustained for longer. There are no significant safety concerns with DVTD regimen
14. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	There are no specific sub groups identified in the Cassiopeia study
The use of the technology	
15. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any	There are no additional visits with this new technology. During first cycle (28 days) patients will be monitored following Daratumumab injections to ensure there are no acute reactions. Patients receive premedication to prevent reactions prior to Daratumumab therapy.

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.)	
16. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	This technology will be given for a fixed number of cycles. Treatment may be stopped earlier due to lack of response or a significant toxicity event.
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?	No

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?	Current induction regimen VTD is unable to induce deep remissions for a proportion of newly diagnosed patients prior to transplant. Addition of Daratumumab is able to increase the proportion of patients who respond and induce deeper responses (MRD negativity) which is sustained for longer
• Is the technology a 'step-change' in the management of the condition?	Induction regimens have incrementally improved outcomes in myeloma over the last 5 years. Addition of Daratumumab is a further improvement to induction regimens prior to autologous stem cell transplant
• Does the use of the technology address any particular unmet need of the patient population?	Enables more patients to go through autologous stem cell transplant
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?	Increased risk of infections noted with addition of Daratumumab, but has been manageable with use of prophylactic antibiotics and close monitoring. Daratumumab induces reactions require monitoring and premedication.
Sources of evidence	

20. Do the clinical trials on the technology reflect current UK clinical practice?	VTD the control arm in Cassiopeia is current standard of care. Hence trial has tested current standard of care against DVTD the test regimen
• If not, how could the results be extrapolated to the UK setting?	Patients in CASSIOPEIA trial receive 4 cycles before transplant and 2 cycles after transplant. We give VTD 6 cycles before transplant.This change can be seamlessly incorporated in our clinical practice.
• What, in your view, are the most important outcomes, and were they measured in the trials?	CR rates, MRD negativity rates, PFS
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	MRD negativity rates are used as a surrogate measure. Previously published meta analysis of induction trials in newly diagnosed MM patients show MRD negativity to be a good surrogate marker for improved OS
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	No
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 technology appraisal guidance [TA311]?	No
23. How do data on real-world experience compare with the trial data?	No real world data available
Equality	
24a. Are there any potential equality issues that should be taken into account when considering this treatment?	Nil
24b. Consider whether these issues are different from issues with current care and why.	
Topic-specific questions	
25. Do you anticipate any issues with integrating consolidation	No

therapy into clinical practice in the NHS?	
26. Is CTd (cyclophosphamide, thalidomide and dexamethasone) used in clinical practice for this indication?	CTD is no longer used since Bortezomib thalidomide and dexamethasone (TA331) has been made available as induction therapy for transplant eligible patients

PART 2 – Technical engagement questions for clinical experts	
Issues arising from technical engagement	
We welcome your response to the questions below, but you do not have to answer every question. If you think an issue that is important to clinicians or patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the professional organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
Key issue 1: Uncertainty in hazard ratios from the company's meta-analysis of the effects of minimum residual disease (MRD) status on survival outcomes	
Key issue 2: Inconsistency in the company's approach for defining and analysing MRD-negative patients	

Key issue 3: Uncertainty in the company's adjustment of progression-free survival (PFS) to capture the effect of a second randomisation to maintenance therapy	
Key issue 4: Plausibility of long-term survival with standard care (autologous stem cell transplant [ASCT] with bortezomib, thalidomide and dexamethasone [BTd] induction and consolidation)	5 yrs - 80%, 10 yrs - 60%, 20 yrs - 20%
Key issue 5: Uncertainty over daratumumab treatment effects on PFS and overall survival (OS)	This requires longer term follow up to resolve as is the case with all newly diagnosed MM trials
Key issue 6: Waning of treatment effects for daratumumab	This is a difficult once due to no long term data (i.e > 5 yrs) with Daratumumab in frontline therapy. Patients receving transplant are younger and more fitter. BTd vs Td data 10 year median fu published (*Lancet Haematol* 2020; 7: e861–73) shows BTd arm continues to show sustained treatment effect 10 years out. It is not inconceivable with higher MRD negative rates on daratumumab arm, similar results could be observed with Daratumumab use

Are there any important issues that have been missed in ERG report?	
PART 3 -Key messages	
27. In up to 5 sentences, please summarise the key messages of your statement: • New technology DVTD has been compared with existing standard of care • New technology DVTD has shown improved clinical outcomes, manageable safety and can be incorporated into clinical practice • DVTD significantly improves MRD negativity rates (deeper response measure) which is a good surrogate for OS • Clinicians and patients will be keen to take up this new technology to improve outcomes in clinical care •	

Thank you for your time.

Please log in to your NICE Docs account to upload your completed document, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

NICE National Institute for Health and Care Excellence

☐ **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.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments **5:00pm, 11 January 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response 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.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. 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.

- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- 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 replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, 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 during engagement 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.

About you

Your name	**NEIL RABIN**
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**UK MYELOMA FORUM**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**NONE**

NICE National Institute for Health and Care Excellence

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Uncertainty in hazard ratios from the company's meta-analysis of the effects of minimum residual disease (MRD) status on survival outcomes	**NO**	No further comments to add
Key issue 2: Inconsistency in the company's approach for defining and analysing MRD-negative patients	**YES**	The company submission (CS) states that MRD-negativity was determined regardless of response and that this is inconsistent with the International Myeloma Working Group (IMWG) definition of MRD negativity which requires a complete response. We would support using MRD-negativity data despite their being serological evidence of disease (VGPR), due to the long half life of a serum paraprotein, interference from Daratumumab and development of oligoclonal bands post ASCT. Patients who are MRD negative will going into a Complete Response (CR) over time. It would be reasonable for follow up data to demonstrate this deepening response.
Key issue 3: Uncertainty in the company's adjustment of progression-free survival (PFS) to capture the effect of a second randomisation to maintenance therapy	**NO**	No further comments to add

Key issue 4: Plausibility of long-term survival with standard care (autologous stem cell transplant [ASCT] with bortezomib, thalidomide and dexamethasone [BTd] induction and consolidation)	**YES**	Regarding adjustments based on landmark analysis and the comments made the concerns mentioned by the ERG. We would support the CS as clinically plausible. Given the improvements in clinical care and access to novel therapies it is reasonable to expect better outcomes for our patients than was published in the BTD NICE submission (TA311).
Key issue 5: Uncertainty over daratumumab treatment effects on PFS and overall survival (OS)	**NO**	This requires longer term follow up to resolve as is the case with all newly diagnosed MM trials
Key issue 6: Waning of treatment effects for daratumumab	**YES**	This is difficult due to the lack of long-term data (i.e > 5 yrs) with Daratumumab in the frontline setting. Patients receiving transplants are younger and fitter. Published data with BTd compared to Td, with a 10 year median follow up, (*Lancet Haematol* 2020; 7: e861–73) shows a sustained effect of BTd therapy at 10 years. It is conceivable that the improved MRD rate seen with the addition of Daratumumab (D-BTd) may show similar (if not better) improvements at 10 years.

NICE National Institute for Health and Care Excellence

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response
Additional issue 1: Insert additional issue	Please indicate the section(s) of the ERG report that discuss this issue	YES/NO	Please include your response, including any new evidence, data or analyses, and a description of why you think this is an important issue for decision making
Additional issue 2: Insert additional issue	Please indicate the section(s) of the ERG report that discuss this issue	YES/NO	Please include your response, including any new evidence, data or analyses, and a description of why you think this is an important issue for decision making
Additional issue N: Insert additional issue			[INSERT / DELETE ROWS AS REQUIRED]

NICE National Institute for Health and Care Excellence

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the company's base-case ICER
Insert key issue number and title as described in the ERG report	Briefly describe the company's original preferred assumption or analysis	Briefly describe the change(s) made in response to the ERG report	Please provide the ICER resulting from the change described (on its own), and the change from the company's original base-case ICER
..	..	..	[INSERT / DELETE ROWS AS REQUIRED]
Company's preferred base case following technical engagement	Incremental QALYs: [QQQ]	Incremental costs: [£££]	Please provide the revised company base-case ICER resulting from combining the changes described, and the change from the company's original base-case ICER

NICE National Institute for Health and Care Excellence

Patient expert statement and technical engagement response form

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Thank you for agreeing to give us your views on this treatment 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.

About this Form

In **part 1** we are asking you to complete questions about living with or caring for a patient with the condition.

In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.

The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a patient perspective could help either:

- resolve any uncertainty that has been identified
  or
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.
-

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

If you have any questions or need help with completing this form please email the public involvement team via pip@nice.org.uk (please include the ID number of your appraisal in any correspondence to the PIP team).

Please return this form by **5pm** on **11 January 2021**

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Please use this questionnaire with our hints and tips for patient experts. You can also refer to the Patient Organisation submission guide. **You do not have to answer every question** – they are prompts to guide you. There is also an opportunity to raise issues that are important to patients that you think have been missed and want to bring to the attention of the committee. The text boxes will expand as you type.

Important information on completing this expert statement

- 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 want 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 15 pages.

PART 1 – Living with or caring for a patient with multiple myeloma and current treatment options	
About you	
1. Your name	
2. Are you (please tick all that apply):	X a patient with multiple myeloma? **X** a patient with experience of the treatment being evaluated? ☐ a carer of a patient with multiple myeloma? **X** a patient organisation employee or volunteer? ☐ other (please specify):
3. Name of your nominating organisation.	Myeloma UK
4. Has your nominating organisation provided a submission? Please tick all options that apply.	☐ No, (please review all the questions below and provide answers where possible) **X** Yes, my nominating organisation has provided a submission **X** I agree with it and **will be** completing a patient expert statement ☐ Yes, I authored / was a contributor to my nominating organisations submission ☐ I agree with it and **do not wish to** complete this statement ☐ I agree with it and **will be** completing

5. How did you gather the information included in your statement? (please tick all that apply)	**X** I am drawing from personal experience. ☐ I have other relevant knowledge/experience (e.g. I am drawing on others' experiences). Please specify what other experience: ☐ I have completed part 2 of the statement **after attending** the expert engagement teleconference ☐ I have completed part 2 of the statement **but was not able to attend** the expert engagement teleconference **X** I have not completed part 2 of the statement
Living with the condition	
6. What is your experience of living with multiple myeloma? If you are a carer (for someone with multiple myeloma) please share your experience of caring for them.	One day in 2012 I woke up with double vision. I was 45, with 2 children (7 and 9) and working for a local authority children's services department as a commissioner. I was diagnosed with a plasmacytoma at the brain stem. This was treated with radiotherapy. I had further course of radiotherapy in the same year to treat another bone lesion in my spine. I continued to work until 2014 when I was diagnosed with multiple myeloma. I chose to take early retirement so I could concentrate on looking after myself and my family. I have had 3 treatments: - Velcade, thalidomide and dexamethasone followed by a autologous stem cell transplant (ASCT) in 2015 - Ixazomib, thalidomide and dexamethasone followed by ASCT in 2017 (ACCoRD trial) - Daratumumab, velcade and dexamethasone 2020

	I currently on maintenance treatment (monthly Daratumumab and dexamethasone). Whilst this might sound gruelling, my disease has been relatively well controlled and for the most part I've lived a normal life, especially during periods of remission! I don't feel like I've missed out on anything. I've attended all my children's parents evenings, shows and sporting events. We've had lots of wonderful holidays, birthdays and christmases together. It's been great. The most difficult part of living with myeloma is knowing that relapse is inevitable.
Current treatment of the condition in the NHS	
7a. What do you think of the current treatments and care available for multiple myeloma on the NHS? 7b. How do your views on these current treatments compare to those of other people that you may be aware of?	The best thing about myeloma treatment is the range of options now available. It gives me hope that I can continue to live a relatively normal life for years to come. I have found the best treatments are the least invasive. By this I mean those that can be given in tablet form or subcutaneous injections. Such treatments are quick and easy and don't intrude on your daily activities so much. I think is goes without saying that any treatment that gives a longer remission, especially those that have an end point (like the one under review), are great. In my experience, once you have recovered, life can really get back to normal. Time in remission, treatment free, is a gift.
8. If there are disadvantages for patients of **current NHS treatments** for multiple myeloma (for example how the treatment is given or taken, side effects of treatment etc) please describe these	I've found all the treatments relatively easy to tolerate (especially daratumumab) The main side effect I've experienced is fatigue which I attribute to dexamethasone as it interrupts sleep.

Advantages of this treatment	
9a. If there are advantages of daratumumab over current treatments on the NHS please describe these. For example, the impact on your Quality of Life, your ability to continue work, education, self-care, and care for others? 9b. If you have stated more than one advantage, which one(s) do you consider to be the most important, and why? 9c. Does daratumumab help to overcome/address any of the listed disadvantages of current treatment that you have described in question 8? If so, please describe these.	9a. I've had all the treatments under review but not as a quadruplet. These are the advantages that I can see: Longer, treatment free remission I believe adding daratumumab to VTD at induction is unlikely to add to much in terms of additional side effects (I have been receiving dara monthly for 8 cycles with no noticeable side effects) 9b. Longer, treatment free remission 9c. Longer, treatment free remission
Disadvantages of this treatment	
10. If there are disadvantages of daratumumab over current treatments on the NHS please describe these? For example, are there any risks with daratumumab? If you are concerned about any	No

potential side affects you have heard about, please describe them and explain why.	
Patient population	
11. Are there any groups of patients who might benefit more from daratumumab or any who may benefit less? If so, please describe them and explain why. Consider, for example, if patients also have other health conditions (for example difficulties with mobility, dexterity or cognitive impairments) that affect the suitability of different treatments	If I could have been confident of a long, treatment free remission after my first ASCT I might have considered staying in work
Equality	
12. Are there any potential equality issues that should be taken into account when considering multiple myeloma and daratumumab? 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 More general information about the Equality Act can and equalities issues can be found at https://www.gov.uk/government/publications/easy-read-the-equality-act-making-equality-real and https://www.gov.uk/discrimination-your-rights.	
Other issues	
13. Are there any other issues that you would like the committee to consider?	

PART 2 – Technical engagement questions for patient experts	
Issues arising from technical engagement	
We welcome your response to the questions below, but you do not have to answer every question. If you think an issue that is important to patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the patient organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
14a. Are the comparators in the company submission used in the NHS for treating multiple myeloma? 14b. Is the assessment tool used in the clinical trial appropriate for assessing the severity of multiple myeloma? 14c. What are the main benefits of daratumumab for	

patients? If there are several benefits please list them in order of importance. Are there any benefits of daratumumab that have not been captured? 14d. What are the benefits of daratumumab for carers?	
15. Are there any important issues that have been missed in ERG report?	

PART 3 -Key messages

16. In up to 5 sentences, please summarise the key messages of your statement

- I understand this drug combination will offer myeloma patients a longer, treatment free, remission at first line treatment and I would have been so grateful for this which is why I fully support this proposal.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

.....

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.

.....

NICE National Institute for Health and Care Excellence

Technical engagement response form

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments **5:00pm, 11 January 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response 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.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. 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.

- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- 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 replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, 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 during engagement 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.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Myeloma UK**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**None**

NICE National Institute for Health and Care Excellence

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: Uncertainty in hazard ratios from the company's meta-analysis of the effects of minimum residual disease (MRD) status on survival outcomes	**NO**	Myeloma UK notes that the company is to provide an updated systematic literature review (SLR) and meta-analysis of the effects of MRD status on survival outcomes. We would like to emphasise that the **achievement of MRD-negative status following treatment was associated with a significant improvement in progression-free survival and overall survival.**[1] This has been demonstrated across both clinical trial settings and through studies looking at real world data.[2]
Key issue 2: Inconsistency in the company's approach for defining and analysing MRD-negative patients	**NO**	Myeloma UK notes that this could be clarified by the company when they provide an update of the meta-analysis in Issue 1. In the ERG report it was stated that "**the reported rates of MRD negative response from the trial exceed the reported rates of complete response and stringent complete response (≥CR). The CS notes that this is due to a lag in**

[1] Munshi NC, Avet-Loiseau H, Rawstron AC, et al. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. *JAMA Oncol.* 2017;3(1):28–35. doi:10.1001/jamaoncol.2016.3160

[2] Real-world sustained minimal residual disease (MRD) negativity using NGS in multiple myeloma. Audrey Demaree, Anthony Hewitt, Lik Wee Lee, and Benjamin Eckert Journal of Clinical Oncology 2020 38:15_suppl, e19280-e19280

		the decay of serum paraprotein (required for complete response) compared to clearance of malignant cells in the bone marrow (required for MRD) ." Myeloma UK would agree with the company's submission that there is a difference between MRD status and serological response – some patients who are measured as having very good partial response (VGPR) could possibly move into a complete response (CR) or stringent complete response (sCR).
Key issue 3: Uncertainty in the company's adjustment of progression-free survival (PFS) to capture the effect of a second randomisation to maintenance therapy	**NO**	No comment – company to supply updated analysis when Part 2 of the trial is unblinded.
Key issue 4: Plausibility of long-term survival with standard care (autologous stem cell transplant [ASCT] with bortezomib, thalidomide and dexamethasone [BTd] induction and consolidation)	**NO**	Myeloma UK notes that the issue of modelling long term survival in myeloma. The ERG report highlights the "**high uncertainty over the plausibility of survival estimates over the modelled time horizon of over 40 years because of a lack of long-term data for the population of interest."** We would emphasise that patients living with myeloma for 40 years is exceptionally rare. There is a clinical assumption that 5% of patients will reach 20 years of survival with myeloma.
Key issue 5: Uncertainty over daratumumab treatment effects on PFS and overall survival (OS)	**NO**	No Comment

Key issue 6: Waning of treatment effects for daratumumab	**NO**	Myeloma UK notes that clinical opinion on the plausibility of persistence of treatment effects has been requested to resolve the issue. We would agree with the assessment put forward from the UKMF Technical engagement response.

NICE National Institute for Health and Care Excellence

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response
Additional issue 1: Insert additional issue	Please indicate the section(s) of the ERG report that discuss this issue	**YES/NO**	Please include your response, including any new evidence, data or analyses, and a description of why you think this is an important issue for decision making
Additional issue 2: Insert additional issue	Please indicate the section(s) of the ERG report that discuss this issue	**YES/NO**	Please include your response, including any new evidence, data or analyses, and a description of why you think this is an important issue for decision making
Additional issue N: Insert additional issue			[INSERT / DELETE ROWS AS REQUIRED]

NICE National Institute for Health and Care Excellence

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	**Company's base case before technical engagement**	**Change(s) made in response to technical engagement**	**Impact on the company's base-case ICER**
Insert key issue number and title as described in the ERG report	Briefly describe the company's original preferred assumption or analysis	Briefly describe the change(s) made in response to the ERG report	Please provide the ICER resulting from the change described (on its own), and the change from the company's original base-case ICER
..	..	..	[INSERT / DELETE ROWS AS REQUIRED]
Company's preferred base case following technical engagement	**Incremental QALYs: [QQQ]**	**Incremental costs: [£££]**	**Please provide the revised company base-case ICER resulting from combining the changes described, and the change from the company's original base-case ICER**

CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR Systematic Reviews Programme on behalf of NICE

Daratumumab in combination for untreated multiple myeloma when stem cell transplant is suitable [ID1510]

Evidence Review Group's critique of the company's response to technical engagement

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors Joanne Lord, Neelam Kalita, Lois Woods, Lorna Hazell, David Scott, Geoff Frampton,

Correspondence to Geoff Frampton

Date completed **30 January 2021**

[redacted]

1. Introduction

This document is the ERG's critique of the response by the company (Janssen-Cilag Ltd) to technical engagement. The ERG received the company's response on 13 January 2021.

The company has responded to each of the Key Issues raised in sections 1.4 and 1.5 of the ERG report and three other issues raised in section 1.6 of the ERG report. The company has provided additional cost-effectiveness analyses to address some of the issues, summarised in their TE response with more detail and scenarios in an appendix.

In this report we present the following:

- ERG critique of the company's response to each of the Key Issues for engagement (section 2) and additional issues raised by the ERG (section **Error! Reference source not found.**).
- Additional cost-effectiveness scenarios produced by the ERG (Appendix 2)

2. Key issues for engagement

Issue 1 – Uncertainty in hazard ratios (HRs) from the company's meta-analysis of the effects of minimal residual disease (MRD) status on survival outcomes

Company response	ERG comments
In their Appendix "ERG feedback" the company have provided clarification on the methods used to conduct their updated MRD meta-analysis, which followed an approach reported by Guyot et al. 2012. Graphs are presented which compare the KM curves constructed by the company against the original KM curves reported in the studies.	• The ERG are satisfied that the meta-analysis methodology has been correctly applied. • The company have not updated the meta-analysis with the latest data cut from CASSIOPEIA. The CASSIOPEIA data in the meta-analyses are based on a median follow-up of 29.2 months whereas those in the landmark analysis and model survival extrapolations are based on a median follow up of 44.5 months. However, this additional data is unlikely to make much difference to the meta-analysis or cost-effectiveness results. • Overall, the ERG agree that the company's estimated KM curves provide a good fit to the original data of studies included in the SLR. An

	exception is that for the Rawstron 2013 study, the OS curve appears to have a poor fit at the tail end but given the low numbers at risk this is unlikely to be of importance. • These curves resolve uncertainty in how HRs were obtained for all but three studies. The source of HR and/or confidence intervals remain unclear for Cohen 2016, Paiva 2008 and Hahn 2019 (Appendix 1 in ERG Report Addendum; SLR critique).
The studies varied in how they defined the start time of PFS and OS assessments, whether from inclusion (n=4), start of therapy or ASCT (n=8), from MRD assessment (n=4), or no definition was given (n=5). The company clarified that they assume that the variation in assessment timepoints would not be an effect modifier for the relative effect of MRD, although a rationale for this assumption is not given.	• The ERG are uncertain about the plausibility of the company's assumption. We conducted subgroup analyses which show that HRs differ according to the timepoint from which the PFS and OS assessments begin (Appendix 1). • For PFS, HRs range from 0.29 (95% CI 0.18 to 0.46) when PFS was measured from inclusion, to 0.53 (95% CI 0.43 to 0.65) when PFS was measured from MRD assessment. For OS, HRs range from 0.48 (95% CI 0.39 to 0.60) when PFS was measured from start of therapy or ASCT, to 0.68 (95% CI 0.55 to 0.84) when PFS was measured from time of MRD assessment (Appendix 1). These results suggest that when PFS and OS are measured from a later timepoint they give higher (less favourable) HRs for the effect of MRD negativity on survival.
The company provided a risk of bias assessment for 8 studies in their updated SLR for which a risk of bias assessment had not previously been provided. The company report that they judged two studies to have serious risk of bias and excluded these from	• The company's criteria for determining bias risk are inconsistent across studies. The ERG do not fully agree with the assessments. • The ERG had previously assessed all studies for their key risks of bias as reported in an Addendum to the ERG Report (SLR critique). We concluded that in most cases the risk of bias was unclear, but sensitivity analyses suggested this would have

the PFS and OS meta-analyses in a sensitivity analysis.	relatively little influence on PFS and OS HRs from the meta-analyses. The company's updated analyses do not alter this conclusion.
Studies which reported MRD status regardless of response have been included as a subgroup in the PFS and OS meta-analyses	• The ERG agree with the company assessments, with the exception that one study by Gu et al. was classified as 'at least VGPR' but we were unable to verify this. It appears that all 104 patients recruited had a MRD assessment so this study might be better classified as 'any response'. However, adding this study to the sub-group for 'any response' has minimal impact on the HRs for PFS and OS.

ERG conclusion

- The company's clarification on their meta-analysis methods has reduced uncertainty in the estimation of MRD HRs for the majority of studies.
- However, some uncertainty in HRs obtained from the meta-analysis remains due to heterogeneity of the included studies, as noted in the ERG's SLR critique (ERG Report Addendum).
- Subgroup analyses suggest that studies' results depend on the timing of the survival assessment and there is an indication that studies which assessed survival from later timepoints provide less favourable HRs for the effect of MRD negativity on survival. Although the subgroup analysis showed wide variation in the MRD HR estimates, we note that economic results are not very sensitive to these parameters (see ERG report Table 39).

Issue 2 – Inconsistency in the company's approach for defining and analysing MRD-negative response

Company response	ERG comments
CASSIOPEIA trial estimates of post-consolidation MRD negativity rates (OR for DBTd versus BTd) are 'broadly consistent' between MRD negativity measured according the IMWG definition (MRD- with CR or better) and MRD negativity measured regardless of response	We note that this information has already been presented in CS Appendix L.4. • The confidence interval for the OR with the IMWG definition of MRD negativity (2.06, 95% CI: 1.56 to 2.72) overlaps that with the MRD regardless of response definition (2.27; 95% CI 1.78 to 2.90). • However, the absolute rates of MRD negativity are substantially lower with the IMWG definition (33.7% for DBTd and 19.9% for BTd) than with the MRD regardless of response definition (63.7% for DBTd and 43.5% for BTd). • We report a scenario analysis, which shows that the lower absolute rates of MRD negativity with the IMWG definition increase the ICER by over £7,000 per QALY (Table 2 in Appendix 2). This analysis is only illustrative, as the model relies on results from the CASSIOPEIA landmark analysis that are not consistent with the IMWG definition of MRD negativity (the company has not provided this analysis). See Issue 5 below.
Subgroup analysis of survival by MRD status from the updated meta-analysis gives 'broadly consistent' results with the original meta-analysis when the meta-analysis is restricted to estimates with MRD status defined regardless of response as when all studies are included.	• The HRs (MRD- versus MRD+) for the MRD regardless of response subgroup are: PFS: [redacted] (n=9) OS: [redacted] (n=7) • In comparison, the overall HRs for all studies included in the updated meta-analysis (which contained a mixture of studies with different definitions of MRD negativity) are: PFS: [redacted] (n=20) OS: [redacted] (n=11)

	• We agree with the company that the HRs for the MRD regardless of response subgroup are broadly similar to those of the overall meta-analysis including all studies. The concordance between these analyses is closer for OS than for PFS, but confidence intervals are wide, so it is unclear whether this is a meaningful difference.
The company's base case has been updated so that a consistent definition of MRD (regardless of response) is applied in the MRD meta-analysis, as well as in the CASSIOPEIA MRD response (ITT population) and landmark analyses. This has a negligible impact on the ICER.	• Not surprisingly, the above change has little impact on the ICER: £17,957 per QALY for the revised base case compared to £17,842 per QALY with the original meta-analysis. • We agree these the three data sources, which are the key drivers of ICERs (MRD meta-analysis, rates of MRD negativity at post-consolidation, and the landmark analysis), are now consistently defined according to MRD regardless of response. • However, we do not have an analysis with a consistent definition of MRD according to IMWG criteria, so it is not possible to compare the impact of the MRD definitions on ICERs. In particular, we note that the company have not provided results for a landmark analysis of CASSIOPEIA following an IMWG definition of MRD negativity. This is important, as treatment effects from the landmark analysis are the most influential parameters in the economic model.
A similar scenario exploring the cost-effectiveness results applying the IMWG definition for MRD negativity was not possible because no studies identified in the SLR update reported OS results based on the IMWG definition.	• The company are correct: there are no studies available that meet the IMWG definition for OS (only 4 studies meet the IMWG definition for PFS).

ERG conclusion:

- There does not appear to be a clear consensus on which of the MRD definitions (per IMWG criteria or regardless of response) is the most clinically appropriate.
- Three key clinical outcomes influence the economic model: the proportion MRD-negative in CASSIOPEIA, the HRs from the landmark analysis, and the HRs from the MRD meta-analysis. The company have addressed an inconsistency in the definition of these outcomes such that they are now all based on the definition of MRD when assessed regardless of response.
- Absolute MRD negativity rates from the CASSIOPEIA trial are much lower with the IMWG definition, which increases the ICER for daratumumab by over £7,000 per QALY when applied to the company's revised base case analysis.
- Given the sensitivity of ICERs to MRD negativity rates in CASSIOPEIA, we believe that it is important that the company also provide a sensitivity analysis for the landmark analysis with the IMWG definition of MRD negativity.
- We accept that it is not possible to obtain an HR (MRD positive versus MRD negative) for OS from the meta-analysis according to IMWG criteria, as no studies in the SLR reporting OS met the IMWG MRD definition.

Issue 3 – Uncertainty in the company's adjustment of PFS to capture the effect of second randomisation to maintenance therapy

Company response	ERG comments
The revised company base case includes updated landmark analysis for PFS and OS with censoring of patients re-randomised to daratumumab maintenance Updated PFS and OS with IPW adjustments to latest (August 2020) CASSIOPEIA data cut are consistent with earlier data cuts and indicate minimal impact of the second randomisation.	• The company's updated analysis increases follow up (from median 29.2 months to 44.5 months) in the updated landmark analysis (Company TE response Table 5 and Figures 3 and 4). • Reported results from the updated ITT analysis with IPW adjustment show minimal impact of the second randomisation on HR estimates (DBTd vs. BTd) for PFS and OS (company TE response Tables 1 and 2). • However, we note that the ITT IPW results are inconsistent with the updated and censored landmark analysis (company TE response Table 5), which are key inputs to the economic model. For PFS, the updated analysis with IPW adjustment increases the ITT population HR from 0.495 to [redacted] (company TE response Table 1). However, the updated landmark analysis with censoring decreases the HR for the MRD+ subgroup from [redacted] and for the MRD- subgroup from [redacted] (company TE response Table 5).
IPW with pre-specified weights independent of time was requested by the FDA (for simplicity ease of verification. Other methods were not considered.	• Whilst the company justify their approach for using IPW, their response does not explain why other adjustment methods were not considered. The ERG would have liked to have seen an alternative adjustment methodology as a scenario analysis.
Piecewise HR by study phase (CS B.2.6.2) demonstrates consistent benefit for DBTd compared to BTd, which increased over time. Hence, use	• The piecewise analysis shows that hazards are not proportional. • The approach to test for proportional hazards seems ad hoc, choosing to split the curves by trial phase which does not seem to fit visually with the

of IPW is not invalidated by early violation of the PH assumption.	shape of hazards in the log-log plot in Figure 3 of the company's Clarification Response document. Over the course of the trial, the log hazards repeatedly converge then diverge again. The ERG believe the company should have followed a more formal procedure for accounting for non-proportional hazards, such as fitting cubic splines or fractional polynomials. • HRs remain uncertain due to the apparent violation of the proportional hazards assumption.

ERG conclusion:

- There is an inconsistency in estimated treatment effects obtained using censoring and IPW adjustment approaches. The reason for this is unclear, but it is likely that, although censoring removes potential bias due to any direct effects of the maintenance treatment, it is susceptible to selection bias if there are prognostic differences between people who undergo re-randomisation and those who do not. We therefore consider that the censored landmark analysis used in the economic model is unreliable, and that an analysis of the landmark data would be more appropriate. (See Issue 5 below)
- Uncertainty remains regarding whether the proportional hazards assumption has been adequately met. A more robust evaluation of this would be appropriate, e.g. consideration could be given to fitting models that are not sensitive to the proportional hazards assumption.

Issue 4 – Plausibility of long-term survival with standard care (ASCT with BTd induction and consolidation)

Company response	ERG comments
Survival analysis updated from revised landmark analysis (August 2020 data cut), with censoring for maintenance therapy.	• The updated survival analysis provides an additional 15 months of follow up, though censoring will have reduced power after the re-randomisation to maintenance treatment. • Censoring avoids potential bias due to any effects of daratumumab maintenance, but this analysis is susceptible to selection bias if the patients who were re-randomised to maintenance were different to those who were not. As noted in Issue 3 above, HRs from the updated landmark analysis with censoring are not consistent with the updated ITT HRs with IPW adjustment (see further discussion in Issue 5 below). • Fitted survival curves for extrapolation of PFS and OS in the economic may also be susceptible to selection bias due to censoring of patients who were randomised to daratumumab maintenance.
Extrapolation of PFS and OS for BTd MRD+ performed in accordance with NICE DSU TSD 14 (see Appendix A). Exponential distribution was selected for both PFS and OS, and scenario with Weibull for PFS (and exponential for OS), as in the original CS.	• For PFS, the exponential, Weibull and Gompertz extrapolations for BTd MRD+ have similar goodness-of-fit statistics (AIC and BIC) and are broadly consistent with predictions from clinical experts. However, the Gompertz and Weibull clearly have a better visual fit than the exponential (company TE response Appendix A Tables 3 and 4 and Figure 2). • OS data is less mature. All distributions for BTd MRD+ have a similar visual fit and similar AIC statistics, although the exponential has a lower BIC and gives predictions that are closest to clinical estimates (Company Appendix A Tables 1 and 2 and Figure 1). We therefore consider that the

	company's choice of exponential for the OS extrapolation is reasonable, but that a Weibull extrapolation, which gives higher long-term OS predictions, should also be considered.
Updated OS and PFS outcomes predicted by the model for the overall cohort (BTd MRD- and MRD+), are presented in Figure 1 and Figure 2 respectively with a comparison of survival predictions against the original model presented in Table 3.	• The exponential model for PFS in the overall cohort for standard care does not show a good fit to the KM data (Company TE response Figure 2). The model underestimates PFS during the first three years and then overestimates PFS. Weibull or Gompertz PFS extrapolations give a better overall fit to trial data. • The exponential extrapolation underestimates OS for the overall standard care cohort (Company TE response Figure 1). This suggests that the Weibull may be more appropriate.
To help validate long-term survival predictions for standard of care, Janssen performed a naïve comparison of survival rates (PFS and OS) from the updated model with rates reported from a range of sources (Table 4).	• PFS and OS estimates from the company's revised base case model are similar to updated results from the PHS cohort at 3 years. • Three-year OS estimates from the revised company base case model are also very similar to those from the GIMEMA trial BTd arm. Model OS predictions are similar to GIMEMA results at 5 and 10 years. Modelled PFS was considerably lower at 3, 5 and 10 years than in GIMENA. This may be due to differences in the trial protocols or patient characteristics (see company TE response Appendix C). • Other data sources cited in Table 4 are less useful. The US retrospective cohort only included 51 patients who received first line BTd induction prior to ASCT and the company note that this data was subject to high censoring after five years. The ONS data is not directly relevant, as it includes both transplant-eligible and transplant ineligible patients.

ERG conclusion:

- The ERG has concerns about the use of censored data for the landmark analysis and for fitting survival extrapolations.
- The company's choice of exponential to extrapolate PFS does not produce a good fit to the CASSIOPEIA KM. We suggest that a Weibull extrapolation (as in the Company's original base case) is more appropriate. We also present a scenario with a Gompertz distribution for PFS (Table 2 below).
- We agree with the use of an exponential for OS, although this appears to underestimate survival in CASSIOPEIA. It also gives lower survival predictions than in long-term follow-up of the GIMEMA trial, though this may relate to differences in trial protocols. For comparison we report a scenario with a Weibull extrapolation for OS, which gives higher long-term survival estimates (Table 2).

Issue 5 – Uncertainty over daratumumab treatment effects on PFS and OS (landmark analysis)

Company response	ERG comments
The treatment effect for DBTd is driven by the depth of post-consolidation response achieved with this quadruplet combination and is founded on biological plausibility. Extensive evidence demonstrates improved survival outcomes with deeper responses.	• We acknowledge that it is plausible that treatments that induce a deeper response within (as well as between) MRD response categories are associated with better survival, and that there is evidence in the literature to support this. • Nevertheless, evidence for these additional effects for daratumumab from the CASSIOPEIA trial is subject to uncertainty. This is important because the cost-effectiveness results are very sensitive to these parameters.
Revised landmark analysis with CASSIOPEIA August 2020 data-cut (additional 15.3 months follow-up) and censoring of patients re-randomised to daratumumab maintenance	• The updated and censored landmark analysis has not resolved uncertainty over the additional treatment effects for daratumumab. • The estimated effects for PFS (HR for DBTd vs. BTd) [redacted].

therapy. (Company TE response Table 5 and Figures 3 and 4)	Although, as argued in Issue 3, these results are inconsistent with the IPW ITT results reported in company TE response Table 1. • For OS, the estimated HRs [redacted].
Clear trade-off with censoring for maintenance: risk of bias is eliminated at the cost of precision. Arguably, IPW results suggest uncertainty over OS is minimised with original landmark analysis despite shorter median follow-up.	• The company explain the OS result as a consequence of the smaller effective sample size due to censoring. This is likely to be a factor, but as argued in Issue 3, censoring also introduces risk of a selection bias. Note that the ERG did not suggest censoring as the solution to potential bias from maintenance treatment. An adjusted landmark analysis would have been more appropriate.
Additional scenario with HRs from the original landmark analysis for OS and updated landmark analysis for PFS for the comparison of DBTd versus BTd.	• This scenario represents cherry-picking of results.
ERG conclusion: • The updated landmark analysis lends support to the treatment effect on PFS ([redacted]). • Despite additional follow-up, [redacted]. • We note that the company has not responded to our request for a sensitivity analysis of the landmark analysis with an IPW definition of MRD negativity.	

Issue 6 – Waning of treatment effects for daratumumab

Company response	ERG comments
The key question for clinicians is therefore whether deeper responses are expected to result in a fundamental shift in a patient's prognosis, or whether the survival benefit associated with deeper responses is expected to wane over time.	• We agree that this is the key question.
With median follow-up approaching four years, the revised analysis continues to demonstrate a relative benefit in favour of DBTd, with no evidence to suggest a possible waning of relative benefit over time.	• As argued above, it is difficult to draw conclusions from the updated landmark analysis, due to problems with the use of censoring.
Additional evidence is presented from Part 2 of CASSIOPEIA over two years of follow up for: sustained MRD negativity (Figure 5 and Figure 6); increasing response (≥CR) without maintenance therapy (Table 6); MRD negative conversion (Table 7); and PFS2 (Table 8).	• These analyses show promising indications that measures of response and delayed progression for at least 2 years after consolidation treatment.
GIMEMA study with median follow-up of 10-years showed a persistent relative benefit of BTd versus Td for PFS, and a statistically significant improvement in OS.	• This trial provides evidence for prolonged survival benefit for patients treated at first line with double ASCT and BTd induction and consolidation.
ERG conclusion: • The company provides supportive indirect evidence that treatments that are associated with 'deeper' measures of response are associated with delayed progression and improved survival from the literature.	

APPENDICES

Appendix 1 Heterogeneity in HRs (MRD- versus MRD+) for PFS and OS in relation to the timing of the start of survival assessments (random effects model)

Survival analysis assessment time	PFS	OS
From start of study (inclusion)	**HR=0.29 (0.18 to 0.46)** Cohen 2016 Luoma 2019 Parrondo 2019 Rossi 2018	No studies
From start of therapy or ASCT	**HR=0.34 (0.27 to 0.43)** Bakkus 2004 Chakraborty 2017 Clark 2018 Gu 2018 Paiva 2008 Popat 2017 Rawstron 2002	**HR=0.48 (0.39 to 0.60)** Bakkus 2004 Chan 2019 Chakraborty 2017 Gu 2018 Paiva 2008 Rawstron 2002
From time of MRD assessment	**HR=0.53 (0.43 to 0.65)** CASSIOPEIA Rawstron 2013 Sololev 2016 Sololev 2018	**HR=0.68 (0.55 to 0.84)** CASSIOPEIA Rawstron 2013
Not reported or unclear	**HR=0.28 (0.15 to 0.54)** Garifullin 2019 Hahn 2019 Paiva 2020 Ribolla 2020 Schinke 2017	**HR=0.66 (0.37 to 1.19)** Hahn 2019 Ribolla 2020 Schinke 2017

Appendix 2 ERG validation and additional cost-effectiveness analysis

The ERG reproduced the results for the company's original base case (ERG report Table 33) and revised base case (Company TE response Appendix D Table 8) using the updated company model. Table 1 below shows the cumulative effect of changes that the company made in their TE response (each row incorporates changes from previous rows). The updated landmark analysis (August 2020 data cut) and update to the MRD meta-analysis have a negligible impact on the ICER. The change from Weibull to exponential PFS extrapolation in the BTd MRD+ subgroup increases the ICER by £2,130 per QALY gained.

Table 1 Cumulative change from company's original base case to revised base case (deterministic analysis with confidential PAS for daratumumab)

Individual scenarios on the base case	Treatment	Total costs (£)	Total QALYs	ICER (£/QALY)
Company's original base case	BTd	[redacted]	[redacted]	-
	DBTd			[redacted]
+ CASSIOPEIA data cut (August 2020)	BTd			-
	DBTd			£15,815
+ MRD HR from updated meta-analyses	BTd			-
	DBTd			£15,827
+ Exponential PFS for BTd MRD positive	BTd			-
	DBTd			£17,957
Company's revised base case (Jan 2021)	BTd			-
	DBTd			£17,957

The company reported that probabilistic analysis gave a higher ICER than the deterministic analysis: £21,474 per QALY for the revised base case (Table 9, Company TE response Appendix D). The ERG got a similar result on re-running the PSA: £20,689 per QALY gained (71% probability that DBTd has an ICER below £30,000 per QALY). We also replicated the company's deterministic sensitivity analysis (DSA) and scenario analysis (Figure 5 and Table 10, Company TE response Appendix D). The DSA results are similar to those in the original company submission, showing that the model is most sensitive to the effects of treatment on survival estimated from the CASSIOPEIA landmark analysis (HRs for OS, DBTd versus BTd in MRD- and MRD+ subgroups). The most influential company scenarios are 1B (survival extrapolations for DBTd MRD- estimated using HRs from the MRD meta-analysis) and waning of treatment effects (HR=1 after 5 years in the MRD- subgroup). Selected scenario analyses conducted by the ERG are reported in Table 2 below. These include assumptions from the ERG's preferred analysis (ERG report Table 40) and selected

scenarios to explore the impact of new information in the company's TE response. The ICER is most sensitive to the assumption of no additional effect on overall survival for DBTd compared with BTd, based on the lack of significance in the updated landmark analysis (company TE response Table 5). The next most influential scenario is that of waning of additional treatment effects on PFS and OS after 5 years. The ICER is also moderately sensitive to the lower post-consolidation rates of MRD negativity with the IMWG definition (as reported in the company's TE response to Key Issue 2).

Table 2 Additional scenarios conducted by the ERG on company's revised base case (deterministic analysis with confidential PAS for daratumumab)

Individual scenarios on the base case	Treatment	Total costs (£)	Total QALYs	ICER (£/QALY)
Company revised base case (Jan 2021)	BTd DBTd	[redacted]	[redacted]	- £17,957
Mean age of population at the start of induction				
PHE dataset: [redacted] years	BTd DBTd	[redacted]	[redacted]	- £20,643
Response rates				
IMWD definition of MRD negativity (DBTd 33.7% and BTd 19.9%)	BTd DBTd	[redacted]	[redacted]	- £25,173
Extrapolation of baseline survival curves				
Weibull PFS for BTd MRD+	BTd DBTd	[redacted]	[redacted]	- £15,827
Gompertz PFS for BTd MRD+	BTd DBTd	[redacted]	[redacted]	- £16,419
Weibull OS for BTd MRD+	BTd DBTd	[redacted]	[redacted]	- £20,847
Treatment effects (DBTd versus BTd)				
Loss of effects after 5 years (OS and PFS HR=1 MRD+ & MRD-)	BTd DBTd	[redacted]	[redacted]	- £29,787
No effects on OS (HR=1 MRD+ & MRD-)	BTd DBTd	[redacted]	[redacted]	- £67,905
Resource use and costs				
Daratumumab IV formulation	BTd DBTd	[redacted]	[redacted]	- £20,562
No PBd at 3L	BTd DBTd	[redacted]	[redacted]	- £18,778

Finally, in Table 3 we present the cumulative effect of four key changes that the ERG thinks should be applied to the company's revised base case analysis:

- **Weibull distribution for PFS MRD+ subgroup**. The Weibull gives a better fit to the CASSIOPEIA KM data (company Appendix A Figure 2), . See discussion in Issue 4 above.
- **Loss of additional daratumumab treatment effects estimated from the landmark analysis after five years**. This reflects uncertainty over these additional effects, which are applied on top of effect of daratumumab in inducing and consolidating an MRD negative response after ASCT. We note the wide confidence intervals (including 1) around the HR estimates for OS from the updated landmark analysis, as well as the lack of evidence for the persistence of these additional effects after the current four years of follow-up.
- The **higher mean baseline age of patients ([redacted] years)** observed in the Public Health England dataset is reflective of UK practice. The company argues that modelling an older cohort is inconsistent with the use of outcomes for the younger trial population in the CASSIOPEIA trial (mean 56.6 years). However, this is not necessarily the case, as advice to ERG is that eligibility for ASCT is based on fitness rather than age. And the purpose of an economic model is to predict real-world cost-effectiveness, rather than cost-effectiveness in a trial context. The corollary of this point is that we have reconsidered our previous preference for assuming the cost of intravenous daratumumab as used in the trial, rather than the subcutaneous formulation, which is likely to be preferred by clinicians and patients.
- **No use of PBd at third line**. Instead, we assume the costs of Ld for all patients at third line. This reflects advice from ERG experts about current practice. We acknowledge the company's argument that when lenalidomide is used earlier in the treatment pathway, other regimens (such as cyclophosphamide or dexamethasone) may be used instead as a bridging strategy to fourth-line treatments. We note, however, that these alternatives are of similar or lower cost than Ld. The use of less expensive alternatives would have the effect of further increasing the ICER.

Table 3 Cumulative impact of ERG's preferred assumptions

(deterministic analysis with confidential PAS for daratumumab)

Individual scenarios on the base case	Treatment	Total costs (£)	Total QALYs	ICER (£/QALY)
Company base case (revised)	BTd DBTd	[redacted]	[redacted]	- £17,957
+ Weibull extrapolation for PFS MRD+	BTd DBTd	[redacted]	[redacted]	- £15,827
+ Loss of effects after 5 years (PFS and OS HR=1 MRD+ &MRD-)	BTd DBTd	[redacted]	[redacted]	- £31,128
+ Mean age [redacted] years (PHE dataset)	BTd DBTd	[redacted]	[redacted]	- £32,886
+ No PBd at 3L (100% Ld)	BTd DBTd	[redacted]	[redacted]	- £34,846
ERG updated preferred model	BTd DBTd	[redacted]	[redacted]	- £34,846