# CONFIDENTIAL UNTIL PUBLISHED

## Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

### Daratumumab in combination with bortezomib and dexamethasone for treating relapsed or refractory multiple myeloma

**Produced by** Southampton Health Technology Assessments Centre

**Authors** Neelam Kalita, Research Fellow (Health Economics), SHTAC
Karen Pickett, Research Fellow, SHTAC
Joanne Lord, Professorial Fellow, SHTAC
Geoff Frampton, Senior Research Fellow, SHTAC
G. Lily Yao, Associate Professor (Health Economics), Southampton University
Joanna Picot, Senior Research Fellow, SHTAC

**Correspondence to** Dr. J. Picot
Southampton Health Technology Assessments Centre
University of Southampton
First Floor, Epsilon House
Enterprise Road, Southampton Science Park
Southampton SO16 7NS

www.southampton.ac.uk/shtac

**Date completed** 03 May 2018
Updated following Company factual error check response 23 May 2018

**Source of funding**: This report was commissioned by the NIHR HTA Programme as project number 16/56/05.

**Declared competing interests of the authors**

None from the authors and none from the clinicians who provided advice.

**Acknowledgements**

We are very grateful to Dr J. Bird, Consultant Haematologist, Bristol Haematology Unit, Bristol Haematology and Oncology Centre and Dr. C. Kallmeyer, Consultant Haematologist, United Lincolnshire Hospitals NHS Trust who offered clinical advice and comments on the draft report. We would also like to thank: Karen Welch, Information Scientist, SHTAC, for appraising the literature search strategies in the company's submission, and for running searches where necessary; and Dr Emma Loveman Senior Reviewer / Partner, Effective Evidence LLP, for providing a quality assurance review of the draft ERG report.



**Rider on responsibility for report**

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

**This report should be referenced as follows:**

Kalita, N; Pickett, K; Lord, J; Frampton, G; Yao, GL; Picot, J. Daratumumab in combination with bortezomib and dexamethasone for treating relapsed or refractory multiple myeloma: A Single Technology Appraisal. SHTAC 2018.

**Contributions of authors**

Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Karen Pickett critically appraised the clinical

effectiveness systematic review and drafted the report; Joanne Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Geoff Frampton critically appraised the clinical effectiveness systematic review and drafted the report; G. Lily Yao critically appraised the economic evaluation, and contributed to drafting the report; Joanna Picot critically appraised the clinical effectiveness systematic review, drafted the report, project managed the review and is the project guarantor.

**Word count: 42,311**

TABLE OF CONTENTS



LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse event |
| AIC | Akaike information criterion |
| ASCO | American Society of Clinical Oncology |
| ASCT | Autologous stem cell transplant |
| ASH | American Society of Hematology |
| Bd | Bortezomib and dexamethasone |
| BMI | Body mass index |
| CAA | Confidential commercial access agreement |
| CADTH | Canadian Agency for Drugs and Technologies in Health |
| Cd | Carfilzomib in combination with dexamethasone |
| CEAC | Cost-Effectiveness acceptability curve |
| CI | Confidence interval |
| CIC | Commercial in confidence |
| CR | Complete response |
| CRD | Centre for Reviews and Dissemination |
| CrI | Credible interval |
| CS | Company submission |
| CSR | Clinical study report |
| DBd | Daratumumab in combination with bortezomib and dexamethasone |
| DOR | Duration of response |
| DSU | Decision support unit |
| ECOG | Eastern Cooperative Oncology Group |
| EFS | Event-free survival |
| EHA | European Hematology Association |
| EORTC-QLQ-C30 | European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire |
| EQ-5D-5L | EuroQol Five Dimensions Questionnaire |
| ERG | Evidence Review Group |
| EU | European Union |
| HMRN | Haematological Malignancy Research Network |
| HR | Hazard ratio |
| HRQoL | Health-related quality of life |

| | |
|---|---|
| HTA | Health technology assessment |
| ICER | Incremental cost effectiveness ratio |
| IMWG | International Myeloma Working Group |
| IPCW | Inverse probability of censoring weights |
| IPE | Iterative Parameter Estimation |
| IRR | Infusion-related reactions |
| ISPOR | International Society for Pharmacoeconomics and Outcomes Research |
| ISS | International staging system |
| ITT | Intent-to-treat |
| IV | Intravenous |
| IWRS | Interactive web response system |
| KM | Kaplan-Meier |
| Ld | Lenalidomide and dexamethasone |
| MIMS | Monthly index of medical specialities |
| MM | Multiple myeloma |
| MOD | Mitozantrone, vincristine, dexamethasone |
| MRD | Minimal residual disease |
| NA | Not applicable |
| NE | Not evaluable |
| NHS | National Health Service |
| NHSE | National Health Service England |
| NICE | National Institute for Health and Clinical Excellence |
| NIHR | National Institute for Health Research |
| NMA | Network meta-analysis |
| NR | Not reported |
| OR | Odds ratio |
| ORR | Overall response rate |
| OS | Overall survival |
| PAS | Patient Access Scheme |
| PBAC | Pharmaceutical Benefits Advisory Committee |
| Pd | Pomalidomide and dexamethasone |
| PD | Progressed disease |
| PFS | Progression-free survival |

| | |
|---|---|
| PI | Proteasome inhibitor |
| PPS | Post-progression survival |
| PR | Partial response |
| PSA | Probabilistic sensitivity analysis |
| PSS | Personal social services |
| QALY | Quality-adjusted life year |
| RCT | Randomised controlled trial |
| RPSFTM | Rank Preserving Structure Failure Time Models |
| RRMM | Relapsed/refractory multiple myeloma |
| SC | Subcutaneous |
| sCR | Stringent complete response |
| SD | Standard deviation |
| SE | Standard error |
| SHTAC | Southampton Health Technology Assessments Centre |
| SMC | Scottish Medicines Agency |
| TDE | Time dependent surrogate outcomes |
| TEAE | Treatment emergent adverse event |
| TFI | Treatment-free interval |
| TTD | Time to treatment discontinuation |
| TTNT | Time to next therapy |
| TTP | Time to disease progression |
| TTR | Time to response |
| UK | United Kingdom |
| US | United States |
| VAd | Vincristine, doxorubicin and dexamethasone |
| VAS | Visual analogue scale |
| VGPR | Very good partial response |

## SUMMARY

### Scope of the company submission

The company's submission (CS) includes a narrower patient group than that specified in the NICE scope. This is because the company's decision problem population is adults with relapsed or refractory multiple myeloma (RRMM) who have received one previous treatment (i.e. second-line patients) whereas the NICE scope is adults with RRMM with no limitation by the number of lines of previous treatment. The company's rationale for focussing on second-line patients is that daratumumab in combination with bortezomib and dexamethasone (DBd) "*offers the greatest benefit to patients on second-line treatment*". Two of the three comparators listed in the NICE scope for second-line treatment [bortezomib-based therapy and carfilzomib in combination with dexamethasone (Cd)] are only used in current clinical practice for those who are bortezomib naive but the company have not distinguished in their decision problem between those second-line patients who received bortezomib as a first-line therapy and those who are bortezomib naive. The company have confirmed in response to clarification question B9 that DBd is being positioned as a second-line treatment option regardless of prior bortezomib exposure status.

### Summary of submitted clinical effectiveness evidence

The company's systematic review of clinical effectiveness identified one relevant randomised controlled trial (RCT) of DBd.

- The CASTOR trial (phase III, open label, multicentre, superiority trial) compared DBd versus bortezomib and dexamethasone (Bd)

The ERG believes the company has identified all the relevant RCTs of DBd.

There are no head-to-head RCTs of DBd versus bortezomib alone, Cd, or combination chemotherapy which are the comparators defined in the company's decision problem for second-line treatment. Therefore the company conducted a network meta-analysis (NMA) to perform an indirect treatment comparison. The company's systematic review identified a further two RCTs that were considered for inclusion in the NMA. One, the ENDEAVOR trial, compared Cd versus Bd and the other, an RCT conducted by Phillips and colleagues[1] compared the chemotherapy regimens of vincristine, doxorubicin, and dexamethasone (VAd) versus mitozantrone, vincristine and dexamethasone (MOD). The company states the combination chemotherapy regimens examined by Phillips and colleagues[1] are not currently used for the

treatment of RRMM (the study having been carried out between 1986 and 1992) and in the absence of a common comparator, it was not feasible to connect this study to the other two trials (CASTOR and ENDEAVOR) in the NMA. Consequently an NMA was carried out to enable a comparison of DBd versus Cd using the CASTOR and ENDEAVOR trials.

To be enrolled in the CASTOR RCT, patients had to have received at least one prior line of therapy (there was no upper limit) and in the ENDEAVOR RCT enrolled patients had received one to three previous treatments. Consequently only a proportion of the participants in both trials are relevant to the company's decision problem which has focussed on second-line patients. Second-line patients account for 47.2% of those enrolled in the CASTOR trial and 50% of those in the ENDEAVOR trial. In both the CASTOR and ENDEAVOR RCTs randomisation was stratified by the number of previous lines of treatment and subgroup data for second-line participants from both RCTs were available for progression-free survival (PFS) and overall survival (OS) which are two of the five clinical effectiveness outcomes that contribute data to the economic model.

The CS reports the effects of DBd treatment across a range of outcomes relevant to the NICE scope and the company decision problem, which are summarised below.

PFS is the primary outcome of the CASTOR RCT. For the subgroup of second-line patients median PFS was approximately 18 months longer in the DBd arm than in the Bd arm and this is a statistically significant improvement (HR = 0.23, 95% CI 0.16 to 0.33, $p<0.0001$). In the whole trial population the improvement in PFS was not as great (9.7 months) in favour of DBd (HR 0.32, 95% CI 0.25, 0.40, $p<0.0001$).

Results from the NMA to indirectly compare DBd versus Cd gave PFS hazard ratios in favour of DBd for both the subgroup of second-line patients and the whole trial population, and in both cases the probability that DBd was the best treatment was estimated at 100%. The hazard ratio for Cd from the NMA is used in the economic model. The company also conducted NMA sensitivity analyses of PFS efficacy of DBd compared with Bd and compared with Cd among bortezomib naïve second-line patients. The resulting hazard ratios favoured DBd, with probabilities of this being the best treatment of 100% versus Bd and 97.2% versus Cd.

OS is a secondary outcome for the CASTOR trial and the OS data are immature so a median OS is not available for either study arm for the second-line patient subgroup. However, the proportion of deaths in the DBd arm is lower than in the Bd arm (20.5% versus 35.4% respectively), and a statistically significant difference in favour of DBd is reported (HR 0.50, 95% CI 0.30 to 0.84, p=0.008). In the whole trial population the improvement in OS is not statistically significant (HR 0.77, 95% CI 0.57 to 1.04, p=0.0884). At the 26.9 month median follow-up reported in the CS there had been 179 deaths in total in the whole trial population, 32.7% in the DBd arm versus 39.3% in the Bd arm. Final OS analysis will occur after 320 deaths.

The NMA for OS in the subgroup of second-line patients returned a hazard ratio for DBd vs Cd of 0.60 [95% credible interval (CrI) 0.33 to 1.10] and a probability of DBd being better than Cd of 95%. For the whole trial population, in the OS NMA, the probability of DBd being better than Cd was just 55.8% and the hazard ratio was 0.97 (95% CrI 0.68 to 1.39). The hazard ratio for Cd from the NMA is used in the economic model. The company conducted NMA sensitivity analyses of OS for DBd versus Bd and versus Cd among bortezomib naïve second-line patients. The resulting hazard ratios favoured DBd, with probabilities of this being the best treatment of 94.2% versus Bd and 76.8% versus Cd.

Although CASTOR was a multicentre, multinational trial, there were no UK trial centres. The CS stresses that during the trial some participants received treatments after disease progression (in both the DBd and Bd arms) that are not available in England. Furthermore the proportion of participants in receipt of treatment after disease progression differed between the trial arms. Consequently the company made an adjustment [using the inverse probability of censored weights (IPCW) methodology] to the overall survival outcome from the CASTOR trial which aimed to reduce the bias introduced by a higher proportion of participants in the Bd arm than in the DBd arm receiving subsequent treatments not available in England.

The effect of adjusting OS to account for subsequent treatments not available in England was a [redacted] in the HR for OS (i.e. a [redacted] in the risk of death for those in the DBd group). In the subgroup of second line-patients the adjusted HR for OS is [redacted] (95% CI [redacted]). The DBd and Bd OS estimates in the subgroup of second-line patients adjusted for the use of subsequent therapies not available in England were used in the base case of the economic model.

The other outcomes from the CASTOR RCT that contribute data to the economic model are time to treatment discontinuation (TTD), adverse events and health-related quality of life (specifically the EQ-5D-5L).

TTD data come from a post-hoc analysis conducted to inform the economic model. DBd was associated with a 56% reduction the risk of treatment discontinuation at 26.9 month of follow-up compared with Bd.

TEAEs are summarised for the total CASTOR trial population and data on eight adverse events (neutropenia, anaemia, thrombocytopenia, lymphopenia, pneumonia, fatigue, peripheral neuropathy and hypertension) which occurred in at least 5% of the patients in either arm of the trial as a grade 3 event or higher were included in the economic model. No new safety signals were identified by the company.

HRQoL outcomes [European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLC-C30) and EQ-5D-5L] were not reported separately for the subgroup of second-line patients. In the total trial population there was no significant different between the two groups at any time point (from baseline to week 24).

Among the remaining secondary outcomes that did not contribute data to the economic model the results were reported separately for the subgroup of second-line patients only for MRD, response and TTP. The results were in favour of the DBd group.

In addition to the subgroup of second-line patients who meet the company's decision problem, subgroup analyses for PFS, TTP, ORR and rate of VGPR or better were conducted on the whole CASTOR trial population. Among the subgroups of bortezomib-naive and bortezomib-experienced patients outcomes favoured the DBd treated group with results being numerically better in the bortezomib naive subgroup than the bortezomib experienced subgroup for all four outcomes. In contrast, as noted above, the NMA sensitivity analysis results were slightly less favourable in the 2LBN group than in the 2L group as a whole. A consistent effect in favour of DBd was observed across the other subgroups tested.

## Summary of submitted cost effectiveness evidence

The company's submission includes a review of published cost-effectiveness evidence and a new economic model developed for this appraisal. The model compares the cost-effectiveness of DBd with Bd or Cd in patients with RRMM and one prior treatment.

*Review of published economic evidence*

The company conducted a systematic search for published cost-effectiveness evidence, but did not identify any studies relevant to the decision problem.

The ERG updated the company's search and identified two recent papers that report cost-effectiveness studies comparing DBd with Bd in patients with RRMM: Carlson et al. (2018) and Maise et al. (2018). Both studies were conducted from a US perspective, so the costs and cost-effectiveness results are not relevant for the UK NHS. However, the methods, data sources and estimated health outcomes from these studies provide a useful crosscheck for the face validity of the company's model.

The company compared key features of their economic analysis with methods and results from seven previous NICE technology appraisals of treatments for RRMM. We note that the recent NICE appraisal of carfilzomib for previously treated multiple myeloma (TA457) is particularly relevant to the current decision problem, as it compared Cd with Bd for patients with one prior therapy not including bortezomib. We draw comparisons with the TA457 committee preferred assumptions and 'most plausible' Incremental Cost-Effectiveness Ratio (ICER) estimate for Cd versus Bd in our critique of the company model.

*The company's economic analysis deviates from the NICE scope for this appraisal, but otherwise it follows NICE reference case methods*

The company's economic evaluation does not fully address the NICE scope: the analysis is restricted to the second-line population and it does not include combination chemotherapy as a comparator. Otherwise, the methods of economic evaluation are consistent with the NICE reference case. The company states that they adopted an NHS perspective for costing, rather than the NHS and local authority funded personal social services (PSS) perspective specified for the reference case. However, no significant PSS costs are omitted from the company's model.

*The structure and assumptions of the submitted model are reasonable*

The company's submitted model has three main health states, pre-progression, post-progression and death, with the pre- and post-progression states subdivided into 'on' and 'off' treatment stages. Patients enter the model in the pre-progression state at the start of second-line treatment. This stops on disease progression, or earlier if, for example, the patient experiences an adverse event. After disease progression, a proportion of patients start other treatments. Subsequent relapses and lines of treatment are not modelled explicitly, but costs are added for a defined bundle of other treatments and effects on survival are reflected in the OS curves. This limits the ability of the model to assess the cost-effectiveness of treatment sequencing, but is a reasonable simplification.

The company uses a partitioned survival approach to estimate the rates at which patients move between the modelled health and treatment states. This relies on a set of PFS, OS and TTD curves for each treatment, estimated from clinical trial data. Implementation of the partitioned survival approach entails some assumptions and logical constraints (e.g. that PFS cannot exceed OS). The model sets the number of deaths that occur pre-progression as a fixed proportion (14.6%) of the number progressions. The rationale for this assumption is not clear, but it does not cause an unrealistic imbalance in mortality pre/post progression. The model also includes a check that mortality rates in the modelled cohort cannot be lower than would be expected in people of the same age and gender-mix in the general population.

Additional model assumptions include:

- The initial age of the cohort is 63.3 years and 58.7% are male.
- A time horizon of 30 years (effectively lifetime given the initial age)
- The model cycle length is 1 week, with a half-cycle correction
- Treatment effects are extrapolated over the time horizon using survival curves fitted to trial data, with no additional 'waning' of effects over time

The ERG consider that the 3-state model structure is reasonable and appropriate for the decision problem. The partitioned survival approach does entail assumptions, but we are not concerned that these will have biased results.

*Parametric survival curves were fitted to CASTOR data for patients with one prior treatment, but did not exclude patients with prior exposure to bortezomib*

To extrapolate beyond the trial period, the company employed parametric survival analysis. Six candidate functions (exponential, Weibull, log-normal, log-logistic, Gompertz and generalised gamma) were fitted to individual-level data from the CASTOR trial to estimate PFS, OS and TTD survival curves for DBd and Bd. The company concluded that the assumption of proportional hazards does not hold for these outcomes in CASTOR, and so fitted separate curves to the trial arms. The ERG considers that the proportional hazards assumption cannot be assumed to hold for PFS, but that evidence is more equivocal for OS.

The survival curves were fitted using data for the second-line (2L) subgroup of patients with one prior therapy. Curves were not estimated for the second-line bortezomib naïve (2LBN) subgroup for whom Bd and Cd are currently available in England. Treatment effects tended to be more favourable in the 2L than 2LBN subgroup. Although these differences were small and not statistically significant, the use of data for the 2L group will tend to bias ICER estimates in favour of DBd.

*Adjustment of OS for use of subsequent treatments not available in England is appropriate for the economic evaluation, but is likely to be subject to confounding*

For OS, the fitted curves were adjusted for subsequent treatments used in the trial but not available in England. The adjustment was made using the IPCW method. The ERG considers that this method is appropriate in principle, but that it may give biased results if important covariates are omitted. The company's backward stepwise approach to covariate selection led to a very parsimonious reduced model, containing only two covariates (baseline ISS staging and race). We consider that it is highly likely that there are unmeasured confounders. However, cost-effectiveness results are not sensitive to the inclusion or omission of the subsequent treatment adjustment.

*The process of curve selection was not well justified and the choice of OS curves exaggerates the projected survival advantage for DBd based on immature data*

The company states that they based their choice of parametric functions for OS, PFS and TTD on two factors: statistical fit and clinical plausibility of the projections, informed by expert input from an advisory board. The company argue that long-term plausibility should be given more weight than statistical goodness of fit, because of the short follow-up time in CASTOR. They

compare the parametric curve projections against long-term external data on survival for Bd. Their rationale for the choice of a different, and much more optimistic, functional form for DBd survival was to cite opinion on 'the transformational nature' of the treatment. The company notes the higher rate of MRD negativity with DBd than with Bd in CASTOR and argue that this is associated with prolonged OS.

The ERG considers that the company's selection of OS curves (Gompertz for Bd and log-logistic for DBd) gives an overly optimistic prediction of the survival gain from DBd that is not warranted given the immature trial data. We suggest more conservative assumptions, with the same functional form for DBd and comparators, selected for plausibility of 5 and 10-year outcomes. In the recent NICE TA of carfilzomib (TA457), the committee concluded that the Weibull distribution was supported for Bd and Cd by OS data from ENDEAVOR and external validation for the second-line comparison in patients without prior bortezomib. We therefore consider that the Weibull distribution for OS provides a consistent foundation for modelling Bd in this current appraisal, and that there is no reason to prefer a different function for DBd.

*The ERG agrees with the company's choice of survival curves for PFS and TTD*

The company assessed log hazard plots of the CASTOR data and concluded that the proportional hazards assumption does not hold. They therefore fitted independent PFS curves for DBd and Bd. Of the six fitted curves, the company choose the Gompertz distribution for DBd and Bd in their base case, arguing that it provides a balance of fit to the trial data with reasonable long-term projections. The ERG agrees with these conclusions.

A similar approach was taken to select TTD curves for DBd and Bd. The company fitted independent parametric functions to the CASTOR data, as the proportional hazards assumption does not hold. They noted that in this case the statistical fit of the parametric functions was similar, so chose to use the same function as for PFS (Gompertz) for consistency. This recognises the likely correlation between PFS and TTD. We agree with this approach.

*OS and PFS curves for Cd are estimated from ENDEAVOR results, but are not adjusted for the duration of bortezomib treatment, which leads to underestimation of the effectiveness of Cd*

For Cd, the PFS and OS curves were estimated by applying hazard ratios to the fitted curves for Bd. This entails a proportional hazards assumption for the comparison of Cd with Bd. The

company does not discuss the evidence for or against this assumption, but we consider it appropriate based on committee conclusions in the carfilzomib appraisal (TA457).

There is a the discrepancy between the length of treatment approved in the marketing authorisation for bortezomib (24 weeks), as used in CASTOR, and the ongoing use of bortezomib until progression in ENDEAVOR. The effect of this will be to underestimate the the relative effect of Cd compared with Bd and hence to overestimate the effect of DBd compared with Cd. This is acknowledged in the CS as a source of bias, but no attempt is made to address it in the quantitative modelling. This issue was considered in detail in the carfilzomib appraisal (TA457), and analyses of ENDEAVOR data provided estimates of the relative effect on PFS and OS of shortening the duration of bortezomib treatment. The ERG makes use of these adjustments in additional analysis.

The TTD curve for Cd was modelled using a proportional hazard of 0.477 compared with Cd PFS, based on ENDEAVOR. We agree with these assumptions.

*Health state utilities are estimated from CASTOR, but are uncertain due to poor reporting*

Health state utilities were estimated from the CASTOR trial, but the methods of analysis and results are poorly reported in the CS. In particular, it is not stated whether the EQ-5D analysis was restricted or adjusted for 2L patients. In the factual accuracy check, the company stated that utilities in the model were based on 2L patients only. The ERG is unable to verify this as we have not seen any results or description of methods used for this subgroup analysis.

The company mapped the EQ-5D-5L scores to EQ-5D-3L values, using a 'crosswalk' method reported by van Hout and colleagues. Pre-progression utility, estimated as the average of all measured utilities before the date of progression, using repeated measures mixed-effects modelling, was assumed to be same for all patients in DBd and Bd arms. ERG agrees with this approach.

The CS reports that utility in the progressed state was estimated by using the last recorded EQ-5D data collected prior to progression. However, in their response to the factual accuracy check, the company state that this was incorrect and that post-progression utility was actually estimated using a linear mixed model from any utility measured after the date of progression. The ERG cannot assess the quality or accuracy of this analysis as we have not seen details of

the methods or results. We are also concerned that the company uses similar utility estimates for pre- and post progression states which seems unrealistic.

The ERG addresses these issues in the additional analyses.

*In general, company's approach to costing is appropriate and consistent with related NICE guidance, albeit with a few errors in estimation*

Costs and resources associated with drug acquisition, drug administration and co-medication, subsequent treatment, follow-up monitoring and care, adverse events and terminal care were included in the company's cost-effectiveness analyses. These were in line with previous TAs (including TA457). Overall, the costs inputs and sources used were appropriate although the ERG identified two errors in the model. First, the cost of dexamethasone was estimated incorrectly which the company corrected as part of their clarification response. Second, costs associated with administration and co-medication for bortezomib were included in the model after the end of treatment cycles (i.e. 24 weeks). Although these costs were low their inclusion overestimated the total costs of Bd slightly. This is corrected in the ERG's additional analyses.

*Company's base case results*

In the original company's base case, DBd was more expensive and more effective compared with Bd as well as Cd. The ICER for DBb vs Bd was £41,633 and £7,180 for DBd vs Cd, respectively (Table 40).

**Table 1 Cost effectiveness: company base case (list prices)**

| | Total costs (£) | Total LYG | Total QALYs | Pairwise (DBd vs comparator) | | | Full Incremental ICER |
|---|---|---|---|---|---|---|---|
| | | | | Incremental cost (£) | QALYs gained | ICER (£ per QALY gained) | |
| Bd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £41,633 | - |
| Cd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £7,180 | Ext. dom. |
| DBd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £41,633 |

Ext. dom, extended dominance

*A range of uncertainty analyses were conducted by the company, but they have been selective in the scenarios they present*

The company performed a range of deterministic- , probabilitistic- and scenario analyses to assess the methodological as well as parameter uncertainty of their base case analyses. The ERG agrees with their assumptions for DSA and PSA and their results, in general. However, we identified an error in the scenario relating to longer subsequent treatment duration which the company corrected in their response to clarification question. Further, we view that the company did not explore the full range of survival functions or the impact of changing OS or PFS functions for more than one treatment at a time. To address this issue, we conducted a wider range of scenarios in ERG additional analyses.

## Commentary on the robustness of submitted evidence

### Strengths

- The ERG believes that the company have identified all the key studies of DBd and potential comparators for the population of second-line RRMM patients specified in their decision problem. One RCT provides evidence for the effectiveness of DBd versus Bd for people with RRMM and one RCT provides evidence for the effectiveness of Cd versus Bd allowing an indirect comparison of DBd with Cd for selected outcomes (including PFS and OS).
- The model structure is consistent and follows the conventional design for cancer appraisals.
- The perspective of the analysis aligns with the NICE guide to the methods of Technology Appraisal
- The model uses a time horizon of 30 years which is a fair approximation of the life time horizon, given the median age of 63 years for the CASTOR population.
- PFS and TTD are modelled appropriately
- The model uses appropriate sources for costs and resource use and in line with other technology appraisals
- The model cycle length adequately captures differences between dosing schedules regularly used in RRMM.
- Discounting and half cycle correction are applied correctly and aligns with NICE guide to the methods of Technology Appraisal

### Weaknesses and Areas of uncertainty

- The company does not present evidence specific to patients who have already received two lines of therapy or more.
- The company does not provide any comparison of DBd with combination chemotherapy
- OS data used in the analysis is immature. The company's choice of log-logistic for DBd and Gompertz for the Bd arm to extrapolate OS curves, overestimates the cost-effectiveness of DBd vs Bd and Cd. The company did not provide sufficient justification for the selection of curves.
- The company did not provide any evidence for the subgroup of 2LBN patients, aside from the NMA sensitivity analyses for PFS and OS. The company's cost-effectiveness analyses are for 2L patients without distinguishing between 2L and 2LBN patients. This does not align with clinical practice in England which differs for 2LBN patients and 2L patients who have previously received bortezomib.
- The company does not adjust for the difference in treatment duration for bortezomib in the ENDEAVOR and CASTOR trials. Whilst bortezomib is administered until disease progression in ENDEAVOR, in CASTOR the drug is administered for 24 weeks. In our view it is appropriate to make an adjustment to account for this, which is in line with the NICE committee's conclusion in TA457.
- The ERG has concerns relating to the company's reporting of methods used to derive utilities for the model:
    - The CS does not state whether the EQ-5D analysis in CASTOR was restricted or adjusted for 2L patients. In the factual accuracy check, the company state that it was, but the ERG have not seen any results or description of methods for this subgroup analysis.
    - The CS states that estimated utility for progressed disease from the CASTOR analysis was based on the last recorded EQ-5D data collected prior to progression. It is not plausible that this utility will reflect average utility across the post-progression period. In their response to the factual accuracy check, the company state that in fact, post-progression utility was estimated using a linear mixed model from any utility measured after the date of progression. The ERG cannot assess the quality or accuracy of this analysis as we have not seen details of the methods or results.
    - The company's estimates for pre-progression and post –progression utilities are similar, which we view as unrealistic.
- The ERG agrees with the company's overall approach to estimating resource use and costs. However, the company incorrectly included administration- and co-medication costs for

bortezomib after the end of treatment cycles (i.e. 24 weeks) thereby overestimating the total costs of bortezomib slightly.

### Summary of additional work undertaken by the ERG

The ERG conducted a number of scenario analyses. Our preferred assumptions, alongside the scenarios are presented in Table 2.

**Table 2 ERG's preferred assumptions and scenarios**

| Aspect of the model | ERG Preferred assumptions | ERG scenarios |
|---|---|---|
| OS curves | DBd: Weibull<br>Bd: Weibull | We test the effect of assuming proportional hazards to model DBd in relation to the independently fitted Weibull curve for Bd<br><br>KM data with parametric tail |
| Patient group | Second line Bortezomib-naïve RRMM (2LBN) | We test the effect removing the 2LBN adjustment |
| PFS curves | Persistence of OS and PFS benefits | Test effect of modelling DBd in relation to Bd curve.<br><br>KM data parametric tail. |
| Treatment effect | Persistence of OS and PFS benefits | Assume loss of survival benefit for Cd and DBd, relative to Bd, from 5 to 20 years |
| | HR for Cd vs. Bd adjusted for 24-week duration of treatment with bortezomib, as in TA457 | Analysis is conducted without the adjustment for Bd treatment duration |
| Utilities | Utilities as in TA457: ENDEAVOR data mapped to EQ-5D with Proskorovsky algorithm | Other sources of utilities: CASTOR (company base case); Carlson et al.[2]; and van Agthoven[3]. |
| Subsequent treatment | DBd and Bd OS adjusted for treatments not available in England (IPCW method) Cd not adjusted | Use the unadjusted survival model |
| Mortality | Deaths from the PFS state estimated using fixed ratio of deaths to progression (0.146) | Vary ratio of deaths to progression |

The results of the ERG's preferred assumptions are presented in Table 3. Our preferred assumptions increase the ICER for DBd vs Bd significantly. When DBd is compared with Cd, Cd is dominated as it is more expensive and less effective. The following additional scenarios (Table 4) were performed on the ERG's preferred base case model.

**Table 3 Cost-effectiveness: ERG preferred base case (list prices)**

| | Total costs (£) | Total LYG | Total QALYs | Pairwise (DBd vs comparator) | | | Full Incremental ICER |
|---|---|---|---|---|---|---|---|
| | | | | Incremental cost (£) | QALYs gained | ICER (£ per QALY gained) | |
| Bd | [redacted] | | | | | £93,061 | - |
| Cd | | | | | | DBd dominates | - |
| DBd | | | | - | - | - | £93,061 |

**Table 4 Additional ERG scenarios**

| Aspect of the model | ERG scenarios |
|---|---|
| **Patient population** | All second line (2L): includes patients with prior bortezomib |
| **Treatment effects** | OS and PFS extrapolation:<br>• For PFS, KM data till 12 months, then Gompertz<br>• For OS, KM data till 27 months, then Weibull |
| | No adjustment of OS for subsequent treatment |
| **Persistence of effects** | Waning for OS: HR=1 for DBd vs. Bd and Cd vs. Bd after:<br>• 5 years<br>• 10 years<br>• 20 years |
| **Time horizon** | No waning with model time horizon of:<br>• 5 years<br>• 10 years<br>• 20 years |
| **Utilities** | Source of health state utilities (PF and PD):<br>• CASTOR trial (company base case)<br>• van Agthoven 2004<br>• Carlson et al. 2018 |
| **Mortality** | Vary ratio of pre-progression deaths to progression:<br>• 5%,<br>• 10%<br>• 20% |
| **Resource use and costs** | Longer subsequent treatment duration:<br>• 15 months |
| | Subsequent treatment mix based on expert feedback<br>• 100% for Lenalidomide and dexamethasone (Ld) |
| | Allow vial sharing (no wastage) |
| | No dose intensity considered (costs for 100% of dose) |

# 1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to NICE from Janssen on the clinical effectiveness and cost effectiveness of daratumumab in combination with bortezomib and dexamethasone (DBd) for relapsed or refractory multiple myeloma. It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the company by the ERG and NICE on 19 March 2018. A response from the company via NICE was received by the ERG on 05 April 2018 and this can be seen in the NICE committee papers for this appraisal.

# 2 BACKGROUND

## 2.1 Critique of company's description of underlying health problem

### Aetiology

The CS provides a clear and accurate, albeit brief, overview of MM and its aetiology (CS section B.1.3.1). MM starts with the proliferation of abnormal blood plasma cells within the bone marrow and the overproduction of abnormal immunoglobulin fragments (referred to as myeloma proteins or M proteins). Accumulation of these within the bone, blood and multiple organs leads to serious complications including hypercalcaemia, renal impairment, anaemia, bone disease and, less frequently, increased blood viscosity, infections, thrombosis and extramedullary disease (tumours which form outside of the bone marrow).

MM is a heterogeneous disease in terms of its clinical course and prognosis, with considerable variability among patients. Clinical outcomes, including OS, are influenced by several prognostic factors, including the cancer's stage and cytogenetic profile, and the number of prior treatments. A key feature of MM is that patients have multiple relapses, each of which is associated with a reduction in the degree and duration of response to treatment. All surviving patients therefore eventually relapse from, or become refractory to, existing treatments. As a result, prognosis worsens with each successive relapse and each line of treatment (as depicted in CS Figure 1).

As explained in the CS, relapsed or refractory MM (RRMM) is defined by the IMWG[4] as a disease that is non-responsive while on salvage therapy or progresses within 60 days of last

treatment in patients who have achieved a minimum response or better at some previous point before then progressing (CS section B.1.3.1.1). The company state that epidemiology of RRMM is assumed to be similar to that of MM (CS section B.1.3.1.2), and clinical experts advising the ERG agreed.

**Prevalence and incidence**

MM is a rare cancer, accounting for approximately 1% of all cancers and 15-20% of all blood cancers worldwide. In England, 4,632 people were diagnosed with MM in 2015, representing 2% of all new cancer cases. Incidence rates have increased in the UK by 17% in the last decade and rates are expected to continue to increase during the next two decades.[5] The CS states that the increase largely reflects the changing prevalence of (unspecified) risk factors and improvements in diagnosis (CS section B.1.3.1.2). According to Cancer Research UK, risk factors for myeloma are not well understood, but the prevalence of myeloma is higher in older people, males and black people.[5]

**Survival and mortality**

The CS reports OS rates for MM but not specifically for RRMM (CS section B.1.3.1). The company state that global survival has more than doubled, as a result of improvements in MM treatment, increasing from approximately 3 years of survival before 1998 to ≥6 years after 2006 (references cited[6-8]). The ERG notes that the data cited by the company are from a small number of specific trials rather than a broad range of international studies, some of which were on specific populations (e.g. with smouldering myeloma[6]). However, according to Cancer Research UK, myeloma survival in the UK has quadrupled in the past 40 years.[5] The CS states that a recent review of anti-myeloma therapies[8] reported a median survival of 6.1 years and 8-year overall survival (OS) of 57%. The ERG notes that these long-term data are taken from two specific trials included in the review, and that the review[8] focused on trials in which patients had received a novel MM treatment following autologous stem cell transplant (ASCT). According to the latest data available from Cancer Research UK (reference 27 cited by the CS),[5] the 5-year and 10-year survival rates for adults with MM in England and Wales in 2010-2011 were approximately 47% and 33%. The latest mortality data from Cancer Research UK show that there were 2,928 deaths from MM in the UK in 2014.

**Impact of MM on patients and carers**

The CS summarises the effect of MM on patients (CS section B.1.3.1.3), citing evidence from a systematic review (abstract)[9] which found that symptoms and complications of MM reported by patients, as a result of the disease and its treatment, are worse than those for other blood cancers. Among the problems experienced by MM patients are weakness, fatigue, bone pain, weight loss, confusion, excessive thirst, and constipation. Clinical experts advising the ERG also highlighted infections, peripheral neuropathy, bruising/bleeding, diarrhoea, indigestion, low mood and insomnia as problems that patients may experience as a result of MM and/or its treatment.

The prognosis and symptom burden of patients worsens with each relapse, due to the progressive nature of the disease and the cumulative adverse effects of treatment (CS Figure 2).[10] The CS cites a survey showing that patients in their first treatment-free interval (TFI) relative to other phases and those experiencing a longer TFI had significantly better HRQoL scores.[11] The CS, and other authors,[10] argue that prolonging earlier remissions is therefore key to improving patients' HRQoL.

A targeted search for information on the experiences and preferences of patients with MM found that patients' main treatment preferences are to achieve: lasting remission; maximum life expectancy; improved emotional quality of life; further treatment options; and TFIs (CS Figure 3).

The CS briefly summarises the effect of MM on patients' caregivers (CS section B.1.3.1.3), noting that the emotional impact of caring for someone with MM may affect the carer's ability to work. The CS cites a study which found that 49% of MM patients' partners reported anxiety and 14% reported depression, and that carers reported having an unmet need for help with managing the side effects and complications of MM therapy.[12] The ERG concurs that patients' carers may suffer negative impacts of MM and its treatment, but there have been few studies on this.

## 2.2 Critique of company's overview of current service provision

Current service provision (CS section B.1.3.2) is mentioned only briefly by the company. The CS emphasises that the available treatments for RRMM in England are limited compared to

those available in Scotland, Wales, other European countries and the United States (CS Figure 4). Currently, the only NICE-recommended treatments for second-line patients with RRMM are bortezomib monotherapy (NICE Technology Appraisal Guidance 129) and carfilzomib in combination with dexamethasone (Cd) (NICE Technology Appraisal Guidance 457) but both of these treatments are only available for bortezomib-naïve patients. In the case of bortezomib monotherapy this is because of an NHS England (NHSE) restriction to the bortezomib naive group and in the case of Cd due to the recommendation made by NICE. In particular, no second-line (2L) triple therapies are recommended in England, and there are no therapies currently available upon relapse for bortezomib-experienced patients. Clinical experts advising the ERG agreed in general with the company's depiction of treatment availability and the clinical pathway in England in CS Figure 4 and CS Figure 5 respectively, but noted that, in practice, other therapies that are not currently recommended by NICE may be given to patients. For example, one expert stated it is widespread practice to add cyclophosphamide to bortezomib and dexamethasone (Bd) for second-line patients who are bortezomib-naïve (2LBN), and patients who have relapsed after first-line bortezomib-based therapy may receive cyclophosphamide, thalidomide and dexamethasone (CTD). The ERG and clinical experts consulted agree that in clinical practice Bd is used in preference to bortezomib monotherapy as a 2L therapy (as the company indicates in footnote 2 for CS Figure 5).

The company state (and clinical experts advising the ERG agreed) that most of the clinical management of MM takes place in the outpatient setting and that the bulk of care is informal and provided by caregivers. The CS section on current service provision does not refer to any other aspects of NHS infrastructure, staffing or costs (CS section B.1.3.2). However, elsewhere the CS states that administration of IV treatments requires an outpatient visit that may include additional nursing and pharmacist preparation time; and that administration of subcutaneous bortezomib requires an outpatient visit with a specialist cancer nurse (CS section B.3.5.3). The CS does not comment on whether administration of DBd in the NHS would require additional staff training or resource allocation compared to the administration of other RRMM therapies. The ERG assumes that NHS service provision for treating 2L RRMM patients with DBd would follow the general service organisation for MM as specified in the current NICE Guideline on myeloma diagnosis and management [NG35].[13] Clinical experts advising the ERG concurred, and one commented that administration of DBd requires day unit time and staff training, and the first infusion in particular requires several hours spent in the day unit.

## 2.3 Critique of company's definition of decision problem

**Population**

The population described in the decision problem is adults with RRMM who have had one previous treatment (i.e. 2L patients). This population is appropriate for the NHS. However, it is narrower than the population specified in the NICE scope, which is adults who have had at least one previous treatment (i.e. not limited to the number of previous lines of treatment). The company justifies the use of DBd as a 2L therapy on the basis that there is an unmet need for 2L triple therapies, and because this is where they believe DBd would provide the greatest clinical benefit and cost effectiveness (CS section B.1.1). We note that the key phase III pivotal trial (CASTOR) which informs the CS includes adult patients with RRMM, not limited by the number of previous lines of treatment. As a consequence of the company's focus on 2L patients, only a subgroup of the patients randomised in the CASTOR trial (47.2%) is relevant to the company's decision problem. Furthermore the ERG note that the company have not distinguished between those 2L patients who received bortezomib as a first-line therapy and those who are bortezomib naïve; this is relevant because the comparators for 2L treatment (bortezomib and carfilzomib) are only available in current clinical practice for those who are bortezomib naive as noted above in section 2.2. In response to clarification question B9 the company confirmed that DBd is being positioned as a 2L treatment option regardless of prior bortezomib exposure status.

**Intervention**

The intervention as described in the decision problem is daratumumab (Darzalex®), in combination with bortezomib and dexamethasone (DBd) (i.e. triple therapy), which is consistent with the intervention specified in the NICE scope and the marketing authorisation. According to the marketing authorisation (as specified in the Summary of Product Characteristics, SmPC[14]), DBd triple therapy is licensed "f*or the treatment of adult MM patients who have received at least one prior treatment*" (CS Table 2). However, (as noted above) the population specified in the decision problem (2L patients) is narrower than this.

DBd is a relevant triple therapy for use in the NHS in the 2L setting where, as the company point out, recommended treatment options are very limited in England (CS Figures 4 and 5).

The dosing of daratumumab specified by the company (CS Table 2) is consistent with the SmPC.[14] Daratumumab is administered as an intravenous infusion at a dose of 16 mg per kg

body weight weekly in weeks 1-9 (cycles 1-3), every three weeks in weeks 10-24 (cycles 4-8), and every four weeks from week 25 onwards. Each cycle is 21 days.

The CS and the SmPC[14] do not explicitly specify the recommended posology of bortezomib and dexamethasone for co-administration with daratumumab. In the CASTOR trial, bortezomib and dexamethasone were administered as follows, which clinical experts advising the ERG agreed reflects NHS clinical practice (CS Tables 5 and 8):

- Bortezomib: 1.3 mg per $m^2$ body area, administered subcutaneously twice weekly on days 1, 4, 8 and 11 for eight 21-day cycles (cycles 1-8).
- Dexamethasone: 20mg given orally on days 1, 2, 4, 5, 8, 9, 11 and 12 in each of the first eight treatment cycles (total dose of 160mg per cycle). In weeks when the patient received daratumumab, dexamethasone was administered as a 20mg intravenous (IV) infusion instead of orally, prior to the daratumumab infusion. Some adjustments were permissible to the dexamethasone dosing for patients who were older than 75 years, underweight (body mass index [BMI] <18.5), had poorly controlled diabetes mellitus or prior intolerance or adverse event (AE) to steroid therapy (CS Table 8).

According to the SmPC,[14] several concomitant medications are recommended for administration with daratumumab to reduce the risk of infusion-related reactions (IRR):

- 1-3 hours pre-infusion: corticosteroid (long-acting or intermediate-acting), antipyretic (oral paracetamol) and antihistamine (oral or IV diphenhydramine or equivalent);
- Post-infusion: low-dose oral methylprednisone or equivalent on the day after daratumumab infusion.

The CS states that methylprednisone is given on the first and second days after infusion, and that post-infusion medications may be discontinued after >4 infusions if there are no major IRRs (CS Table 46).

### Comparators

The company has limited their decision problem to the 2L setting, whereas the NICE scope lists comparators for second-, third- and fourth-line settings (CS Table 1).

The comparators specified in the company's decision problem are consistent with those specified in the NICE scope for 2L patients (bortezomib with or without dexamethasone [the NICE scope specifies 'bortezomib-based therapy']; Cd; or combination chemotherapy).

The company's decision problem does not include the following comparators which are specified in the NICE scope for third-line and fourth-line patients:

- 3rd line: lenalidomide with dexamethasone; or panobinostat with bortezomib and dexamethasone;
- 4th line: pomalidomide with dexamethasone; panobinostat with bortezomib and dexamethasone; or daratumumab monotherapy (available under the Cancer Drugs Fund; CDF)

## Outcomes

The outcomes specified in the company's decision problem are standard outcomes for assessing the clinical effectiveness and safety of cancer interventions and include all of those listed in the NICE scope [OS, progression-free survival (PFS), response rates, adverse effects of treatment, and HRQoL).

In addition to the outcomes stated in the NICE scope, the decision problem includes minimal residual disease (MRD) as an indicator of response to therapy, and time to treatment discontinuation (TTD) which is an input parameter in the company's economic model. The ERG agrees that these are appropriate and clinically meaningful outcomes.

### Economic analysis

The decision problem does not specify the economic analysis.

### Subgroups

The NICE scope specifies that, if evidence allows, subgroups based on the number of lines of previous therapy will be considered. The company's decision problem does not explicitly describe any subgroups. However, as noted above, the company focuses on the 2L setting, which is a subgroup of the RRMM population specified in the NICE scope (and a subgroup of the population included in the pivotal RRMM phase III CASTOR trial). The CS also reports further subgroup analyses which were pre-specified in the CASTOR trial.

**Equality considerations**

The CS states that there are no equality issues arising in relation to DBd (CS section B.1.4). The ERG, and clinical experts advising the ERG, are not aware of any equality issues.

# 3 CLINICAL EFFECTIVENESS

## 3.1 Critique of company's approach to systematic review

### 3.1.1 Description of company's search strategy

The company submission comprises the following five systematic literature searches:

- Clinical Evidence (CS Appendix D). Run on 11th December 2017.
- Adverse Reactions (CS Appendix F). A previous search conducted for the NICE appraisal of daratumumab monotherapy was extended and run for the period July 2016 to September 2017.
- Cost Effectiveness (CS Appendix G). Run for the period 1st March 2005 to 22nd August 2017.
- Health Related Quality of Life (CS Appendix H). Run for the period 1st March 2005 to 22nd August 2017.
- Cost & Healthcare Resource Identification, Measurement & Valuation (CS Appendix I). Run for the period 1st March 2005 to 22nd August 2017.

An acceptable range of databases was used in all five searches. The strategies employed an appropriate balance of descriptors and suitably truncated free text, with sets combined correctly. The searching syntax appears to be of sound quality, and appropriate search filters have been applied in all searches. PRISMA charts are provided for each search.

Extensive supplementary grey literature searching was undertaken for all five searches. Key conferences searched included: the American Society of Clinical Oncology (ASCO), American Society of Hematology (ASH), European Hematology Association (EHA), International Myeloma Working Group (IMWG), International Society for Pharmacoeconomics and Outcomes Research (ISPOR) with additional searches performed on Embase to identify conference material. Conference proceedings were searched for the period 2013-2017 (except ASH searches, 2013-2016). Other sources searched included: English language technology appraisals conducted by the Pharmaceutical Benefits Advisory Committee (PBAC), Canadian Agency for Drugs and

Technologies (CADTH), National Institute for Health and Care Excellence (NICE) and Scottish Medicines Agency (SMC). Clinicaltrials.gov was examined to identify ongoing trials.

The clinical effectiveness search focused on the health condition with a clinical trial search filter appended. There was no specification of individual drugs; searches for these were therefore broad, with a filter applied at the eligibility screening step. The Adverse Reactions search modified a search used originally in the NICE appraisal of daratumumab monotherapy (the original search method is documented in CS Appendix F.6). The modified search included DBd as an intervention, included a wider range of comparators, and specified the population as having ≥1 prior therapy (CS Appendix F Table 1). This updated search for adverse reactions was limited to a 14-month period up to September 2017 and was intended to complement safety data identified in the Clinical Evidence search. The cost, quality of life and healthcare resource identification searches used the condition (multiple myeloma) with appropriate search filters appended.

The ERG identified some minor issues (e.g. only the Adverse Reactions search documented the number of hits per line; there is a typographical error in the NOT operator in all searches; and the company does not provide a justification for searching conferences only after 2012) but these are not considered to have resulted in any key evidence having been missed.

The ERG did not re-run the full clinical effectiveness search, given that this is nearly up-to-date, but we consulted abstracts presented at ASH 2017, checked for ongoing trials at WHO ICTRP and NIHR UKCTG, and ran a Delphis overarching database search to identify any late 2017 and 2018 papers. No additional comparative clinical effectiveness studies or ongoing studies likely to report additional data in the next 12 months were identified.

The cost-effectiveness search was updated by the ERG in March 2018, run on Embase Medline and Econlit. All supplementary search results carried out by the ERG Information Specialist were checked independently against the inclusion criteria by two researchers and two publications relevant to the decision problem were identified. These are discussed further in section 4.2.2 of this report.

An ERG search of the Health Utilities Database in March 2018 did not identify any further relevant unique studies on health utilities.

In summary, the company's searches are well documented, generally comprehensive and do not appear to have missed any key relevant studies.

### 3.1.2 Statement of the inclusion/exclusion criteria used in the study selection.

The company reports the inclusion and exclusion criteria for the systematic literature review in the CS (CS Appendix D, Table 10). References were screened against the criteria to identify both direct and indirect evidence: that is, to find studies comparing the efficacy of DBd directly with specified 2L comparator treatments, and to identify trials comparing the stated 2L comparator treatments which could potentially be used in a network meta-analysis (NMA). Table 5 summarises the systematic review eligibility criteria.

**Table 5 Summary of the CS systematic review eligibility criteria**

| | |
|---|---|
| **Population** | People with RRMM who had received "*at least one prior line of therapy*" (CS Appendix D Table 10) |
| **Intervention** | DBd |
| **Comparators** | Licensed or treatments under investigation for RRMM, including:<br>• bortezomib-based therapy<br>• Cd<br>• combination chemotherapy<br>Or if they compared any of the specified comparators to each other |
| **Outcomes** | Studies needed to measure and report extractable data for the following outcomes:<br>• OS<br>• PFS<br>• TTP<br>• tumour response outcomes [CR, stringent complete response (sCR), VGPR, partial response (PR), minimal response, progressive disease and stable disease]<br>• HRQoL<br>• AEs |
| **Study design** | Randomised controlled trials (RCTs) |

| | |
|---|---|
| **Publication type and/or date** | Full publications of studies indexed in the literature in or prior to 1994 were excluded. Only conference abstracts presented in 2012 or afterwards, from named conferences, were included (section 3.1.1). [a]<br><br>References reporting more than one RCT or pooled analyses were excluded, but no rationale for this is provided. The company subsequently explained that this was applicable in cases where it was not possible to separate and extract data from the individual RCTs (clarification response A13). [b]<br><br>The ERG agrees these eligibility criteria for publication type and date are reasonable. |

Source: Parts of CS Appendix D Table 10 & Appendix section D.1.2.

Note. The company does not specify treatment setting as an inclusion criterion nor place any limits on inclusion relating to the quality of the RCTs. We agree that this is appropriate for ensuring that all relevant evidence was identified (the quality of included trials was assessed subsequently).

The inclusion criteria reflect the licensed indication for daratumumab when used as a triple therapy in combination with bortezomib and dexamethasone (although drug doses and regimens are not specified in the criteria). The criteria, however, do not wholly reflect the decision problem presented in the CS or the NHS position, in the following notable ways:

**Eligible population:**

- The patient population eligibility criterion specifies people who had received "*at least one*" previous line of treatment (CS Appendix D Table 10), whereas the decision problem states patients eligible for 2L treatment were of interest. This means that whilst RCTs on patients receiving subsequent therapy lines were eligible, data from these patients do not inform the decision problem.

**Eligible comparators:**

- The comparator eligibility criteria do not fully reflect NICE's[15 16] and NHS England's[17] restriction of the use of Cd and bortezomib-based therapy, respectively, to 2LBN patients. The company's systematic review did not restrict eligible studies according to first-line therapy received. This means RCTs of 2LBN and 2L bortezomib-experienced

patients were eligible for inclusion in the review (and the company is positioning DBd for the treatment of both these patient populations; clarification response B9). The range of eligible comparators is wider than that specified in the decision problem in two ways: First, the company specified bortezomib-based therapy as an eligible comparator, but this is wider than the decision problem, which more specifically states bortezomib with or without dexamethasone is a comparator of interest. Second, RCTs of "*licensed or treatments under investigation for RRMM*" could be included (CS Appendix D, Table 10). While these criteria are wider than the decision problem, we agree that they are appropriate for the purposes of the review because the company used the criteria to identify both indirect and direct evidence to inform their NMA. It is appropriate to include studies of comparators outside the decision problem if they contribute evidence to the network.

Overall, we consider the systematic review inclusion and exclusion criteria were adequate for identifying evidence broadly relevant to the decision problem and NICE's scope.

The CS includes a PRISMA flow diagram showing the number of studies included and excluded at each main stage of the review (CS Appendix D, Figure 1). The diagram provides the reasons for excluding references, along with the number excluded for each reason [all references excluded at full-text screening (n=214) are listed in CS Appendix D Table 12]. The company excluded two references because they reported "*extended treatment*" studies (CS Appendix D, Figure 1). At the clarification stage, the company explained that these involved mixed study designs that were not solely RCTs (clarification response A15). As such, we agree that these exclusions are appropriate.

The flow diagram states that 45 references were included in the review, but does not specify how many RCTs these reported. CS Appendix D.1.2.3 clarifies that three RCTs that investigated 2L treatments were included, and the 45 references describing them are listed in CS Appendix D Table 11. These RCTs were CASTOR,[18] ENDEAVOR[19] and Phillips 1995, and they are described further in the following section of this report (3.1.3).

### 3.1.3 Identified studies

The company's systematic review identified three RCTs: CASTOR,[18] ENDEAVOR[19] and Phillips 1995. CASTOR was the only RCT that evaluated DBd. This was a head-to-head trial of DBd versus Bd in adults with documented relapsed or refractory multiple myeloma. The CASTOR trial was sponsored by Janssen Research and Development (i.e. by the company).[18] The CASTOR trial meets the systematic review inclusion criteria, but only addresses the decision problem's focus on 2L patients through pre-specified subgroup analyses of these patients (around half the trial population: 45.7% of the patients in the DBd arm and 48.6% in the Bd arm had received one prior line of therapy). The remaining trial patients had received two or more prior therapies.

ENDEAVOR and Phillips 1995 investigated comparator treatments for 2L patients and were considered by the company for inclusion in their NMA. Further details about these latter two RCTs and their eligibility for inclusion in the MTC are provided in section 3.1.7 of this report. ENDEAVOR was sponsored by Onyx Pharmaceuticals, Inc. (an Amgen subsidiary).

The company provides a publication for the CASTOR RCT (Palumbo 2016, Doc B ref 67) which reports on methods and results from the interim analysis conducted after approximately 60% of the planned events for the final analysis. The data cut off for this interim analysis was January 11th 2016 when the median follow-up period was 7.4 months. The CS reports results from a later analysis point (median follow-up 26.9 months) for which there is no publication. In addition the CASTOR Clinical Study Report (CSR) is also provided (Doc. B ref. 68)

An RCT report is also provided for the ENDEAVOR RCT (Doc. B ref 77, Dimopoulos et al 2016) which contributes data to the NMA. In response to clarification question A7 the company clarified that data were also included from the second interim analysis of ENDEAVOR (Dimopoulos 2017) which is not cited separately.

The company considered a second RCT, Phillips 1995, for possible inclusion in their NMA but an incorrect reference is cited (the CS cites ref. 73 but this is not Phillips 1995). Data from Phillips 1995 are not used in the CS as this RCT was not included in the NMA.

**Non-randomised studies**

The company did not include any non-randomised studies in the review of DBd or in the indirect comparison. The company conducted additional searches, though, for non-randomised studies that provided AE data additional to that provided in the CASTOR trial. It is unclear if the data from the non-randomised studies contributes to the evidence presented in the CS. In response to clarification question A10 the company supplied full details for the 20 studies from which data were extracted which included many non-randomised studies. Data from these 20 studies were not presented in the CS but the company notes that no new safety signals were identified for any of the treatments of interest.

**Overview of the CASTOR trial**

We have provided summary details of the CASTOR trial in Table 6. The treatment regimens, and doses of each drug, used in the DBd and Bd treatment arms match those specified in the SmPC for daratumumab[14] and the SmPC for bortezomib (when it is combined with dexamethasone).[20] That is, the treatment regimens of the individual drugs are in line with their licensed indications. We note, however, that the SmPC for bortezomib states that it is to be administered by intravenous injection,[21] but in both arms of the CASTOR trial it was administered subcutaneously. It is unclear if these different modes of administration might have an impact on outcomes – this is not discussed in the CS, CSR or trial protocol. A published RCT[22] that we found compared these modes of administering bortezomib in patients with relapsed multiple myeloma and reported that subcutaneous administration provided non-inferior efficacy to intravenous injection, and reduced AEs. The ERG's clinical experts advised us that bortezomib is now administered subcutaneously in practice. Intravenous injection would only be used in the rare instance of a severe local reaction to the drug. The experts agreed that toxicity is higher with intravenous injection than subcutaneous administration. Therefore, subcutaneous administration of bortezomib in the CASTOR trial reflects how bortezomib is administered in practice.

**Table 6 Summary of the CASTOR RCT**

| Design, patient population, length of follow-up and N (%) of 2L patients | Intervention: DBd | Comparator: Bd |
|---|---|---|
| *Design:* Phase III, open-label, multicentre RCT carried out in 16 countries (no UK centres).<br><br>*Patient population:* Adults with documented evidence of relapsed or refractory multiple myeloma, as assessed against IMWG criteria.<br><br>*Key eligibility criteria:*<br>To be eligible for the trial, patients had to have received at least one prior line of treatment, and to have achieved at least a partial response to the previous treatment. ECOG PS of 0, 1 or 2.<br><br>N randomised = 498 (251 DBd; 247 Bd).<br><br>*Median length of follow-up:*<br>Interim data reported in the CS, at a median follow-up of 26.9 months.<br><br>*N (%) of 2L patients:*<br>DBd: 122 (48.6); Bd: 113 (45.7) | Daratumumab: 16mg/kg of body weight weekly for three 21-day cycles, then every 3 weeks for cycles 4 to 8 and thereafter every 4 weeks<br><br>Bortezomib: 1.3mg/m$^2$ body surface area, subcutaneously, twice weekly on days 1, 4, 8, and 11, for up to eight 21-day cycles (cycles 1 to 8)<br><br>Dexamethasone: total dose of 80mg weekly in 2 out of 3 weeks of a 21-day cycle (on days 1, 2, 4, 5, 8, 9,11 and 12), for up to 8 cycles | Bortezomib: 1.3mg/m$^2$ body surface area, subcutaneously, twice weekly on days 1, 4, 8, and 11 for up to eight 21-day cycles (cycles 1 to 8)<br><br>Dexamethasone: total dose of 80mg weekly in 2 out of 3 weeks of a 21-day cycle (on days 1, 2, 4, 5, 8, 9,11 and 12), for up to 8 cycles |

Source: CS Tables 5 and 6 and the trial publication.[18] The full list of eligibility criteria is given in CS Table 6.

**CASTOR trial patient eligibility criteria**

The population of interest in this appraisal includes people with refractory MM. However patients refractory to bortezomib or another proteasome inhibitor (that is, patients who had progression of disease while receiving PI therapy or within 60 days of ending PI therapy) were not eligible to enter the CASTOR trial (CS Table 6). That is, patients in the 2L subgroup who had previously had bortezomib were only eligible to take part in the trial if they had relapsed after bortezomib treatment. The ERG's clinical experts advised that this exclusion criterion was reasonable and reflects clinical practice. This is also a minor point, however, as the ERG

believes the most relevant 2L patient population for this appraisal is 2LBN patients. We also note that to be eligible for the trial "*Patients must have achieved a response (PR [partial response] or better) to at least one prior regimen*" (CS Table 6). Clinical expert advice to the ERG is that patients who have not had a response of PR or better (i.e. those with refractory disease) with a non-proteasome inhibitor treatment would be offered treatment with agents from another group (proteasome inhibitor or immunomodulatory drug, depending on the resistance pattern) in clinical practice. Expert advice was also that those with refractory disease in practice would likely to be refractory to bortezomib, as most patients are treated with this at first-line. Therefore, we suggest this exclusion criterion does not fully reflect clinical practice, as it excludes the minority of patients seen in practice who were refractory to a non-proteasome inhibitor first-line treatment. To be included, all patients were also required to have "*documented relapsed multiple myeloma with measurable disease in the serum and/or urine as defined by the IMWG [International Myeloma Working Group] criteria*" (CS section B.2.3.1.1). The CS does not state if these criteria are used in clinical practice in England to assess disease progression, so it is unclear how well the trial patient population reflects those seen in clinical practice. The ERG's clinical experts advised that these criteria are used in clinical practice, but that this inclusion criterion excludes a minority of patients with non-secretory myeloma who in theory could be monitored with serial bone marrow examinations. Expert advice is that there are few treatment options for these patients.

**CASTOR baseline characteristics**

The CS states that the CASTOR trial[18] patient baseline characteristics for the whole trial population were well balanced between the treatment arms (p. 34). We mostly agree with this, but note that the trial paper reports that the difference in the proportions of patients receiving previous immunomodulatory drug therapy in the DBd arm (71.3%) versus in the Bd arm (80.2%) was statistically significant (p = 0.02). The ERG's clinical experts believe that this difference is unlikely to impact on the treatment effect. The CS does not report the baseline characteristics for the 2L treatment subgroup, but the company provided these in their clarification response A6. We have reproduced part of the company's table below (Table 7) and the full table is available in clarification response A6. We note that the 2L patient baseline characteristics were well balanced between treatment arms, with some exceptions, including the patients' prior treatment and refractory to last line of therapy statuses. The proportion of patients who received prior lenalidomide was higher in the comparator group (Bd 29.2% vs DBd 12.3%, difference 16.9%). Proportionally more patients in the Bd arm (22.1%) had been refractory prior to therapy

to an immunomodulatory drug therapy than in the DBd arm (11.5%), and refractory to lenalidomide specifically (Bd 15.9% vs. DBd 4.9%). The company indicate that refractoriness to lenalidomide and the type of prior treatment received are prognostic factors (CS sections B.1.3.1.1 and B.2.5.2). However, The ERG's clinical advisors stated the differences were unlikely to impact on the treatment effect (see section 3.1.4 for discussion regarding the potential impact of these differences on risk of bias).

Clinical expert advice to the ERG is that the 2L patients' baseline characteristics are representative of the patients seen in clinical practice, except the trial patients were on average slightly younger and proportionally more had had prior exposure to lenalidomide at first-line treatment than patients treated in practice. Furthermore, patients treated in practice do not receive anthracycline.

**Table 7 Key characteristics of patients in the 2L treatment subgroup of the CASTOR trial**

| Population characteristic | | DBd (n=122) | Bd (n=113) | Total (n=235) |
|---|---|---|---|---|
| Age, years, mean (SD) [range] | | 62.6 (9.83) [30-84] | 64.2 (9.88) [40-85] | 63.3 (9.87) [30-85] |
| Male, n (%) | | 74 (60.7) | 64 (56.6) | 138 (58.7) |
| Race: | White | 104 (85.2) | 99 (87.6) | 203 (86.4) |
| | Asian | 10 (8.2) | 8 (7.1) | 18 (7.7) |
| | Black or African American | 4 (3.3) | 2 (1.8) | 6 (2.6) |
| | Other, unknown or not reported | 4 (3.3) | 4 (3.5) | 8 (3.4) |
| Weight, kg, mean (SD) [range] | | 78.3 (17.4) [45-135] | 77.6 (14.9) [45-132] | 77.9 (16.2) [45-135] |
| Baseline ECOG score, n (%): | 0 | 57 (46.7) | 56 (49.6) | 113 (48.1) |
| | 1 | 58 (47.5) | 51 (45.1) | 109 (46.4) |
| | 2 | 7 (5.7) | 6 (5.3) | 13 (5.5) |
| Time from MM diagnosis, years, mean (SD) [range] | | 3.6 (2.8) [0.7-14.9] | 3.6 (2.5) [0.6-18.1] | 3.6 (2.7) [0.6-18.1] |
| Any cyto-genetic abnormality, n (%) | Standard risk | 70 (76.9) | 67 (84.8) | 137 (80.6) |
| | High risk | 21 (23.1) | 12 (15.2) | 33 (19.4) |
| | Del17p | 13 (14.3) | 6 (7.6) | 19 (11.2) |
| | T(4;14) | 5 (5.5) | 5 (6.3) | 10 (5.9) |
| | T(14;16) | 3 (3.3) | 4 (5.1) | 7 (4.1) |

| Population characteristic | DBd (n=122) | Bd (n=113) | Total (n=235) |
|---|---|---|---|
| Risk stratification: [a] High risk | 7 (5.7) | 4 (3.5) | 11 (4.7) |
| Standard risk | 73 (59.8) | 66 (58.4) | 139 (59.1) |
| Low risk | 11 (9.0) | 9 (8.0) | 20 (8.5) |
| Not done | 31 (25.4) | 34 (30.1) | 65 (27.7) |
| Prior ASCT | 76 (62.3) | 66 (58.4) | 142 (60.4) |
| Prior radiotherapy | 28 (23.0) | 24 (21.2) | 52 (22.1) |
| Prior cancer-related surgery | 13 (10.7) | 20 (17.7) | 33 (14.0) |
| Prior protease inhibitor, n (%) | 65 (53.3) | 59 (52.2) | 124 (52.8) |
| Bortezomib | 62 (50.8) | 57 (50.4) | 119 (50.6) |
| Carfilzomib | 1 (0.8) | 2 (1.8) | 3 (1.3) |
| Ixazomib | 2 (1.6) | 0 | 2 (0.9) |
| Prior IMiD, n (%) | 72 (59.0) | 81 (71.7) | 153 (65.1) |
| Lenalidomide | 15 (12.3) | 33 (29.2) | 48 (20.4) |
| Thalidomide | 58 (47.5) | 48 (42.5) | 106 (45.1) |
| Refractory to IMiD only, n(%) | 14 (11.5) | 25 (22.1) | 39 (16.6) |
| Refractory to last line of therapy, n (%) | 18 (14.8) [b] | 25 (22.1) [b] | 43 (18.3) [b] |
| Refractory to Lenalidomide | 6 (4.9) | 18 (15.9) | 24 (10.2) |
| Refractory to Thalidomide | 8 (6.6) | 7 (6.2) | 15 (6.4) |

Source: Reproduction of data in Table 4 in clarification response A6.
[a] Risk stratification is based on three factors: International Staging System (ISS); presence of chromosomal abnormalities of t(4; 14), del17 or del17p by fluorescence in situ hybridisation (FISH) or Karyotype testing and age.
[b] Most of these patients were refractory to lenalidomide or thalidomide.

**2L patients' prior exposure to bortezomib**

The company reports in clarification response A6 that 50.6% of the patients in the 2L subgroup had previously received bortezomib (50.4% in the Bd arm and 50.8% in the DBd arm). The company indicates that this is in line with bortezomib usage in the multiple myeloma population at first-line treatment in the UK, and our clinical experts concurred with this. In clinical practice in England, however, bortezomib is not recommended for use at second-line treatment in patients who have previously received it.[16 17] Therefore, around half the patients in the 2L trial subgroup do not provide data relevant to clinical practice guidance in England. The company does, however, provide estimates of PFS and OS in the post-hoc 2LBN subgroup for DBd versus BD and DBd versus Cd in NMA sensitivity analysis.

**Summary of key limitations of the evidence provided by the CASTOR trial**

In summary, we consider the CASTOR trial evidence has limitations in terms of addressing the decision problem and reflecting clinical guidance in England in the following ways:

- The trial provides evidence relevant to the decision problem through subgroup analysis only. 2L patients formed 47.2% (n = 235/498) of the trial population.
- 50.6% of the patients in the 2L subgroup had previously received bortezomib. Therefore, around half the patients in the 2L trial subgroup do not provide data relevant to clinical practice guidance in England, which states bortezomib to be used only in bortezomib-naïve patients. Estimates of OS and PFS for this population are available in the CS from a post-hoc NMA sensitivity analysis.
- The trial patient eligibility criteria do not fully reflect the 2L patients seen in clinical practice in England, as they excluded the following two patient groups: one, patients who had been refractory to prior treatment with a non-proteasome inhibitor; and, two, patients with non-secretory myeloma. Both these groups form a minority of the patients treated in practice.

**ERG's appraisal of whether all relevant studies were included in the review**

The ERG believes the company has included all relevant studies in the review. The ERG's searches did not identify any additional evidence (see section 3.1.1).

**Ongoing studies**

The CS lists three ongoing trials that are expected to provide additional clinical evidence in the next 12 months, and details the populations and interventions investigated in these. None of the trials are relevant to the company's decision problem or to NICE's scope for this appraisal. NICE and the ERG asked the company for additional information about the ongoing studies' designs, including the number of comparator arms, completion dates and whether interim will become available before the completion dates and this information was provided in response to clarification question A11. This information confirmed that none of the trials are relevant to the company's decision problem.

The ERG searched for ongoing trials and no additional relevant RCTs expected to provide evidence in the next 12 months were identified.

### 3.1.4 Description and critique of the approach to validity assessment

The company has provided a risk of bias assessment for the CASTOR trial based on standard NICE criteria (CS Table 12). The company's and ERG's judgements are shown in Table 8 (full details including explanation for the judgments is available in Appendix 1). Although we agree with most of the company's judgements, there are some differences.

The ERG considers that all the CASTOR trial outcomes are at an unclear risk of selection bias in the whole trial population and the 2L treatment subgroup, due to an imbalance in the proportion of patients who received lenalidomide as a first-line therapy, but the direction of the bias is unclear, since the reasons why patients received prior lenalidomide (as opposed to receiving other possible therapies), are not reported. Additionally, in the 2L subgroup, proportionally more patients in the Bd arm were refractory to their previous treatment, including to lenalidomide. The company indicate that the type of prior treatment received and refractoriness to lenalidomide are prognostic factors (CS sections B.1.3.1.1 and B.2.5.2). However, The ERG's clinical advisors stated the imbalances observed between trial arms for these factors were unlikely to impact on the treatment effect. Due to these differing statements we are unclear if these differences could potentially lead to less favourable efficacy results for 2L patients in the Bd arm compared with the DBd arm.

The ERG considers that the HRQoL outcomes in CASTOR are at high risk of performance bias due to the lack of blinding. However, as indicated in Table 8, we believe the lack of blinding is unlikely to have introduced bias for other outcomes which are less subjective. We are unclear about the risk of selection bias. For the remaining domains of bias (detection bias, attrition bias and reporting bias) we agree with the company that the risk of bias is likely to be low. An exception is that we consider the risk of attrition bias for HRQoL outcomes to be unclear, due to lack of clarity on how missing data were handled.

**Table 8 Company and ERG assessments of risk of bias**

| Question | Risk of bias | |
|---|---|---|
| Was randomisation carried out appropriately? | Company | Low |
| | ERG | Low |
| Was the concealment of treatment allocation adequate? | Company | Potential risk of bias as open label design could have influenced investigator's assessment of PFS events |
| | ERG | Probably low |
| Were the groups similar at the outset of the study in terms of prognostic factors? | Company | Low |
| | ERG | Unclear |
| Were the care providers, participants and outcome assessors blind to treatment allocation? | Company | Low, as an IDMC reviewed the data |
| | ERG | Low for OS and TTD |
| | | Probably low for response outcomes and PFS |
| | | High for HRQoL |
| Were there any unexpected imbalances in drop-outs between groups? | Company | Low |
| | ERG | Low risk, provided that outcomes are interpreted in the context of the expected imbalance |
| Is there any evidence to suggest that the authors measured more outcomes than they reported? | Company | Low |
| | ERG | 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? | Company | Low |
| | ERG | Low for time-to-event outcomes |
| | | Low for response outcomes |
| | | Unclear for HRQoL outcomes |

Source: Company risk of bias judgements from CS Table 12

### 3.1.5 Description and critique of company's outcome selection

We have tabulated the CASTOR trial outcomes reported by the company in the CS and we indicate which were used in the economic model (Table 9). The outcomes match all those specified in the NICE final scope and the company's decision problem. As shown in the table, the company also included a number of other outcomes which were not specified in the final

scope or decision problem. None of these were used in the company's economic model. We did not identify any relevant outcomes from the trial that were not reported in the CS.

**Table 9 CASTOR trial[18] outcomes reported in the CS**

| Outcome specified in the scope and decision problem (in bold)/in the decision problem only (non-bold) | Outcomes reported in the CS (CASTOR trial) | Whole trial | 2L patients subgroup | NMA Whole trial | NMA 2L patients | Used in economic model (2L patients) |
|---|---|---|---|---|---|---|
| **PFS** | PFS (primary outcome) | Yes | Yes | Yes | Yes | Yes [a] |
| **OS** | OS | Yes | Yes | Yes | Yes | Yes [b] |
| | OS adjusted for subsequent treatment | Yes | Yes | No | No | Yes [c] |
| **Response rates**, including MRD negativity | PR | Yes | Yes | No | No | No |
| | CR, and CR or better | Yes | Yes | CR or better | CR or better | No |
| | sCR | Yes | Yes | No | No | No |
| | Rate of VGPR, and VGPR or better | Yes | Yes | VGPR or better | VGPR or better | No |
| | Overall response rate (ORR) | Yes | Yes | Yes | Yes | No |
| | DOR | Yes | No | No | No | No |
| | Time to response (TTR): time to first response; time of VGPR of better, time to CR or better | Yes | No | No | No | No |
| | Minimal residual disease (MRD)/MRD negative status | Yes | Yes | No | No | No |
| | % of patients achieving clinical benefit (ORR + minimal response) | Yes | No | No | No | No |
| **AEs** | AEs (safety and tolerability) | Yes | No | Yes [d] | Yes [d] | Yes [e] |
| **HRQoL** | EORTC QLQ-C30 | Yes | No | No | No | No |
| | EQ-5D-5L | Yes | No | No | No | Yes |
| Time to treatment discontinuation (TTD) | Time to treatment discontinuation (TTD) (post-hoc analysis) | Yes | Yes | No | No | Yes |
| *Outcomes not specified in scope or decision problem* | TTP | Yes | Yes | No | No | No |
| | Time to subsequent anticancer therapy | Yes | No | No | No | No |
| | PFS on the subsequent line of therapy | Yes | No | No | No | No |
| | Best M-protein response | Yes | No | No | No | No |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Best response to first subsequent anticancer therapy | Yes | No | No | No | No |

[a] PFS data for DBd and Bd taken from the CASTOR trial. For carfilzomib PFS is modelled by applying the hazard ratio from the NMA to the reference Bd curve from CASTOR. [b] Unadjusted OS data used in a scenario analysis. [c] OS data for DBd and Bd in the basecase are taken from the CASTOR trial and adjusted for use of subsequent therapies not available in England. For carfilzomib OS is modelled by applying the hazard ratio from the NMA to the reference Bd curve from CASTOR. [d] specifically the following 9 outcomes: discontinuation due to adverse events; grade 3+ AEs; Grade 3+ anaemia; Grade 3+ diarrhoea; Grade 3+ fatigue; Grade 3+ leukopenia; Grade 3+ neutropenia; Grade 3+ peripheral neuropathy; Grade 3+ thrombocytopenia. [e] Grade 3 or higher AEs with data taken directly from the CASTOR whole trial population for DBd and Bd and from the ENDEAVOR trial for carfilzomib.

**ERG's assessment of the appropriateness of the outcomes selected**

The outcomes selected by the company are appropriate to this appraisal, but we have identified a few caveats, as we detail next.

PFS was the primary outcome in the CASTOR trial. It was defined as the time from patient randomisation to either disease progression or death (whichever came first). The CS and trial paper[18] state disease progression was assessed using a computerised algorithm and in accordance with the International Myeloma Working Group (IMWG) criteria, which the ERG's clinical advisors confirmed are used in clinical practice in England. The CS does not state, however, if there was a review of PFS or any of the other time to event or response outcomes. The trial paper states that a previous trial using the computer algorithm found "strong concordance" (Palumbo et al., 2016, p. 756) with the results derived from the independent review committee in that trial. The trial paper[18] furthermore states that all responses, including progressive disease, were verified by another assessment, but it is unclear how this was conducted (by trial clinicians, a central independent review or other means).

The company included MRD as an outcome, as they argue it is a more sensitive measure of disease burden than CR. The CS states (CS Section B.2.3.1.4) the IMWG guidelines recommend that MRD assessment is considered at the end of each treatment phase among patients who have had a CR. The MRD outcome is assessed through measuring residual tumour cells in the bone marrow. The response is then categorised according to IMWG criteria. The CS states that MRD negative status is associated with longer OS and PFS (CS p. 50 and p. 102). A reference we found also states this.[23] The MRD outcome results from the CASTOR trial are not used directly in the economic model, but the results are used to justify the company's choice of OS extrapolation curve (i.e. its estimate of OS beyond the trial period) which is used to inform the economic model. It is unclear from the CS if this outcome is measured in clinical

practice in England and if it is used to inform treatment or monitoring decisions. It is therefore unclear how clinically relevant it is. Clinical expert advice to the ERG is that it is a prognostic factor, but that it does not inform treatment or monitoring decisions.

The company includes an outcome called '% of patients achieving clinical benefit', which is defined as ORR plus minimal response. NICE and the ERG asked the company in clarification question A4 if this is a standard definition of clinical benefit and, if so, the source of the definition. The company explained that the use of this definition stemmed from an American Society of Hematology and US Food and Drug Administration workshop and several trials in relapsed &/or refractory MM have used this outcome definition. Clinical experts advising the ERG commented that, at second line treatment, this outcome is of limited relevance; as treatment becomes increasingly more palliative, minimal response becomes more acceptable.

**HRQoL measures**

HRQoL was measured in the CASTOR trial using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ-C30) and the EuroQol Five Dimensions Questionnaire (EQ-5D-5L). The company's economic model uses data from the EQ-5D, but not the EORTC-QLQ-C30.

EQ-5D-5L is a validated, generic measure of HRQoL and is NICE's preferred measure of HRQoL in adults.[24] The EQ-5D-5L is a more recent version, and NICE states it can be used in its technology appraisals (with appropriate mapping to EQ-5D-3L, as the 5D version utilities are yet to be validated) (NICE, 2013).[24] Results for the EQ-5D-5L utility score and VAS are reported in the CS.

The EORTC-QLQ-C30 is a generic, self-administered measure of quality of life in cancer patients that has been validated among people with multiple myeloma.[25 26] The instrument has the following scales: Global Health Status/QOL subscale, functional scales (physical, role, cognitive, emotional and social), and symptom scales (fatigue, pain, nausea and vomiting). It also measures financial difficulties, loss of appetite, insomnia, dyspnoea, constipation and diarrhoea. A higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.[26] A brief overview of the results for all the subscales is reported in the CS.

**Summary**

The outcomes selected by the company are appropriate and match the NICE scope, the company's decision problem and NICE's preferred method for measuring HRQoL.[24] The '% of patients achieving clinical benefit' outcome is of limited relevance in the 2L treatment setting in clinical practice, though.

### 3.1.6 Description and critique of the company's approach to trial statistics

The CS reports results for all of the outcome measures specified in the NICE scope. Outcomes were either defined in the CS or in the statistical analysis plan (Doc B ref. 70). The results presented are from an interim analysis with median follow-up of 26.9 months. Key efficacy results are time-to-event data presented in the CS predominantly in terms of events and proportions (n/N, %) alongside medians and hazard ratios (HRs) (with the associated 95% confidence intervals). Kaplan-Meier (K-M) survival curves are presented for PFS, OS and TTD and (except for TTD) these show the number of patients at risk at each time interval. Response rates are presented as proportions and percentages (n/N, %) alongside odds ratios (with associated 95% confidence intervals). P-values are presented for comparisons between groups.

*Sample size & statistical procedures used*

The primary outcome in the CASTOR RCT was PFS. The power calculation is reported in the CS (Table 10 and section B.2.4.1). Approximately 480 participants (randomised 1:1, i.e. 240 per group) were required to achieve 295 PFS events which would provide 85% power to detect a reduction of 30% in the risk of either progression or death (i.e. HR for DBd vs Bd of 0.70) with a log-rank test (two-sided alpha=0.05). The ERG notes that the required sample size was achieved (DBd N = 251; Bd N = 247) with 480 participants receiving study treatment (DBd n = 243; Bd n = 237). The statistical analysis plan states that the primary analysis was a stratified log-rank test for comparison of the PFS distribution in the ITT population between DBd and Bd. The hazard ratio and its two-sided 95% confidence interval was estimated using a stratified Cox regression model with treatment as the only explanatory variable. Three stratification factors were used in the analyses: ISS staging (I, II, and III); number of prior lines of therapy (1 vs 2 or 3 vs >3); and prior bortezomib treatment (no vs yes).

Long-term follow-up is to continue for OS until 320 deaths have been observed at which point the study will have approximately 80% power to detect a 27% reduction in the risk of death (HR

for DBd vs Bd of 0.73) with a log-rank test (two-sided alpha=0.05). This calculation takes into consideration an annual dropout rate of 5%.

The statistical analysis plan specified that if the testing of the primary endpoint, PFS, were statistically significant then secondary endpoints would be sequentially tested (using a hierarchical testing approach to control for Type I error rate) in the following order:

- TTP
- Rate of VGPR or better
- Overall response rate
- MRD negative rate
- OS.

*Analysis populations*

The CS reports on three patient analysis population: the ITT population, the response-evaluable population and the safety population.

The ITT population includes all randomised trial participants. The analyses of PFS (primary endpoint) and all other time-to-event efficacy endpoints are based on this population (CS section B.2.4.1). Based on information reported in the CS the ERG agrees believes that ITT analysis was used for the time-to-event outcomes. Although not explicitly stated it appears, from CS Appendix sections D.3.2.10 and D.3.2.11 that and HRQoL outcomes were also analysed using the ITT population (this is discussed further below in the context of the handling of missing data).

The response-evaluable population is defined as trial participants who have a confirmed diagnosis of multiple myeloma and measurable disease (defined in CS Table 6) at baseline or the screening visit, who received at least one administration of study drug, and who have had at least one post-baseline disease assessment (CS section B.2.4.1). The analyses of the major secondary endpoints of ORR, rate of VGPR or better and the duration of and time to response outcomes are based on the response-evaluable population.

The safety population is not defined in the CS but the Statistical Analysis Plan[27] states that the safety population comprised participants who received at least one administration of any study

treatment (partial or complete). Safety analysis grouping was according to treatment actually received.

*Censoring and missing data*

Censoring rules were the same for PFS and TTP and these are presented in (CS Table 10):

- if subsequent anticancer therapies for MM were started without disease progression, participants were censored at the last disease assessment before the start of subsequent therapies.
- if patients withdrew consent before disease progression they were censored at the last disease assessment before withdrawal of consent.
- if patients were lost to follow-up they were censored at the last disease assessment before being lost to follow-up
- if disease progression had not occurred and the patient was still alive at the cut-off date for assessment they were censored at the last disease assessment
- if the patient had no post-baseline disease assessment they were censored at randomisation.

For OS, data were censored at the date the subject was last known to be alive if vital status was unknown or the patient was alive at the time of data analysis.

The CS states that, unless otherwise specified, no data imputation has been applied for missing safety and efficacy evaluations, and the CS does not mention imputation further. However, HRQoL results presented in CS Appendix D Tables 28 and 29 appear to have incorporated imputation for missing data under a missing at random assumption (the table captions refer to the analysis being by ITT and with data missing at random, which implies the missing values must have been imputed), but no method of imputation is described. In response to a clarification request the company has provided sensitivity analyses based on multiple imputation for EORTC QLC-C30 scores (clarification A22b) and for EQ-5D scores (clarification A23b) to test the assumption that these HRQoL data were missing at random. However, the clarification response does not explain the base case imputation approach used to derive the HRQoL results reported in CS Appendix D Tables 28 and 29. Furthermore, the description of the sensitivity analyses in the clarification responses is superficial and a clear rationale for the choice of the sensitivity analysis scenario is not given. We agree that multiple imputation may reduce the risk of bias compared to a complete-cases analysis, but this depends on the amount of data missing, the reasons for the data being missing, and on the use of an appropriate

analysis method.[28] We also agree that sensitivity analysis on the missing at random assumption is good practice. Ideally, a multiple imputation analysis should report the number of observations which are missing in each study group, the reasons why the data are missing, any differences in the characteristics of missing and non-missing subjects, and a comparison of the results of the multiple imputation against a complete cases analysis.[28 29] This information is either not provided, or unclear, in the CS and clarification responses. The ERG is therefore uncertain whether the analyses of HRQoL presented by the company in CS Appendix sections D.3.2.10 and D.3.2.11 account appropriately for the missing data and, hence, whether they are unbiased.

*Adjustment of OS in consideration of post-progression therapies not available in England.*
As described earlier (cross ref to appropriate section) the CASTOR trial was an international trial but there were no trial centres in the UK. Therefore some participants in the trial received treatments, after disease-progression, that they would not have had access to if they had been located in England. Additionally, because patients in the Bd arm experienced earlier disease progression than patients in the DBd arm a higher proportion of Bd treated patients received post-progression therapies (43% Bd versus 17% DBd).

A number of adjustment methods are available that aim to reduce the bias that can be introduced to the effect estimate for OS by treatment switching in cancer trials. These have been reviewed in NICE Technical Support Document 16.[30] We note that treatment switching of interest in cancer trials is often specifically from a control arm to an active treatment arm[30] whereas the company's current application of an adjustment method adjusts for switching in both active treatment arms (i.e. DBd and Bd) to address the problem that some patients in each arm received a post-progression therapy unavailable in England.

The method used for adjustment by the company is the Inverse Probability of Censoring Weights (IPCW), as described in CS Appendix D.3.2.14. The IPCW method is one of four possible "complex" adjustment approaches that could be used to adjust for treatment switching, each of which has different strengths and limitations.[30] The company justifies their selection of IPCW on the grounds that the three other approaches [Rank Preserving Structure Failure Time Models (RPSFTM), Iterative Parameter Estimation (IPE), and Two-stage methods] would not be suitable for use in this case, for several reasons which are clearly stated in CS Appendix section D.3.2.14.1 and are consistent with criteria suggested by the NICE Technical Support

Document.[30] The ERG agrees with the company's choice of the IPCW methods to adjust for treatment switching and we note that the IPCW method has been applied to adjust for switching from a control arm to an active treatment arm in NICE appraisals TA215, TA377 and TA432.

The IPCW method as applied by the company involves censoring those patients who switched to a post-progression treatment that would not have been available in England, at the point of the switch. However, this censoring is potentially informative (i.e. the patients whose data are censored may differ from other patients in an informative way). To address this, the observations from the remaining patients (i.e. those who switched to a post-progression treatment that could be received in England and those who had not yet switched) are weighted such that they effectively account for the patients with similar characteristics who have been censored. The aim of the weighting is to remove any censoring-related selection bias. We assume that patients who had not yet experienced disease progression or who had not yet started a post-progression therapy would not have been weighted to account for any censored patients, given that their prognostic characteristics would be different, although this is not explicit in the CS. The weights have not been provided in the submission and therefore we have not been able to check their plausibility.

The main steps of the IPCW adjustment method were: reformatting of individual patient data from the CASTOR trial to create panel data (one record per individual per time interval); estimation of the time-dependent weights using two logistic regression models; and then running Cox regression analyses on the censored panel data to estimate the IPCW hazard ratio for OS. The estimation of the weights for the IPCW is described briefly in CS Appendix D.3.2.14.2, and CS Appendix D Table 34 lists 17 baseline covariates and 12 time varying covariates which were used in the two logistic regression models that estimated the probability of each patient switching. Although the methods reported by the company appear generally appropriate, the description of the methods is rather brief and general, and no examples of the input/output data at each of the analysis steps have been provided. The ERG is therefore unable to comment on whether the methods have been applied correctly.

In common with other statistical methods of adjustment, the IPCW method involves some assumptions and has some limitations.[30] A key assumption for the IPCW method is that there are no unmeasured confounders. The CS states that the validity of the IPCW method relies on

this assumption and states that baseline characteristics and time varying covariates were identified from data collected in the trial and from expert clinical opinion.

For logistic regression model 1 the full model included all the baseline and time-varying covariates listed in CS Appendix D Table 34. A reduced model 1 was derived using a stepwise variable selection method with a significance level of 15%. This stepwise variable selection method is described in more detail in the response to clarification question A19. The ERG believes that the IPCW-adjusted hazard ratios for OS were obtained from a final reduced model which included just two covariates (baseline ISS staging and race, p.74 of CS appendix D). Although not explicitly stated in the CS, we assume that the other variables in the model were deleted because they had not produced an F statistic at the predefined significance level of 15% (as described in the response to clarification question A19). We consider that it is highly likely that there are unmeasured confounders.

*Subgroups*

Subgroup analyses are reported in CS section B.2.7. However, it should be noted that the CS focuses on the subgroup of 2L patients and hence only the results for this subgroup are reported in the main body of the CS. Subgroups based on the number of previous lines of therapy are listed in the NICE scope under 'Other considerations'. Outcomes from other subgroup analyses are reported in CS Appendix E.

Pre-planned subgroup analyses were conducted for the outcomes of PFS, TTP, ORR and rate of VGPR or better. Analysis of OS by subgroup is not planned until final analysis, but an analysis based on the data from the 26.9 month median follow-up is presented for 2L patients as per the company's decision problem. A subgroup analysis of data for 2L patients is also presented in CS Section 2.7.7 for the post-hoc TTD outcome which informs the economic model.

Three of the subgroups were randomisation stratification factors in the CASTOR trial. These were ISS disease stage (stage I, II, or II), the number of previous lines of therapy (1 vs. 2 or 3 vs. >3), and previous treatment with bortezomib (no vs. yes).

In response to clarification question A25 the company confirmed that type 1 error rate was not controlled for in subgroup analyses.

**Summary the company's approach to trial statistics**

In summary, the statistical procedures used in the CASTOR trial are clearly reported in the CS or in the cited Statistical Analysis Plan and are appropriate for the evaluation of a cancer treatment. The randomised sample size is adequate and the ITT population is appropriately defined and used for the efficacy analyses. However the ERG notes that the CS presents evidence from an interim analysis with median follow-up of 26.9 months. The IPCW method used for adjusting for bias in treatment switching appears appropriate (subject to the assumption that there are no unmeasured confounders being met) and has been adequately justified by the company; however, the method is reported rather generally and does not provide sufficient information for the ERG to be certain that the methods were applied correctly.

### 3.1.7 Description and critique of the company's approach to the evidence synthesis

A narrative review of the evidence from the CASTOR[18] RCT is presented in CS Section B.2. As there was only one RCT of DBd versus Bd there is no direct meta-analysis.

There were no RCTs that compared DBd with the other possible comparators in 2L RRMM patients, Cd or combination chemotherapy. The company therefore conducted an NMA to perform an indirect treatment comparison. The company's systematic review identified a further two RCTs that were considered for inclusion in the NMA. One, the ENDEAVOR trial, compared Cd versus Bd and the other, an RCT conducted by Phillips and colleagues[1] compared the chemotherapy regimens of vincristine, doxorubicin, and dexamethasone (VAd) versus mitozantrone, vincristine and dexamethasone (MOD). The CS states that the combination chemotherapy regimens examined by Phillips and colleagues[1] are not currently used for the treatment of RRMM (the study having been carried out between 1986 and 1992) and in the absence of a common comparator, it was not feasible to connect this study to the other two (CASTOR and ENDEAVOR) in the NMA. The ERG agrees with the exclusion of the Phillips and colleagues study. Consequently NMA using a Bayesian framework was carried out to enable a comparison to DBd versus Cd using the CASTOR and ENDEAVOR trials. The ERG notes that the company's search identified two trials of Bd vs cyclophosphamide + Bd which were excluded by the company but not commented on; however the ERG agrees that neither of these studies provided evidence that could have been included in the NMA.

The NMA followed a Bayesian approach, as described by the company in CS Appendix D, section D.3.3. With only two trials in the network a fixed-effect model was fitted which the ERG agrees is appropriate.

CS B.2.9.1 states that "*CASTOR and ENDEAVOR were considered sufficiently comparable for analysis*". However, there are some key differences between the trials (Table 10) and the CS does not explicitly state whether the assumption of similarity has been assessed. Therefore it is not clear whether the assumption of similarity has been met.

**Table 10 Comparative summary of key differences between trial methodologies**

| **Trial number** | CASTOR | ENDEAVOR |
|---|---|---|
| **Eligibility criteria for participants** | 1. Subject must have received at least 1 prior line of therapy with no upper limit<br>2. Excluded patients refractory to bortezomib<br>3. Bortezomib administered subcutaneous<br>4. Bd treatment duration limited until disease progression, unacceptable toxicity or up to eight cycles | 1. One to three previous treatments<br>2. Requirement for patients to have a left ventricular ejection fraction of at least 40%<br>3. Requirement for patients to have creatinine clearance of at least 15 mL/min.<br>4. Bortezomib administered intravenous or subcutaneous<br>5. Bd treatment duration limited until disease progression or unacceptable toxicity with no upper limit on the number of cycles |
| **Participant characteristics** | 5. 66% of patients with prior exposure to bortezomib | 6. 54% patients with prior exposure to bortezomib |

Source: Reproduction of CS Table 18 with minor amendments to formatting

Patient characteristics for the two trials are compared with key differences tabulated (Appendix D.2.4 Table 17). The ERG has reorganised this table to aid comparison between the two trials (Table 11) but notes that these characteristics are for the whole trial population, characteristics for the 2L subgroup from both trials are not reported. The CS states (p.65) that baseline characteristics were similar for age, sex, high cytogenetic risk, Eastern Cooperative Oncology

Group (ECOG) performance status and ISS Stage. Differences in the patient eligibility criteria regarding creatinine clearance and LVEF were not expected to have an impact on the comparison of trials. The potential impact of differences in bortezomib administration is described and the indirect comparison is conducted despite this. The company states these differences are likely to overestimate the efficacy of DBd compared with Cd. For each comparison only one trial was available and therefore issues of homogeneity and consistency were not applicable in this case.

**Table 11 Characteristics of participants in the studies across treatment groups**

| | CASTOR (N=498) | | ENDEAVOR (N=929) | |
|---|---|---|---|---|
| | DBd (N=251) | Bd (N=247) | Cd (N=464) | Bd (N=465) |
| Median age, years (range) | 64.0 (30–88) | 64.0 (33–85) | 65 (35–89) | 65 (30–88) |
| Male sex, % | 54.6% | 59.5% | 52% | 49% |
| ECOG Performance Status | 0: 42.4% | 0: 47.0% | 0: 48% | 0: 50% |
| | 1: 52.5% | 1: 45.3% | 1: 45% | 1: 44% |
| | 2: 5.2% | 2: 7.7% | 2: 7% | 2: 6% |
| | >2: 0% | >2: 0 | | |
| ISS Stage | I: 39.0%<br>II: 37.5%<br>III: 23.5% | I: 38.9%<br>II: 40.5%<br>III: 20.6% | I: 44%<br>II–III: 56% | I: 44%<br>II–III: 56% |
| Median number of prior lines of treatment (range) | 2 (1–9) | 2 (1–10) | 2 (1-2) | 2 (1-2) |
| Number of prior lines of treatment | 1: 48.6%<br>2: 27.9%<br>3: 14.7%<br>>3: 8.8% | 1: 45.7%<br>2: 30.0%<br>3: 13.0%<br>>3: 11.3% | 1: 50%<br>2: 34%<br>3: 16% | 1: 50%<br>2: 31%<br>3: 19% |

Source: CS Appendix D Table 17

The company conducted a sensitivity analysis for their NMA (CS section B.2.9.5) for the subgroup of 2LBN patients within the subgroup of 2L patients. This sensitivity analysis was conducted because the NICE recommendation for Cd is in 2LBN patients only. For this sensitivity analysis OS data from the CASTOR study were not adjusted for the use of post-progression therapies not available in England. Although not explicitly stated in the CS the ERG

assumes that this is because data were not available to make a similar adjustment for the ENDEAVOR study. The ERG agrees that it was appropriate to conduct this sensitivity analysis.

The ERG has used the indirect treatment comparison (ITC) software available from the Canadian Agency for Drugs and Technologies in Health (CADTH)[31] (which uses the Bucher method and extensions of this method) as a rapid check of some of the company's NMA outcomes and we obtained almost identical results.

## 3.2 Summary statement of company's approach

Table 12 provides the ERG's quality assessment appraisal of the company's systematic review of clinical effectiveness. As the table shows, the systematic review meets four of the five criteria of a good quality systematic review, and partly meets the remaining criterion. The CASTOR trial is generally appropriately summarised in the CS, but we note that the HRQoL and AE results are provided for the whole trial population rather than the 2L subgroup that is relevant to the decision problem.

**Table 12 Quality assessment (CRD criteria) of the company's clinical effectiveness review**

| CRD quality item | ERG comments |
|---|---|
| 1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question? | Yes. Broad inclusion and exclusion criteria are reported, to identify RCTs comparing the 2L therapies specified in the NICE scope. Although the eligibility criteria are wider than the decision problem (see section 3.1.2 of this report), they address the company's review questions about the efficacy and safety of daratumumab and relevant comparators in RCTs involving patients with RRMM. |
| 2. Is there evidence of a substantial effort to search for all relevant research? i.e. all studies identified | Yes (please see section 3.1.1 of this report for our critique of the company's searches). A wide range of electronic databases and other sources was searched. |
| 3. Is the validity of included studies adequately assessed? | Yes. The quality of the CASTOR trial was assessed using standard NICE criteria (CS Section B.2.5.1). However, the company only provided a one-word overall rating of "*moderate*" quality for the ENDEAVOR trial, without explaining how this was derived (CS Appendix D Section D.3.4). The company subsequently provided on request a full risk of bias assessment for ENDEAVOR (clarification response A21). |
| 4. Is sufficient detail of the individual studies presented? | Yes. Sufficient information is provided about the CASTOR and ENDEAVOR trials' methodology and statistical analyses. |
| 5. Are the primary studies summarised appropriately? | Partly. Results of the studies are on the whole appropriately summarised, but we note HRQoL and AE results are provided for the whole trial population rather than for the 2L subgroup relevant to the decision problem. The CS gives |

| | |
|---|---|
| | limited patient baseline characteristics from the ENDEAVOR trial and does not provide baseline characteristics for the 2L subgroups from either the CASTOR or ENDEAVOR trials (patient characteristics of the whole ENDEAVOR trial population and the 2L subgroup of the CASTOR trial population were subsequently provided in clarification responses A17 and A6). Patient characteristics for the 2L subgroup in ENDEAVOR were not available. |

The processes for inclusion and exclusion screening are detailed in CS Appendix D (Sections D.1.2.1 and D.1.2.2). Two independent reviewers conducted title and abstract screening. One reviewer conducted full text screening, with a second reviewer (not specified whether they were independent) checking that excluded studies should have been excluded, and a third reviewer was consulted if there were disagreements. Two independent reviewers conducted data extraction (as is recommended in systematic reviewing[32]), with an additional final check of extracted data (detailed in CS Appendix D, D.1.3).

**Relevance of evidence to the company's decision problem**

We consider that the submitted evidence does not fully reflect the company's decision problem. This is because the decision problem focuses on 2L patients, whereas the CASTOR trial[18] included patients with RRMM who had received at least one prior line of therapy (i.e. patients receiving second and subsequent lines of therapy). Relevant data from the CASTOR trial for the 2L population are only available in the CS through pre-specified subgroup analyses. This approach means the statistical analyses are based on smaller sample sizes, which increases uncertainty around the treatment effect estimate and lowers statistical power. The CS provides results for 2L subgroup for some of the outcomes specified in the decision problem, but not for HRQoL or AEs. The submitted evidence also does not reflect the NICE scope or the NHS position, because 2L patients' receipt of DBd and Bd in the CASTOR trial was not restricted by the prior therapy they had received. That is, the trial included both bortezomib naïve and experienced patients when NHS England guidance restricts use of bortezomib at 2L treatment to bortezomib naïve patients.

**Summary**

Overall, the company's systematic review is of a good quality, and there is a low risk of systematic error based on the methods described in the CS. However, the submitted evidence only addresses the company's decision problem focus on 2L patients through subgroup analyses that are at risk of being statistically under-powered. The evidence also does not fully

reflect NICE's final scope or clinical practice guidance in England, because the CASTOR trial included both bortezomib experienced and naïve patients whereas bortezomib-based treatments are only recommended for patients who have not received first-line bortezomib-based treatment.

## 3.3 Summary of submitted evidence

In this section the ERG focuses on the population that matches the decision problem (i.e. the focus is on 2L patients) and the outcomes of the CASTOR trial presented in the CS that match the decision problem and feed into the economic model (with cross-references provided to the economic section of the ERG report). These outcomes are PFS, OS and TTD. CS data have been supplemented with data from other sources (e.g. trial publications and clinical study reports) if necessary. The outcomes from the NMA are also included in this section. AE data, some of which feeds into the model, are presented in section 3.3.7 of this ERG report. EQ-5D-5L data, which also feed into the model, are presented in section 3.3.5.

Outcomes presented in the CS that do not feed into the economic model are summarised in section 3.3.4.

### 3.3.1 Summary of results for PFS (Primary endpoint)

The CS presents the PFS results for the CASTOR trial which was the primary outcome for this trial (subgroup of 2L patients CS B.2.7.1 and whole trial CS B.2.6.1). The whole trial analysis presented in the CS was undertaken when 362 progression events had occurred at a median follow-up of 26.9 months (the power calculation is based on 295 PFS events). The proportion of PFS events occurring in the DBd arm was lower than that in the Bd arm in for both the whole trial and the subgroup of 2L patients which is the focus of the appraisal (Table 13 and Figure 1).

For 2L patients median PFS was approximately 18 months longer in the DBd arm than the Bd arm. The improvement in PFS with DBd was statistically significant [HR = 0.23, 95% confidence interval (CI) 0.16 to 0.33, p<0.0001].

**Table 13 PFS results for the CASTOR trial**

| Parameter | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n=122) | Bd (n=113) | DBd (N=251) | Bd (N=247) |
| Events, n/N (%) | 60/122 (49.2) | 94/113 (83.2) | 158/251 (62.9) | 204/247 (82.6) |
| Median PFS (95% CI), months | 26.22 (21.19, NE) | 7.92 (6.77, 9.03) | 16.72 (13.14, 19.38) | 7.06 (6.21, 7.66) |
| HR, (95% CI) p-value | 0.23 (0.16, 0.33) p<0.0001 | | 0.32 (0.25, 0.40) p<0.0001 | |

Source: Parts of CS Table 13 and CS Table 15

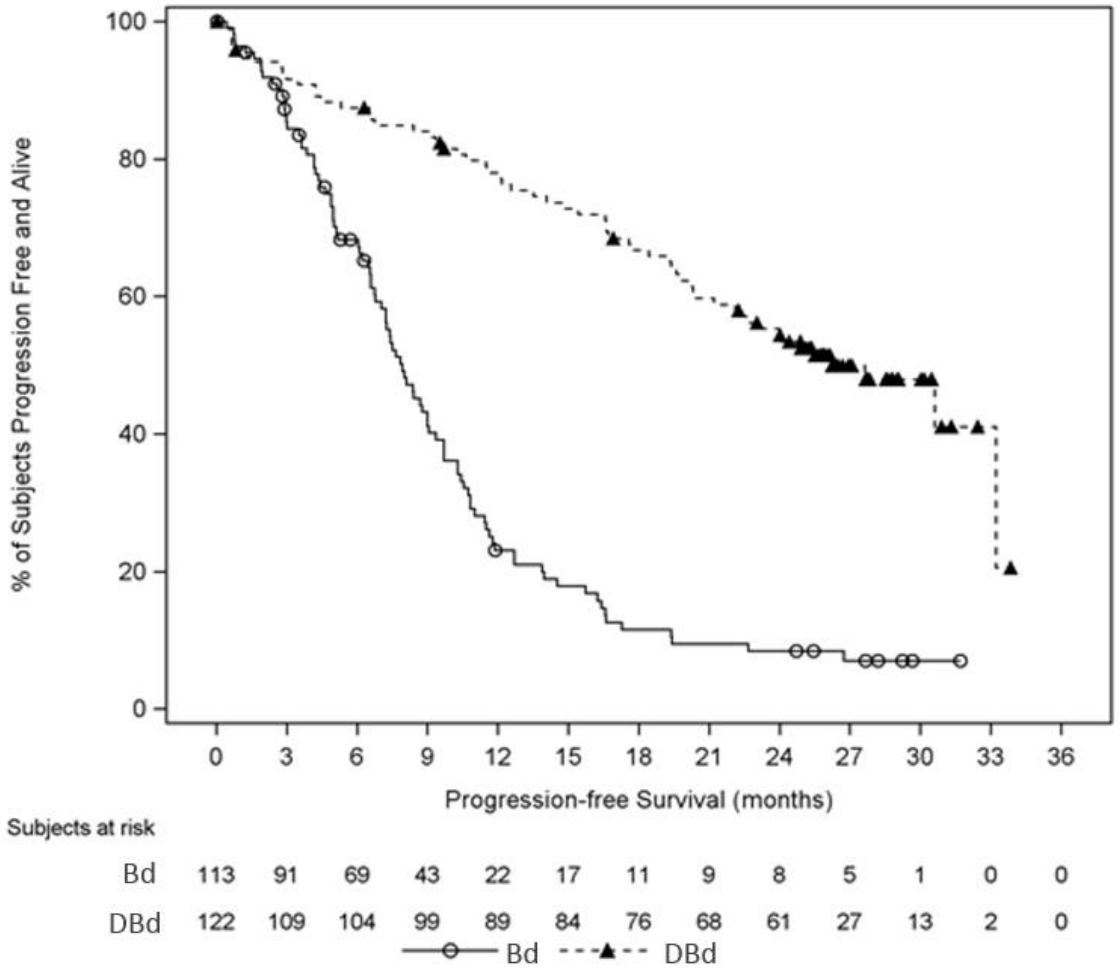



**Figure 1 Kaplan-Meier plot of PFS among 2L patients treated with DBd compared with Bd in the CASTOR trial (median follow-up 26.9 months, CS Figure 14)**

As described earlier (ERG report section 3.1.7) the company conducted an indirect comparison to compare DBd, with Bd and Cd. As shown in Table 14, for both the subgroup of 2L patients and for the total trial population the hazard ratios were in favour of DBd and the probability that DBd was the best treatment was 100%.

**Table 14 NMA results for PFS**

| Comparison | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | **HR (95% CrI)** | **Probability[a]** | **HR (95% CrI)** | **Probability[a]** |
| DBd vs Bd | 0.21 (0.15, 0.30) | 100% | 0.32 (0.25, 0.40) | 100% |
| DBd vs Cd | 0.47 (0.29, 0.75) | 99.9% | 0.60 (0.45, 0.81) | 100% |

a Probability of DBd being better than the comparator

Source: CS Appendix D Table 37 and CS Appendix D Figure 23

The company conducted an NMA sensitivity analysis for the subgroup of 2LBN patients in CASTOR and ENDEAVOR. In the CASTOR trial stratification factors at randomisation included both number of prior lines of treatment (1 versus 2 or 3 versus >3) and prior treatment with bortezomib (no versus yes). In the ENDEAVOR trial stratification factors included number of prior lines of treatment (1 versus 2 or 3) but prior treatment with bortezomib was not a stratification factor.[19] Creating a subgroup of 2LBN patients involves breaking randomisation for both trials. In the ERG's view the stratification of CASTOR by prior lines of treatment and prior treatment with bortezomib may lessen the impact of breaking randomisation. However, as the ENDEAVOR trial was not stratified by prior treatment with bortezomib the risk of bias being introduced between the trial arms may be greater for this trial.

Results of the sensitivity analysis are similar to the analyses of all 2L patients but the hazard ratios are slightly higher (so slightly less advantage for DBd versus the comparator) and the 95% credible intervals (CrIs) wider (Table 15). The hazard ratios were in favour of DBd and in comparison to both Bd and Cd, DBd has the greatest probability of being the better treatment. However, for the comparison with Cd, the 95% CrI now crosses one.

**Table 15 NMA sensitivity analysis for PFS in 2L bortezomib naive patients**

| Comparison | HR (95% CrI) | Probability of DBd being better than the comparator |
|---|---|---|
| DBd vs Bd | 0.23 (0.14, 0.38) | 100% |
| DBd vs Cd | 0.52 (0.27, 1.02) | 97.2% |

Source: Part of CS Table 24

### 3.3.2 Summary of results for OS (secondary outcome)

OS is a secondary outcome of the CASTOR trial and the OS data are currently immature. The final OS analysis will occur after 320 deaths have been observed whereas, at the analysis after a median follow-up of 26.9 months there had been 179 deaths. As a consequence of the immaturity of the data a median OS for either study arm of the subgroup of 2L patients is not yet available (Table 16). It can be observed however that the proportion of deaths in the DBd arm is lower than in the Bd arm (20.5% vs 35.4% respectively), and a statistically significant difference between survival in the study arms in favour of DBd is reported (HR 0.50, 95% CI 0.30 to 0.84, p=0.008) (Table 16 and Figure 2). In the total trial population immaturity of the data also precludes calculation of a median OS for the DBd arm. Although a difference in the proportion of deaths in favour of the DBd group is reported (DBd 32.7% vs Bd 39.3%), as Table 16 shows, the improvement in OS for the total trial population is not statistically significant (HR 0.77, 95% CI 0.57 to 1.04, p=0.0884).

**Table 16 OS results for the CASTOR trial**

| Parameter | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n=122) | Bd (n=113) | DBd (N=251) | Bd (N=247) |
| Events, n/N (%) | 25/122 (20.5) | 40/113 (35.4) | 82/251 (32.7) | 97/247 (39.3) |
| Median OS (95% CI), months | NE (NE, NE) | NE (28.85, NE) | NE (NE, NE) | 31.18 (28.85, NE) |
| HR, (95% CI) p-value | 0.50 (0.30, 0.84) p=0.008 | | 0.77 (0.57, 1.04) p=0.0884 | |

NE: not estimable

Source: Parts of CS Table 13 and CS Table 15

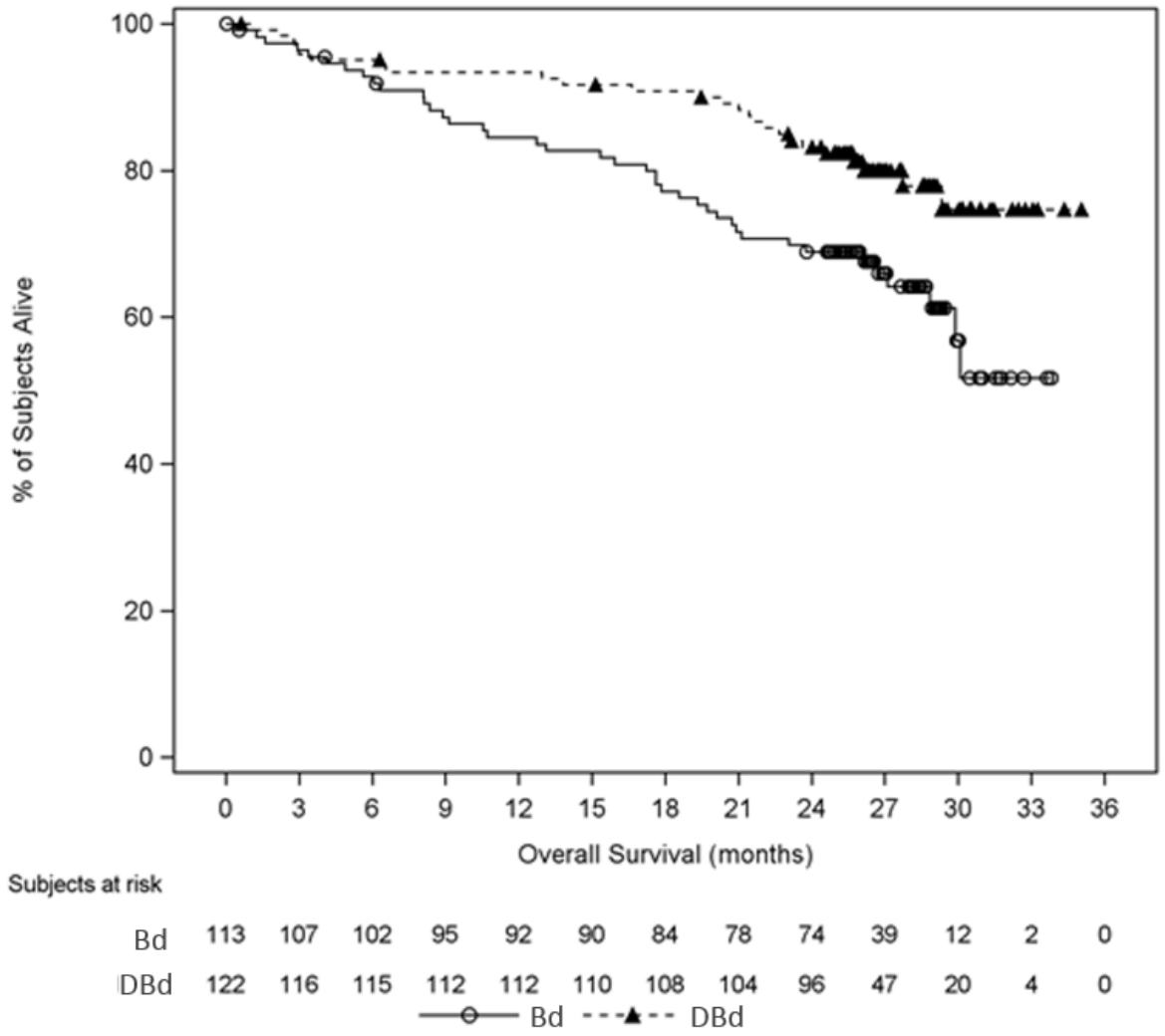



**Figure 2 Kaplan-Meier plot of OS among 2L patients treated with DBd compared with Bd in the CASTOR trial (median follow-up 26.9 months, CS Figure 16)**

The CS states the indirect comparison to compare DBd, with Bd and Cd indicated that DBd had the highest probability of being the best treatment (Table 17). However, the 95% CrIs for the hazard ratios were wide and crossed 1 for DBd vs Cd in 2L patients, as well as DBd vs Bd and DBd vs Cd in the total trial population. This reflects greater uncertainty about where the true HR for OS lies, which is unsurprising given the immature nature of the OS data. The company do not comment in detail on the OS NMA results but the ERG notes that for the DBd vs Cd comparison in the subgroup of 2L patients the HR is 0.60 with a probability of DBd being the best treatment of 95%, whereas in the total trial population the HR is close to 1.00 at 0.97 and the probability of DBd being the best treatment has dropped to 55.8%.

**Table 17 NMA results for OS**

| Comparison | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | HR (95% CrI) | Probability[a] | HR (95% CrI) | Probability[a] |
| DBd vs Bd | 0.50 (0.30, 0.84) | 99.6% | 0.77 (0.57, 1.04) | 95.6% |
| DBd vs Cd | 0.60 (0.33, 1.10) | 95.0% | 0.97 (0.68, 1.39) | 55.8% |

a Probability of DBd being better than the comparator

Source: Combines data from CS Appendix D Table 39 and CS Appendix D Figure 24

The company conducted an NMA sensitivity analysis for the subgroup of 2LBN patients in CASTOR and ENDEAVOR in the same way as reported for PFS (Section 3.3.1).

Results of the OS sensitivity analysis are similar to the analyses of all 2L patients but, similarly to the PFS sensitivity analysis, the hazard ratios are slightly higher (so slightly less advantage for DBd versus the comparator) and the 95% CrIs wider (Table 18). The 95% CrI crosses 1 for both the DBd vs Bd and the DBd vs Cd comparisons. Nevertheless the hazard ratios are in favour of DBd and DBd has the greatest probability of being the better treatment.

**Table 18 NMA sensitivity analysis for OS in 2L bortezomib naive patients**

| Comparison | HR (95% CrI) | Probability of DBd being better than the comparator |
|---|---|---|
| DBd vs Bd | 0.53 (0.24, 1.17) | 94.2% |
| DBd vs Cd | 0.66 (0.22, 1.98) | 76.8% |

Source: Part of CS Table 24

As described in CS Section B.2.5.2 some participants in the CASTOR trial received therapies not available in the England when they experienced disease progression and started their next line of therapy. To reduce bias in the effect estimate for OS caused by the use of therapies not available in England and bias caused by the greater proportion of patients in the Bd arm of the trial who received a post-progression therapy, the company adjusted the OS data using the IPCW method. The effect of the adjustment was [redacted] in the HR for OS [redacted] in comparison to the unadjusted value. In the subgroup of 2L patients the adjusted HR is [redacted] (95% CI [redacted], [redacted]) representing a [redacted] reduction in the risk of death for the DBd arm in comparison to the Bd arm (Table 19) whereas the unadjusted HR reported above represents a 50% reduction in the risk of death for DBd versus Bd. In response to clarification question A1 the company provided a Kaplan-Meier plot showing the observed and adjusted OS

curves for 2L patients on the same plot ([redacted]). The OS estimate in the subgroup of 2L patients adjusted for the use of subsequent therapies not available in England was used in the base case of the economic model.

**Table 19 OS adjusted for subsequent treatments not available in England for the CASTOR trial**

| Parameter | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n not reported) | Bd (n not reported) | DBd (N not reported) | Bd (N not reported) |
| HR (95% CI) | [redacted] ([redacted], [redacted]) | | [redacted] ([redacted], [redacted]) | |

Source: CS sections B.2.6.9 and B.2.7.6

Source: clarification response A1

**Figure 3 Kaplan-Meier plots of OS for observed and adjusted data in 2L patients treated with DBd compared with Bd in the CASTOR trial (median follow-up 26.9 months)**

An NMA based on OS adjusted for therapies received after progression that are not available in England could not be conducted because these data are not available from the ENDEAVOR trial of Cd vs Bd.

### 3.3.3 Summary of results for time to treatment discontinuation (post-hoc analysis)

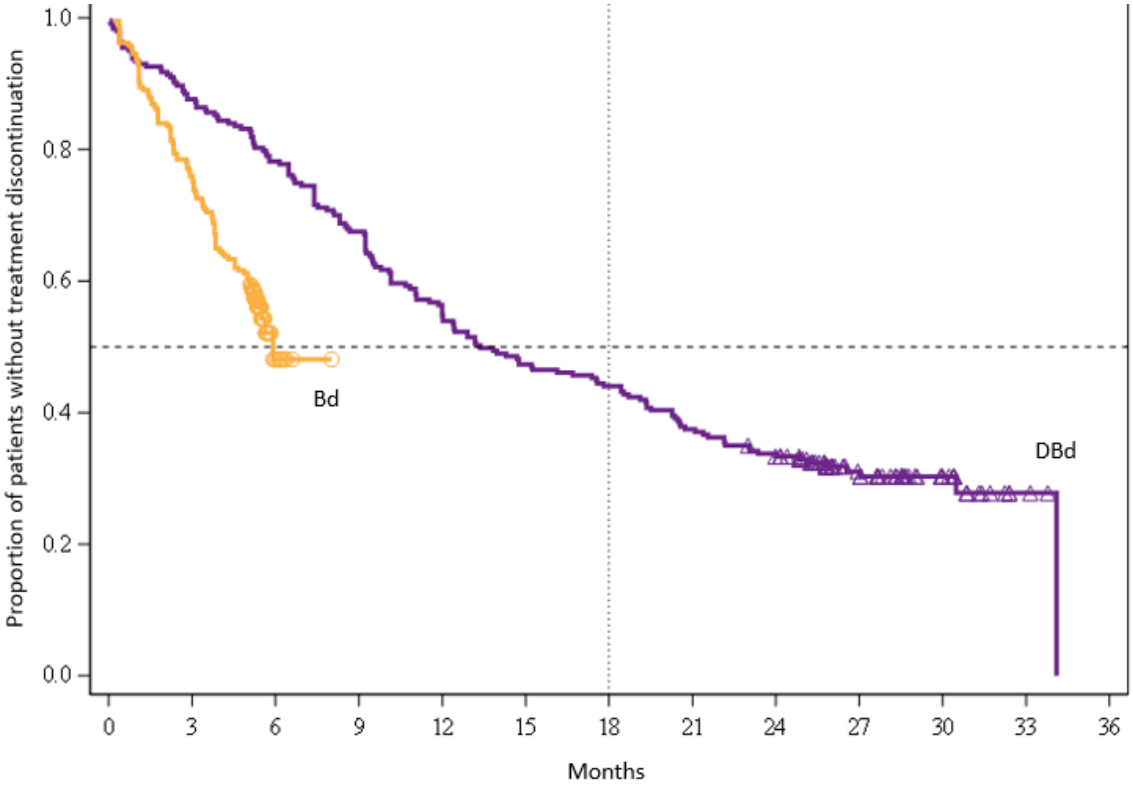
TTD was a post-hoc analysis conducted at 26.9 months follow-up. It should be noted that Bd treatment (either as part of DBd or as the comparator treatment) has a defined endpoint (after eight 21-day cycles) whereas the daratumumab component of DBd was administered until disease progression or unacceptable toxicity. In 2L patients the CS reports that DBd treatment was associated with a 56% reduction in the risk of treatment discontinuation compared with Bd (HR 0.44, 95% CI 0.26 to 0.76, p<0.0031) (Table 20 and Figure 4).

**Table 20 TTD results for the CASTOR trial (post-hoc analysis)**

| Parameter | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n=122) | Bd (n=113) | DBd (N=251) | Bd (N=247) |
| HR, (95% CI) p-value. Data from CS[a] | 0.44 (0.26, 0.76) p<0.0031 | | 0.40 (0.29, 0.57) p= not reported | |
| | | | | |
| Events, n/N (%)[b] | 67/119 | 41/111 | 169/243 | 104/237 |
| Median TTD[b] (95% CI), months | 23.98 | Not evaluable | 13.37 | 5.91 |
| HR, (95% CI)[b] p-value | 0.407 (0.240, 0.691) p = 0.0009 | | 0.385 (0.275, 0.540) p < 0.0001 | |

[a] The CS cites reference 66[33] as the source of these TTD data but data in the cited reference are slightly different and are shown in the bottom section of the table.

[b] Data from the reference cited by the CS.[33] Although not explicitly stated, from the number of patients included in the analyses, the data in the reference appear to be from analyses of patients who received the study treatment.




Source: CS Figure 17

**Figure 4 Kaplan-Meier plots for time to treatment discontinuation in 2L patients treated with DBd compared with Bd in the CASTOR trial (median follow-up 26.9 months)**

An indirect comparison to compare DBd with Cd could not be undertaken for the TTD outcome because TTD was not reported for the ENDEAVOR RCT.

### 3.3.4 Summary of other secondary outcomes reported for the CASTOR trial

Secondary outcomes reported in the CS that were not used in the economic model were: MRD-negative rate; outcomes for response to treatment and duration of and time to response; time to disease progression (TTP); use of subsequent treatment; time to subsequent anticancer therapy; PFS on the subsequent line of therapy; best M-protein response; and best response to first subsequent anticancer therapy (Table 22 and Table 23).

#### 3.3.4.1 Minimal residual disease

A higher MRD-negative rate was apparent in the DBd group at the 26.9 month follow-up for both the subgroup of 2L patients and the total trial population.

**Table 21 MRD-negative rate results for the CASTOR trial (at median 26.9 months follow-up)**

| MRD-negative rate at $10^{-5}$ threshold | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n=122) | Bd (n=113) | DBd (N=251) | Bd (N=247) |
| n/N | 20/122 | 3/113 | 30/251 | 4/247 |
| % MRD (95% CI) | 16.4 (10.3, 24.2) | 2.7 (0.6, 7.6) | 12.0 (8.2, 16.6) | 1.6 (0.4, 4.1) |
| OR (95% CI), p-value from CS | 7.19 (2.07, 24.92) p = 0.00018 | | 8.25 (2.86, 23.78) p = 0.000001 | |

Source: Parts of CS Tables 13 and 15

### 3.3.4.2 Response outcomes

ORR, sCR, CR and VGPR rates were in favour of DBd in the subgroup of 2L patients and in the whole trial population. Among the subgroup of 2L patients a CR or better was obtained by 43% of the DBd group in comparison to 15% of the Bd group (p<0.0001) and a VGPR or better by 77% of DBd group in comparison to 42% of the Bd group (p< 0.0001) (Table 22).

**Table 22 Response rates results for the CASTOR trial at median follow-up 26.9 months (response-evaluable population)[1]**

| Outcome | Subgroup of 2L patients | | Total trial population | |
|---|---|---|---|---|
| | DBd (n=119) | Bd (n=109) | DBd (n=240) | Bd (n=234) |
| ORR (sCR + CR + VGPR + PR), % | 91.6 (n=109) | 74.3 (n=81) | 84.6 (n=203) | 63.2 (n=148) |
| | OR, 3.97; 95% CI, 1.74, 9.04; p < 0.0007 | | OR, 3.60; 95% CI, 2.24, 5.81; p < 0.0001 | |
| CR or better (sCR + CR), % | 42.9 (n=51) | 14.7 (n=16) | 30.0 (n=72) | 9.8 (n=23) |
| | OR, 4.95; 95% CI, 2.38, 10.27; p < 0.0001 | | OR, 4.67; 95% CI, 2.65, 8.25; p <0.0001 | |
| VGPR or better (sCR + CR + VGPR), % | 76.5 (n=91) | 42.2 (n=46) | 62.9 (n=151) | 29.1 (n=68) |
| | OR 4.19; 95% CI, 2.33, 7.53; p < 0.0001 | | OR, 4.94; 95% CI, 3.23, 7.55; p <0.0001 | |

| | | | | |
|---|---|---|---|---|
| sCR, % | 14.3 | 4.6 | 9.6 (n=23) | 2.6 (n=6) |
| CR, % | 28.6 | 10.1 | 20.4 (n=49) | 7.3 (n=17) |
| VGPR | 33.6 | 27.5 | 32.9 (n=79) | 19.2 (n=45) |
| PR, % | 15.1 | 32.1 | 21.7 (n=52) | 34.2 (n=80) |
| % of patients achieving clinical benefit (ORR + minimal response), % | No analysis of this outcome in this subgroup is presented | | 87.5 (n=210) | 71.8 (n=168) |

Source: CS Tables 13 and 15; CS Figure 15; and CS Appendix D Table 21
[1]Results from the intention-to-treat analyses of these outcomes are reported in CS Appendix D Section D.3.2.4 for the whole trial population. Intention-to-treat analysis results for the 2L patient subgroup are not reported in the CS or CS appendices. Therefore, we have presented the results for the response-evaluable population here. Results for the response-evaluable and ITT populations are comparable (CS Appendix D.3.2.4).

Median time to disease progression was almost 20 months longer in the DBd group (27.63 months versus 8.02 months in the Bd group; HR 0.21, 95% CI 0.15 to 0.32), p<0.0001) (Table 23).

**Table 23 TTP for the CASTOR trial at median follow-up 26.9 months**

| **Parameter** | **Subgroup of 2L patients** | | **Total trial population** | |
|---|---|---|---|---|
| | **DBd** | **Bd** | **DBd** | **Bd** |
| n/N (%) | 56/122 (45.9) | 88/113 (77.9) | 137/251 (54.6) | 192/247 (77.7) |
| Median months (95% CI) | 27.63 (22.11, not evaluable) | 8.02 (7.23, 9.69) | 18.60 (15.80, 21.19) | 7.23 (6.41, 8.02) |
| HR (95% CI), p-value | 0.21 (0.15, 0.32), p<0.0001 | | 0.27 (0.21, 0.35), p<0.0001 | |

Source: Parts of CS Table 13 and CS Table 15

Outcomes where data were not presented separately for the subgroup of 2L patients have not been included in this ERG report (DOR, time to response, time to subsequent anticancer therapy, PFS on the subsequent line of therapy, best M-protein response in serum, best M-protein response in urine, best response to first subsequent anticancer therapy).

### 3.3.5 Summary of Health related quality of life

Health-related quality of life outcomes were not reported separately for the subgroup of 2L patients which is the focus of the company's decision problem. However, for the total trial

population the company states that baseline values for the EQ-5D-5L utility score and VAS score were comparable for the DBd and Bd groups with no significant differences over time for "*most time points*" (CS B.2.6.13). Similarly for the EORTC-QLQ-C30 measure, baseline values for all subscales were comparable for the DBd and Bd groups and there was no significant difference between the two groups at any time point.

Additional information on HRQoL data from CASTOR as used in the economic model can be found in ERG report section 4.3.5.

### 3.3.6 Sub-group analyses results

The CS decision problem already focuses on the subgroup of 2L patients within the CASTOR trial and results for this subgroup have been presented in the sections above. As stated in Section 3.1.6 of this report, subgroup analyses of the whole trial population were pre-planned and conducted for the outcomes of PFS, TTP, ORR and rate of VGPR or better and results of these analyses are reported in CS Appendix E. The subgroups covered a range of 15 demographic, clinical and risk variables (16 for PFS), including the number of prior lines of therapy and prior bortezomib treatment.

As can be observed from CS Appendix E Figures 1 to 5 some of the subgroups were small and consequently had wide 95% CIs which in some cases cross the line of no effect in the forest plots. Overall, however, a consistent effect is seen in favour of DBd across the subgroups tested.

In particular data are presented for the subgroups of patients who had not received prior bortezomib treatment (only patients who have not received prior bortezomib are currently eligible for treatment with Bd or Cd as a 2L therapy) (Table 24). As Table 24 shows, the results consistently favoured DBd across the outcomes presented for this subgroup. The CS does not present data for PFS, TTP, ORR or VGPR or better for 2LBN patients (such data would effectively be subgroup analyses of the existing 2L patient subgroup from CASTOR), aside from the NMA sensitivity analysis for PFS (ERG report section 3.3.1) and OS (ERG report section 3.3.2).

**Table 24 Subgroup analyses by prior bortezomib therapy**

| Outcome | Prior bortezomib | DBd (N=251) | | Bd (N=247) | | Hazard Ratio (95% CI) |
|---|---|---|---|---|---|---|
| | | Event/N | Median (95% CI) | Event/N | Median (95% CI) | |
| PFS | Yes | 116/162 | 12.1 | 135/164 | 6.7 | 0.34 (0.26, 0.45) |
| | No | 42/89 | 27.6 | 69/83 | 7.5 | 0.28 (0.18, 0.43) |
| TTP | Yes | 101/162 | 13.9 (11.5, 17.8) | 126/164 | 7.1 (6.2, 8.1) | 0.29 (0.22, 0.40) |
| | No | 36/89 | 30.6 (23.2, NE) | 66/83 | 7.5 (6.1, 9.4) | 0.24 (0.15, 0.37) |
| | | | | | | |
| **Outcome** | **Prior bortezomib** | **DBd (N=240)** | | **Bd (N=234)** | | **Odds Ratio (95% CI)** |
| | | **n/N** | **% (95% CI)** | **n/N** | **% (95% CI)** | |
| ORR | Yes | 125/154 | 81.2 (74.1, 87.0) | 91/153 | 59.5 (51.3, 67.3) | 3.23 (1.86, 5.63) |
| | No | 78/86 | 90.7 (82.5, 95.9) | 57/81 | 70.4 (59.2, 80.0) | 4.72 (1.85, 12.04) |
| VGPR or better | Yes | 86/154 | 55.8 (47.6, 63.8) | 34/153 | 22.2 (15.9, 29.6) | 4.89 (2.89, 8.27) |
| | No | 65/86 | 75.6 (65.1, 84.2) | 34/81 | 42.0 (31.1, 53.5) | 4.90 (2.39, 10.03) |

Source: CS Appendix E

In addition to the pre-planned subgroup analyses reported above, the CS also reports on the MRD-negative rate in patients defined as high-risk [having at least one high-risk cytogenetic abnormality including del17p, t(4:14) or t(14:16)]. It is presumed that this is a post-hoc subgroup analysis. MRD-negativity in the subgroup of high-risk patients only occurred in the DBd arm of the trial (18.2% at the $10^{-5}$ threshold).

### 3.3.7 Summary of adverse events

The CS presents an overview of the treatment-emergent AEs (TEAEs) for the safety population at the median 26.9 month follow-up. The majority of patients had experienced at least one TEAE (DBd 99.2%, Bd 95.4%) (Table 25). The proportions of TEAEs leading to discontinuation or to death were similar between the trial arms but a greater proportion of participants in the DBd arm experienced Grade 3/4 TEAEs and serious TEAEs. The CS does not indicate what types of TEAEs resulted in discontinuation or death.

**Table 25 Summary of TEAEs at median 26.9 months of follow-up (CASTOR safety population).**

| | DBd (n=243) | Bd (n=237) |
|---|---|---|
| Any TEAE, n (%) | 241 (99.2) | 226 (95.4) |
| Grade 3/4 TEAE, n (%) | 197 (81.1) | 149 (62.9) |
| Serious TEAE, n (%) | 123 (50.6) | 81 (34.2) |
| TEAE leading to discontinuation, n (%) | 23 (9.5) | 22 (9.3) |
| TEAEs leading to death, n (%) | 15 (6.2) | 14 (5.9) |

Source: Data reproduced from CS Table 25

The most frequently reported TEAEs (≥20%) are presented in Table 26. A more detailed summary of TEAEs is provided in CS Table 26.

**Table 26 Most frequently reported TEAEs**

| TEAEs (≥20%) | DBd (n=243) | Bd (n=237) |
|---|---|---|
| Common haematologic AEs | | |
| Thrombocytopenia, n (%) | 60% | 44% |
| Anaemia, n (%) | 28% | 32% |
| Common nonhaematologic AEs | | |
| Peripheral sensory neuropathy, n (%) | 50% | 38% |
| Fatigue, n (%) | 22% | 25% |
| Cough, n (%) | 28% | |
| Diarrhoea, n (%) | 35% | 22% |
| Constipation, n (%) | 22% | |
| Upper respiratory tract infection, n (%) | 33% | |
| Back pain, n (%) | 20% | |

Source: This is a modified and reduced version of CS Table 26

The clinical experts consulted by the ERG indicated that the toxicity risks differed between DBd, Bd, and Cd (the latter not part of the CASTOR trial but included in the NMA). One expert highlighted the risk of infusion related reaction and cytopenia for DBd whereas for Cd risks were cardiac, pulmonary and renal. The other expert also highlighted cardiac AEs as a particular concern for carfilzomib but stated daratumumab is relatively well tolerated as is subcutaneously administerd bortezomib.

Infusion-related reactions (a commonly expected AE) were not reported separately in the IA2 data cut on which the CS is based. However data on IRRs are available from the median 7.4 months follow-up. The company state these data are likely to be representative of all the IRRs expected to be observed because the vast majority of IRRs occurred on day 1 of the first infusion of daratumumab (Table 27). Note that in the CASTOR trial bortezomib was delivered subcutaneously and therefore data on IRRs only applies to the DBd arm who received daratumumab by intravenous infusion.

**Table 27 Infusion-related reactions at 7.4 months follow-up**

| | **DBd (n=243)** | |
|---|---|---|
| | All grades | Grade 3 |
| Total number of participants with infusion related reactions, n (%) | 110 (45.3%) | 21 (8.6%) |
| Number of subjects with IRRs in more than one infusion | 4 (1.6%) | 1 (0.4%) |
| Number of subjects with IRR at first infusion | 108 (44.4%) | Not reported |
| Number of subjects with IRR at second infusion | 2 (0.8%) | Not reported |
| Number of subjects with IRR at subsequent infusion | 4 (1.7%) | Not reported |

Source: Data from Tables 6 and 7 provided in response to clarification question A9

The company indicated in response to clarification question A10 that no new safety signals were identified from their systematic literature review for AE data from non-randomised studies which is reported in CS Appendix F (with further data presented in the response to clarification question A10).

## 3.4 Summary

The company's decision problem defines a population narrower than that described in the NICE scope for this appraisal because it is limited to adults with RRMM who had received only one previous treatment (i.e. 2L patients). In contrast the NICE scope allows for patients who have received two or more prior therapies (at third-line and beyond). In other respects (intervention, comparators, and outcomes) the company's decision problem meets the NICE scope, albeit with comparators limited to those relevant to 2L patients. However two of the three comparators for 2L treatment, bortezomib-based therapy and Cd, are only available for use in current clinical practice for patients who are bortezomib naive. The company does not distinguish between 2LBN and 2L bortezomib-experienced patients in their decision problem.

The systematic review of clinical effectiveness evidence in the CS includes one RCT of DBd versus Bd (CASTOR). CASTOR is an open-label trial that enrolled patients who had received at least one prior line of therapy (with no upper limit). Therefore only a proportion of the CASTOR trial (47.2%) are 2L patients relevant to the company's decision problem.

The CASTOR RCT was judged in the CS to be of moderate methodological quality (CS Appendix D section D.3.4). Overall, CASTOR appears to have been well conducted. The clinical efficacy outcomes reported in the CS that contribute data to the economic model are PFS, OS, TTD, some AE data and HRQoL data. The results presented come from an interim analysis with a median follow-up of 26.9 months. The CS also reports outcomes that do not contribute to the economic model which have been summarised in the ERG report.

There is no direct evidence comparing DBd when used as a 2L therapy to any of the 2L comparators listed in the NICE scope and included in the company's decision problem (i.e. bortezomib-based therapy other than Bd, Cd and combination chemotherapy). Therefore the company searched for evidence to include in an NMA. The evidence identified for combination chemotherapy (one trial) could not be connected to the network and involved chemotherapy regimens not used in current practice. The ENDEAVOR trial was identified which provides evidence on Cd vs Bd and this was included in the network (which in essence is a simple indirect comparison of two RCTs, DBd vs Bd vs CD), to allow indirect comparison of DBd vs Cd. The ENDEAVOR trial contributing data to the NMA was judged by the company to be of moderate methodological quality.

With only two trials included in the NMA, a fixed-effect NMA using a Bayesian framework was carried out. For the subgroup of 2L patients, outcomes assessed by NMA were PFS, OS, ORR, VGPR or better, and CR or better. NMA sensitivity analyses were conducted for the subgroup of 2LBN patients for the outcomes of PFS and OS only.

PFS was the primary outcome of the CASTOR trial and a statistically significant benefit was observed for 2L patients in the DBd group where median PFS was approximately 18 months longer than in the Bd group (median PFS DBd 26.22 months versus Bd 7.92 months, HR = 0.23, 95% CI 0.16 to 0.33, p <0.0001). The hazard ratios from the fixed-effect NMA for 2L patients were in favour of DBd versus both Bd and Cd with a reported 100% probability that DBd was the best treatment. The NMA sensitivity analysis for 2LBN patients produced similar results but with slightly higher hazard ratios and wider 95% CrIs such that the 95% CrI for the DBd vs Cd hazard ratio crossed one (DBd vs Bd HR 0.23 95% CrI 0.14 , 0.38; DBd vs Cd HR 0.52 95% CrI 0.27, 1.02).

OS is a secondary outcome of the CASTOR trial but the data are immature and median survival has not yet been reached. Among 2L patients there is a difference in the proportion of deaths in favour of DBd (DBd 20.5% vs Bd 35.4%) which is statistically significant (HR 0.50, 95% CI 0.30, 0.84; p=0.008). The fixed-effect NMA in 2L patients indicated that DBd had the highest probability of being the best treatment (99.6% vs Bd and 95.0% vs Cd) but the DBd vs Cd 95% CrI crossed one. The NMA sensitivity analysis for 2LBN patients also indicated DBd had a high probability of being the best treatment but the 95% CrIs for both comparisons (DBd vs Bd and DBd vs Cd) crossed one (DBd vs Bd HR 0.53 95% CrI 0.24 , 1.17; DBd vs Cd HR 0.66 95% CrI 0.22, 1.98).

After disease progression, some participants in the CASTOR trial received a therapy not available in England. Furthermore, at the data analysis point more participants in the Bd arm than in the DBd arm had progressed. These factors could have introduced bias into the OS estimate. To reduce any bias the company adjusted the OS data using the IPCW method which involves censoring those patients who received a post-progression therapy that would not be available in England and weighting outcomes from the remaining patients with similar characteristics to account for the patients who have been censored. In the adjusted OS analysis for the subgroup of 2L patients there was a [redacted] reduction in the risk of death for the

DBd arm in comparison to the Bd arm whereas the unadjusted value was a 50% reduction in the risk of death. The adjusted OS value is used in the base-case of the economic model.

TTD data contribute to the economic model and come from a post-hoc analysis. DBd was associated with a 56% reduction the risk of treatment discontinuation at the median 26.9 months of follow-up compared with Bd. Bd has a maximum treatment duration (eight 21-day cycles) whereas the daratumumab component of DBd can be administered until disease progression (or unacceptable toxicity).

Other clinical efficacy outcome reported in the CS (MRD, response, duration of and time to response, TTP, best M-protein response, time to subsequent anticancer therapy, PFS on the subsequent line of therapy, and best response to first subsequent anticancer therapy) were not all reported separately for the subgroup of 2L patients. However, for those that were (MRD, response, and TTP) results were in favour of the DBd group.

HRQoL outcomes (EORTC-QLQ-C30) and EQ-5D-5L) were not reported separately for the subgroup of 2L patients. In the total trial population there was no significant difference between the two groups at any time point (from baseline to week 24).

The subgroup of 2L patients is the focus of the company's decision problem population but other subgroup analyses of the CASTOR trial were pre-planned and conducted for four outcomes (PFS, TTP, ORR and rate of VGPR or better). Included in the subgroup analysis of the whole CASTOR trial data were the subgroups of BN and bortezomib-experienced patients. Outcomes favoured the DBd treated group in both subgroups with results being numerically better in the BN subgroup than the bortezomib experienced subgroup for all four outcomes. It is not clear to the ERG whether in the subgroup of 2L patients, the patients who are bortezomib naive would have better outcomes than those who are bortezomib experienced (the NMA sensitivity analysis demonstrated slightly less favourable results in the bortezomib naive group in comparison to all 2L patients). Across all the other subgroups tested a consistent effect favour of DBd is observed.

TEAEs are summarised and although there are some differences in the types of TEAEs experienced and the proportions of patients experiencing them in the DBd and Bd groups the company states that no new safety signals were identified from their systematic review of AE

data. Data on eight AEs (neutropenia, anaemia, thrombocytopenia, lymphopenia, pneumonia, fatigue, peripheral neuropathy and hypertension) which occurred in at least 5% of the patients in either arm of CASTOR as a grade 3 event or higher were included in the economic model.

Overall, the ERG has identified the following key limitations of the evidence presented in the CS:

- Interim trial data are presented and the OS data are currently immature.
- No trials directly comparing DBd with Cd were identified, so only an indirect comparison was possible through an NMA containing two trials.
- Limited evidence was presented for the subgroup of 2LBN patients; the group eligible to receive Cd and bortezomib-based treatment at 2L. The company states they are positioning DBd for the treatment of both 2LBN and 2L bortezomib-experienced patients.
- We considered the 2L subgroup results to be at an unclear risk of selection bias. A greater proportion of patients in the BD arm received prior lenalidomide and were refractory to their previous treatment than in the DBd arm at baseline. We were unclear if the differences observed would bias the results, as our clinical experts advised that the differences were unlikely to be important, yet the company stated in the CS that these are prognostic factors.
- The CASTOR trial patient eligibility criteria do not fully reflect the 2L patients seen in clinical practice in England, as they excluded the following two patient groups: one, patients who had been refractory to prior treatment with a non-proteasome inhibitor; and, two, patients with non-secretory myeloma. Both these groups form a minority of the patients treated in practice.

# 4 COST EFFECTIVENESS

## 4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

i) a review of published economic studies and full economic evaluations of treatments for patients with RRMM (CS B.3.1 and Appendix G).

ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of DBd is compared with Bd and Cd for patients with one prior treatment for RRMM (CS B.3.2).

## 4.2 Review of published economic evaluations

### 4.2.1 Critique of company review

See section 3.1.1 of this report for discussion of the company's search strategy, including their search for published cost-effectiveness evidence. We consider that the search was well-documented and comprehensive. However, as it only included papers published before 22 August 2017, we updated the search.

The inclusion and exclusion criteria for the company's systematic review are listed in Table 10 in Appendix G of the CS, page 15. The inclusion criteria state that economic studies and full economic evaluations of pharmacological treatments under investigation, other licensed pharmacological treatments and standard of care/best supportive care were included. Table 10 reports that studies indexed before March 2005, not relevant to the UK perspective, not reporting QALY or ICER data that could be extracted and abstracts from conferences other than those in the grey literature search were excluded. In addition, it is stated in the text that articles based on expert opinion, commentary, letters, editorials, reviews, studies with no abstracts or articles written in a language other than English were excluded.

59 studies were identified by the company from screening 151 titles and abstracts. Of these 57 studies were excluded, mainly because they did not report a study design (n=25) or outcomes (n=8) of interest, or that they were not relevant to the UK perspective (n=14). Results from the 14 studies judged not relevant to the UK perspective are summarised in Table 13 of CS Appendix G.[34-48] The ERG agrees that these 14 publications are not directly relevant to the UK NHS, although we cannot verify the appropriateness of the other exclusions as the references

are not provided in the CS. One of the publications excluded as not relevant to the UK perspective - an abstract by Carlson et al. (2017) – reports a cost-utility analysis of treatments for RRMM that is relevant to the current decision problem.[45]

The PRISMA diagram (Figure 1 on page 17 of Appendix G) shows that 2 publications were included in the review: 1 primary and 1 secondary publication. The secondary publication is not cited or discussed further. The primary paper (Brown et al. 2013)[46] related to a UK cost-utility analysis that used a discrete event simulation model to evaluate a combination of lenalidomide and dexamethasone (Ld) compared with dexamethasone alone in patients with MM after failure of first-line therapy (Table 12, CS Appendix G, page 20). We agree with the company's conclusion that the technologies in this study are not relevant to the decision problem.

### 4.2.2 ERG update of review

The ERG update search found two publications relevant to the decision problem:

- Carlson et al. 2018.[2] This is a full report of an abstract excluded by the company, and relates to a 2016 health technology assessment prepared by the Institute for Clinical and Economic Review (ICER) for the Midwest Comparative Effectiveness Public Advisory Council (Midwest CEPAC).[49] The 2016 ICER report did not include DBd, due to a lack of evidence at the time, but the Carlson et al. 2018 paper does.
- An abstract by Maiese et al.[38] reporting a cost-effectiveness analysis based on CASTOR data. The analysis was funded by Janssen.

Both of these analyses used a US perspective, so the costs and ICERs are not relevant for the UK. However, the model methods, input parameters and health outcomes provide a useful cross-check for the face validity of the company's economic analysis. Models developed for previous NICE appraisals for RRMM also provide a useful comparison for the company's submitted model. The CS includes an overview of methods and results from appraisals (CS section B.3.2.2.1 Table 29, pages 87 to 90). Analyses conducted for the recent NICE appraisal for Carfilzomib (TA457)[16] are particularly relevant, as they assess the cost-effectiveness of the two comparators Cd and Bd in the subgroup of interest (MM patients who have had one previous therapy, not including bortezomib). We present a brief summary of the Carlson, Maiese and NICE TA Carfilzomib models here and then compare with the company's results in our validation section (Section 4.4.1 of this report).

**Carlson et al. (2018) cost-effectiveness analysis**

Carlson et al.[2] compared the cost-effectiveness of treatments for RRMM: seven treatments at second-line, including DBd and Bd (but not Cd); and eight treatments at third-line. They used partitioned survival model with three health states: progression-free survival (PFS); progressed disease (PD); and death. The PFS state was divided into 'on' and 'off' treatment stages, and costs for subsequent treatments in the PD state were included. The analysis used a 1-week model cycle, a lifetime horizon and a 3% discount rate for costs and health outcomes, with a half-cycle correction. Costs were estimated from a health sector perspective in 2016 US dollars. Second-line health utilities were based on estimates submitted to ICER by Amgen, derived from analysis of the ASPIRE trial: 0.82 (0.78 to 0.88) for PFS on treatment; 0.84 (0.82 to 0.97) for PFS off treatment; and 0.65 (0.62 to 0.74) for PD. A disutility of 0.08 (0.07 to 0.08) was used for adverse events, based on NICE TA427[50]).

PFS was modelled by fitting parametric survival curves to KM data for a baseline comparator, Ld in second and third line settings and then applying HRs for other treatments compared with Ld from an NMA. The Weibull distribution was selected for the Ld base case curve, based on measures of model fit and face-validity of the extrapolations. Although the authors preferred a random-effects NMA model, they had to use fixed-effects to obtain the statistically significant results observed in the trials, because the network relied mainly on single-study connections. The method of Ouwens et al.[51] was used to test whether the proportional hazards assumption was violated for NMA results. For the trials including daratumumab in the NMA (CASTOR[18] and POLLUX[52]), results for the whole trial populations were used, as sub-group analyses for second and third line treatment were not available. This means that results for DBd at second-line are conservative and of limited use for comparison with the results from the company model described below.

Due to the lack of OS data for some comparisons and susceptibility to bias from treatment crossover and differences in subsequent treatments, OS curves were derived using an estimated relationship between OS and PFS. This was based on a meta-regression by Felix et al. 2013[53] (see section 4.4.1 below for our critique of methods to estimate OS from PFS). Felix et al. reported an increase in median OS of 2.45 months (95% CI 1.71 to 3.20) for every 1 month increase in median PFS. Carlson et al. used this result to approximate a HR for OS compared with PFS of 0.41 (1/2.45), which they applied to the fitted PFS curves for each treatment. This HR was varied in sensitivity analysis and tested in a scenario with an estimate of

the OS/PFS relationship derived only the Carlson et al. review of RCTs for RRMM: 3.27 month increase in median OS for each month increase in median PFS.

**Maiese et al. cost-effectiveness study[38]**

This analysis was based on CASTOR trial data. A decision analytic model was used to estimate the cost-effectiveness of DBd compared with Bd for patients previously treated for MM and in a subgroup of patients with only one prior treatment. The model follows a partitioned survival approach, with three health states (PFS, PD and death). Parametric PFS curves were fitted to individual patient data from CASTOR, with the Weibull reported as giving the best fit for the first-relapse subgroup. OS was estimated from PFS using the surrogate relationship estimated by Felix et al.[53] as in the Carlson et al. study.[2] Analysis was conducted over a lifetime horizon, with a 3% annual discount rate applied to costs and outcomes. Utility data was taken from the literature (source not cited).

**Model submitted for NICE appraisal of carfilzomib (TA457)[16]**

The model submitted to NICE for the appraisal of carfilzomib for previously treated MM included an analysis focussed on the comparison of Cd with Bd for patients with one previous treatment, not including bortezomib.[16] This aligns with the scope in this current appraisal. The carfilzomib model was also a partitioned survival model with PFS, PD and death states. It used parametric PFS and OS curves fitted to ENDEAVOR trial data, with adjustment for the subgroup of interest. The analysis followed NICE reference assumptions with an NHS and PSS perspective, 3.5% annual discount rate for costs and effects, lifetime horizon (40 years), 28-day cycle with a half cycle correction. The ERG who worked on the carfilzomib appraisal were critical of the company's original approach to survival modelling, utility estimation and assumption of ongoing Bd treatment. After publication of an Appraisal Consultation Document, a new version of the model was submitted following committee preferred assumptions:

- Jointly-fitted Weibull models for PFS and OS (assumption of proportional hazards);
- Bortezomib treatment limited to 8 3-week treatment cycles (24 weeks), with capping of costs and adjustment of efficacy for the shorter duration of treatment than in the ENDEAVOR trial;
- Utility data mapped from EORTC quality of life observations from ENDEAVOR to EQ-5D values, using the Proskorovsky et al. (2014) algorithm without adjustment for treatment.[54]

## 4.3 Critique of the company's submitted economic evaluation

### 4.3.1 NICE reference case

**Table 28 NICE reference case requirements**

| NICE reference case requirements: | Included? | Comment |
|---|---|---|
| Decision problem: As per the scope | No | Model is restricted to second-line. It does not address use of daratumumab at third or fourth line, as in the scope. |
| Comparator: As listed in the scope | No | Combination chemotherapy not included. Model includes patients with prior bortezomib exposure, who are not currently eligible for Bd or Cd. |
| Perspective on costs: NHS and PSS | Yes | Company specify NHS perspective. But end of life costs are similar to estimates with local authority-funded social care.[55] |
| 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 | |
| Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers | Yes | |
| Type of economic evaluation: Cost utility analysis with fully incremental analysis | Yes | |
| Synthesis of evidence on outcomes: Based on a systematic review | Yes | |
| Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared | Yes | |
| Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of health related quality of life. | Yes | |
| Source of data for measurement of health related quality of life: Reported directly by patients and/or carers. | Yes | |
| Source of preference data: Representative sample of the UK population | Yes | |
| Equity considerations: An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit. | Yes | |
| Discount rate: 3.5% pa for costs & health effects | Yes | |

### 4.3.2 Modelled decision problem

The company models the cost-effectiveness of DBd compared with Bd and Cd in a population of adults with multiple myeloma who have received one prior therapy. Combination chemotherapy is not included in the model. The company state that they made every effort to identify clinical effectiveness evidence for chemotherapies used in clinical practice at second-line.

The population in the model is narrower than that in the licensed indication and NICE scope, reflecting the company's target positioning of DBd at 2L, where they argue it will have the greatest effect (CS section B.3.2.1). However, the model does not distinguish 2L patients who have been treated previously with bortezomib from those who have not. As we discussed in section 2.3 above, clinical practice guidance differs for these groups in England: consequently Cd and Bd are only recommended for use in the 2LBN group. The company notes that the patent for bortezomib is due to expire in May 2019, anticipating a subsequent fall in price and lifting of the NHS England restriction on repeat use of bortezomib at first and second line (CS B.2.5.2). We acknowledge that this may be correct, but for the moment suggest that the comparison of DBd with Bd and Cd is only appropriate for the 2LBN subgroup.

NMA results for the 2L and 2LBN subgroups are presented in section B.2.9.5 of the CS. The company advise that the 2LBN analysis should be treated with caution, as it is post hoc. This is true and there is also greater uncertainty over the results than in the pre-planned second-line subgroup analysis. It is reassuring to note that the estimated effects of DBd are similar in the 2L and 2LBN subgroups – although the latter are slightly less favourable for DBd, suggesting some bias in the company's ICER estimates (see Table 29). We consider this issue in ERG analysis (4.4.3).

**Table 29. Summary of NMA results**

| DBd vs: | HR [95% Credible Interval] | | | | | |
|---|---|---|---|---|---|---|
| | Base case (ITT popn) | | Subgroup analyses (2L) | | Sensitivity analyses (2LBN) | |
| | PFS | OS | PFS | OS | PFS | OS |
| Bd | 0.32 [0.25, 0.40] | 0.77 [0.57, 1.04] | 0.21 [0.15, 0.30] | 0.5 [0.30, 0.84] | 0.23 [0.14, 0.38] | 0.53 [0.24, 1.17] |
| Cd | 0.60 [0.45, 0.81] | 0.97 [0.68, 1.39] | 0.47 [0.29, 0.75] | 0.60 [0.33, 1.10] | 0.52 [0.27, 1.02] | 0.66 [0.22, 1.98] |

Key: 2L, second-line; 2LBN, second-line bortezomib naive.

Source: adapted from CS A.8 Table 5

### 4.3.3 Model structure and assumptions

The company describes the structure and key features of their model in CS section B.3.2.2 (page 82). Figure 5 below reproduces the company's illustration of the model structure. It is a partitioned survival model of conventional design for cancer appraisals. The model has three main health states: pre-progression; post-progression; and death. The pre- and post-progression states are subdivided into 'on' and 'off' treatment stages. The arrows illustrate how the cohort moves between health states over time.

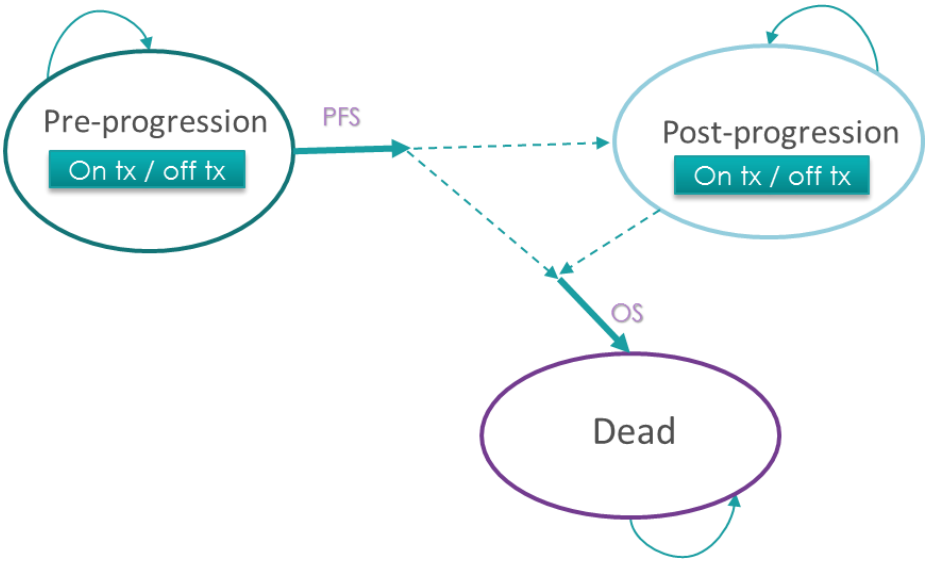



**Figure 5. Model diagram (reproduced from Figure 19, CS B.3.2.2)**

OS = overall survival; PFS = progression-free survival; Tx = treatment

Dotted lines represent the fact the transitions between health states are not directly tracked, but proportions of patients in each health state are calculated through the partition approach at each time point.

Patients enter the model in the pre-progression state at the start of second-line treatment. From there, they may stop treatment (e.g. due to an adverse event), experience disease progression or die. After progression, a proportion of patients have further treatment. Subsequent relapses and lines of treatment are not modelled explicitly, but are captured implicitly through costs for a bundle of subsequent treatments with effects on survival reflected in the OS curves. The model is run three times, once for each treatment (DBd, Bd and Cd) and costs and health outcomes (life years and QALYs) are estimated for each treatment based on the average time that members of the cohort spend in the different health states, cost and utility parameters.

Movement of the cohort between the health and treatment states is calculated using a partitioned survival approach.[56] This relies on a set of PFS, OS and TTD curves for each treatment and some additional assumptions described below.

First, the proportion of patients in the three main health states is calculated at each time point.

- *Death*: calculated as 1 minus the minimum of OS and survival in the general population (from the life table for the same age and gender mix as the cohort). This ensures that as patients age, their death rate is no lower than if they had not had MM.
- *Pre-progression*: the minimum of PFS and OS. This prevents progression exceeding survival in any sensitivity or scenario analyses.
- *Post-progression*: the residual of the cohort who are not pre-progression or dead.

Patients in the pre and post-progression states are divided by treatment status:

- For pre-progression, the proportion receiving second-line treatment at each time point is calculated as the minimum of TTD and the total who are progression free at that time. Thus, patients are assumed to stop second-line treatment at disease progression.
- For post-progression, the proportion of patients on subsequent treatment is defined by two parameters: a fixed proportion who have further treatment after progression; and a median duration of subsequent treatments (9 months), which is converted to a constant probability of discontinuation per model cycle.

However, to calculate post-progression treatment costs it is also necessary to estimate the number of new progressions in each period. This is not defined in the partitioned survival approach without an additional assumption. As illustrated by the dotted arrows in Figure 5 above, to calculate the number of progressions one needs to know the number of deaths from the pre-progression state. But the OS curve only gives the total number of deaths, both pre and post-progression. The company tackles this problem by fixing the number of pre-progression deaths as a constant proportion of the number of progressions in each model cycle (CS B.3.3.1.3 pages 104 to 105). The proportion of pre-progression deaths to progressions (14.6%) was estimated from CASTOR and is assumed the same for all treatment arms and unchanging over time. The ERG has concerns about this method, as it may lead to unrealistic mortality rates pre and post-progression. We check this and discuss this further in section 4.4.3 below.

The cost of treatment after progression is calculated using a bundle of therapies used in CASTOR that are available in England: Ld, pomalidomide and dexamethasone (Pd) and daratumumab monotherapy, but not the three-drug combinations (PBd and ILd) which are also now recommended by NICE at third or fourth-line. The effect of subsequent treatment on survival is embedded in the OS curves. In their base case, the company use DBd and Bd OS curves for the second line subgroup from CASTOR, adjusted for treatments not available in England (see section 3.1.6 above). This adjustment was not possible for Cd, as individual patient data were not available from the ENDEAVOR trial.

The model also includes treatment costs and health effects of adverse events associated with the second-line (but not subsequent) treatments. These are applied as a one-off cost and QALY loss in the first period.

The company summarise key features of the model in comparison with previous NICE appraisals of treatments for RRMM in Table 29 (CS B.3.2.2.1). We repeat the company's justification for their assumptions alongside ERG comments in Table 30.

**Table 30 Other model features**

| Factor | Company justification | | ERG comments |
|---|---|---|---|
| Summary of analytic methods | Partitioned survival model | Best use of available data, minimum amount of assumptions and captures the novel mechanism of action of daratumumab. | We agree with 3-state model structure and partitioned survival approach. This is common in RRMM appraisals, and given immaturity of OS data, estimation of post-progression survival would be highly uncertain. |
| Patient population | Adult patients with multiple myeloma who have received one prior therapy. | Population identical to the second-line population included in the CASTOR phase III clinical study | Analysis for whole scope population would have been preferable, but we acknowledge the superior outcomes for the 2L subgroup. Given NHS England and NICE restrictions, the population should also have been restricted to patients without prior bortezomib (2LBN subgroup). |
| Perspective | NHS | Aligns with NICE guide to the methods of technology appraisal | Local authority funded social care should also be included. However, end of life care costs in model are similar to Nuffield Trust estimates that include these costs.[55] |

| Factor | Company justification | | ERG comments |
|---|---|---|---|
| Time horizon | 30 years | Given the median age of 63 years for CASTOR population, 30 years is a fair approximation of a lifetime time horizon | We agree with the use of a lifetime horizon. But extrapolation of treatment effects over this very long period is not reasonable given immaturity of survival data. |
| Waning of treatment effect | No, independently fitted curves | Due to mechanism of action of daratumumab, which results in fundamental change to OS | Long-term effect on survival is not yet demonstrated. We use more conservative assumptions about persistence of effects and test in scenario analysis. |
| Model cycle length | 1 week | Adequately captures differences between dosing schedules regularly used in RRMM (3 or 4 weeks) | Agree |
| Half-cycle correction | Applied | | Agree |
| Source of utilities | EQ-5D-5L data from trial, mapped to EQ-5D-3L; van Agthoven (2004) tested in scenario analysis | Allows utility calculation from the exact population from which efficacy data were derived. Aligns with NICE position statement | Trial-based EQ-5D data valued using the van Hout cross-walk approach is theoretically preferable. But we have concerns over the CASTOR EQ-5D analysis (see 4.3.5 below). We therefore prefer the health state utility values from the mapping of ENDEAVOR EORTC data to EQ-5D, as used in TA457.[16] |
| Discount | 3.5% | Align with NICE guide to the methods of technology appraisal | Agree |

The model requires three main sets of input parameters:

- **Clinical inputs** that govern rates of disease progression (PFS) and mortality (OS), as well as the duration of treatment (TTD) and incidence of adverse events;
- **Utilities** for pre and post-progression health states and disutility and adverse events;
- **Resource use and costs** for second-line and subsequent treatments, monitoring and medical follow-up and end of life care.

Base case values and sensitivity ranges for model parameters are shown in CS Appendix N.

### 4.3.4 Treatment effectiveness and extrapolation

The key parameters driving clinical effectiveness in the model are survival functions of PFS, OS and TTD for the three included treatments. The company approach is described in CS section B.3.3. We outline this below, before discussing the choice of PFS, OS and TTD curves in more detail.

#### 4.3.4.1 Summary and critique of approach to fitting survival curves

To extrapolate beyond the trial period, the company employed parametric survival analysis. Six candidate functions (exponential, Weibull, log-normal, log-logistic, Gompertz and generalised gamma) were fitted to individual-level data from the CASTOR trial to estimate PFS, OS and TTD survival curves for DBd and Bd. The company concluded that proportional hazards does not hold for these outcomes in CASTOR, and so fitted separate curves to the trial arms. They do not present any curve parameters or model results for jointly-fitted PFS, OS or TTD curves. The ERG considers that the proportional hazards assumption cannot be assumed to hold for PFS, but that evidence is more equivocal for OS – see below for further discussion.

The survival curves were all fitted using data for the second-line (2L) subgroup of patients with one prior therapy. Curves were not estimated for the second-line bortezomib naïve (2LBN) subgroup for whom Bd and Cd are currently available in England. Given relative effects in the 2L subgroup compared with the 2LBN subgroup (Table 29), the fitted curves will tend to bias results in favour of DBd. We attempt to adjust for this in additional ERG analysis (4.4.3 below).

For OS, the fitted curves were adjusted for subsequent treatments used in the trial but not available in England. The adjustment was made using the IPCW method, as described in section 3.1.6 above. The ERG considers that this method is appropriate in principle, but that results may be biased if important covariates are omitted or conversely if the model is over-fitted, with the inclusion of too many covariates with weak explanatory power. The company's model has an option to include or exclude subsequent treatment adjustment of OS. We test the impact of excluding the adjustment in ERG scenario analysis.

The company states that they based their choice of parametric functions for each curve on two factors: statistical fit, as measured by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) statistics; and clinical plausibility of the projections, informed by expert input from an advisory board. The company argue that long-term plausibility should be

given more weight than statistical goodness of fit, because of the short follow-up time in CASTOR (median 26.9 months). The ERG broadly agrees with this, although we note that achieving a good fit during the trial period is also important for face validity of results and accurate estimation of the ICER.

In response to a clarification question, the company present two other methods for estimating PFS and OS curves:

- **Piecewise modelling**, using Kaplan-Meier (KM) data directly for the trial period, followed by a parametric 'tail' to extrapolate over the long term. The facility to use this KM+tail approach was built into a revised version of the company model. The company attached the tail at the end of the KM data: when at least 10 patients remained at risk, except for PFS with DBd for which only 3 patients remained at risk. The last horizontal segment of the KM curve was continued until it met the parametric curve. The ERG believes that the KM+tail approach is useful to test the effect of a more accurate fit to the trial data, but only if this does not add to uncertainty or give implausible predictions. We question the use of KM data with so few patients remaining at risk. A threshold of at least 20 patients remaining at risk in each arm is more conventional, and we test this more conservative threshold in ERG scenario analysis.
- **Fractional polynomials**. The company also tested a range of fractional polynomial curves to see if they could give a better fit to CASTOR PFS and OS data. They concluded that for PFS, many of the curves provided a similar fit to the parametric curves, and that the best-fitting curves (as measured by minimum Deviance Information Criteria) have a higher probability of progression in the first few months that was not observed in the trial and is not clinically plausible. For OS, the company argues that the approach is not effective at reducing uncertainty for the economic analysis, as the confidence intervals around the fractional polynomial curves are wide, and the best-fitting fractional polynomial curves give unrealistic estimates of survival: with no patients surviving beyond 4 years. We agree with this conclusion and do not pursue the fractional polynomial approach in further ERG analysis.

For Cd, the PFS and OS curves were estimated by applying hazard ratios from the NMA to the fitted curves for Bd. This entails a proportional hazards assumption for the comparison of Cd with Bd. The company does not discuss the evidence for or against this assumption, but we

consider it appropriate based on committee conclusions in the recent carfilzomib appraisal (TA457), which apply to the same patient group (2LBN) and source of data (ENDEAVOR).

Another committee consideration from TA457 that is relevant for this current appraisal relates to the duration of bortezomib treatment. The TA457 committee noted the discrepancy between the length of treatment approved in the marketing authorisation for bortezomib (24 weeks), as used in CASTOR, and ongoing use until progression in ENDEAVOR. Following consultation, the carfilzomib manufacturer submitted a revised model truncating the cost of bortezomib at 24 weeks and adjusting PFS and OS estimates from ENDEAVOR for this shorter duration of treatment, which the committee accepted. However, in this current appraisal, the model applies costs for bortezomib for 24 weeks (in both DBd and Bd arms) but uses unadjusted PFS and OS hazard ratios from ENDEAVOR. The effect of this omission will be to underestimate the cost-effectiveness of Cd compared with Bd and hence to overestimate the cost-effectiveness of DBd compared with Cd. We attempt to adjust for this factor in additional ERG analysis (section 4.4.3 below).

The TTD curve for Cd was estimated by applying a hazard ratio to the Cd PFS curve. The hazard ratio for PFS compared with TTD was estimated from ENDEAVOR. We consider this acceptable in the absence of KM data on TTD from ENDEAVOR.

#### 4.3.4.2 Progression free survival curves

The company presents plots of log hazard functions of CASTOR data (Figure 3 CS Appendix L, reproduced in Figure 6 below). These clearly show that the proportional hazards assumption does not hold, as the Weibull log-log cumulative survival curves cross. The company therefore fits independent PFS curves for DBd and Bd. We agree with this approach.

Parameter values and measures of fit for the six PFS survival distributions are shown in Tables 1 and 2 in CS Appendix L. Based on AIC /BIC alone, log-logistic is the best fit for Bd, while Gompertz has the worst fit. However, for DBd Gompertz has the lowest AIC and second-lowest BIC. The company choose the Gompertz distribution for both DBd and Bd in their base case to maintain consistency between the trial arms (as recommended by the NICE Decision Support Unit [57]) and because it provides a good visual fit to the KM data and reasonable long term projections.

We show the CASTOR KM and fitted PFS curves over a 3-year time horizon in Figure 7. The Gompertz provides a good visual fit to the data for DBd but a poor fit for Bd. The generalised gamma provides a better fit to trial data for Bd without greatly compromising the fit to DBd. In terms of long-term projections, all functions predict that a small proportion of patients will remain progression-free after 5 years with Bd. But the functions differ in their long-term predictions for DBd. After 5 years, the Gompertz predicts 9% of patients to be progression-free, whereas the generalised gamma predicts that all patients will have progressed.

For Cd a HR of 0.45 (95% credible interval 0.28 to 0.72) compared with Bd was estimated from the NMA.

**Table 31 Goodness of fit for PFS parametric models (second line)**

| | Bd | | DBd | |
|---|---|---|---|---|
| | **AIC** | **BIC** | **AIC** | **BIC** |
| Exponential | 634.32 | 637.05 | 562.58 | 565.39 |
| Weibull | 625.72 | 631.17 | 564.41 | 570.02 |
| Log-normal | 622.23 | 627.69 | 575.84 | 581.45 |
| Log-logistic | 616.19 | 621.65 | 568.39 | 574.00 |
| Gompertz | 634.77 | 640.23 | 562.50 | 568.11 |
| Generalized gamma | 621.60 | 629.78 | 563.43 | 571.84 |

Source: CS Table 30

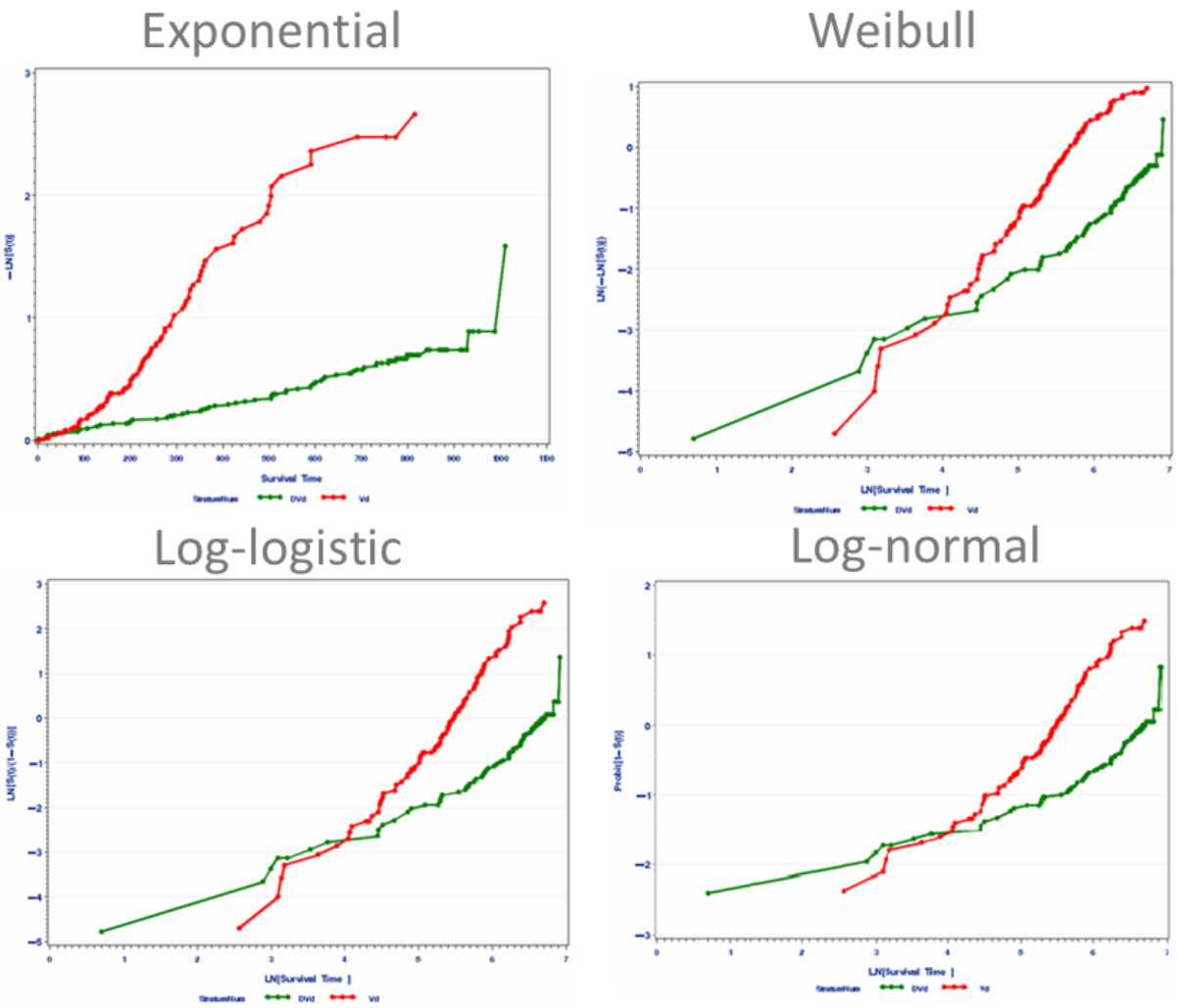
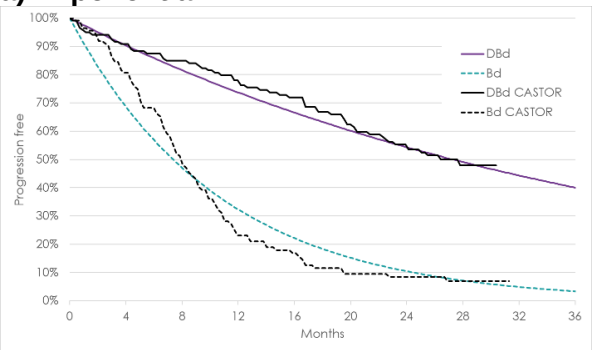
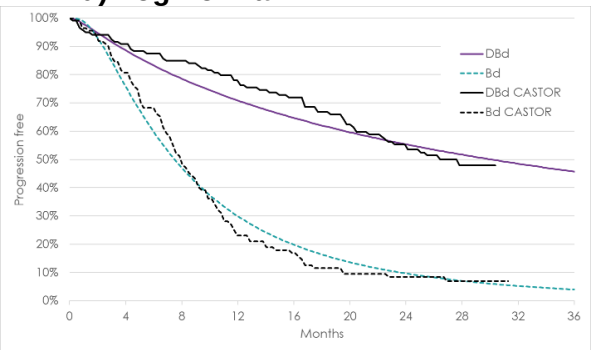



**Figure 6 Log hazard plots for PFS for DBd versus Bd (second line)**
(Reproduced from CS Appendix L Figure 3)

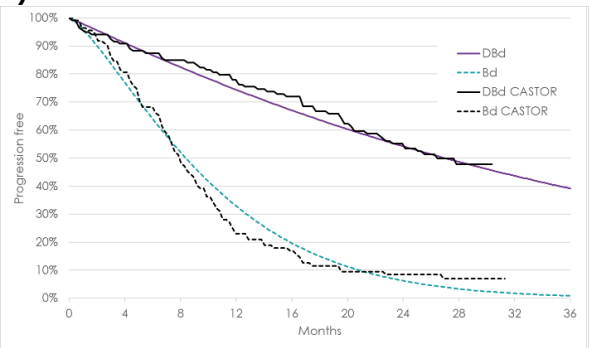
**a) Exponential**



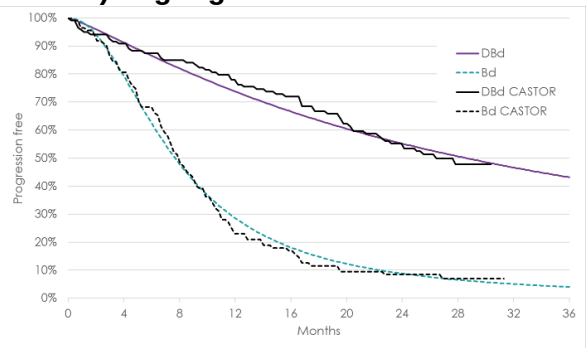

**d) Log-normal**



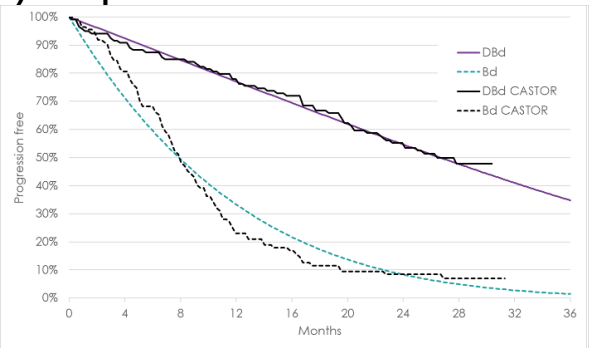

**b) Weibull**




**e) Log-logistic**




**c) Gompertz**




**f) Generalised gamma**

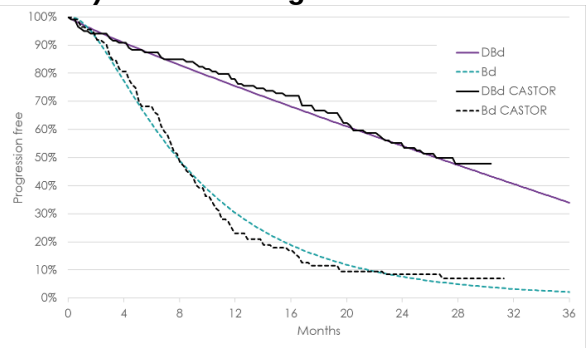



**Figure 7 Progression free survival (CASTOR; second-line)**

#### 4.3.4.3 Overall survival curves

The company examined the OS log hazard curves from CASTOR, concluding that the proportional hazards assumption between the treatment arms does not hold. Consequently, they fit DBd and Bd curves separately. We reproduce the company's log hazard plots in Figure 8 (from CS Appendix L Figure 6). Here the conclusion about proportional hazards is less clear-cut. The DBd curve in particular is 'noisy' crossing over the Bd curve and then back. It is not clear if this reflects real changes in relative treatment effects, or if it is a chance finding, due to the small sample sizes and immature data.

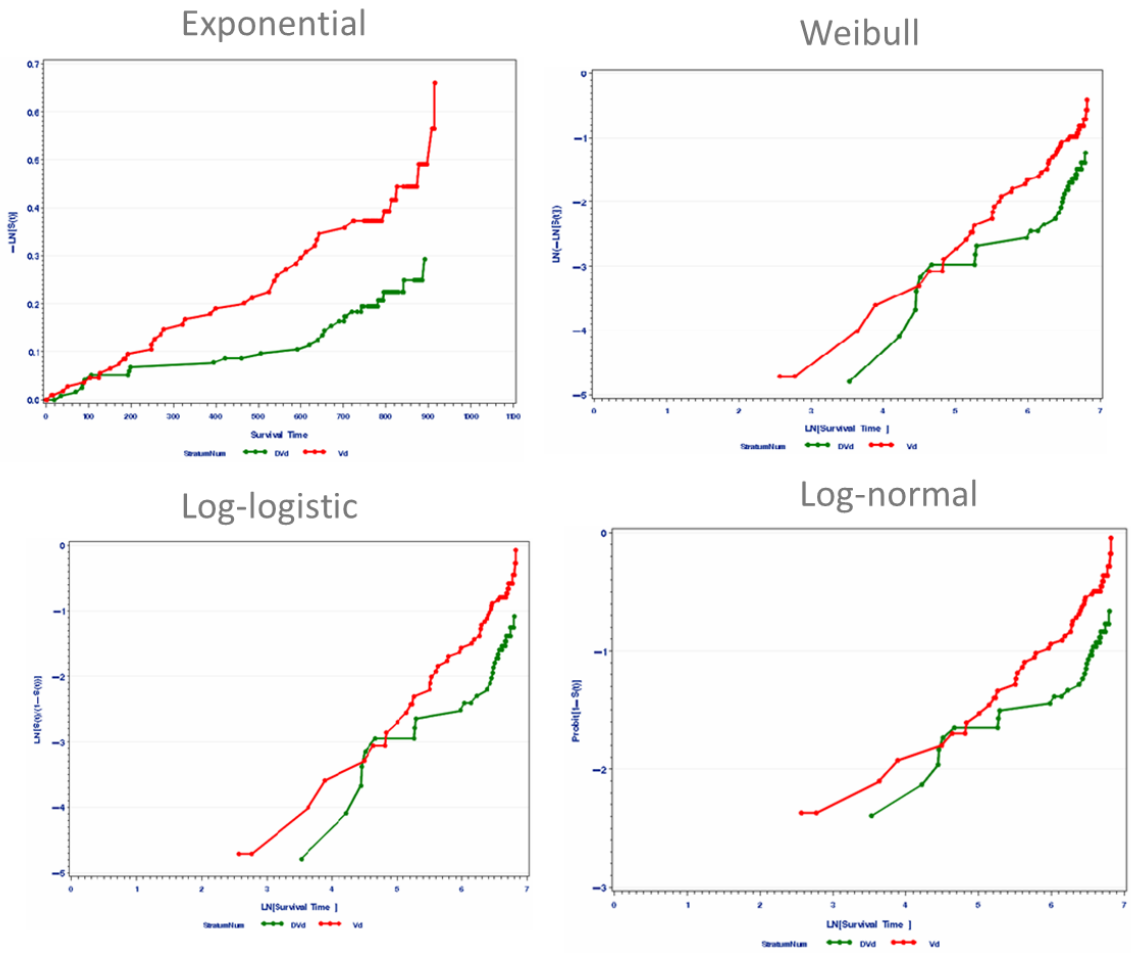



**Figure 8 Log hazard plots for adjusted OS for DBd versus Bd (second line)**
(Reproduced from CS Appendix L Figure 6)

Goodness of fit statistics for OS in the 2L subgroup and adjusted for subsequent treatments not available in England are reproduced in Table 32 below (from CS Table 33). There is little difference between the AIC/BIC values for different functions. This may indicate, as the company suggest, that all the curves are well-matched to the data, or that they are equally

ill-matched. Despite the similar fits to trial data, the six survival functions give a very wide range of long-term survival predictions.

For Bd, the exponential has the best statistical fit, followed by Gompertz and Weibull. The company choose the Gompertz for their base case, based on the comparison of predicted median and mean survival with external data and shape of the mortality hazard curve predicted by experts (Table 34 and Figure 24 CS B.3.3.1.2). For DBd, the company note that the Gompertz and exponential have the best statistical fit, closely followed by the Weibull and log-logistic. They also note that clinical expert opinion supported the Weibull based on predicted survival at 5 and 10 years. However, the company choose the log-logistic curve for their base case, based on opinion about the "*transformational nature*" of DBd as a novel therapy. They note the higher rate of MRD negativity with DBd than with Bd and argue that this is associated with prolonged OS.

For Cd, a HR of 0.83 is applied to the Bd curve (95% credible interval: 0.45 to 1.52).

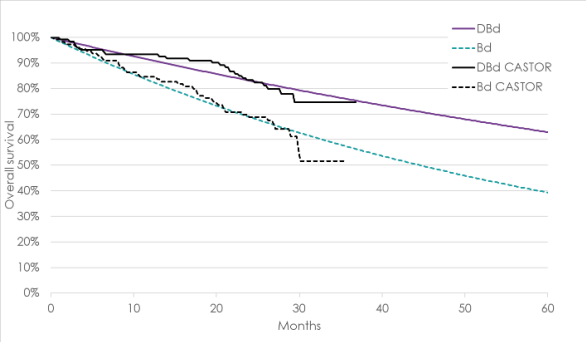
**Table 32 Goodness of fit for adjusted OS parametric models (second line)**

| | **Bd** | | **DBd** | |
|---|---|---|---|---|
| | AIC | BIC | AIC | BIC |
| Weibull | 302.80 | 308.20 | 261.30 | 266.90 |
| Log-normal | 304.50 | 309.90 | 263.50 | 269.10 |
| Log-logistic | 303.20 | 308.70 | 261.90 | 267.50 |
| Exponential | 301.50 | 304.20 | 260.00 | 262.80 |
| Generalized gamma | 304.50 | 312.70 | 262.60 | 271.00 |
| Gompertz | 302.60 | 308.00 | 259.30 | 264.90 |

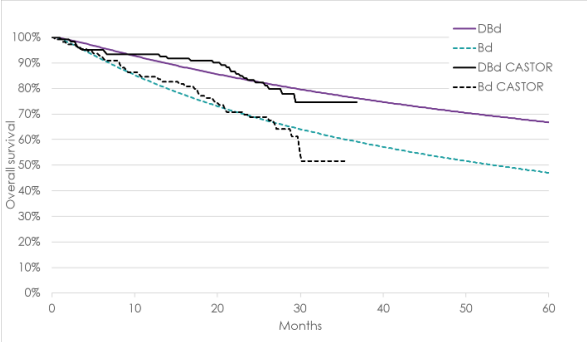
Source: CS Table 33

The comparative fit of the different Bd and DBd curves is illustrated in Figure 9. The ERG considers that the company's selection of curves (Gompertz for Bd and log-logistic for DBd) gives a very optimistic prediction of the survival gain from DBd that is not warranted given the immature OS data. We suggest more conservative assumptions, with the same functional form for DBd and comparators, selected for plausibility of 5 and 10-year outcomes. In the recent NICE TA of carfilzomib (TA457),[16] the committee concluded that the Weibull distribution was supported by OS data from ENDEAVOR and external validation for the second-line comparison in patients without prior bortezomib. We therefore conclude that the Weibull distribution for OS provides a consistent foundation for modelling Bd in this current appraisal, and that there is no reason to prefer a different function for DBd.

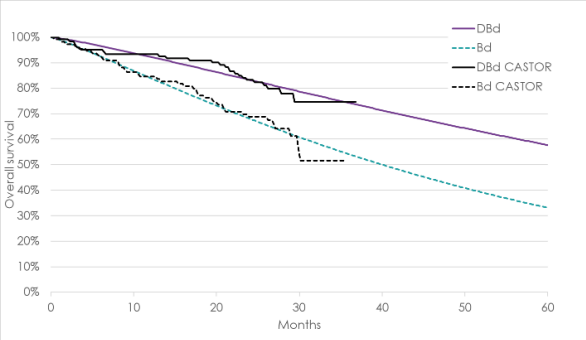
**a) Exponential**



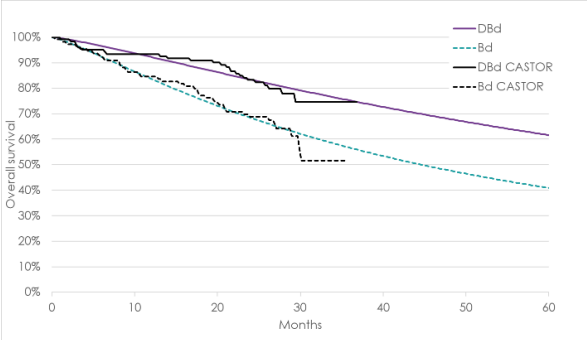

**d) Log-normal**



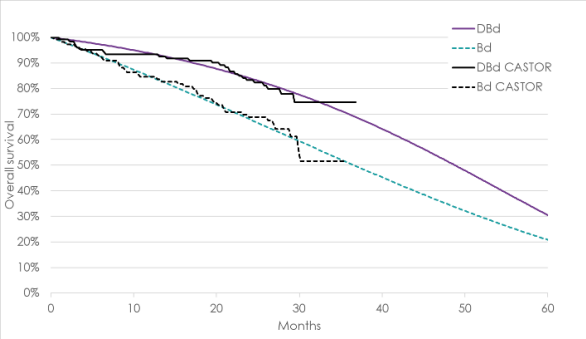

**b) Weibull**




**e) Log-logistic**




**c) Gompertz**




**f) Generalised gamma**

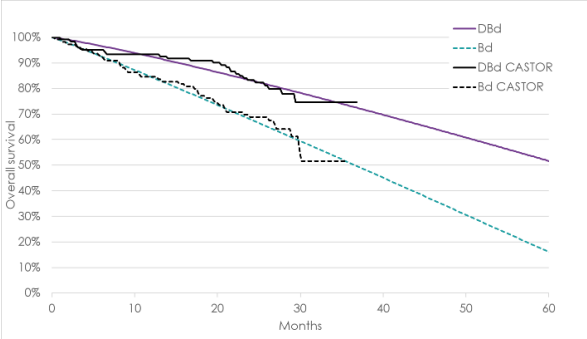



**Figure 9 Overall survival adjusted for subsequent treatment (CASTOR; second-line)**

#### 4.3.4.4 Time to treatment discontinuation curves

TTD curves are assigned to DBd and Bd based on analysis of individual patient data from CASTOR. The log hazard plots are shown in Figure 10. In this case it is not surprising that the proportional hazards assumption is not met, because Bd had a maximum treatment duration of 24 weeks in CASTOR, whereas daratumumab continued until progression or discontinuation for other reasons. The company fitted independent parametric functions to the two study arms (2L subgroup). The model parameters and fit statistics are shown in Table 9 and 11 of CS Appendix L. We summarise the AIC/BIC results in Table 33. Differences between the models are small and the company chose to use the same function as for PFS (Gompertz) for consistency. This recognises the likely correlation between PFS and TTD. The TTD curve for Cd was modelled using a proportional hazard of 0.477 compared with PFS, based on ENDEAVOR. We agree with these assumptions.

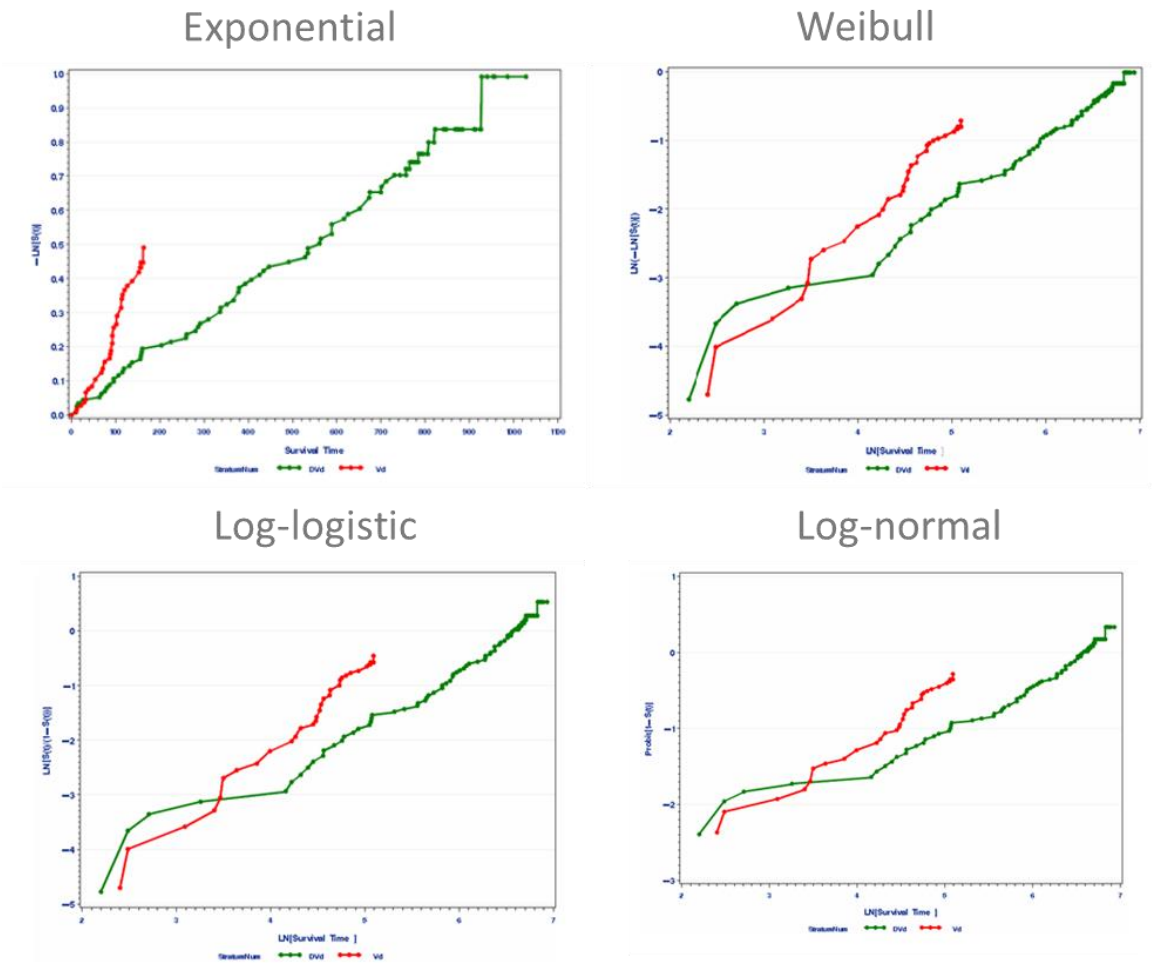



**Figure 10 Log hazard plots for TTD for DBd versus Bd (second line)**

Source: CS Appendix L Figure 10

**Table 33 Goodness of fit for adjusted TTD parametric models (second line)**

| | Bd | | DBd | |
|---|---|---|---|---|
| | AIC | BIC | AIC | BIC |
| Weibull | 285.91 | 291.32 | 610.00 | 615.56 |
| Log-normal | 284.99 | 290.41 | 616.29 | 621.85 |
| Log-logistic | 285.11 | 290.53 | 612.60 | 618.16 |
| Exponential | 288.69 | 291.40 | 608.03 | 610.81 |
| Generalized gamma | 286.89 | 295.02 | 610.88 | 619.22 |
| Gompertz | 288.33 | 293.75 | 609.59 | 615.15 |

Source: Table 33

#### 4.3.4.5 Mortality assumptions

As noted above, the partitioned survival model requires additional assumptions to estimate the numbers of death and progression events per model cycle from the PFS state. The company assumes a constant ratio of death versus progression events, set at 14.56% based on CASTOR data. The model also includes an option to define a fixed mortality rate from the PFS state, set at 0.136% per week. In addition, the model includes a check to ensure that the mortality rate for the modelled cohort is not less than that in the general population from UK Life Tables (adjusted for age and gender-mix).

### 4.3.5 Health state utilities

The company model used health state utility values estimated from the CASTOR trial. As per the trial protocol, utility values obtained from EQ-5D-5L were assessed at the first day of each treatment cycle, end of treatment, post-treatment week 8 and post treatment week 16. A schema of the utility scores from the trial is reproduced from CS Figure 28 in Figure 11 below.

Methods of analysis are outlined in CS B.3 and D.3.2.11, but the description is sparse. To adhere to NICE guidelines, the company mapped the EQ-5D-5L scores to EQ-5D-3L values, using a 'crosswalk' method reported by van Hout and colleagues.[58] Methods of handling missing EQ-5D data are described in response to clarification question A23.

The company state that there was no statistically significant difference in the utility scores between the two treatment arms. The company assumed the same utility for all patients in pre-progression health state across the two treatment arms: 0.7280 (95% CI 0.7062 to

0.7497). This was calculated as the average of all measured utilities before the date of progression, using repeated measures mixed-effects modelling.

For patients in the progressed health state, the company use an estimated utility of 0.6950 (95% CI 0.6511 to 0.7389). They state in the CS that this was derived from utility at last observation as a proxy (CS page 109). However, in their response to the factual accuracy check, the company state that:

> "Post-progression utility was defined as any utility measured after the date of progression. Eligible patients were patients who progressed and whose progression was not death. Average utility was calculated using a linear mixed model because of multiple measurements per subject. A subject random intercept was used to handle the different values for each individual." (Factual Accuracy Check, Issue 8)

The ERG cannot assess the quality or accuracy of this analysis as we have not seen details of the methods or results.

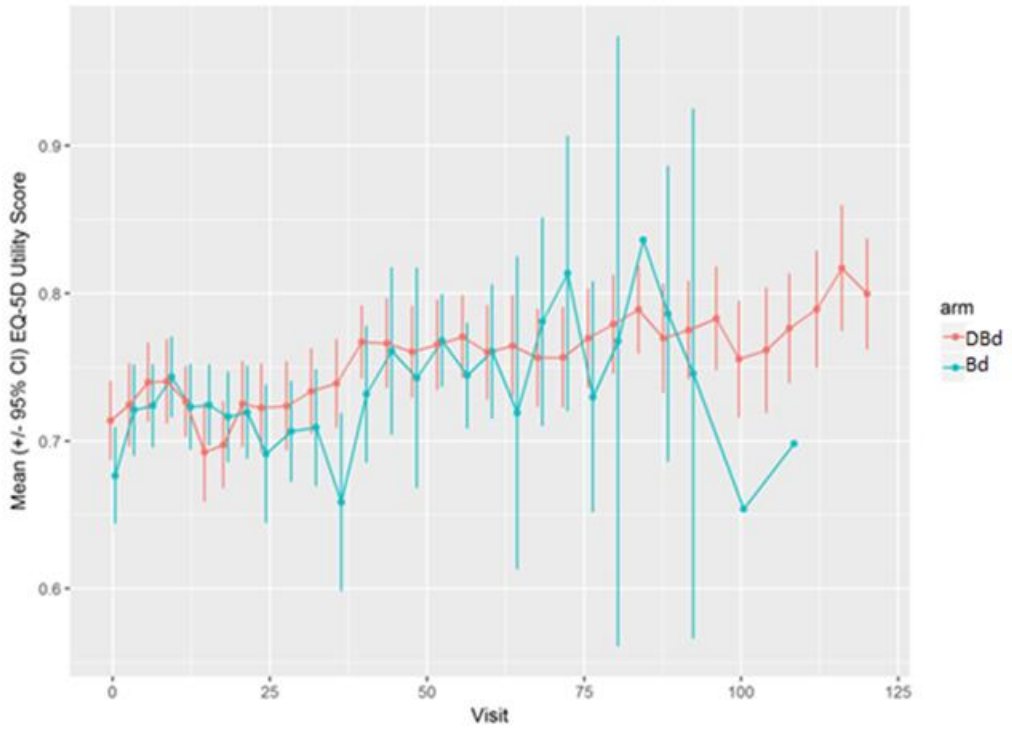



**Figure 11 EQ-5D-5L utility score-CASTOR (reproduced from CS Figure 28)**

The ERG is satisfied with the methods used to estimate pre-progression utility from the CASTOR EQ-5D data, but we are concerned about poor reporting of methods used to derive

the estimate of post-progression utility. We also note the similarity of the pre and post-progression utility estimates, which seems unrealistic.

The company conducted a systematic literature review to identify HRQoL studies for patients with RRMM who had failed at least one prior therapy. The review methods are described in CS Appendix H. Of 1,031 studies identified, 10 publications were included: six primary and four secondary. The primary studies are summarised in CS Table 39 and CS Appendix H Table 12. None of these six studies informed the utility parameters in the company's cost-effectiveness analyses. The ERG updated the search on ScHARR Health Utilities Database in March 2018 which yielded four papers, but these did not provide utility estimates relevant to this appraisal. However, the cost-effectiveness analysis reported by Carlson et al.[2] that we discuss in Section B.2.2 above, reported utility estimates provided to them from an analysis of ASPIRE trial data by Amgen (data submitted to ICER).

The ERG compared the utility values reported by previous technology appraisals in MM submitted to NICE with the values used by the company. Details of the utility values are presented below in Table 34. The ERG view is that the patient population in TA457 is the same as the relevant population in the current appraisal: RRMM patients with one prior therapy excluding bortezomib. In response to the ACD consultation, Amgen provided additional estimates of utility from the ENDEAVOR trial: mapped from EORTC data using a mapping algorithm from Proskorovsky et al (2014)[54] and assuming no difference in utility between treatments.

**Table 34 ERG's comparison of the utility values used in previous NICE TAs**

| Source | Treatments | Patient group | PFS | PD | Source |
|---|---|---|---|---|---|
| **CASTOR** | DBd and Bd | Not specified | 0.728 | 0.695 | **CS B.3.4.2** |
| **ASPIRE** | CLd and Ld | MM patients relapsed with 1 prior treatment | 0.82 on treatment<br>0.84 off treatment | 0.65 | Carlson et al.[2] |
| TA457[16] | Cd | MM patients if they have had only 1 previous therapy, which did not include bortezomib | Treatment cycle 1-2: 0.737<br>Later cycles: 0.741 | 0.638 | Amgen Carfilzomib ACD response December 2016 |
| TA129[15] | Bortezomib (monotherapy) | People who are at first relapse having had 1 prior therapy and who have undergone, or are unsuitable for bone marrow transplantation | 0.81 | 0.64 | van Agthoven et al. 2004.[3] |
| TA171[59] | Ld | MM patients who have received 2 or more prior therapies | Complete response, partial response and stable disease states: 0.82 | 0.64 | |

| Source | Treatments | Patient group | PFS | PD | Source |
|---|---|---|---|---|---|
| | | | After 2 years, patients whose disease had not progressed had a utility of 0.77 | | |
| TA380[60] | PBd | MM patients who have received at least 2 prior therapies including bortezomib and an immunomodulatory agent | Bd: 0.725 Pano+Bd: 0.706; No treatment: 0.762 | Len+dex and last line treatment: 0.64 | |
| TA427[50] | Pd | MM adults with 3rd or subsequent relapse | 0.76 | 0.62 | MM-003 |

On balance, the ERG prefers the mapped utility values from the ENDEAVOR trial as used in the NICE carfilzomib appraisal (TA457) to the CASTOR EQ-5D estimates used in the company model, because the latter do not reflect adequately disutility for progressed disease. Use of the same values as in TA457 also has the advantage of aligning the present analysis with previous committee conclusions for the same patient group and comparator treatments. We consider alternative sources of utility estimates in scenario analysis: Carlson et al.[2] and van Agthoven et al. 2004.[3]

### 4.3.6 Adverse events

The company included disutility and treatment costs for grade 3 or 4 adverse events with reported incidence of at least 5% of the patients in any treatment arms in CASTOR. Table 35 below (reproduced from CS Table 40) outlines the AEs included in the company's economic model along with the cumulative probabilities of AE occurrence during the treatment period.

The company applied one-off utility decrements due to AEs at the start of treatment. These decrements were based on treatment specific rates of AEs, AE duration and associated disutility, which were obtained from published literature. A summary of the AE disutilities are reproduced from CS Table 41 in Table 35.

Costs of treating the included adverse events were also included in the economic model, applied as a one-off cost at the start of treatment. These costs were based on the National Schedule of Reference Costs for Year 2016-17.

**Table 35 AEs included in the cost-effectiveness analysis**

| Adverse Event | Reported incidence during treatment period | | | Duration (Days) | Disutility | Cost per AE |
|---|---|---|---|---|---|---|
| | DBd [a] | Bd [a] | Cd [b] | | | |
| Neutropenia | 11.8% | 3.6% | 0.9% | 13.2 | -0.145 | £1,580 |
| Anaemia | 10.1% | 9.0% | 12.9% | 10.7 | -0.310 | £1,112 |
| Thrombocytopenia | 42.0% | 20.7% | 6.5% | 14.1 | -0.310 | £1,447 |
| Lymphopenia | 7.6% | 3.6% | 4.3% | 15.5 | -0.065 | £1,362 |
| Pneumonia | 11.8% | 9.0% | 6.5% | 12.0 | -0.190 | £1,690 |
| Fatigue | 3.4% | 4.5% | 6.0% | 14.6 | -0.115 | £878 |
| Peripheral neuropathy | 7.6% | 6.3% | 2.2% | 8 | -0.065 | £1,190 |
| Hypertension | 5.0% | – | 10.3% | 0 | 0 | £584 |
| **Mean QALY loss** | **0.0078** | **0.0044** | **0.0028** | | | |
| **Mean cost** | **£1,359** | **£772** | **£559** | | | |

Source: CS Tables 40, 41 and 53
a) From CASTOR study; b) From ENDEAVOR study

### 4.3.7 Resource use and costs

The economic model included the following costs:

- Drug acquisition
- Drug administration and co-medication
- Subsequent treatment
- Follow up monitoring and care
- Adverse events; and
- Terminal care

**Drug acquisition costs**

The company's base case uses the list prices for all drugs, as shown in Table 36 below. The company's base case model is built to include Patient Access Scheme (PAS) and confidential commercial access agreements (CAAs) for comparator and subsequent therapies, but no PAS discounts are included in results reported in the CS. We present results including all available PAS/CAA agreements in a confidential addendum to this report.

**Table 36 Drug acquisition costs for company base case**

| Drug | Drug units (vials or capsules) per pack | Strength | Price per Pack | Source |
|---|---|---|---|---|
| Daratumumab 100 mg | 1 | 100 mg | £360.00 | MIMS UK Drug Database. Access date: Jan 9, 2018. |
| Daratumumab 400 mg | 1 | 400 mg | £1,440.00 | |
| Pomalidomide | 21 | 4 mg | £8,884.00 | |
| Carfilzomib | 1 | 60 mg | £1,056.00 | |
| Bortezomib | 1 | 3.5 mg | £762.38 | |
| Thalidomide | 28 | 50 mg | £298.48 | |
| Lenalidomide | 21 | 25 mg | £4,368.00 | |
| Dexamethasone | 50 | 8.0 mg | £120.03 | |

Source: CS Table 43

The costs of the drugs are informed by dosing of the treatment regimens (CS Table 42). The dosing is, in turn, dependent on patient characteristics including body weight (mean 77.94 Kg) and/or body surface area (1.87 $m^2$). For DBd and Cd, the CS obtained the relevant information on dosing from CASTOR whereas dosing information for Cd was obtained from ENDEAVOR. The company base case includes an allowance for drug wastage, assuming vial sharing is not practised. The model also includes assumptions about dose intensity – the proportion of doses received by the patient before treatment discontinuation (CS Table 44). Due to lack of available data on dose intensity for Cd, the company assumed the same dose intensity as for daratumumab. We agree with this assumption.

**Table 37 Dose intensity**

| Dose Intensity | Component 1 | BOR | DEX | Source |
|---|---|---|---|---|
| DBd | 93.8% | 81.7% | 87.3% | CASTOR |
| Bd | | 87.2% | 90.9% | CASTOR |
| Cd | 93.8% | | 87.3% | Assumption: same as daratumumab |

Source: CS Table 44

**Drug administration and co-medication costs**

Drug administration costs included in the economic model are reproduced from CS Table 45 and summarised in Table 38. The unit costs were sourced from National Schedule of Reference Costs for Year 2016-17. For co-medications, the company sourced the prices from MIMS UK Database (CS Tables 46 and 47). The company assumed that bortezomib would be administered by subcutaneous injection twice weekly. As noted in TA457, clinical opinion suggests that in practice weekly subcutaneous injection is preferred to twice-weekly intravenous infusion as specified in the marketing authorisation, because this is associated with fewer adverse reactions.

**Table 38 Drug administration costs**

| Mode of Administration | Unit Cost | Source |
|---|---|---|
| 1st daratumumab infusion | £385.99 + £3.10 | SB14Z Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance – Daycase and Regular Day/Night admissions + blood sample prior 1st infusion |
| Each subsequent daratumumab Infusion | £205.09 | SB15Z - Deliver Subsequent Elements of a Chemotherapy Cycle - Outpatient |
| Each IV administration of carfilzomib | £205.09 | SB15Z - Deliver Subsequent Elements of a Chemotherapy Cycle - Outpatient |
| Each SC administration of bortezomib | £82.09 | N10AF – Specialist Nursing, Cancer Related, Adult, Face to face |
| Oral drug initiation | £163.82 | SB11Z – Deliver Exclusively Oral Chemotherapy - Outpatient |
| % of SC administration of bortezomib on daratumumab infusion days | 0% | On days when daratumumab is also administered, no additional cost is assumed |

Source: CS Table 45. National Schedule of Reference Costs - Year 2016-17 - NHS trusts and NHS foundation trusts (chemotherapy and community health services)

**Subsequent treatment costs**

The model included costs associated with subsequent treatments, using a simple approach wherein a proportion of patients who discontinued from the initial modelled treatment continued to a basket of potential treatment options. This basket consisted of treatments

which were received by patients in CASTOR, with adjustment for treatments not available in England. The proportion of patients receiving subsequent treatment was obtained from CASTOR for DBd and Bd (70% for DBd and 96% for Bd). For Cd, the company assumed the lower of the proportions observed for DBd and Bd (i.e. 70%). The economic model assumed the same duration of subsequent treatment (9 months) for each RRMM treatment. The distribution of subsequent treatment per treatment arm is presented in Table 39. The treatment acquisition costs of subsequent therapies are the same as the list prices, as in Table 36 above.

**Table 39 Distribution of subsequent treatments**

| | After DBd | After Bd | After Cd |
|---|---|---|---|
| **% continuing to subsequent treatment** | **70%** | **96%** | **70%** |
| Daratumumab monotherapy | 0.0% | 56.2% | 56.2% |
| Ld | 64.7% | 31.5% | 31.5% |
| Pd | 35.3% | 12.3% | 12.3% |
| Cd | 0.0% | 0.0% | 0.0% |

Source: CS Table 48

**Routine follow up**

The economic model analysed costs associated with routine follow-up care for each health state: progression free and progressed disease. The CS assumed same routine follow up care per health state for all the comparators. Previous NICE TAs in MM (ie. TA228 and TA338) were used to inform the types and frequencies of medical resource use. These are detailed in CS Tables 51-52. Costs of treating the included adverse events were also included in the economic model (see Table 35 above). Further, the company included a one-time cost of £7,920 for terminal care at death. This was in line with TA457.

In summary, the ERG considers that the company's approach to costing is appropriate and consistent with related NICE guidance.

### 4.3.8 Model validation

The company describe their approach to model validations in CS section B.3.9. They state that they conducted a range of checks of internal validity, using a checklist spreadsheet to document specific tasks and results. The face validity of the model predictions was assessed by two advisory board meetings with clinical experts and consultation with a UK health economist.

They also report model predictions against observed data from the trials used as data sources and against external evidence. These comparisons showed that:

- **Median PFS** (Table 32 CS B.3.3.1.1): Modelled estimates for second-line DBd and Bd (independently fitted Gompertz) were close to observed values from CASTOR. But the modelled estimate for Cd was very different to the ENDEAVOR value (16.2 months versus 22.2 months respectively). The company argues this is due to differences between the two trials in patient populations and Bd treatment schedule and duration.
- **Median OS** (Table 34 CS B.B.1.2): Median results not yet reached for OS. The comparison with external estimates of survival under standard treatment (Bd) are relatively poor. For example, the Gompertz curve fitted to the Bd 2L subgroup data from CASTOR gave a prediction of 36.8 months for median OS, compared with 19.2 months for an observational HMRN cohort and 24.5 months using the surrogate relationship between PFS and OS (as estimated by Dimopoulos et al. 2017).[61] The company note that the HMRN cohort are older with a poorer prognosis than the trial populations (clarification response to question B2).
- **Visual fit of OS** (Figure 23 CS B.3.3.1.2). This figure shows the fitted parametric curves against selected long-term OS data sources. Comparison with observed survival from Orlowski et al. 2016[62] shows a fair fit to the Weibull function fitted to CASTOR Bd ITT data. We discuss external evidence for selection of model curves further in section 4.4.1 below

### 4.3.9 Cost effectiveness results

Results of the company's base case model are presented as incremental cost effectiveness ratios (ICERs) for DBd compared with Bd and Cd (CS Table 56). Disaggregated costs, LYs and QALYs for each treatment are shown in CS Table 55. At list price for all drugs, the company reported an incremental cost per QALY gained of £41,633 for DBd compared to Bd and £7,180 for DBd compared to Cd (Table 40). In this analysis, Cd is extendedly dominated by DBd and Bd. Thus, the relevant ICER for the full incremental analysis is £41,633 for DBd versus Bd. However, we note that these analyses are conducted at list prices for all drugs, so do not reflect agreed discounts that are available within the NHS.

**Table 40 Cost effectiveness: company base case (list prices)**

| | Total costs (£) | Total LYG | Total QALYs | Pairwise (DBd vs comparator) | | | Full Incremental ICER |
|---|---|---|---|---|---|---|---|
| | | | | Incremental cost (£) | QALYs gained | ICER (£ per QALY gained) | |
| Bd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £41,633 | - |
| Cd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £7,180 | Ext. dom. |
| DBd | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | £41,633 |

Ext. dom, extended dominance.

We present results including PAS price discounts for daratumumab and other comparator and subsequent treatments in a confidential addendum to this report.

### 4.3.10 Assessment of uncertainty

The CS reported a range of sensitivity and scenario analyses to assess structural and parameter uncertainties. The results of these analyses are summarised below.

**Probabilistic sensitivity analysis**

The company conducted a probabilistic sensitivity analysis (PSA) on their base case to assess parameter uncertainty. Assumptions used to characterise uncertainty are described in CS B.3.8.1. Whilst the CS reported that beta distributions were assigned to AE disutilities, the economic model used normal distributions. This is unlikely to have an impact on results. Otherwise, the ERG considers that the parameters and their assigned distributions are appropriate and correctly implemented. PSA results are summarised in CS Table 58 and diagrammatically presented as scatterplots in CS Figure 30 and Cost-Effectiveness Acceptability Curves (CEACs) in CS Figure 31. The PSA results are similar to the deterministic base case results. The CS stated that at a willingness to pay threshold of £30,000 per QALY, the probability of DBd being cost-effective compared to Cd was 92%, but only 19% relative to Bd.

**Deterministic sensitivity analysis**

The company reports parameters and ranges included in their Deterministic Sensitivity Analysis (DSA) in CS Table 59. They state that where information was available, parameters were varied using confidence intervals or published ranges. Otherwise, upper and lower bounds were varied by ±20% of the mean base case value. The ERG considers that all relevant parameters were included in the DSA, with appropriate ranges and we could replicate the results. We observed a few inconsistencies in parameters included in the DSA within the economic model and those reported in the CS (e.g. unit costs for monitoring tests

are listed in CS Table 59 but excluded from DSA in the model), but none of these differences impact on conclusions. The company present DSA results as tornado plots (CS Figures 32 and 33). The tornado plots show that the model was most sensitive to parameters defining the OS curves for Bd and DBd. The very wide ranges illustrate the considerable impact of uncertainty over the prediction of OS for Bd and DBd. Note that parameterisation of the OS curve for Bd impacts on the cost-effectiveness of DBd compared with Cd because the OS curve for Cd is calculated by applying a hazard ratio to the OS curve for Bd. Other parameters including mean body weight (which determines treatment dosing and hence cost) and DBd treatment duration and PFS also influence the base case ICERs, but to a lesser extent.

**Scenario analyses**

The company conducted a range of scenario analyses to assess the impact of structural uncertainties over their base case assumptions. We noted inconsistencies in the ICERs reported in the CS and the values that we obtained from the scenarios relating to longer subsequent treatment duration. The company corrected this in response to Clarification question B5. A summary of the company's scenarios, along with their justifications and the results obtained are presented in Table 41. To note, this table uses the corrected ICER for the longer subsequent treatment scenario mentioned above. The company concluded that most assumptions and alternative scenarios had relatively little impact on the results. They noted that shortening the time horizon had the greatest impact, followed by assuming a Weibull function for DBd OS. The ERG considers that the company has been very selective in the scenarios that they present. In particular, they do not explore the full range of survival functions or the impact of changing OS or PFS functions for more than one treatment at a time. We present a wider range of scenarios in ERG additional analysis below.

**Table 41 Company scenario analyses (list prices) (CS Tables 61 to 63)**

| Scenario and cross reference | Brief rationale | ICER (DBd vs Bd) | ICER (DBd vs Cd) |
|---|---|---|---|
| **Company base case** | | £41,633 | £7,180 |
| **Unadjusted overall survival** | To assess the results by using direct observations of OS from CASTOR | £43,650 | £7,488 |
| **Different survival curves** | | | |
| • PFS exponential | Due to relatively short follow up of the trial data, to extrapolate long term data more emphasis was given to statistical fits and clinical validity | £43,188 | £7,063 |
| • DBd OS Weibull | | £63,066 | £10,631 |
| • DBd OS extrapolation | | £45,011 | £7,811 |
| • Bd OS Weibull | | £49,278 | £8,376 |
| • TTD exponential | | £40,942 | £7,020 |
| **Longer subsequent treatment duration** | | | |
| • 13 months * | Subsequent treatment is an important element in the economic analysis as MM patients receive multiple lines of treatment which influence the costs as well as health benefits | £41,959 | £8,774 |
| • 15 months * | | £42,326 | £9,675 |
| **Patients continuing subsequent treatment** (85.71%) | | £46,174 | £6,883 |
| **Health state utilities** (van Agthoven) | To assess the impact of using utility values from previous models and technology assessments | £42,515 | £7,538 |
| **Generic Price for Bortezomib** | Bortezomib to lose its patent in 2019 which potentially could lead to a decrease in price by 50% | £42,507 | £7,180 |
| **Different time horizons** | | | |
| • 5 years | To test structural assumptions | £195,480 | £18,202 |
| • 10 years | | £77,061 | £12,333 |
| • 15 years | | £55,120 | £9,051 |
| • 20 years | | £46,910 | £7,860 |
| • 25 years | | £43,142 | £7,354 |
| **Allow vial sharing** | | £40,222 | £13,586 |
| **Dose intensity option off** | | £45,072 | £8,634 |
| **Discounting** | | | |
| • Utilities 0%; Costs 0% | To demonstrate the effect of discounting | £30,303 | £6,469 |
| • Utilities 1.5%; Costs 0% | | £36,355 | £7,816 |
| • Utilities 6.0%; Costs 0% | | £58,330 | £12,821 |
| • Utilities 0%; Costs 1.5% | | £29,127 | £5,632 |
| • Utilities 1.5%; Costs 1.5% | | £34,943 | £6,804 |
| • Utilities 6%; Costs 1.5% | | £56,064 | £11,161 |
| • Utilities 0%; Costs 6% | | £26,343 | £3,788 |
| • Utilities 1.5%; Costs 6% | | £31,604 | £4,577 |
| • Utilities 6%; Costs 6% | | £50,706 | £7,508 |

Source: CS Table 62. * ERG corrected values (Company response to clarification question B5)

## 4.4 Additional work undertaken by the ERG

### 4.4.1 ERG model validation

#### 4.4.1.1 Model verification procedures

We conducted a range of manual checks to verify model inputs, calculations and outputs ('white box' tests). This included:

- Cross-checking of all parameter inputs against values in the CS and cited sources;
- Tracing input parameters from entry cells to the "Parameters" sheet, survival curve and 'Model Engine' (the Markov trace) sheets;
- Use of PFS, OS and TTD results to estimate the distribution of the cohort by health state and the numbers of events over time in the 'Model Engine' sheet
- We checked QALY and cost calculations in the Model Engine;
- And the links from the total costs and outcomes from the Model Engine back to the results tables
- We checked all model outputs against results cited in the CS, including the base case, PSA and DSA and we manually ran scenarios.

In addition, to check the company curve fitting and coding of the survival function in the spreadsheet, we re-estimated exponential and Weibull survival functions based on KM estimates of survival data from the company submitted model. We linearised the survival functions and used least squared methods to estimates the parameters.

#### 4.4.1.2 Relationship between PFS and OS

Felix et al.[53] conducted a meta-regression analysis to estimate the relationship between median OS and median time dependent surrogate outcomes (TDE); which included time to progression (TTP); progression-free survival (PFS); and event-free survival (EFS). OS and TDE results and study-level covariates were extracted from 153 prospective observational and experimental studies with 230 study arms (22,696 patients) identified from a systematic review. Two-stage least squares regression was used to control for endogeneity (confounding) and heteroskedasticity (non-constant variance) in the study-level data. The authors' preferred model specification was adjusted for year of publication, demographics (age and gender), patient type (newly diagnosed MM, RRMM, or mixed), type of TDE (TTP, PFS or EFS) and included censored outcomes. Results for alternative model specifications ranged from 1.82 to 2.64 (compared with 2.45 in the preferred specification), all with overlapping confidence intervals. Trials reporting results for patients with RRMM reported a higher median survival than trials in patients with newly diagnosed MM (p=0.06), though it is

unclear whether the surrogate relationship differs by stage of MM. The model was reasonably accurate at predicting median OS for first-line treatments, with 13 out of 16 observed values lying within the 95% confidence interval for the prediction for RCTs comparing thalidomide-based treatments (although the confidence intervals were wide). The model was less good at predicting OS in RCTs for RRMM: with 3 out of 4 observed values lying outside the prediction interval.

In the CS, the company use an estimate of the relationship between OS and PFS as supporting evidence to validate their model results. They rely on results from an abstract by Dimopoulos et al. (2017), which criticised the previous analysis for combining studies of newly-diagnosed and relapsed and refractory MM and noted that several new RCTs of treatments for RRMM are now available.[61] Dimopoulos et al. analyse OS, PFS and co-variates from 22 RCTs (including 7,884 RRMM patients) identified through a systematic review. After adjusting for median age, sex and publication year, median OS was estimated to be 3.10 months (95% CI 2.20 to 4.00) longer for each additional month of median PFS. This study is only reported in abstract form, without detailed description of methods or declaration of interests. And there is no information about the accuracy of the model predictions. The ERG therefore considers that the predicted relationship between PFS and OS from the Dimopoulos et al.[61] and Felix et al.[53] studies provide weak evidence for model validation.

#### 4.4.1.3 Comparison with long-term survival data

The company present a comparison of OS curves fitted to the Bd arm of CASTOR against long-term survival data (CS Figure 23). Sources of external data included over 8 years of follow up from a trial of bortezomib-based therapies for patients with RRMM with one or more previous lines of therapy (Orlowski et al. 2016)[62] and 5-year follow up from the Bd arm of the PANORAMA 1 trial in patients with RRMM and 1 to 2 previous treatments (San-Miguel et al. 2016).[63] We reproduce KM curves from these two studies and the Bd arm of CASTOR alongside the company's parametric curves fitted to Bd CASTOR data in Figure 12. This illustrates the wide spread of long-term projections from the fitted parametric curves. For comparison, Cancer Research UK statistics indicate 47% of newly diagnosed MM patients surviving to 5 years and 33% surviving to 10 years. One might expect rather worse survival rates for patients starting second-line treatment after relapse. Two clinical experts advising the ERG suggested 5-year survival rates in this population of 50% and 40% and 10-year rates of 15% and 20%. We conclude that the Weibull curve gives the best fit to the long-term external data. The NICE TA457 committee came to a similar conclusion regarding the validity of extrapolations from the ENDEAVOR trial.

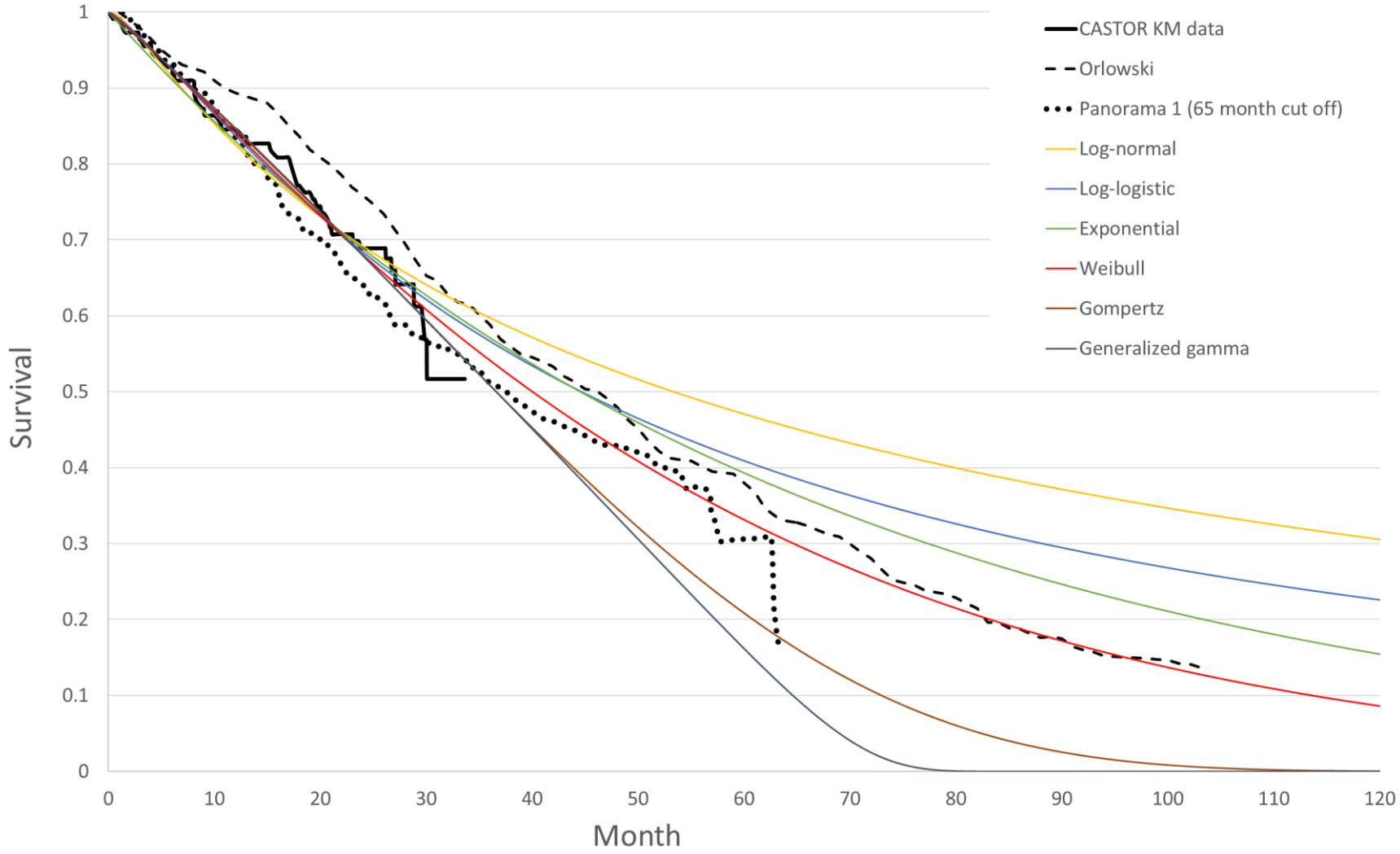



**Figure 12 Bd OS projections compared with external data (adapted from CS Figure 23)**

#### 4.4.1.4 Comparison with other model outcomes

In section 4.2.2 above we described three other models that provide outcome estimates for patients treated with at least one of the comparators (DBd, Bd or Cd) after one prior therapy for MM. These provide sources for cross-validation of results from the company base case analysis. There are methodological differences between the models, as well as differences in the decision problem. Modelled estimates of mean discounted progression-free and overall life years and QALYs are summarised in Table 42.

**Table 42 Comparison of modelled outcomes**

| Treatment | Outcome | Discounted mean (years) [a] | | | |
|---|---|---|---|---|---|
| | | **Carlson et al. 2018** [2] | **Maiese et al. 2017** [38] | **TA457** (preferred assumptions) | **Submitted model** (base case) |
| **Bd** | PFS | 1.83 | 0.80 | | 0.83 |
| | LY | 3.73 | 1.90 | 3.34 | [redacted] |
| | QALY | 2.74 | 1.55 | 2.20 | [redacted] |
| **Cd** | PFS | | | | 1.57 |
| | LY | | | 5.87 | [redacted] |
| | QALY | | | 3.96 | [redacted] |
| **DBd** | PFS | | 2.56 | | 2.32 |
| | LY | | 5.74 | | [redacted] |
| | QALY | | 4.78 | | [redacted] |

a Discount rate 3.5% per year for TA457 and company base case, 3% per year for Carlson and Maiese results.

The most relevant analysis for the current appraisal is the final version from the recent NICE TA of Carfilzomib (TA457), reflecting the committee's preferred assumptions. This appraisal relates to the subgroup of patients with RRMM after one prior therapy not including Bortezomib and the results are adjusted for a limited duration of bortezomib treatment (24 weeks), rather than ongoing treatment. Compared with the final TA457 model, the company's base case produces similar mean LY and QALY estimates for Bd, but lower estimates for Cd. This contributes to a very high estimated ICER for Cd vs. Bd from the company's base case model that does not align with the 'most plausible' estimate from TA457 (even accounting for the Carfilzomib PAS discount). We address this inconsistency in additional ERG analysis below: aligning our preferred version of the company model to reflect committee preferences and the ICER estimate from TA457.

Results from the base case model as submitted are also more favourable than the company funded analysis reported by Maiese et al.

### 4.4.2 ERG corrections to company analyses

We identified a few errors in the company's original model: see Table 43 below. The company corrected issues 1 and 3 as responses to the clarification questions and submitted a revised model for their base case. The ERG implemented the correction in the costs estimation of Bd (Problem 2 in Table 43) in this version of the company's model. The ERG corrections resulted in a small increase in the ICERs (see Table 44).

**Table 43 ERG corrections to company model**

| Aspect of model | Problem | ERG Correction |
|---|---|---|
| Cost calculations | 1. Costs of dexamethasone: CS reported cost as £120.03 for 50X8mg; instead of £200 for 10x40mg | The company corrected their base case model as response to clarification question B3. |
| | 2. Costs of BD: The model includes administration and co-medication costs of BD beyond the duration of treatment administration which is for 24 weeks in CASTOR | Recoded column CQ and DD in sheet 'Drug Cost calculations' |
| Scenario analysis | 3. Scenarios relating to longer subsequent treatment duration of 13 months and 15 months gave base case ICERs for DBd vs Bd and DBd vs Cd | Corrected in company's response to clarification question B5. |

**Table 44 Cost-effectiveness: ERG corrected company base case (list prices)**

| | Total costs (£) | Total LYG | Total QALYs | Pairwise (DBd vs comparator) | | | Full Incremental ICER |
|---|---|---|---|---|---|---|---|
| | | | | Incremental cost (£) | QALYs gained | ICER (£ per QALY) | |
| Bd | [redacted] | | | | | £42,190 | - |
| Cd | | | | | | £7,188 | Ext. dom. |
| DBd | | | | - | - | - | £42,190 |

Ext. dom, extended dominance.

Sensitivity and scenario analysis results for the ERG corrected version of the company's model are shown below in Figure 13 to Figure 16, Table 45 and Table 46.

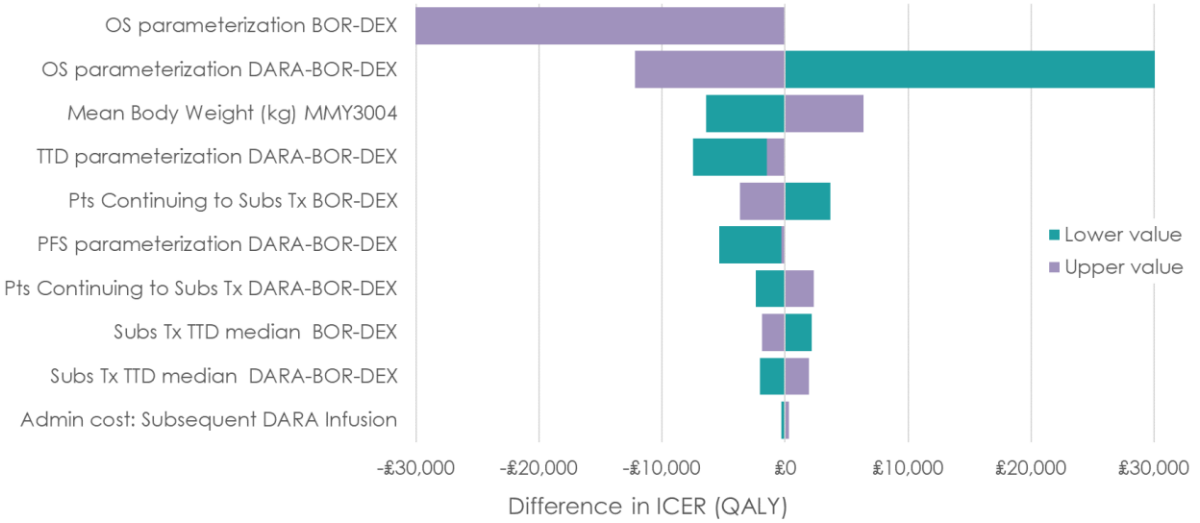



**Figure 13 Tornado plot DBd vs Bd: ERG corrected company base case (list prices)**

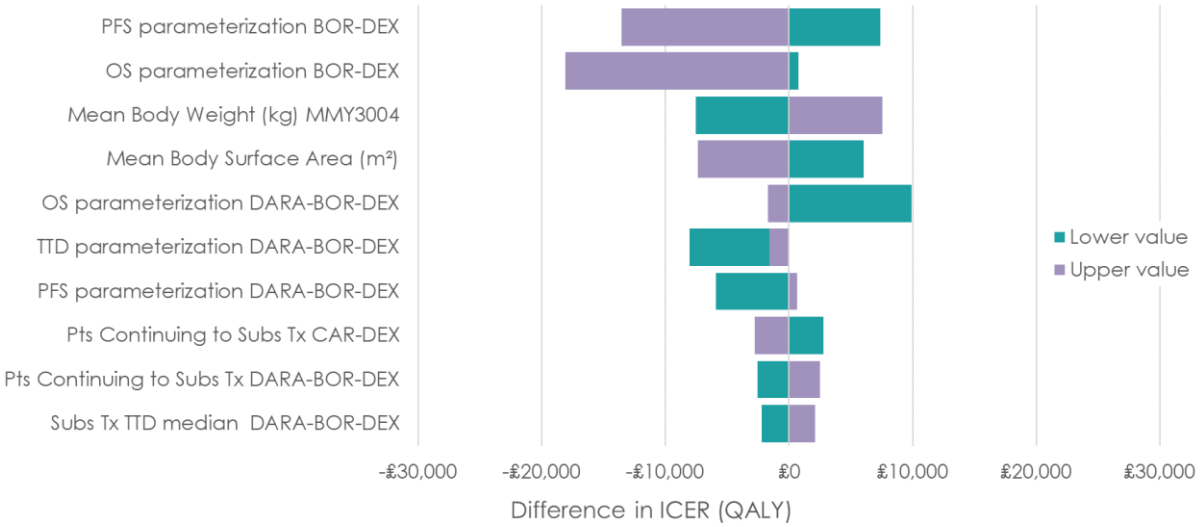



**Figure 14 Tornado plot DBd vs Cd: ERG corrected company base case (list prices)**

**Table 45 PSA results from ERG corrected company base case (list prices)**

| Comparator | Mean LY | Mean QALY | Mean Total cost |
|---|---|---|---|
| DBd | [redacted] | | |
| Bd | | | |
| Cd | | | |

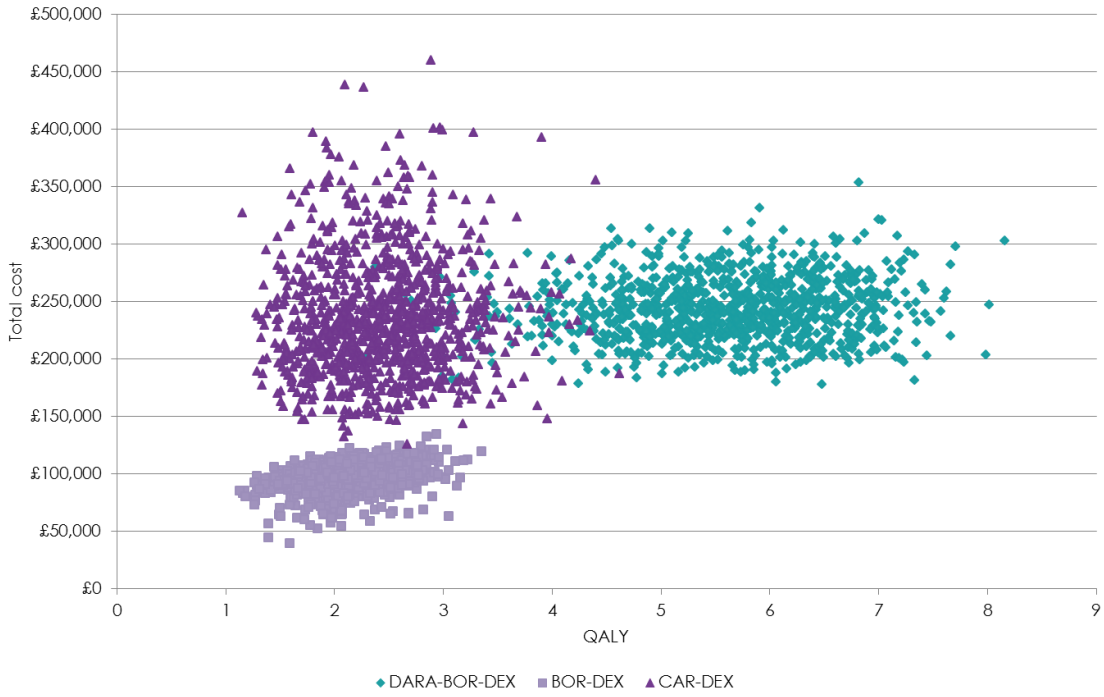



**Figure 15 Scatter plots for ERG corrected company base case (list prices)**

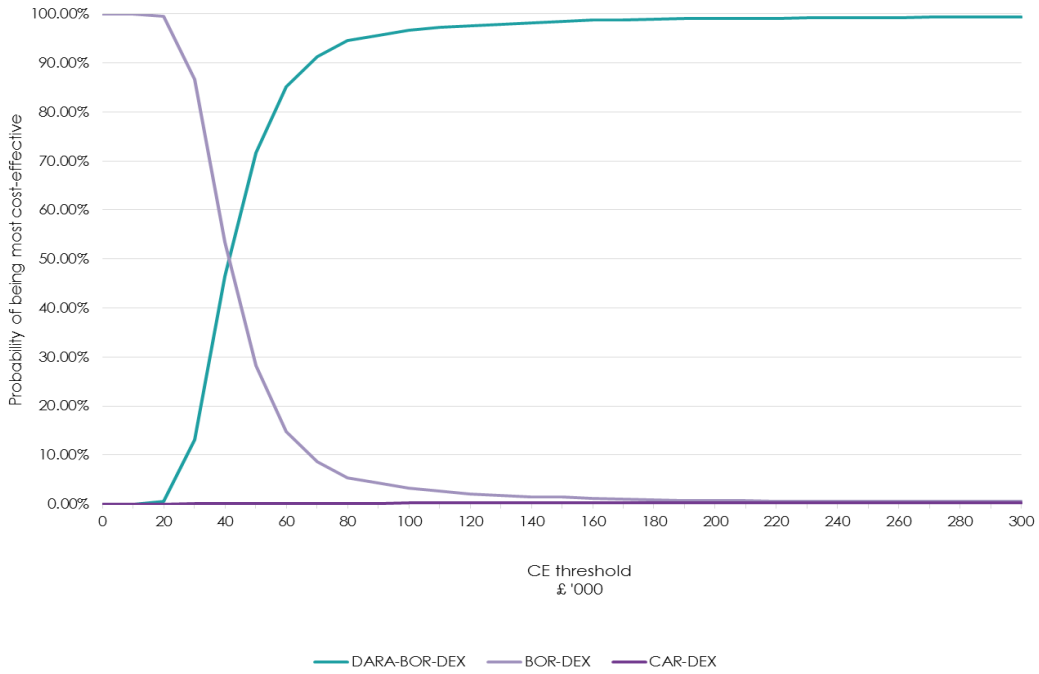



**Figure 16 CEACs for ERG corrected company base case (list prices)**

**Table 46 Scenario analysis: company base case, ERG corrected (list prices)**

| Scenarios | | ICER (DBd vs Bd) | ICER (DBd vs Cd) |
|---|---|---|---|
| Company base case (ERG corrected) | | £42,190 | £7,188 |
| PFS curves | DBd: Exponential / Bd: Exponential | £43,739 | £7,066 |
| | Piecewise: KM up to 12 months | £41,545 | £7,182 |
| OS curves | DBd: Weibull / Bd: Gompertz | £63,919 | £10,642 |
| | DBd: Exponential / Bd: Gompertz | £45,615 | £7,821 |
| | DBd: log-logistic / Bd: Weibull | £49,945 | £8,385 |
| | DBd: Weibull / Bd: Weibull | £84,534 | £14,978 |
| | Piecewise: KM up to 27 months | £42,049 | £7,996 |
| | Unadjusted for subsequent treatment | £44,236 | £7,496 |
| TTD curves | DBd: Exponential / Bd: Exponential | £42,040 | £7,027 |
| Utilities | van Agthoven | £43,085 | £7,546 |
| | ENDEAVOR mapped (TA457) | £44,565 | £7,700 |
| | Carlson et al. 2018 | £42,470 | £7,435 |
| Subsequent treatment | Duration 13 months | £42,529 | £8,799 |
| | Duration 15 months | £42,903 | £9,708 |
| | Patients continuing 85.7% | £46,750 | £6,900 |
| | 100% Ld at third line | £46,311 | £9,356 |
| Bortezomib price | Generic after end of patent | £44,635 | £7,188 |
| Vial sharing | Allow (no wastage) | £40,779 | £13,594 |
| Dose intensity | Do not apply (100% of dose) | £45,633 | £8,637 |
| Time horizons | 5 years | £198,365 | £18,182 |
| | 10 years | £78,119 | £12,349 |
| | 15 years | £55,868 | £9,063 |
| | 20 years | £47,543 | £7,869 |
| | 25 years | £43,722 | £7,363 |
| Persistence of OS effect | Hazards equal after 5 years | £150,843 | £25,108 |
| | Hazards equal after 10 years | £76,431 | £13,205 |
| | Hazards equal after 20 years | £47,811 | £8,168 |
| Mortality from PF state | Ratio of deaths to progressions 5% | £41,455 | £7,045 |
| | Ratio of deaths to progressions 10% | £41,840 | £7,120 |
| | Ratio of deaths to progressions 20% | £42,422 | £7,274 |
| Discounting | Utilities 0%; Costs 0% | £30,690 | £6,477 |
| | Utilities 1.5%; Costs 0% | £36,818 | £7,826 |
| | Utilities 6.0%; Costs 0% | £59,073 | £12,836 |
| | Utilities 0%; Costs 1.5% | £29,506 | £5,638 |
| | Utilities 1.5%; Costs 1.5% | £35,399 | £6,812 |
| | Utilities 6%; Costs 1.5% | £56,796 | £11,174 |
| | Utilities 0%; Costs 6% | £26,706 | £3,792 |
| | Utilities 1.5%; Costs 6% | £32,039 | £4,582 |
| | Utilities 6%; Costs 6% | £51,405 | £7,516 |

### 4.4.3 ERG preferred analysis and scenarios

Table 47 lists the ERG's preferred assumptions, and further explanation is given in Table 48.

**Table 47 ERG's preferred modelling assumptions**

| | Company base case | ERG preferred assumptions |
|---|---|---|
| Patient group | • Second line (2L) | • 2L Bortezomib-naïve (2LBN) |
| OS curves | • DBd: Log-logistic<br>• Bd: Gompertz<br>• Cd: Proportional hazard assumption; HR vs Bd | • Bd: Weibull<br>• DBd and Cd: HR vs. Bd (proportional hazard assumptions) |
| PFS curves | • DBd: Gompertz<br>• Bd: Gompertz<br>• Cd: Proportional hazard assumption; HR vs Bd | • Bd: Gompertz<br>• DBd and Cd: HR vs. Bd (proportional hazard assumptions) |
| TTD curves | • DBd: Gompertz<br>• Bd: Gompertz<br>• Cd: Proportional hazard to PFS for 2LBN | • DBd: Gompertz<br>• Bd: Gompertz<br>• Cd: Proportional hazard to PFS for 2LBN |
| Utilities | • EQ-5D-5L from CASTOR mapped to EQ-5D-3L using van Hout crosswalk procedure | • Utility values from TA457: ENDEAVOR EORTC data mapped to EQ-5D using Proskorovsky algorithm |
| Treatment effectiveness | • HR for Cd vs Bd for PFS and OS from ENDEAVOR not adjusted for duration of bortezomib treatment | • Adjustment of the HR for Cd vs Bd for PFS and OS to reflect 24 week bortezomib treatment in practice (from TA457) |
| Subsequent treatment | • OS adjusted for subsequent treatments used in trials but not available in England | • Same as company model |

**Table 48 ERG preferred assumptions and scenarios**

| Aspect of the model | Company base case | ERG Preferred assumptions | ERG scenarios | Reason for analysis |
|---|---|---|---|---|
| OS curves | DBd: log-logistic<br>Bd: Gompertz | DBd: Weibull<br>Bd: Weibull | We test the effect of assuming proportional hazards to model DBd in relation to the independently fitted Weibull curve for Bd<br><br>KM data with parametric tail | We prefer the Weibull distribution for OS because:<br>• In TA457, the committee preferred assumptions included Weibull curves for OS from the ENDEAVOR trial, jointly fitted for Cd and Bd and with an assumption of proportional hazards.<br>• Validation against external data (Orlowski et al. 2016[62]) indicates that the Weibull gives realistic long-term predictions of survival with under current treatment.<br>• The Weibull provided a good visual fit for Bd and DBd KM curves from CASTOR and CS Table 33 showed that Weibull had the second lowest AIC and BIC values for model fit<br>• CS Appendix L Figure 6 lends support to the assumption of proportional hazards between the two arms of DBd and Bd.<br>• We consider it appropriate to use the same survival function for both treatment arms.<br>For scenario analysis, we use the KM data with a cut-off point where number of people at risk is at least 20: 27 months, when 39 and 47 patients are still at risk in Bd and DBd arms respectively. The aim of this analysis is to test the impact of a better fit to trial data. |
| Patient group | Second line RRMM (2L) | Second line Bortezomib-naïve RRMM (2LBN) | We test the effect removing the 2LBN adjustment | We adjust the OS and PFS curves for DBd relative to the fitted curves for Bd using the reported hazard ratios from the company's NMA for the 2LBN subgroup (presented in CS Table 24). |
| PFS curves | DBd: Gompertz<br>Bd: Gompertz | Same as company | Test effect of modelling DBd in relation to Bd curve.<br><br>KM data parametric tail. | We agree that the Gompertz function provides the best fit to CASTOR PFS data for both treatment arms. CS Appendix L Figure 3 indicated that proportional hazard assumption does not hold for DBd vs Bd.<br><br>KM data up to a point with at least 20 people still at risk in both arms: 12 months, 22 patients remain in the Bd arm and 89 in the DBd arm. |

| Aspect of the model | Company base case | ERG Preferred assumptions | ERG scenarios | Reason for analysis |
|---|---|---|---|---|
| Treatment effect | Persistence of OS and PFS benefits | Same as company | Assume loss of survival benefit for Cd and DBd, relative to Bd, from 5 to 20 years | Given the immature OS data from CASTOR, we test a more conservative approach and explore scenarios where the mortality hazards for DBd and Cd equal to those of Bd after 5, 10 and 20 years. |
| | No adjustment of Cd vs. Bd hazard ratios from ENDEAVOR for shorter duration of Bd treatment in practice (24 weeks) | HR for Cd vs. Bd adjusted for 24-week duration of treatment with bortezomib, as in TA457 | Analysis is conducted without the adjustment for Bd treatment duration | To align with the NICE committee's conclusion in TA457 that adjustment should be made for the treatment duration of bortezomib in ENDEAVOR (until progression) compared with marketing authorisation (24-weeks). This adjustment also addresses the difference in Bd treatment duration in the two trials used for the indirect comparison of DBd with Bd. To apply this adjustment, we use the relative increase in the hazards when bortezomib is stopped after 24 weeks: 1.36 for PFS and 1.46 for OS, reported in Carfilzomib ERG review of additional evidence (April 2017). |
| Utilities | EQ-5D-5L from CASTOR mapped to EQ-5D-3L using van Hout crosswalk algorithm | Utilities as in TA457: ENDEAVOR data mapped to EQ-5D with Proskorovsky algorithm | Other sources of utilities: CASTOR (company base case); Carlson et al.[2]; and van Agthoven[3]. | We are concerned about the poor reporting of methods used to estimate utilities from CASTOR EQ-5D data. Details of the methods of analysis and results have not been supplied. We also consider the small difference between pre and post-progression utilities to be implausible. ENDEAVOR utilities also provide consistency with values used in the recent appraisal of carfilzomib (TA457). |
| Subsequent treatment | DBd and Bd OS adjusted for treatments not available in England (IPCW method) Cd not adjusted | Same as company | Use the unadjusted survival model | It is appropriate to adjust OS for differences in subsequent treatments given in the trial and current provision in England. The CS adjusted OS curve estimates for DBd and Bd based on CASTOR data. However, no adjustment was made for Cd arm, due to the lack of individual level data from ENDEAVOR. This inconsistency could bias results, so we test the effect of using unadjusted survival data for all the three arms. |
| Mortality | Deaths from the PFS state estimated using fixed ratio of deaths to progression (0.146) | Same as company | Vary ratio of deaths to progression | Exploratory analysis |

The impact of the ERG preferred assumptions on cost-effectiveness results at list prices are shown in Table 49. This table includes step-by-step changes made to the company's base case model, which cumulative comprise the ERG preferred analysis. The change that has the biggest impact is the adoption of Weibull distributions for OS, which doubles both pairwise ICERs. Adding the assumption of proportional hazards for DBd vs. Bd on its own makes little difference to the results. Adjusting for the 2LBN population causes a modest increase in the ICERs, while adjusting the Cd vs. Bd HR estimated from ENDEAVOR for the shorter duration of Bd treatment makes Cd relatively less cost-effective. Finally, the change to health state utilities makes little difference to estimated cost-effectiveness.

**Table 49 Cumulative effect of ERG preferred assumptions (at list prices)**

| | Comparator | Total cost | Total QALYs | Pairwise ICERs (DBb vs comparators) |
|---|---|---|---|---|
| **Company base case (ERG corrected)** | DBd | [redacted] | | |
| | Bd | | | £42,190 |
| | Cd | | | £7,188 |
| **OS curves** | **Independent Weibull curves for DBd and Bd** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £84,534 |
| | Cd | | | £14,978 |
| **PFS and OS** | **+ DBd estimated relative to Bd** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £83,565 |
| | Cd | | | £17,759 |
| **Adjustment for 2LBN subgroup** | **+ HR for DBd and Cd vs. Bd from 2LBN subgroup analysis** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £91,816 |
| | Cd | | | £19,225 |
| **Adjustment for 24-week Bd treatment** | **+ HR for Cd vs Bd adjusted for 24-week Bd (as in TA457)** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £91,816 |
| | Cd | | | DBd dominates |
| **Utilities** | **+ ENDEAVOR mapped (same utilities as in TA457)** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £93,061 |
| | Cd | | | DBd dominates |
| **ERG preferred analysis** | **All the above assumptions** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £93,061 |
| | Cd | | | DBd dominates |

Results of the ERG's preferred analysis are detailed in Table 50. The pairwise ICERs are £92,071 for DBd vs Bd; whereas DBd dominates Cd (Cd is estimated to be more expensive and less effective than DBd at list prices).

**Table 50 Cost-effectiveness: ERG preferred base case (list prices)**

| | Total costs (£) | Total LYG | Total QALYs | Pairwise (DBd vs comparator) | | | Full Incremental ICER |
|---|---|---|---|---|---|---|---|
| | | | | Incremental cost (£) | QALYs gained | ICER (£ per QALY gained) | |
| Bd | [redacted] | | | | | £93,061 | - |
| Cd | | | | | | DBd dominates | - |
| DBd | | | | - | - | - | £93,061 |

**Table 51 Additional ERG scenarios**

| Aspect of the model | ERG scenarios |
|---|---|
| **Patient population** | All second line (2L): includes patients with prior bortezomib |
| **Treatment effects** | OS and PFS extrapolation:<br>• For PFS, KM data till 12 months, then Gompertz<br>• For OS, KM data till 27 months, then Weibull |
| | No adjustment of OS for subsequent treatment |
| **Persistence of effects** | Waning for OS: HR=1 for DBd vs. Bd and Cd vs. Bd after:<br>• 5 years<br>• 10 years<br>• 20 years |
| **Time horizon** | No waning with model time horizon of:<br>• 5 years<br>• 10 years<br>• 20 years |
| **Utilities** | Source of health state utilities (PF and PD):<br>• CASTOR trial (company base case)<br>• van Agthoven 2004<br>• Carlson et al. 2018 |
| **Mortality** | Vary ratio of pre-progression deaths to progression:<br>• 5%,<br>• 10%<br>• 20% |
| **Resource use and costs** | Longer subsequent treatment duration:<br>• 15 months |

| Aspect of the model | ERG scenarios |
|---|---|
| | Subsequent treatment mix based on expert feedback<br>• 100% for Lenalidomide and dexamethasone (Ld) |
| | Allow vial sharing (no wastage) |
| | No dose intensity considered (costs for 100% of dose) |

We performed a range of additional scenario analyses on the ERG preferred base case: listed in Table 51, with results in Table 52.

Under this list price analysis, DBd dominates Cd in all the scenarios except one (scenario when vial sharing is allowed). The ICERs for DBd vs Bd ranged between £83,858 (scenario: 2L patient population) and £248,621 (scenario: model time horizon of 5 years). However, these results do not take account of PAS discounts that are available in the NHS for daratumumab, carfilzomib and bortezomib, and subsequent treatments. It is therefore not possible to draw conclusions about cost-effectiveness from the results presented in the body of this report. We present a confidential addendum that repeats our analyses with costs accounting for all agreed PAS and CAA price discounts.

We do note however, that with PAS discounts, the ICER for the comparison of Cd with Bd under the ERG preferred analysis closely mirrors the TA457 committee's 'most plausible' ICER. This is not surprising because the two analyses share the same population (2LBN) and build from the same clinical evidence for the comparison of Cd with Bd (the ENDEAVOR trial). We have also adopted the TA457 committee's preferred assumptions relating to the choice of OS curves (Weibull for both comparators), the adjustment of ENDEAVOR results for the duration of bortezomib treatment and source of health state utilities, because we think they are equally relevant to this current appraisal.

**Table 52 Scenarios analysis: ERG preferred base case (list prices)**

| Scenario | Comparator | Total cost | Total QALY | Pairwise ICER (DBb vs comparator) |
|---|---|---|---|---|
| **ERG preferred base case** | DBd | [redacted] | | |
| | Bd | | | £93,061 |
| | Cd | | | DBd dominates |
| **Patient population** | ***2L (no adjustment of DBd relative to Bd for PFS and OS)*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £84,804 |
| | Cd | | | DBd dominates |

| Scenario | Comparator | Total cost | Total QALY | Pairwise ICER (DBb vs comparator) |
|---|---|---|---|---|
| **Treatment effects** | ***KM data for OS and PFS and distributions fitted to the tails*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £89,258 |
| | Cd | | | DBd dominates |
| | ***OS not adjusted for subsequent treatments*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £96,440 |
| | Cd | | | DBd dominates |
| **Persistence of effects** | ***Equal OS hazards for DBd, Cd and Bd from 5 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £154,485 |
| | Cd | | | DBd dominates |
| | ***Equal OS hazards for DBd, Cd and Bd from 10 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £110,234 |
| | Cd | | | DBd dominates |
| | ***Equal OS hazards for DBd, Cd and Bd from 20 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £94,656 |
| | Cd | | | DBd dominates |
| **Time horizon** | ***5 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £238,026 |
| | Cd | | | DBd dominates |
| | ***10 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £134,555 |
| | Cd | | | DBd dominates |
| | ***20 years*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £97,279 |
| | Cd | | | DBd dominates |
| **Utilities** | ***Trial based: CASTOR*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £91,816 |
| | Cd | | | DBd dominates |
| | ***Van Agthoven*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £86,595 |
| | Cd | | | DBd dominates |
| | ***Carlson et al.*** | | | |

| Scenario | Comparator | Total cost | Total QALY | Pairwise ICER (DBb vs comparator) |
|---|---|---|---|---|
| | DBd | [redacted] | | |
| | Bd | | | £85,473 |
| | Cd | | | DBd dominates |
| **Mortality** | ***Ratio of death to progression 5%*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £91,370 |
| | Cd | | | DBd dominates |
| | ***Ratio of death to progression 10 %*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £92,296 |
| | Cd | | | DBd dominates |
| | ***Ratio of death to progression 20 %*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £93,651 |
| | Cd | | | DBd dominates |
| **Resource use** | ***Percentage of patients starting subsequent treatment (85.7%)*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £103,003 |
| | Cd | | | DBd dominates |
| | ***Duration of subsequent treatment: 15 months*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £92,045 |
| | Cd | | | DBd dominates |
| | ***Subsequent treatment mix: 100% Ld (expert opinion)*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £102,457 |
| | Cd | | | DBd dominates |
| | ***Allow vial sharing (no wastage)*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £89,878 |
| | Cd | | | £19,489 |
| | ***No dose intensity (100% of patients use full dose)*** | | | |
| | DBd | [redacted] | | |
| | Bd | | | £100,797 |
| | Cd | | | DBd dominates |

# 5 End of life

The NICE end of life treatment criteria are not applicable and are not included in the CS.

# 6 Innovation

The CS makes the case (CS Section B.2.12) that daratumumab is the first and only licensed human monoclonal antibody that binds the CD38 antigen which is expressed on MM tumour cells. CD38 is also expressed on other immune system cell types which are associated with decreased immune function and disease progression. Daratumumab therefore targets MM directly as well as modulating the CD38-expressing immune-suppressing cells. This combination of direct and indirect effects is believed to explain the efficacy of daratumumab.

The CS describes DBd as a step-change in the management of MM making the case that DBd "*offers patients a second chance at a front-line prognosis*". The company believe that DBd will result in a change in the management of RRMM patients from the current situation of limiting relapses to one of prolonging remission.

# 7 DISCUSSION

## 7.1 Summary of clinical effectiveness issues

*The company's decision problem population is narrower than the NICE scope*

The company have limited their decision problem to second-line patients (i.e. those who had received one prior therapy) which is a narrower population than described in the NICE scope. Of the 498 patients randomised in the CASTOR trial (DBd n=251; Bd n=247), just under half were second-line patients (47.2%, DBd n=122; Bd n=113). The whole CASTOR trial population includes patients who had received two prior therapies (28.9%), three prior therapies (13.9%) and more than three prior therapies (10%). Outcomes from subgroup analyses including by number of prior lines of therapy are presented in CS Appendix E for PFS, TTP, ORR and VGPR or better but no economic analysis is presented for these groups.

*Clinical effectiveness evidence for bortezomib-naive second-line patients from the CASTOR trial is not presented*

As described in CS Section B.1.3.2, the only recommended therapies in England for patients who are refractory to or have relapsed after their first-line therapy (i.e. second-line patient) are: a bortezomib-based therapy; Cd; or combination chemotherapy. Bortezomib-based therapy and Cd are currently only available to those second-line patients who have not previously received bortezomib but, aside from NMA sensitivity analyses for PFS and OS, the company does not present separate clinical effectiveness evidence from the CASTOR trial for bortezomib-naive second-line patients. Just under half of the second-line patients in the CASTOR trial had not previously received bortezomib (DBd n= 60; Bd n= 56). Therefore less than a quarter (23%) of the patients within the CASTOR trial, if they were located in England, would meet the criteria for receipt of a bortezomib-based therapy or Cd as their second-line treatment option.

*Combination chemotherapy is not included as a comparator*

All second-line patients (regardless of prior bortezomib exposure) could receive combination chemotherapy as their next therapy. However only one trial of combination chemotherapy was identified by the company's systematic review and this did not reflect current clinical practice nor could it be linked to the NMA. Consequently the CS does not include a comparison of DBd versus combination chemotherapy as a second-line treatment.

*Data for overall survival data are immature*

The CASTOR trial is ongoing and the results presented are from an interim analysis with 26.9 months follow-up. Data are therefore immature and for OS this means that in the subgroup of second-line patients median survival has not been reached in either arm however there is a statistically significant OS benefit (HR for DBd versus Bd OS 0.50, 95% CI 0.30, 0.84; p=0.008). No OS Kaplan-Meier curve has been presented for the subset of bortezomib-naive second-line patients but the NMA sensitivity analysis produced a similar HR as for second-line patients overall but a credible interval that crosses one (HR for DBd versus Bd OS in bortezomib-naive second-line patients 0.53, 95% Cr 0.24, 1.17). Therefore there is some uncertainty regarding the extent of the OS benefit, particularly for bortezomib-naive second-line patients.

## 7.2 Summary of cost effectiveness issues

*Patient population*

The company's analyses are on 2L patients; they do not distinguish between 2L and 2LBN patients. As noted throughout the document, clinical practice in England differs for patients who have received prior bortezomib in comparison those who have not; Cd and Bd are not routinely used for 2LBN patients. Further, TA457 (the most recent appraisal on MM) recommended Cd in for 2LBN patients, based on sub-group of the ENDEAVOR trial. In line with this, we view that it is only appropriate to compare DBd with Bd and Cd for the subgroup of 2LBN patients.

*Extrapolation of OS*

For the base case analysis, the company used the log-logistic function for DBd and the Gompertz function for the Bd arm to extrapolate OS curves beyond the trial period. We do not believe the company provides sufficient justification for the selection of curves. We view that the Weibull distribution provides the best fit for the following reasons:

- In TA457, the committee's preferred assumptions included Weibull curves for OS from the ENDEAVOR trial, jointly fitted for Cd and Bd and with an assumption of proportional hazards.
- Validation against external data (Orlowski et al. 2016) indicates that the Weibull gives realistic long-term predictions of survival under current treatment.
- The Weibull provided a good visual fit for Bd and DBd KM curves from CASTOR, and CS Table 33 showed that Weibull had the second lowest AIC and BIC values for model fit.
- CS Appendix L Figure 6 lends support to the assumption of proportional hazards between the two arms of DBd and Bd.
- We consider it appropriate to use the same survival function for both treatment arms.

*Treatment effect of Cd vs Bd*

The company does not adjust for the difference in treatment duration for bortezomib in the ENDEAVOR and CASTOR trials. Whilst bortezomib is administered until disease progression in ENDEAVOR, in CASTOR the drug is administered for 24 weeks. This is in line with the marketing authorisation. This issue was discussed in TA457. To align with the NICE committee's conclusion in TA457, that adjustment should be made for the treatment duration of bortezomib in ENDEAVOR (until progression) compared with marketing authorisation (24-weeks), we view it is appropriate to use a relative increase in the hazards when bortezomib is stopped after 24 weeks (which is: 1.36 for PFS and 1.46 for OS as reported in Carfilzomib ERG

review of additional evidence, April 2017). This adjustment also addresses the difference in Bd treatment duration in the two trials used for the indirect comparison of DBd with Bd.

*Utilities*

We have concerns about poor reporting of methods used to estimate utilities from CASTOR EQ-5D data. The CS did not specify the population for the pre-progression analysis (ITT or 2L) and the company has indicated that they had misreported methods of analysis for the post-progression utilities. Details of the correct methods of analysis and results have not been supplied. In addition, we consider the small difference between pre and post-progression utilities used in the company base case to be implausible. These issues can be addressed by using the estimates from previous appraisal TA457 where utility values of ENDEAVOR are mapped using published algorithm by Proskovsky et al. This provides consistency between the analysis of Bd and Cd in TA457 and this current appraisal.

*Resource use*

In general, we view the company's overall approach to estimating resource use and costs as appropriate and in line with previous TAs (including TA457). However, the company incorrectly included administration costs and co-medication costs for bortezomib after the end of treatment (i.e. 24 weeks), thereby overestimating the total costs of bortezomib slightly. We corrected this issue in our preferred base case. Further, the clinical experts advised the ERG that cost of co-trimoxazole was excluded from the company's analyses. The weekly cost of the drug is £4.22. Whilst we acknowledge this, we view that the inclusion of this cost will have a negligible impact on the overall results and as such this cost is not included within ERG's preferred base case.

# 8 REFERENCES


1. Phillips JK, Sherlaw-Johnson C, Pearce R, et al. A randomized study of MOD versus VAD in the treatment of relapsed and resistant multiple myeloma. *Leukemia & lymphoma* 1995;17(5-6):465-72. doi: 10.3109/10428199509056859 [published Online First: 1995/05/01]
2. Carlson JJ, Guzauskas GF, Chapman RH, et al. Cost-effectiveness of Drugs to Treat Relapsed/Refractory Multiple Myeloma in the United States. *Journal of managed care & specialty pharmacy* 2018;24(1):29-38.
3. van Agthoven M, Segeren C, 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.
4. Rajkumar SV, Harousseau J-L, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Working Group Consensus Panel 1. *Blood* 2011;117:4691-95.
5. Cancer Research UK. Myeloma statistics 2017 [Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma accessed March 2018.
6. Barlogie B, Van Rhee F, Shaughnessy JD, et al. Seven-year median time to progression with thalidomide for smoldering myeloma: partial response identifies subset requiring earlier salvage therapy for symptomatic disease. *Blood* 2008;112:3122-25.
7. Lokhorst HM, van der Holt B, Zweegman S, et al. A randomized phase 3 study on the effect of thalidomide combined with adriamycin, dexamethasone, and high-dose melphalan, followed by thalidomide maintenance in patients with multiple myeloma. *Blood* 2010;115:1113-20.
8. Moreau P, Attal M, Facon T. Frontline therapy of multiple myeloma. *Blood* 2015;125:3076-84.
9. Rizzo M, Xu Y, Panjabi S, et al. A systematic literature review of the humanistic burden of multile myeloma. *Value Health* 2014;17(7):A537.
10. Hulin C, Hansen T, Heron L, et al. Living with the burden of relapse in multiple myeloma from the patient and physician perspective. *Leuk Res* 2017;59:75-84.
11. Acaster S, Gaugris G, Velikova G, et al. Impact of the treatment-free interval on health-related quality of life in patients with multiple myeloma: a UK cross-sectional survey. *Support Care Cancer* 2013;21:599-607.
12. Molassiotis A, Wilson B, Blair S, et al. Unmet supportive care needs, psychological well-being and quality of life in patients living with multiple myeloma and their partners. *Psycho-Oncol* 2011;20:88-97.
13. National Institute for Health and Care Excellence (NICE). Myeloma: diagnosis and management. NICE guideline NG35: NICE, 2016.
14. EMA (European Medicines Agency). Summary of Product Characteristics. Darzalex: EMA, 2016:1-35.
15. National Institute for Health and Care Excellence (NICE). NICE Technology Appraisal Guidance 129: Bortezomib monotherapy for relapsed multiple myeloma 2007 [Available from: https://www.nice.org.uk/guidance/ta129.

16. National Institute for Health and Care Excellence (NICE). NICE Technology appraisal guidance 457: Carfilzomib for previously treated multiple myeloma 2017 [Available from: https://www.nice.org.uk/guidance/ta457 accessed January 2017.
17. NHS Commissioning Board. National Cancer Drugs Fund List. NHS England, 2013.
18. Palumbo A, Chanan-Khan A, Weisel K, et al. Daratumumab, bortezomib, and dexamethasone for multiple myeloma. *N Engl J Med* 2016;375:754-66.
19. Dimopoulos MA, Moreau P, Palumbo A, et al. Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study. *Lancet Oncol* 2016;17(1):27-38.
20. EMA (European Medicines Agency). Assessment report. Darzalex. International non-proprietary name: daratumumab: EMA, 2017:1-105.
21. EMA (European Medicines Agency). Summary of Product Characteristics. Velcade., 2012:1-104.
22. Moreau P, Pylypenko H, Grosicki S, et al. Subcutaneous versus intravenous administration of bortezomib in patients with relapsed multiple myeloma: a randomised, phase 3, non-inferiority study. *Lancet Oncol* 2011;12(5):431-40.
23. Moreau P, Zamagni E. MRD in multiple myeloma: more questions than answers? *Blood Cancer Journal* 2017;7(639)
24. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. *NICE, London* 2013
25. Osborne TR, Ramsenthaler C, Siegert RJ, et al. What issues matter most to people with multiple myeloma and how well are we measuring them? A systematic review of quality of life tools. *European Journal of Haematology* 2012;89:437-57.
26. Cankurtaran ES, Ozalp E, Soygur H, et al. Understanding the reliability and validity of the EORTC QLQ-C30 in Turkish cancer patients. *European Journal of Cancer Care* 2008;17:98-104.
27. Janssen. [Data on file] MMY3004 Statistical analysis plan, 2016.
28. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338(b2393)
29. Rezvan PH, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol* 2015;15:30:1-14.
30. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Sheffield, UK: ScHARR, University of Sheffield, 2014:1-57.
31. CADTH. ITC Software 2009 [Available from: https://www.cadth.ca/resources/itc-user-guide/download-software-win-xp.
32. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. University of York: Centre for Reviews and Dissemination, 2008.
33. Janssen. [Data on File] MMY3004, IA2 Analysis, 2018.

34. Vandewalle B, Félix J, Almeida J, et al. Cost-effectiveness of lenalidomide-plus-dexamethasone in multiple myeloma patients who have received at least one prior therapy: a South Korean perspective. *Value in Health* 2014;17(7):A736.
35. Shen Z, King J, Kaura S, et al. Economic analysis of lenalidomide/dexamethasone (LEN/DEX) vs. Bortezomib (BORT) for the treatment of Relapsed/Refractory Multiple Myeloma (RRMM) in China. *Clinical Lymphoma, Myeloma and Leukemia* 2015;15:e307.
36. Roy A, Kish JK, Bloudek L, et al. Estimating the Costs of Therapy in Patients with Relapsed and/or Refractory Multiple Myeloma: A Model Framework. *American health & drug benefits* 2015;8(4):204-15.
37. Möller J, Nicklasson L, Murthy A. Cost-effectiveness of novel relapsed-refractory multiple myeloma therapies in Norway: lenalidomide plus dexamethasone vs bortezomib. *Journal of medical economics* 2011;14(6):690-97.
38. Maiese EM, Dimova M, Baio G, et al. Cost per median overall month of survival in multiple myeloma patients with≥ 3 lines of therapy or were double refractory: American Society of Clinical Oncology, 2016.
39. Cost-effectiveness of lenalidomide plus low-dose dexamethasone for the treatment of relapsed or refractory multiple myeloma in china. European Hematology Association; 2017.
40. Jakubowiak AJ, Campioni M, Benedict Á, et al. Cost-effectiveness of adding carfilzomib to lenalidomide and dexamethasone in relapsed multiple myeloma from a US perspective. *Journal of medical economics* 2016;19(11):1061-74.
41. Jakubowiak A, Majer IM, Houisse I, et al. Economic Evaluation of Carfilzomib+ Dexamethasone (Kd) Vs Bortezomib+ Dexamethasone (Vd) in Relapsed or Refractory Multiple Myeloma (R/RMM): Am Soc Hematology, 2016.
42. Hornberger J, Rickert J, Dhawan R, et al. The cost-effectiveness of bortezomib in relapsed/refractory multiple myeloma: Swedish perspective. *European journal of haematology* 2010;85(6):484-91.
43. Fragoulakis V, Kastritis E, Psaltopoulou T, et al. Economic evaluation of therapies for patients suffering from relapsed-refractory multiple myeloma in Greece. *Cancer management and research* 2013;5:37.
44. Durie B, Binder G, Pashos C, et al. Total cost comparison in relapsed/refractory multiple myeloma. *Journal of medical economics* 2013;16(5):614-22.
45. Cost-effectiveness of drugs to treat relapse/refractory multiple myeloma in the US. Value in Health; 2017. Elsevier Science Inc 360 Park Ave South, New York, NY 10010-1710 USA.
46. Brown RE, Stern S, Dhanasiri S, et al. Lenalidomide for multiple myeloma: cost-effectiveness in patients with one prior therapy in England and Wales. *The European Journal of Health Economics* 2013;14(3):507-14.
47. Borg S, Nahi H, Hansson M, et al. Cost effectiveness of pomalidomide in patients with relapsed and refractory multiple myeloma in Sweden. *Acta Oncologica* 2016;55(5):554-60.
48. Alsaid N, McBride A, Agarwal AB, et al. Cost effectiveness of carfilzomib (CAR), ixazomib (IXA), elotuzumab (ELO), or daratumumab (DAR) with lenalidomide and dexamethasone (LEN+ DEX) vs LEN+ DEX in relapsed/refractory multiple myeloma (R/R MM): American Society of Clinical Oncology, 2017.

49. Ollendorf D, Chapman R, Khan S. Treatment options for relapsed or refractory multiple myeloma: effectiveness, value, and value-based price benchmarks. Final evidence report. Institute for Clinical and Economic Review. 2016
50. National Institute for Health and Care Excellence (NICE). NICE Technology Appraisal Guidance 427: Pomalidomide for multiple myeloma previously treated with lenalidomide and bortezomib. , 2017.
51. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Research synthesis methods* 2010;1(3-4):258-71.
52. Dimopoulos MA, Oriol A, Nahi H, et al. Daratumumab, Lenalidomide, and Dexamethasone for Multiple Myeloma. *N Engl J Med* 2016;375(14):1319-31. doi: 10.1056/NEJMoa1607751 [published Online First: 2016/10/06]
53. Félix J, Aragão F, Almeida JM, et al. Time-dependent endpoints as predictors of overall survival in multiple myeloma. *BMC cancer* 2013;13(1):122.
54. Proskorovsky I, Lewis P, Williams CD, et al. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health and quality of life outcomes* 2014;12(1):35.
55. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust, 2014.
56. Woods B, Sideriis E, Palmer S, et al. NICE DSU technical support document 19: partitioned survival analysis for decision modelling in health care. A critical review: Decision Support Unit, ScHARR, University of Sheffield, 2017.
57. Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. NICE Decision Support Unit. 2013, 2013.
58. 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. doi: 10.1016/j.jval.2012.02.008
59. National Institute for Health and Care Excellence (NICE). NICE Technology Appraisal Guidance 171: Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy 2009 [Available from: https://www.nice.org.uk/guidance/ta171.
60. National Institute for Health and Care Excellence (NICE). NICE Technology Appraisal Guidance 380: Panobinostat for treating multiple myeloma after at least 2 previous treatments 2016 [Available from: https://www.nice.org.uk/guidance/ta380.
61. Dimopoulos M, Sonneveld P, Nahi H, et al. Progression-Free Survival as a Surrogate Endpoint for Overall Survival in Patients with Relapsed or Refractory Multiple Myeloma. *Value in Health* 2017;20(9):A408.
62. Orlowski RZ, Nagler A, Sonneveld P, et al. Final overall survival results of a randomized trial comparing bortezomib plus pegylated liposomal doxorubicin with bortezomib alone in patients with relapsed or refractory multiple myeloma. *Cancer* 2016;122(13):2050-56.
63. San-Miguel JF, Hungria VTM, Yoon S-S, et al. Overall survival of patients with relapsed multiple myeloma treated with panobinostat or placebo plus bortezomib and dexamethasone (the PANORAMA 1 trial): a randomised, placebo-controlled,

phase 3 trial. *The Lancet Haematology* 2016;3(11):e506-e15. doi: 10.1016/S2352-3026(16)30147-8

# 9 APPENDICES

**Appendix 1 Company and ERG assessments of risk of bias**

| Question | Response and explanation | | Risk of bias |
|---|---|---|---|
| Was randomisation carried out appropriately? | Company | Yes, randomisation was carried out as per the pre-specified randomisation method; patients were randomised using a central interactive web response system (IWRS). | Low |
| | ERG | Yes. The IWRS used a computer-generated randomization schedule.[20] Randomization was stratified by ISS at screening (I, II, or III), number of prior lines of therapy (1 vs 2 or 3 vs >3) and prior lenalidomide/bortezomib treatment (no vs yes). | Low |
| Was the concealment of treatment allocation adequate? | Company | CASTOR was open label. Concealment of treatment was not practical in CASTOR owing to the different dosing schedules. Potential bias was mitigated by use of an IDMC that was masked to treatment allocated | Potential risk of bias as open label design could have influenced investigator's assessment of PFS events |
| | ERG | Unclear. Details of the IWRS and whether it concealed allocation are not reported in the CS, trial publication[18] or CSR. The company's response here refers to blinding, instead of allocation concealment. | Probably low |
| Were the groups similar at the outset of the study in terms of prognostic factors? | Company | Yes, demographic and baseline characteristics were well balanced between the two treatment groups with no categories having a difference of ≥10% | Low |
| | ERG | Unclear. A higher proportion of the Bd group received prior lenalidomide in the whole trial population and the 2L subgroup. In the 2L subgroup a greater proportion of patients in the Bd arm were refractory to last line of therapy and refractory to lenalidomide specifically. The ERG is unclear whether the noted imbalances present a risk of selection bias (we discuss the reasons for this in more detail in the text in this section). | Unclear |
| Were the care providers, | Company | No, CASTOR was open label and only Janssen were blinded to the results. | Low, as an IDMC |

| | | | |
|---|---|---|---|
| participants and outcome assessors blind to treatment allocation? | | | reviewed the data |
| | ERG | No. CASTOR was open label. Review of outcomes by a blinded IDMC would reduce the risk of detection bias but would not reduce the risk of performance bias. Outcomes that are objective (OS) or time-limited (TTD) would be unlikely to incur bias. Subjective outcomes (HRQoL) are at high risk of bias where blinding is lacking. The CS and trial publication[18] state that response outcomes and PFS were assessed by a computer algorithm based on uniform response criteria recommendations, but the method of data input to the algorithm is not reported. | Low for OS and TTD<br>Probably low for response outcomes and PFS<br>High for HRQoL |
| Were there any unexpected imbalances in drop-outs between groups? | Company | No, of the 498 patients randomised (251 in the DBd group and 247 in the Bd group), 480 received study treatment: 243 patients received DBd and 237 patients received Bd (see Section B.2.4.4) | Low |
| | ERG | No. The proportions of treated patients who discontinued treatment differed between the trial arms, but this was expected. 70% of treated patients in the intervention arm and 44% in the comparator arm discontinued treatment, primarily reflecting withdrawal due to progressive disease (54% in the intervention arm and 25% in the comparator arm). This is explained by a difference in the treatment protocol between the intervention arm (treated until progression) and the comparator arm (fixed treatment duration) (CS section D.3.1). Treatment withdrawals due to other reasons were not unbalanced (15% in the intervention arm and 18% in the comparator arm).<br><br>The proportion of randomised patients who discontinued the study was 10.3 percentage points higher in the comparator arm, mainly due to slight imbalances in the rate of deaths (6.2 percentage points higher in the comparator arm) and withdrawals by the patient (3.7 percentage points higher in the comparator arm). | Low risk, provided that outcomes are interpreted in the context of the expected imbalance |
| | Company | None | Low |

| | | | |
|---|---|---|---|
| Is there any evidence to suggest that the authors measured more outcomes than they reported? | ERG | No. The trial protocol does not list any outcomes that are not reported in the CS. | 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? | Company | 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 |
| | ERG | Time-to-event outcomes: Yes. ITT analysis was defined as including all randomised patients. Missing data were accounted for by censoring rules (CS Tables 10 and 11). CS section B.2.4.2 states sensitivity analyses on different censoring assumptions were conducted; results of these are not given in the CS or trial publication but are provided in the CSR. The CSR (section 6.2.1.2) states that the sensitivity analyses on censoring assumptions showed results consistent with the primary analysis. | Low for time-to-event outcomes |
| | | Response outcomes: Yes. The company presented results for the total trial ITT and response-evaluable populations in the CS. The company, however, only presented results from the response-evaluable population for the 2L patients subgroup. As only 11 patients from the total trial population ITT group were not included in the response-evaluable analyses and the results for the two populations were similar, we consider there to be a low risk of bias from the company presenting only the response-evaluable population results for the 2L subgroup. | Low for response outcomes |
| | | HRQoL outcomes: Unclear. The results presented in CS Appendix D sections D.3.2.10 and D.3.2.11 appear to be ITT analyses with missing data imputed under a missing at random assumption. However, this is not explicitly stated and, as noted in section 3.1.5 of this report, the data imputation method is unclear. We are therefore uncertain whether the HRQoL results presented in the appendix are unbiased. | Unclear for HRQoL outcomes |