NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Carfilzomib for previously treated multiple myeloma [ID934]

The following documents are made available to the consultees and commentators:

1. **Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)**

2. **Consultee and commentator comments on the Appraisal Consultation Document** from:
   - Amgen (company)
   - Myeloma UK
   - Janssen

   'No comment' response received from Department of Health

3. **Comments on the Appraisal Consultation Document from experts:**
   - Karthik Ramasamy – clinical expert, nominated by UK Myeloma Forum

4. **Comments on the Appraisal Consultation Document received through the NICE website**

5. **Evidence Review Group critique of companies ACD response** – prepared by BMJ Group

6. **Response to request for additional information** - prepared by Amgen

7. **Evidence Review Group critique of company's new evidence –** prepared by BMJ Group

8. **Addendum to the ACD response provided by the company** – prepared by Amgen

9. **Evidence Review Group critique of the ACD addendum** – prepared by BMJ Group

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

Confidential until publication

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal

Carfilzomib for previously treated multiple myeloma

Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)

Definitions:

Consultees – Organisations that accept an invitation to participate in the appraisal including the companies, national professional organisations, national patient organisations, the Department of Health and the Welsh Government and relevant NHS organisations in England. Consultees can make a submission and participate in the consultation on the appraisal consultation document (ACD; if produced). All non-company consultees can nominate clinical experts and/or patient experts to verbally present their personal views to the Appraisal Committee. Company consultees can also nominate clinical experts. Representatives from NHS England and clinical commissioning groups invited to participate in the appraisal may also attend the Appraisal Committee as NHS commissioning experts. All consultees have the opportunity to consider an appeal against the final recommendations, or report any factual errors, within the final appraisal determination (FAD).

Clinical and patient experts and NHS commissioning experts – The Chair of the Appraisal Committee and the NICE project team select clinical experts and patient experts from nominations by consultees and commentators. They attend the Appraisal Committee meeting as individuals to answer questions to help clarify issues about the submitted evidence and to provide their views and experiences of the technology and/or condition. Before they attend the meeting, all experts must either submit a written statement (using a template) or indicate they agree with the submission made by their nominating organisation..

Commentators – Commentators can participate in the consultation on the ACD (if produced), but NICE does not ask them to make any submission for the appraisal. Non-company commentator organisations can nominate clinical experts and patient experts to verbally present their personal views to the Appraisal Committee. Commentator organisations representing relevant comparator technology companies can also nominate clinical experts. These organisations receive the FAD and have opportunity to report any factual errors. These organisations include comparator technology companies, Healthcare Improvement Scotland any relevant National Collaborating Centre (a group commissioned by NICE to develop clinical guidelines), other related research groups where appropriate (for example, the Medical Research Council and National Cancer Research Institute); other groups such as the NHS Confederation, the NHS Commercial Medicines Unit, the Scottish Medicines Consortium, the Medicines and Healthcare Products Regulatory Agency, the Department of Health, Social Services and Public Safety for Northern Ireland).

Public – Members of the public have the opportunity to comment on the ACD when it is posted on the Institute's web site 5 days after it is sent to consultees and commentators. These comments are usually presented to the appraisal committee in full, but NICE reserves the right to summarise and edit comments received during consultations, or not to publish them at all, where in the reasonable opinion of NICE, the comments are voluminous, publication would be unlawful or publication would be otherwise inappropriate.

Please note: Comments received in the course of consultations carried out by NICE are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the submissions that NICE has received, and are not endorsed by NICE, its officers or advisory committees.

Comments received from consultees

Consultee	Comment [sic]	Response
Amgen	We have carefully reviewed the Committee's consideration of the evidence for the single technology appraisal (STA) of *carfilzomib for previously treated multiple myeloma* [ID934]. We are extremely disappointed by the conclusions reached and the resulting preliminary guidance not to recommend carfilzomib. The Committee considered there to be a high degree of uncertainty in the cost-effectiveness evidence, which prevented them from being able to recommend carfilzomib. We welcome the Committee's recognition of the clinical need for alternative treatments for previously treated multiple myeloma (MM), their acknowledgement that carfilzomib offers improvements over current treatment options and their agreement with Amgen's proposed place for carfilzomib in the treatment pathway. We are committed to working with NICE to address all of the Committee's concerns. In order to provide the Committee with a high degree of certainty in the cost-effectiveness analysis, we have addressed the following: covariate selection for efficacy estimates, validity of the proportional hazards (PH) assumption, use of different parametric models (in line with Decision Support Unit [DSU] methods), use of utilities directly mapped from trial data, and length of treatment and dosing schedule of bortezomib. A summary of our responses is presented below followed by detailed responses in Sections 1 to 6. We fully anticipate that the consistency of our findings based on the further analyses presented, will sufficiently address the uncertainties identified by	Comment noted. The committee recognised that the company provided revised analyses to address the concerns and uncertainties outlined in the appraisal consultation document (ACD). The committee's consideration of the new analyses are presented in the final appraisal determination (FAD); see section 4.18, 4.19, 4.23 and 4.24 of the FAD.

Consultee	Comment [sic]	Response
	the Committee. We believe that our comprehensive response, together with our [redacted] patient access scheme ([redacted]% confidential PAS discount) to counter any associated residual uncertainty in cost effectiveness, will allow the Committee to recommend carfilzomib	

Comments received from clinical experts and patient experts

Nominating organisation	Comment [sic]	Response
UK Myeloma Forum	• Has all of the relevant evidence been taken into account? Yes • Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence? Yes • Are the provisional recommendations sound and a suitable basis for guidance to the NHS? Provisional recommendations would limit therapy options for patients with relapsed myeloma. Addition of an irreversible reasonably well tolerated proteasome inhibitor, also shown in UK based MUK5 trial (ASH 2015 Abstract No 1840) is likely to improve long-term outcomes for patients. Myeloma is genomically unstable with clonal tiding and additional mutations at relapse (Smith et al British Journal of Haematology, 2015, 171, 881–883). Outcomes for patients with high-risk disease (up to 30% of patients) at relapse are poor with significant management challenges. Carfilzomib in combination with lenalidomide and dexamethasone significantly improves clinical outcomes for high-risk patients in the ASPIRE trial (Blood 2016 128:1174-1180). The current standard of care Lenalidomide and dexamethasone is clearly suboptimal for this group of patients.	Comment noted. The committee considered the benefits of carfilzomib and accepted that there is a clinical need at relapse stages of multiple myeloma; see section 4.1 of the FAD. The committee considered the clinical evidence and accepted that carfilzomib shows a progression-free survival benefit over the comparators; see section 4.5 of the FAD.

Nominating organisation	Comment [sic]	Response
	Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination versus any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity? Nil	
Myeloma UK	Myeloma UK welcomes the opportunity to comment on the NICE ACD on carfilzomib (Kyprolis®) in combination with lenalidomide and dexamethasone or dexamethasone alone. We have a good working relationship with NICE and have absolute confidence in its appraisal methodology and processes. We also understand the difficulties faced by the committee in approving new medicines, particularly in the face of uncertainty. However, we are obviously very disappointed at the decision reached by the NICE appraisal committee on carfilzomib. As myeloma is a complex and individual cancer, clinicians need a range of treatments available at every stage of the disease, to ensure that they are able to treat their patients optimally. The negative decision means that relapsed myeloma patients will face a further delay in accessing carfilzomib, a very effective treatment option, on the NHS. The ACD highlights the appraisal committee's clear acceptance of the clinical case and need for both carfilzomib combinations. In particular, recognising the survival benefits, the clinical and patient need for carfilzomib and the quality-of-life benefits of the treatment to patients. NICE therefore has all the data available to them demonstrating why patients and their carers will benefit from accessing carfilzomib as part of their treatment pathway and agree that a "compelling" clinical case has been made for approval.	Comment noted. The committee considered the benefits of carfilzomib and accepted that there is a clinical need at relapse stages of multiple myeloma; see section 4.1 of the FAD. The committee considered the clinical evidence and accepted that carfilzomib shows a progression-free survival benefit over the comparators; see section 4.5 of the FAD.

Nominating organisation	Comment [sic]	Response
	Analysing the Committee's concerns around cost-effectiveness highlighted in the ACD, we are cautiously optimistic that the challenges with the health economic modelling and uncertainty will be overcome by further clarification and dialogue. We therefore urge NICE and Amgen to collaborate to find a solution that benefits everyone and provides vital access to a new and innovative treatment for myeloma patients on the NHS. We look forward to working with NICE to find a solution for myeloma patients and carers. Please do not hesitate to contact me if we can provide any further information to support the appraisal.	

Comments received from commentators

Commentator	Comment [sic]	Response
Janssen	Janssen is pleased to have the opportunity to comment on the above ACD for carfilzomib. **Sections 4.2 to 4.5** Janssen would like to highlight that in Sections 4.2 to 4.5 (Decision problem and treatment pathway), no mention is made of the 1st line treatment of patients eligible for a stem cell transplant. As noted in the Final scope for this appraisal: "NICE technology appraisal guidance 311 recommends bortezomib as an option, in combination with dexamethasone or with dexamethasone and thalidomide, for the induction treatment of adults with untreated multiple myeloma who are eligible for high-dose chemotherapy with haematopoietic stem cell transplantation." Furthermore, the exclusion of this aspect of the patient journey results in inconsistency between Myeloma appraisals. **Section 4.5** Within the ACD, Section 4.5, page 8, it is stated "carfilzomib with dexamethasone would only replace bortezomib with dexamethasone at second line if people had not had bortezomib therapy at first line (and	Comment noted. The committee was aware that there is a group of people for whom stem-cell transplant may be considered and in which case bortezomib may be used first-line. The committee considered the positioning presented by the company and understood that carfilzomib's place in the pathway would be at second and third line for people who cannot have a stem-cell transplant, for which the comparators were bortezomib in combination with dexamethasone at second line and lenalidomide in combination with dexamethasone at third line; see section 4.3 and 4.4 of the FAD.

Commentator	Comment [sic]	Response
	instead had thalidomide therapy at first line, as the most commonly used regimen; see section 4.4). Janssen would like to highlight that with the inclusion of patients eligible for stem cell transplant, bortezomib is the most commonly used 1st line treatment.	

Comments received from members of the public

Role*	Section	Comment [sic]	Response
NHS professional	1	The ACD should be considered carefully as the clinical data is very impressive and providing Amgen can resolve the economic modelling this technology would help bridge a significant healthcare need in this group of patients who have limited prognosis and QOL on the limited alternative therapies at present available in this setting. It would appear that assumptions about the dosage have been derived by the company from the time on treatment rather than true numbers of cycles which is clearly an error. I am sure you will recieve many comments re this ACD as Carfilzomib clearly can provide a significant improvement in our present treatment armoury.	Comment noted. The committee considered the clinical evidence and accepted that carfilzomib shows a progression-free survival benefit over the comparators; see section 4.5 of the FAD. The committee considered the company's revised analysis; see section 4.18 and 4.19 of the FAD.
NHS professional	1	As the [redacted] trial I would like to firstly say that Amgen/Onyx have been hugely supportive of clinical trials in the U.K. This has allowed many physicians to gain first hand experience of this powerful and effective proteosome inhibitor. The endeavour and aspire studies show considerable efficacy over the current standards of care. In addition it is not associated with a significant risk of neuropathy which is a difficult and frequently disabling irreversible side-effect of both velcade and thalidomide. Also the current nice guideance on single agent velcade is woefully out of date and there is an onus on	Comment noted. The committee considered the clinical evidence and accepted that carfilzomib shows a progression-free survival benefit over the comparators; see section 4.5 of the FAD. The committee considered the company's revised analysis; see section 4.18 and 4.19 of the FAD.

* When comments are submitted via the Institute's web site, individuals are asked to identify their role by choosing from a list as follows: 'patent', 'carer', 'general public', 'health professional (within NHS)', 'health professional (private sector)', 'healthcare industry (pharmaceutical)', 'healthcare industry'(other)', 'local government professional' or, if none of these categories apply, 'other' with a separate box to enter a description.

Role*	Section	Comment [sic]	Response
		Nice to provide better guideance on front and second line therapy. The FAD on second line single agent velcade is out of date as most patients now receive velcade as front line therapy and NICE and NHS England advice on second line therapy are currently at odds. The myeloma community would welcome the availability of carfilizomib in combination with either dexamethasone or dexamethasone/revlimid.	

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Single Technology Appraisal (STA)

Carfilzomib for treating multiple myeloma in people who have received at least one prior therapy

Response to Appraisal Committee Document

Prepared by:

Date: December 2016

File name	Version	Contains confidential information	Date
	1.0	Yes CIC: Highlighted in blue and underlined	1 December 2016

Contents

List of tables

List of figures

1 Executive Summary

We have carefully reviewed the Committee's consideration of the evidence for the single technology appraisal (STA) of *carfilzomib for previously treated multiple myeloma* [ID934].

We are extremely disappointed by the conclusions reached and the resulting preliminary guidance not to recommend carfilzomib. The Committee considered there to be a high degree of uncertainty in the cost-effectiveness evidence, which prevented them from being able to recommend carfilzomib. We welcome the Committee's recognition of the clinical need for alternative treatments for previously treated multiple myeloma (MM), their acknowledgement that carfilzomib offers improvements over current treatment options and their agreement with Amgen's proposed place for carfilzomib in the treatment pathway.

We are committed to working with NICE to address all of the Committee's concerns. In order to provide the Committee with a high degree of certainty in the cost-effectiveness analysis, we have addressed the following: covariate selection for efficacy estimates, validity of the proportional hazards (PH) assumption, use of different parametric models (in line with Decision Support Unit [DSU] methods), use of utilities directly mapped from trial data, and length of treatment and dosing schedule of bortezomib. A summary of our responses is presented below followed by detailed responses in Sections 1 to 6.

We fully anticipate that the consistency of our findings based on the further analyses presented, will sufficiently address the uncertainties identified by the Committee. We believe that our comprehensive response, together with our [redacted] patient access scheme ([redacted]% confidential PAS discount) to counter any associated residual uncertainty in cost effectiveness, will allow the Committee to recommend carfilzomib.

1. **Adjustment for different combinations of covariates yields broadly similar progression-free survival (PFS) and overall survival (OS) treatment effects compared to our original approach, for both carfilzomib/dexamethasone (Cd) versus bortezomib/dexamethasone (Vd) and carfilzomib/lenalidomide/dexamethasone (CRd) versus lenalidomide/dexamethasone (Rd). The step-wise selection method is considered the most appropriate for identification of prognostic factors.**

   Our original submission adjusted for a broad range of clinician-identified prognostic variables. The committee concluded that it would have liked to have seen the rationale for the selection of covariates and plausibility of different combinations of covariates on the efficacy estimates. We have now applied a range of variable selection methods to the clinician-identified covariates including a stepwise selection method and the least absolute shrinkage and selection operator (LASSO) method as suggested by the Evidence Review Group (ERG). These allow for a combination of expert opinion and statistical modelling to select prognostic variables. These covariate-adjustment analyses were based on models fitted separately within each subgroup, aligned with cost-effectiveness analyses which are based on subgroup data.

   - The different methods of covariate selection yield broadly similar PFS and OS treatment effect estimates for both Cd versus Vd and CRd versus Rd in the subgroups of interest, compared to our original approach and are unlikely to have

significantly impacted estimates of the cost effectiveness of Cd and CRd in our original submission.
- The step-wise selection method was considered the most appropriate covariate-adjusted model and was therefore used to identify the prognostic variables to adjust for within subsequent (covariate-adjusted) parametric models.

2. **The proportional hazards (PH) assumption was satisfied for all OS analyses, but for PFS, non-proportionality was observed for a few covariates. However, the treatment effect was broadly consistent, even using analyses which accounted for non-proportionality. Therefore these analyses further support the plausibility of the treatment hazard ratios (HRs) used in the cost-effectiveness analyses in our original submission.**

We fully explored the PH assumption using the best-fitting covariate-adjusted Cox models (stepwise selection method), in order to address the Committee's concern around the validity of the PH assumption for the relevant subgroups in our original approach (which used covariate-adjusted subgroup analysis to derive HRs).

- For OS, these analyses showed that the PH assumption was satisfied for treatment (Cd vs Vd and CRd vs Rd) and all covariates in the models.
- For PFS, the PH assumption was satisfied for treatment (Cd vs Vd and CRd vs Rd), but non-proportionality was observed for a few covariates. This was addressed by fitting piece-wise Cox models and notably the treatment effect (Cd vs Vd and CRd vs Rd) was almost identical in these models.
- The PFS and OS HRs derived from these stepwise variable selection models were not utilised in the revised case cost-effectiveness analyses, given that we have now fitted covariate-adjusted parametric models to both arms of the ENDEAVOR and ASPIRE trials in line with NICE Decision Support Unit (DSU) guidance. The PH assumption for the covariate-adjusted parametric models was assessed separately as described below.

3. **Uncertainty around the survival model has been fully explored by conducting a range of analyses; fitting different parametric models to both arms of the trials and comparing model predictions to corresponding covariate-adjusted trial data. The inverse probability weighted [IPW] Weibull model provided a good statistical fit, clinically plausible estimate, satisfied proportional hazards assumptions for both PFS and OS (in the ENDEAVOR and ASPIRE subgroups of interest) and importantly is aligned with DSU guidance. The treatment effect estimates for PFS and OS from this model are similar to those presented in our original submission which greatly increases confidence in our original approach.**

In order to address the Committee's concerns around separately modelling the different arms in the ENDEAVOR and ASPIRE trials, and to align with NICE DSU guidance, we fitted covariate-adjusted parametric models jointly to both treatment arms of ENDEAVOR and ASPIRE. As specifically requested by the Committee, we have explored the effects of different extrapolation techniques (including exploring a weighted-adjusted covariate model), and have provided a comparison of the model predictions from the most

appropriate modelling method and most plausible parametric distributions to corresponding covariate-adjusted estimates from the trials.

- Of the different extrapolation techniques incorporating covariate adjustment (IPW, corrected group prognosis [CGP] and mean of covariates [MoC]) explored, the IPW-based Weibull model yielded the most clinically plausible projections of PFS and OS for both the ENDEAVOR and ASPIRE subgroups of interest (good statistical fit, an excellent fit to corresponding covariate-adjusted (weighted) trial data, and satisfied proportional hazards assumptions).
- Notably, the IPW-weighted treatment effect estimates (Cd vs Vd and CRd vs Rd) for PFS and OS are similar to the treatment effects estimated for the ENDEAVOR and ASPIRE subgroups of interest in our original submission, which supports the robustness of the treatment effect estimates used to inform the cost-effectiveness model in our original submission.
- Given the above, the IPW-based Weibull model has been used in our revised cost-effectiveness analyses.

4. **The Committee's preferred alternate utility assumptions (using values mapped directly from the trial data) have been incorporated into the revised cost-effectiveness analysis presented (Section 3).**

   We acknowledge that it could be considered a more appropriate option to use utilities mapped directly from trial data, in line with the NICE reference case and the Committee's stated preference. In order to address the Committee's concerns and stated preference, utility data mapped directly from the trials using the mapping algorithm from Proskorovsky et al., 2014 have therefore been used in our revised cost-effectiveness analysis (Section 6).

5. **It is inappropriate to break the dose-efficacy relationship within randomised controlled trial (RCT) evidence. In exploring the Committee's concerns around bortezomib treatment duration it is incorrect only to cap bortezomib costs and not accordingly adjust for efficacy. We believe it is most appropriate to base cost-effectiveness analyses on the dose-efficacy relationship observed in a robust RCT setting (as in our original submission), however we have tried to address the Committee's concerns by adjusting both bortezomib efficacy and capping costs.**

   The Committee expressed concerns about the length of treatment and dosing schedule of bortezomib in the model, which were based on the ENDEAVOR trial, and was considered by the Committee not to be reflective of clinical practice. Although it is inappropriate and uncertain to break the dose-efficacy relationship within the ENDEAVOR RCT, we have tried to address the Committee's concerns.

   - To explore the impact of different bortezomib treatment durations on the cost-effectiveness of Cd versus Vd, we conducted a scenario analysis adjusting for efficacy using data from a recently conducted RCT in a similar patient population where Vd was given for 8 cycles (32 doses). However, we believe it is most appropriate to base our cost-effectiveness analyses on the dose-efficacy relationship observed in a robust RCT setting (as in our original submission).

6. Revised cost-effectiveness analyses

We have updated the cost-effectiveness analyses based on the above considerations. The revised cost-effectiveness results and comparison with the original base case analyses are presented in Table 1 for the Cd versus Vd comparison and Table 2 for the CRd versus Rd comparison.

Table 1. Revised cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib

	Cd	**Vd**	**Incremental value**
Original base case			
Total costs	£121,891	£95,213	£26,678
Total LYG	6.05	4.26	1.79
QALYs	4.28	2.95	1.33
ICER	-	-	£20,044
Using the updated survival model			
Total costs	£117,660	£93,769	£23,891
Total LYG	5.74	4.23	1.51
QALYs	4.09	2.94	1.15
ICER (compared to original base case)	-	-	£20,766
Using directly mapped utilities			
Total costs	£117,660	£93,769	£23,891
Total LYG	5.74	4.23	1.51
QALYs	3.88	2.79	1.09
ICER (compared to original base case)	-	-	£21,137
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£22,009**

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

Table 2. Revised cost-effectiveness results for CRd versus Rd in patients with two prior therapies and no prior lenalidomide

	CRd	Rd	Incremental value
Original base case			
Total costs	£128,654	£95,420	£33,234
Total LYG	6.31	4.93	1.37
QALYs	4.32	3.33	0.99
ICER	-	-	£33,467
Using the updated survival model			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.77	2.95	0.79
ICER (compared to original base case)	-	-	£40,198
Using directly mapped utilities			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.67	2.88	0.79
ICER (compared to original base case)	-	-	£34,404
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£41,429**

CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; Rd, lenalidomide/dexamethasone; QALYs, quality-adjusted life years

Conclusions

We have shown in this response that by applying appropriate alternative approaches to selection of covariates and modelling survival in line with DSU guidance, the revised cost-effectiveness analyses predict outcomes similar to our original approach. Notably, the difference in incremental cost-effectiveness ratios (ICERs) for the comparison of Cd versus Vd in patients who have received one prior therapy and no prior bortezomib was marginal (£20,766/QALY gained compared with £20,044/QALY gained).

For the CRd versus Rd comparison in 3rd line, we would like to highlight that policy level discussions around appraisals for interventions not being cost-effective at zero price are pertinent to this appraisal. In this case, the additional cost of prolonged use of lenalidomide background therapy is a key factor driving up the ICERs. Therefore we urge the Committee to take this into account and to consider scenarios when the additional costs of lenalidomide, have been discounted, resulting in an ICER of £36,455 when using the final revised cost-effectiveness analysis described in Table 2, and £29,671 using the original base case.

We believe that this comprehensive response, together with our [redacted] patient access scheme ([redacted]% confidential PAS discount) to counter any associated residual uncertainty in cost effectiveness, will allow the Committee to recommend carfilzomib.

2 Survival model to estimate long-term effects

2.1 Plausibility of trial covariate-adjusted efficacy estimates

Context

The Appraisal Consultation Document (ACD) states that 'the committee was aware of the limitations and the uncertain outcomes associated with subgroups that were not prespecified. It recognised the company's attempt to counter the uncertainties by adjusting for imbalances in the baseline characteristics with additional covariates by using a Cox proportional hazards model to estimate efficacy (as hazard ratios) of carfilzomib and its comparators. But the committee heard from the ERG that the choice of these covariates was unclear without sufficient justification. The committee noted that the choice of variables to adjust the model should be those that are prognostic of the outcome, including an adjustment for the treatment effect' (ACD Section 4.7, page 9).

The Committee concluded that adjusted hazard ratios for the subgroups were not reliable estimates of the efficacy of Cd versus Vd and CRd versus Rd (ACD Section 4.7, page 9).

To address this uncertainty, the Committee specifically requested additional evidence on 'Plausible efficacy estimates for all comparisons, adjusted by covariates, including a treatment effect, and to explore the plausibility of different combinations of covariates on the efficacy estimates. The covariates to adjust the model, presented with a rationale for why they had been chosen.' (ACD Section 4.12, page 12)

Implementation

Our original submission provided covariate-adjusted estimates of the efficacy of Cd versus Vd and CRd versus Rd in the subgroups of relevance for this appraisal i.e. subgroups aligned with the proposed positioning of Cd and CRd in the treatment pathway in England and Wales. In line with the Committee's request in the ACD, only those variables considered to be prognostic of outcomes in MM were included in these models. An approach informed by clinical expertise was considered preferable to a purely statistical approach based on automated variable selection procedures which does not take account of current knowledge and opinion regarding prognostic factors. With the large number of baseline variables recorded in ASPIRE and ENDEAVOR, a purely statistical approach to variable selection may have resulted in inclusion of variables with no clinical relevance.[1,2] We therefore consulted two leading UK-based haemato-oncologists highly experienced in the treatment of MM and provided them with a list of baseline variables collected in ASPIRE and ENDEAVOR. They were then asked to identify which variables they considered prognostic of outcomes in MM (no distinction was made between PFS and OS). Ten baseline variables were considered prognostic by both clinicians with a further three considered prognostic by one of the clinicians (Table 3).

Table 3. Variables considered prognostic of outcome by at least one of the two clinical experts

Baseline variable	Clinician 1	Clinician 2
Prior lenalidomide	✓	✓
Prior bortezomib	✓	✓
Prior SCT	✓	✓
Number of prior lines of therapy	✓	✓
Age	✓	✓
ECOG status	✓	✓
Creatinine clearance	✓	✓
Time from initial diagnosis	✓	✓
Time from last relapse	✓	✓
Cytogenetic risk status	✓	✓
ISS stage at study entry	✓	-
B2-microglobulin	✓	-
Refractory to last prior treatment	✓	-

ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; SCT, stem cell transplant

In order not to miss any potential prognostic variables, we considered variables identified by at least one of the clinicians (i.e. all variables in Table 3) for use in the covariate-adjusted models. Due to the high proportion of unknown/missing data on cytogenetic risk status (47.3% and 15.5% of patients in the ASPIRE and ENDEAVOR intent to treat [ITT] populations, respectively), this variable was not included in the analysis.

Table 4 shows the final list of variables included in the covariate-adjusted models of the efficacy of Cd versus Vd and CRd versus Rd in the key subgroups of relevance for this submission.

Table 4. Variables included in covariate-adjusted PFS and OS models in key subgroups of relevance (based on clinical expert opinion)

Cd vs Vd - patients with 1 prior therapy and no prior bortezomib exposure (ENDEAVOR)	CRd vs Rd - patients with 2 prior therapies and no prior lenalidomide exposure (ASPIRE)
Prior lenalidomide	Prior bortezomib
Prior SCT	Prior SCT
Age	Age
ECOG status (0, 1-2)	ECOG status (0, 1-2)
Creatinine clearance (<50, 50-<80, ≥80 mL/min)	Creatinine clearance(<50, 50-<80, ≥80 mL/min)
Time from initial diagnosis	Time from initial diagnosis
Time from last relapse	Time from last relapse
ISS stage at study entry (I, II-III)	ISS stage at study entry (I, II-III)
B2-microglobulin (<3.5, ≥3.5 mg/L)	B2-microglobulin (<3.5, ≥3.5 mg/L)

Cd vs Vd - patients with 1 prior therapy and no prior bortezomib exposure (ENDEAVOR)	CRd vs Rd - patients with 2 prior therapies and no prior lenalidomide exposure (ASPIRE)
Refractory to last prior treatment	Refractory to last prior treatment
Cd, carfilzomib/dexamethasone; CRd, carfilzomib, lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; Rd, lenalidomide/dexamethasone; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.	

In order to address the committee's request to explore the impact of different combinations of covariates on efficacy estimates we now provide covariate-adjusted results from a range of analyses. Our original submission adjusted for all clinician-identified variables and the ERG commented on the high number of variables included as well as the potential correlation of some variables. In order to obtain a more parsimonious model, we have now applied variable selection methods to the clinician-identified covariates using a stepwise selection method and the LASSO method suggested by the ERG.[3] This approach allows for a combination of expert opinion and statistical modelling to select prognostic variables. It also allows for different variables to be included in PFS and OS models. The analysis approaches are presented in Table 5.

It should be noted that our original submission approach to estimating covariate-adjusted treatment HRs in the key subgroups of relevance was based on a model using the full trial population data and fitting treatment by subgroup interactions plus all clinician-identified prognostic covariates (analysis approach 1 in Table 5). While this had the advantage of high precision since all available data points were used, it did not necessarily ensure balance on prognostic covariates within the subgroups of relevance. Therefore all subsequent supportive analyses reported within our response are based on models fitted separately within each subgroup (analysis approaches 2 to 5 in Table 5). This also ensures alignment with the cost-effectiveness efficacy models which are based on subgroup data.

Table 5. Analyses performed to obtain covariate-adjusted treatment effects in the key subgroups of relevance

Analysis approach	Comments
1. Original submission covariate-adjusted model Cox proportional hazards model based on full trial population with treatment by subgroup interactions (to enable estimation of treatment effect within the relevant subgroup) and including all clinician-identified covariates	• This was the approach used in our original submission. • Although the model has the advantage of using all available data and therefore increasing precision, it does not necessarily ensure balance on covariates within the key subgroups of relevance. • The ERG commented that the high number of covariates was not adequately justified and made interpretation of bias difficult.[3]
2. Stepwise variable selection model	• This approach is more parsimonious than a model with all clinician-identified covariates.

Analysis approach	Comments
Cox proportional hazards model adjusting for treatment and those clinician-identified covariates retained based on a stepwise variable selection procedure. The model is fitted within key subgroups of relevance. Variables were selected by using a hybrid stepwise selection strategy that considers both forward and backward moves at each step, and selects the "best" of the two. The variable selection strategy was implemented in R (stepAIC package) which uses the AIC criterion to weigh the choices. At each step an add or drop was performed that minimised the AIC score.	• This addresses the ERG concern that some clinician-identified variables are correlated (e.g. ISS and β2-microglobulin).[3]
3. All clinician-identified covariates model Cox proportional hazards model adjusting for treatment and all clinician-identified covariates. The model is fitted within key subgroups of relevance.	• This method is aligned with the approach used in our original submission (Approach 1), • However, the model is now fitted within each relevant subgroup (rather than using the full trial data with treatment by subgroup interactions). This therefore ensures balance on covariates within the key subgroups of relevance.
4. LASSO variable selection model Cox proportional hazards model adjusting for treatment and using the LASSO method[4-6] for variable section based on the clinician-identified covariates. The model is fitted within key subgroups of relevance. Methodology: • In general, the LASSO method balances between two extremes, fitting a regression model and shrinking the coefficient estimates towards zero by introducing a penalty term (λ) for the coefficients. If λ=0, we obtain the "standard" regression estimates, if λ=∞, we obtain all parameters to be zero. • The LASSO procedure attempts to force some regression coefficient estimates to be exactly zero, thus achieving variable selection while shrinking the remaining coefficients toward zero. • Generally speaking, bias increases as λ (amount of shrinkage employed) increases and more coefficients are set to zero. • We used R (glmnet package - Lasso and Elastic-Net Regularized Generalized Linear Models) that allowed the implementation of LASSO in the Cox model framework. We used the λ for the penalty that provides the minimum mean cross-validated error.	• The ERG requested that the LASSO method for variable selection was explored (at clarification stage). • It should be noted that estimated hazard ratios from this method are biased (LASSO trades off unbiasedness with variance) with the bias being in the direction of a lesser treatment effect (i.e. hazard ratio closer to 1).[1] In addition, the available software to implement this analysis does not allow an assessment of the proportional hazards assumption for the covariates.
5. Cox model using LASSO variables	• Allows a comparison of hazard ratios and AIC from LASSO analysis and

Analysis approach	Comments
Cox proportional hazards model adjusting for treatment and those covariates identified in the LASSO analysis (see row above) Model is fitted within key subgroups of relevance	standard Cox proportional hazards model adjusting for the same variables.
AIC Akaike information criterion; ERG, Evidence Review Group; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator.	

A list of covariates retained for each covariate selection method is provided in Appendix A. Efficacy results from these methods are provided in Table 6 (Cd versus Vd) and Table 7 (CRd versus Rd) and show that treatment HRs are broadly similar across the models assessed, and also when compared with our original submission analysis. HRs from the LASSO variable selection model tend to be biased towards the null and are not directly comparable with the other HRs.[1] To allow for an appropriate comparison, we re-ran the Cox model using the covariates identified by the LASSO model.

Model fit statistics (AIC) are presented in Table 6 (Cd vs Vd) and Table 7 (CRd vs Rd) for models where this can be calculated and compared. The model with lowest AIC for PFS and OS was the stepwise variable selection model for both Cd versus Vd (in patients with one prior therapy and no prior bortezomib exposure) and CRd versus Rd (in patients with two prior therapies and no prior lenalidomide exposure). Although it was not possible to compute an AIC for the LASSO variable selection model, a Cox model including only those variables identified by the LASSO variable selection model had worse fit (higher AIC) than the stepwise variable selection model.

The stepwise variable selection models are therefore considered the most appropriate models for both PFS and OS for Cd versus Vd and CRd versus Rd. Given the Committee's stated preference to fit covariate-adjusted parametric models to both arms of the ENDEAVOR and ASPIRE trials in line with NICE DSU guidance (in contrast to our original approach of using covariate-adjusted subgroup analysis derived HRs to estimate PFS and OS curves for Cd and CRd), the PFS and OS HRs derived from these stepwise variable selection models were not utilised in the revised cost-effectiveness model. The stepwise variable selection models were instead used to identify which prognostic variables to adjust for within the covariate-adjusted parametric models described in Section 2.3.

Table 6. PFS and OS treatment hazard ratios and model fits (AIC statistic) from the range of covariate-adjusted analyses Cd vs Vd (ENDEAVOR, - patients with one prior therapy and no prior bortezomib)

	1. Original submission covariate-adjusted model[a]	2. Stepwise variable selection model[b]	3. All clinician-identified covariates model[b]	4. LASSO variable selection model[b]	5. Cox model using LASSO variables[b]
PFS and OS hazard ratios (95% CIs)					
PFS	0.412 (0.286, 0.594)	0.408 (0.267, 0.624)	0.362 (0.231, 0.567)	0.397 (0.272, 0.692)	0.366 (0.235, 0.572)
OS[c]	0.592 (0.392, 0.895)	0.631 (0.384, 1.039)	0.618 (0.373, 1.024)	0.661 (0.388, 1.000)	0.621 (0.377, 1.023)
AIC					
PFS	N/A	883.822	893.576	N/A	890.208
OS[c]	N/A	640.675	648.152	N/A	644.687

[a] based on a full population (ITT) model with treatment by subgroup interactions rather than a model based only on the subgroup. AIC is not presented from this model as it cannot be compared to the other models

[b] based on a model within the relevant subgroup (patients with one prior therapy and no prior bortezomib)

[c] EMA adhoc-analysis data cut-off

AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; ITT, intent to treat; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone.

Table 7. PFS and OS treatment hazard ratios and model fits (AIC statistic) from the range of covariate-adjusted analyses CRd vs Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	1. Original submission covariate-adjusted model[a]	2. Stepwise variable selection model	3. All clinician-identified covariates model	4. LASSO variable selection model	5. Cox model using LASSO variables
PFS and OS hazard ratios					
PFS	0.686 (0.488, 0.966)	0.699 (0.481, 1.02)	0.708 (0.478, 1.047)	0.776 (0.532, 1.000)	0.707 (0.482, 1.038)
OS	0.737 (0.481, 1.129)	0.730 (0.454, 1.174)	0.783 (0.471, 1.299)	0.801(0.503, 1.079)	0.783 (0.471, 1.299)
AIC					
PFS	N/A	1076.913	1087.011	N/A	1079.794

	1. Original submission covariate-adjusted model[a]	2. Stepwise variable selection model	3. All clinician-identified covariates model	4. LASSO variable selection model	5. Cox model using LASSO variables
OS	N/A	675.733	683.207	N/A	681.210

[a] based on a full population (ITT) model with treatment by subgroup interactions rather than a model based only on the subgroup. AIC is not presented from this model as it cannot be compared to the other models

[b] based on a model within the relevant subgroup (patients with two prior therapies and no prior lenalidomide)

AIC Akaike information criterion; CRd, carfilzomib, lenalidomide/dexamethasone; ITT, intent to treat; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; OS, overall survival; Rd, Rd, lenalidomide/dexamethasone.

Full results from the stepwise variable selection models are shown in Table 8 (Cd vs Vd) and Table 9 (CRd vs Rd). As would be expected, these models retained a smaller number of covariates than those based on all clinician-identified covariates for both PFS and OS.

Table 8. Efficacy results from stepwise variable selection model for Cd vs Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Covariate	Cd vs Vd - patients with 1 prior therapy and no prior bortezomib exposure (ENDEAVOR)	
	PFS HR (95% CI)	OS[a] HR (95% CI)
Treatment (Cd vs Vd)	0.408 (0.267-0.624)	0.631 (0.384-1.039)
Prior lenalidomide (yes vs no)	-	-
Prior stem cell transplantation (yes vs no)	-	2.034 (1.125-3.678)
Age (≥65 vs <65)	-	-
ECOG status (1-2 vs 0)	-	2.154 (1.265-3.665)
Creatinine clearance (≥50 - <80 vs other)	-	-
Creatinine clearance (≥80 vs other)	-	0.496 (0.264-0.931)
Time from diagnosis	0.990 (0.983-0.997)	0.986 (0.976-0.997)
Time from last relapse	-	-
ISS stage (II-III vs I)	2.455 (1.590-3.792)	-
B2-microglobulin (≥3.5 vs <3.5 mg/L)	-	2.926 (1.634-5.238)
Refractory to last prior treatment (yes vs no)	-	2.010 (1.158-3.487)

[a] EMA adhoc-analysis data cut-off

Cd, carfilzomib/dexamethasone; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; ISS, International Staging System; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone.

Table 9. Efficacy results from stepwise variable selection model for CRd vs Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Covariate	CRd vs Rd - patients with 2 prior therapies and no prior lenalidomide exposure (ASPIRE)	
	PFS HR (95% CI)	OS[a] HR (95% CI)
Treatment (CRd vs Rd)	0.699 (0.481-1.015)	0.730 (0.454-1.174)
Prior bortezomib (yes vs no)	-	0.569 (0.333-0.973)
Prior stem cell transplantation (yes vs no)	1.448	-

Covariate	CRd vs Rd - patients with 2 prior therapies and no prior lenalidomide exposure (ASPIRE)	
	PFS HR (95% CI)	OS[a] HR (95% CI)
	(0.932-2.250)	
Age (≥65 vs <65)	-	-
ECOG status (1-2 vs 0)	1.578 (1.065-2.339)	-
Creatinine clearance (≥50 - <80 vs other)	-	0.296 (0.142-0.616)
Creatinine clearance (≥80 vs other)	-	0.298 (0.138-0.645)
Time from diagnosis	0.994 (0.988-0.999)	0.993 (0.986-1.000)
Time from last relapse	-	-
ISS stage (II-III vs I)	4.138 (1.964-8.721)	10.111 (4.263-23.986)
B2-microglobulin (≥3.5 vs <3.5 mg/L)	0.523 (0.257-1.063)	0.382 (0.176-0.829)
Refractory to last prior treatment (yes vs no)	-	-

[b] EMA adhoc-analysis data cut-off

CI, confidence interval; CRd, carfilzomib, lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; ISS, International Staging System; PFS, progression-free survival; OS, overall survival; Rd, lenalidomide/dexamethasone.

Conclusion

In order to address the Committee's concerns around the plausibility of the trial covariate-adjusted efficacy estimates, we have explored a range of covariate selection/adjustment methods. These methods (including our original approach) yield broadly similar PFS and OS treatment effect estimates for both Cd versus Vd and CRd versus Rd in the subgroups of interest. Therefore, the choice of selection/adjustment method is unlikely to have significantly impacted estimates of the cost effectiveness of Cd and CRd in our original submission.

Given the Committee's stated preference to fit covariate-adjusted parametric models to both arms of the ENDEAVOR and ASPIRE trials in line with NICE DSU guidance (in contrast to our original approach of using covariate-adjusted subgroup analysis derived HRs to estimate PFS and OS curves for Cd and CRd), the PFS and OS HRs derived from these updated covariate-adjusted subgroup analyses were not utilised in the revised cost-effectiveness model. These analyses were instead used to identify which prognostic variables to adjust for within the covariate-adjusted parametric models described in Section 2.3 (based on the best-fitting stepwise variable selection model).

2.2 Proportional hazards assumptions

Context

The ACD states that 'It was aware the model to extrapolate the carfilzomib arm was based on the subgroup post hoc estimate hazard ratios (see section 4.7), and noted that this assumes the hazard ratios for both arms to be constant over time (benefits of treatment continue until the end of the time horizon or death; proportional hazards). The committee discussed whether this assumption was valid and noted that the Kaplan-Meier estimated curves for the subgroups showed visual points of departure from proportionality (showing non-constant hazards over time).' (ACD Section 4.12, page 11).

The Committee concluded that Amgen had not fully explored the effect of non-proportionality.

Implementation

In our original submission and in response to clarification questions from the ERG, we explored the proportional hazards assumption for treatment and prognostic covariates using a range of methods (log-log plots, Schoenfeld residual plots, interaction tests for treatment by survival time, and Grambsch and Therneau tests for proportionality of hazards [zph test using rank transformation]). These analyses suggested that the proportional hazards assumption for treatment was satisfied for Cd versus Vd in PFS and OS models, although there was evidence of non-proportionality for some covariates. The proportional hazards assumption for treatment also appeared valid for CRd versus Rd in the OS model, but was questionable for PFS, and again there was evidence of non-proportionality for some covariates in both these models. A summary of the Grambsch and Therneau tests for proportionality of hazards as previously provided in response to ERG clarification questions is provided in Appendix B.

We now provide an assessment of the proportional hazards assumption for the best-fitting covariate-adjusted model described in Section 2.1 (stepwise variable selection model) for both ASPIRE and ENDEAVOR. We performed Grambsch and Therneau tests to assess the proportional hazards assumption and if this was violated (p-value < 0.05), a piece-wise Cox model with time-dependent hazard ratios was fitted. The cut-off points for the piece-wise model were determined based on visual assessment of Schoenfeld residual plots (Appendix C). Results from these additional analyses are discussed below.

<u>Cd versus Vd (ENDEAVOR, patients who have received one prior therapy and no prior bortezomib)</u>

The assessment of proportional hazards is shown in Table 10 (PFS) and Table 11 (OS) for this subgroup.

For PFS, the proportional hazards assumption was not violated for treatment. However, there was evidence of non-proportional hazards for the two covariates in the model (time from diagnosis and ISS stage). Therefore, a piece-wise Cox model was also fitted using a 6 month cut-off point for both covariates (based on visual assessment of the Schoenfeld residual plots). In this piece-wise model, the proportional hazards assumption was satisfied for all covariates and for the model as a whole. It is notable that in both the original Cox model and the piece-

wise Cox model, the PFS treatment effects for Cd versus Vd are almost identical (HRs of 0.408 and 0.415, respectively), showing that the treatment effect is robust to any non-proportionality of hazards for the covariates.

For OS, the proportional hazards assumption was not violated for treatment or for any of the covariates.

Table 10. Grambsch and Therneau tests for proportional hazards in covariate-adjusted Cox PH model: Cd vs Vd PFS model (ENDEAVOR, patients with 1 prior therapy and no prior bortezomib exposure)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)	Piece-wise Cox model Hazard ratio (P value of Grambsch and Therneau test)
Treatment (Cd vs Vd)	0.408 (0.781)	0.415 (0.730)
Prior lenalidomide	-	-
Prior stem cell transplantation	-	-
Age (≥65 vs <65)	-	-
ECOG status (1-2 vs 0)	-	-
Creatinine clearance (50-80)	-	-
Creatinine clearance (>80)	-	-
Time from diagnosis	0.990 (0.007)	<6m: 0.978 (0.157) ≥6m: 0.995 (0.699)
Time from last relapse	-	-
ISS stage (II-III vs I)	2.455 (0.037)	<6m: 3.843 (0.520) ≥6m: 1.938 (0.526)
B2-microglobulin (≥3.5 vs <3.5 mg/L)	-	-
Refractory to last prior treatment (yes vs no)	-	-
AIC	**883.822**	**881.688**
Global Grambsch and Therneau test, P value	**0.008**	**0.677**

AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; PFS, progression-free survival; PH, proportional hazards; Vd, bortezomib/dexamethasone.

Table 11. Grambsch and Therneau tests for proportional hazards in covariate-adjusted Cox PH model: Cd vs Vd OS model (ENDEAVOR, patients with 1 prior therapy and no prior bortezomib exposure)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)
Treatment (Cd vs Vd)	0.631 (0.373)
Prior lenalidomide	-
Prior stem cell transplantation	2.034 (0.322)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)
Age (≥65 vs <65)	-
ECOG status (1-2 vs 0)	2.154 (0.867)
Creatinine clearance (50-80)	-
Creatinine clearance (>80)	0.496 (0.897)
Time from diagnosis	0.986 (0.998)
Time from last relapse	-
ISS stage (II-III vs I)	-
B2-microglobulin (≥3.5 vs <3.5 mg/L)	2.926 (0.080)
Refractory to last prior treatment (yes vs no)	2.010 (0.498)
AIC	**640.675**
Global Grambsch and Therneau test, P value	**0.325**
AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; PH, proportional hazards; OS, overall survival; Vd, bortezomib/dexamethasone.	

CRd versus Rd (ASPIRE, patients who have received two prior therapies and no prior lenalidomide)

The assessment of proportional hazards is shown in Table 12 (PFS) and Table 13 (OS) for this subgroup.

For PFS, the proportional hazards assumption was not violated for treatment or most covariates in the model. However, there was evidence of non-proportional hazards for one covariate (time since diagnosis). Therefore, a piece-wise Cox model was also fitted using a 12 month cut-off point for this covariate (based on visual assessment of the Schoenfeld residual plot). In this piece-wise model, the proportional hazards assumption was satisfied for all covariates and for the model as a whole. Again, it is notable that in both the original Cox model and the piece-wise Cox model, the PFS treatment effects for CRd versus Rd are almost identical (HRs of 0.699 and 0.692, respectively), showing that the treatment effect is robust to any non-proportionality of hazards for the covariates.

For OS, the proportional hazards assumption was not violated for treatment or for any of the covariates.

Table 12. Grambsch and Therneau tests for proportional hazards in covariate-adjusted Cox PH model: CRd vs Rd PFS model (ASPIRE, patients who have received two prior therapies and no prior lenalidomide)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)	Piece-wise Cox model Hazard ratio (P value of Grambsch and Therneau test)
Treatment (CRd vs Rd)	0.699 (0.205)	0.692 (0.209)
Prior lenalidomide	1.448 (0.394)	1.542 (0.377)
Prior stem cell transplantation	-	-
Age (≥65 vs <65)	-	-
ECOG status (1-2 vs 0)	1.578 (0.857)	1.594 (0.881)
Creatinine clearance (50-80)	-	-
Creatinine clearance (>80)	-	-
Time from diagnosis	0.994 (0.088)	<12m: 0.985 (0.622) ≥12m: 1.000 (0.769)
Time from last relapse	-	-
ISS stage (II-III vs I)	4.138 (0.825)	4.189 (0.927)
B2-microglobulin (≥3.5 vs <3.5 mg/L)	0.523 (0.275)	0.504 (0.225)
Refractory to last prior treatment (yes vs no)		-
AIC	**1076.913**	**1069.885**
Global Grambsch and Therneau test, P value	**0.049**	**0.275**

AIC Akaike information criterion; CRd, carfilzomib, lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; PFS, progression-free survival; PH, proportional hazards; Rd, lenalidomide/dexamethasone

Table 13. Grambsch and Therneau tests for proportional hazards in covariate-adjusted Cox PH model: CRd vs Rd OS model (ASPIRE, patients who have received two prior therapies and no prior lenalidomide)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)
Treatment (CRd vs Rd)	0.730 (0.271)
Prior lenalidomide	-
Prior stem cell transplantation	0.569 (0.137)
Age (≥65 vs <65)	-
ECOG status (1-2 vs 0)	-
Creatinine clearance (50-80)	0.296 (0.604)

Covariate	Cox PH model Hazard ratio (P value of Grambsch and Therneau test)
Creatinine clearance (>80)	0.298 (0.634)
Time from diagnosis	0.993 (0.295)
Time from last relapse	-
ISS stage (II-III vs I)	10.111 (0.916)
B2-microglobulin (≥3.5 vs <3.5 mg/L)	0.382 (0.389)
Refractory to last prior treatment (yes vs no)	-
AIC	**675.733**
Global Grambsch and Therneau test, P value	**0.265**

AIC Akaike information criterion; CRd, carfilzomib, lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; PH, proportional hazards; OS, overall survival; Rd, lenalidomide/dexamethasone

Conclusion

In order to address the Committee's concern and specific request for additional information, we fully explored the proportional hazards assumption for the best-fitting covariate-adjusted Cox models (stepwise variable selection models) within the relevant subgroups.

- For OS, these analyses showed that the proportional hazards assumption was valid for treatment (Cd vs Vd and CRd vs Rd) and all covariates in the models.
- For PFS, the proportional hazards assumption was valid for treatment (Cd vs Vd and CRd vs Rd), but non-proportionality was observed for a few covariates. This was addressed by fitting piece-wise Cox models and notably the treatment effect (Cd vs Vd and CRd vs Rd) was almost identical in these models, i.e. the treatment effect was robust to any non-proportionality of hazards.

This provides reassurance that the subgroup treatment HRs used in the cost-effectiveness analyses in our original submission are robust.

Given the Committee's stated preference to fit covariate-adjusted parametric models to both arms of the ENDEAVOR and ASPIRE trials in line with NICE DSU guidance (in contrast to our original approach of using covariate-adjusted subgroup analysis derived HRs to estimate PFS and OS curves for Cd and CRd), the PFS and OS HRs derived from these stepwise variable selection models were not utilised in the revised cost-effectiveness analyses. The proportional hazards assumption for the covariate-adjusted parametric models is assessed in Section 2.3.

2.3 Survival modelling and extrapolation

Context

The Committee had concerns around Amgen's approach to survival modelling, which used two separate regression models (for both arms of the trials) to extrapolate the effects over the full model time horizon, with the model to extrapolate the carfilzomib arms based on the subgroup post hoc estimate hazard ratios (ACD Section 4.12, page 11). The Committee further noted that Amgen had had not fully explored the uncertainty of the cost-effectiveness results by fitting all standard distributions (parametric models), exploring different extrapolation methods, and exploring the plausibility of the projection estimates to the observed data from the trials (ACD Section 4.12, page 12).

To address this uncertainty, the Committee specifically requested additional evidence on (ACD Section 4.12, page 12):

- 'The effects of fitting different parametric models, including covariate-adjustments, to both arms of the ENDEAVOR and ASPIRE trials; in line with published technical guidelines, such as NICE DSU Technical Support Document 14.'
- 'The effect of different extrapolation techniques, including exploring a weighted-adjusted covariate model'
- 'An assessment of the resulting predictions from the model and the corresponding covariate-adjusted estimates from the trial'

Implementation

In our original submission, we did not fit parametric models jointly to both treatment arms in ENDEAVOR and ASPIRE because it was necessary to adjust for imbalances in prognostic covariates across study arms within the relevant subgroups, and we felt this was more straightforward to implement within a Cox PH model framework.

The Committee suggested that it would have liked to have seen the effect of fitting different parametric models, including covariate-adjustment, to both arms of the ENDEAVOR and ASPIRE trials, in line with the published technical guideline NICE DSU Technical Support Document 14. To align with NICE DSU guidance, we have now fitted covariate-adjusted parametric models jointly to both treatment arms, though to the best of our knowledge, there is no specific DSU guidance on the most appropriate way to include covariates within such a modelling framework. Therefore, we explored three methods for covariate adjustment:

- Inverse probability weighted (IPW)
- Corrected group prognosis method (CGP)
- Mean of the covariates method (MoC)

To address the Committee's other concerns around the survival model, we have also explored the effect of fitting different parametric distributions using these different extrapolation techniques incorporating covariate adjustment, and have provided a comparison of the model predictions from the most appropriate extrapolation technique and most plausible parametric distributions to corresponding covariate-adjusted estimates from the trials as specifically requested by the Committee.

For each of the different extrapolation techniques incorporating covariate adjustment, we adjusted for the prognostic covariates identified in the best-fitting stepwise variable selection models for PFS and OS for the ENDEAVOR and ASPIRE subgroups described in Section 2.1, rather than applying variable selection procedures within each of the possible parametric survival models (exponential, Weibull, etc). The latter approach was considered excessively complex and has the disadvantage of potentially retaining different covariates for different parametric models. In addition we believe the ERG-suggested LASSO variable selection approach can currently only be implemented within a Cox model.

A summary of the different extrapolation techniques incorporating covariate adjustment we explored is provided in Table 14.

Table 14. Summary of different extrapolation techniques incorporating covariate adjustment

Method	Summary of method
IPW	• With the IPW approach, the treatment effect is estimated in two steps. • In the first step, the covariate distribution is adjusted by reweighting patients using a logistic regression framework. In the logistic regression, the treatment indicator is defined as the dependent variable whereas the covariates identified in the stepwise selection Cox model (Section 2.1) are used as independent variables. With such a logistic regression model, the probability of receiving a particular treatment given the covariates the patient has can be estimated, and by taking the inverse of the estimated probabilities, the patient population is reweighted and imbalances in the included covariates are adjusted for. • In the second step, parametric survival models (e.g. Cox models or parametric survival models) are fitted on the reweighted patient level data without further adjustment. • The IPW approach is similar to the MAIC method, with the exception that patient-level data for baseline covariates are available for both treatment arms instead of just one • The advantage of the IPW method is that after reweighting the patient populations, there is no need for further adjustment, and so the methods proposed in the DSU Technical Support guidance can be directly applied. • Another major advantage of this method is that adjusting for imbalances takes place in the first step (logistic regression), so it is not necessary to explore the proportional hazards assumption for any covariate other than the treatment. In addition, the reweighted trial data (i.e. reweighted Kaplan-Meier curves) and the fitted parametric models can be directly assessed visually.
CGP	• With the CGP method, using survival predictions for each patient given the patient's baseline covariates, the average of the predictions (i.e. a population-averaged value) can be calculated. • An advantage of the CGP approach is that it takes into account the heterogeneity of the patient population for the patient-specific predictions. • The specification of the model must be explicit for each covariate. Proportional hazards should be assumed for each covariate or any non-proportionality should be modelled explicitly which can be challenging regarding the modelling of the time-varying relationship, e.g. selection of the cut-off point for the piece-wise HRs.

	• However, whilst the implementation of the CGP method is technically feasible and not burdensome for a deterministic analysis, conducting probabilistic sensitivity analyses (PSAs) involves either sharing confidential patient-level data within the cost effectiveness model (necessary to automate the PSA) or entering externally simulated survival profiles as hard input parameters. While the first approach is regrettably not feasible, the second approach is not transparent (one would not be able to track the appropriateness of the simulation) and it would increase the size of the electronic file substantially (521 cycles * 1000 simulations values to the Excel model for PFS and OS each)
MoC	• With the MoC method, the mean value of each covariate used in the prediction equation is used to predict PFS and OS for an average patient. • In terms of implementation, using the MoC method for both deterministic analyses and PSAs is transparent and not burdensome, and this method has been claimed to be the most commonly used.[7] • An important limitation of the MoC method is that it can yield skewed survival estimates because the survival of a patient with average baseline characteristics (i.e. 'average patient') might be different from the average of the patient-specific survival estimates. • In addition, mean covariate values between 0 and 1 are assigned to binary variables, which are meaningless on an individual level, and the hazards for a hypothetical average individual rather than a population-average value are estimated. • The MoC method was criticised for its validity in the ERG critique of the manufacturers' submission in the part-review of TA171, where the CGP method was proposed as a better method than MoC for extrapolating survival.[8]
CGP, corrected group prognosis; DSU, decision support unit; ERG, evidence review group; IPW, inverse probability weighted; MAIC, matching-adjusted indirect comparison; MoC, mean of covariates; OS, overall survival; PFS, progression-free survival; TA, technology appraisal.	

Of the different extrapolation techniques incorporating covariate adjustment and different parametric distributions, the IPW-based Weibull distribution yielded the most clinically plausible projections of PFS and OS for both the ENDEAVOR and ASPIRE subgroups of interest. In addition, it provided a generally good statistical fit, an excellent fit to corresponding covariate-adjusted trial data, and satisfied proportional hazards assumptions for both PFS and OS in the ENDEAVOR and ASPIRE subgroups of interest. In contrast, none of the parametric distributions using either the CGP or MoC methods resulted in clinically plausible projections of OS for either the ENDEAVOR or ASPIRE subgroups of interest.

Therefore, we have presented a comprehensive overview of analyses carried out using the IPW method and results from these analyses in the main part of this response document, and provided results for the other extrapolation techniques incorporating covariate adjustment in Appendix D.

Summary of IPW method analyses

In order to fully explore the IPW method in the context of the current appraisal we:

- Fitted joint parametric regression models on the weighted patient-level data from ENDEAVOR and ASPIRE in which the treatment group was the only covariate included.
- Assessed the fit of the standard parametric survival models where we jointly fitted the two treatment arms. Based on the AIC/BIC values and visual assessment of clinical plausibility, we selected the most plausible models for PFS and OS.
- Compared the visual fit of the most plausible parametric model with the corresponding covariate-adjusted estimates from the ENDEAVOR trial data.
- Assessed the proportional hazards assumption for the treatment indicator for the most plausible parametric model based on Grambsch and Therneau tests for proportionality of hazards, visual assessment of Schoenfeld residual plots, and visual assessment of log-log plots

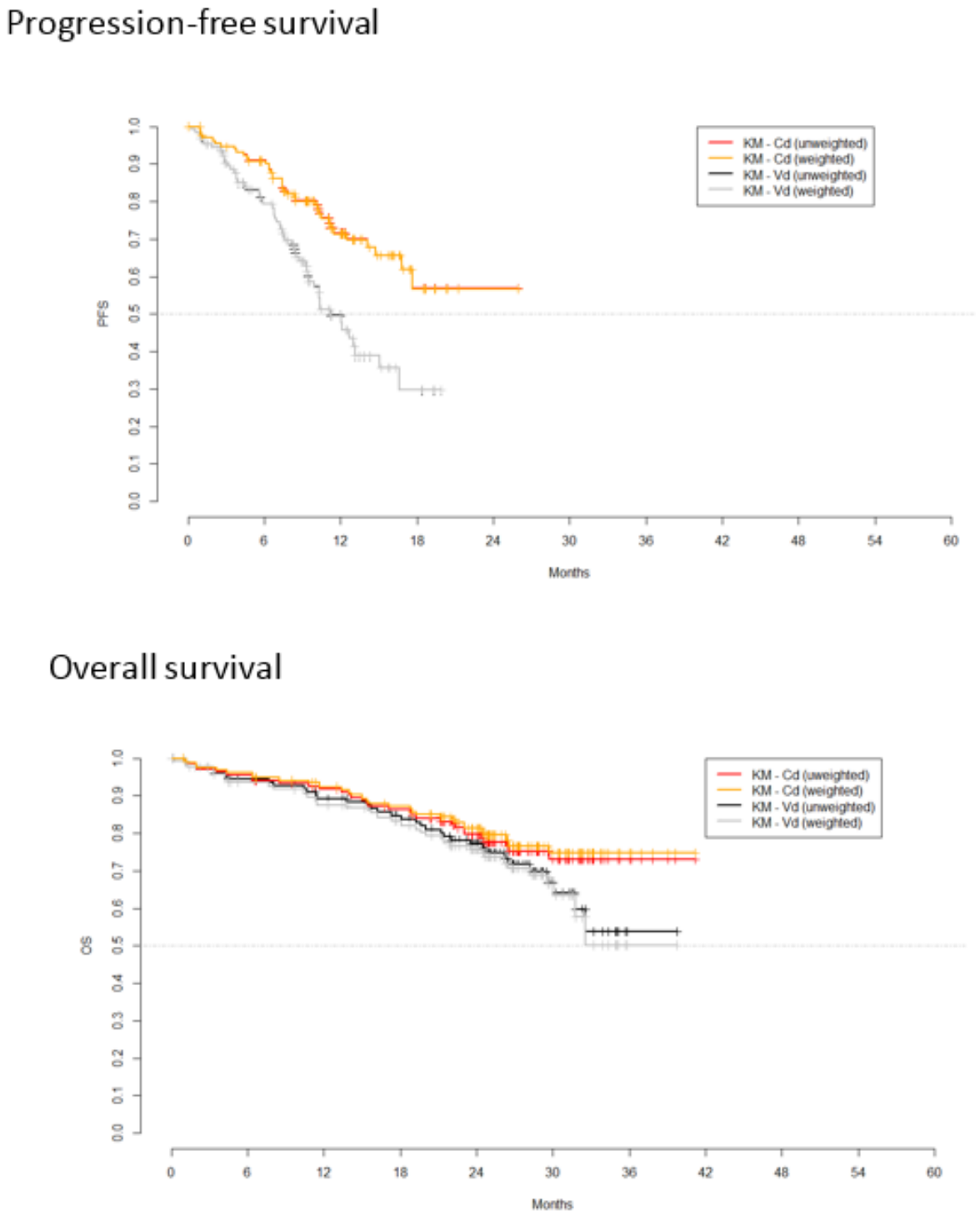
The results of these analyses are described below.

Results of the IPW analyses for Cd versus Vd (patients who have received one prior therapy and no prior bortezomib)

PFS and OS curves from the fitted joint parametric regression model on the weighted patient-level data from ENDEAVOR in which treatment group was the only included covariate are provided in Figure 1. For comparison, unweighted curves are also presented. The Cd:Vd HRs for PFS and OS for the weighted model were 0.452 and 0.640, respectively. For the unweighted model, the Cd:Vd HRs for PFS and OS were 0.449 and 0.741, respectively.

Notably, the weighted Cd:Vd HRs for PFS and OS (PFS: 0.452; OS: 0.640) are similar to the HRs estimated for the ENDEAVOR subgroup of interest in our original submission (PFS: 0.412; OS: 0.592), which supports the robustness of the treatment effect estimates used to inform the cost-effectiveness model in our original submission.

Figure 1. PFS and OS Kaplan-Meier curves from the joint parametric regression model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

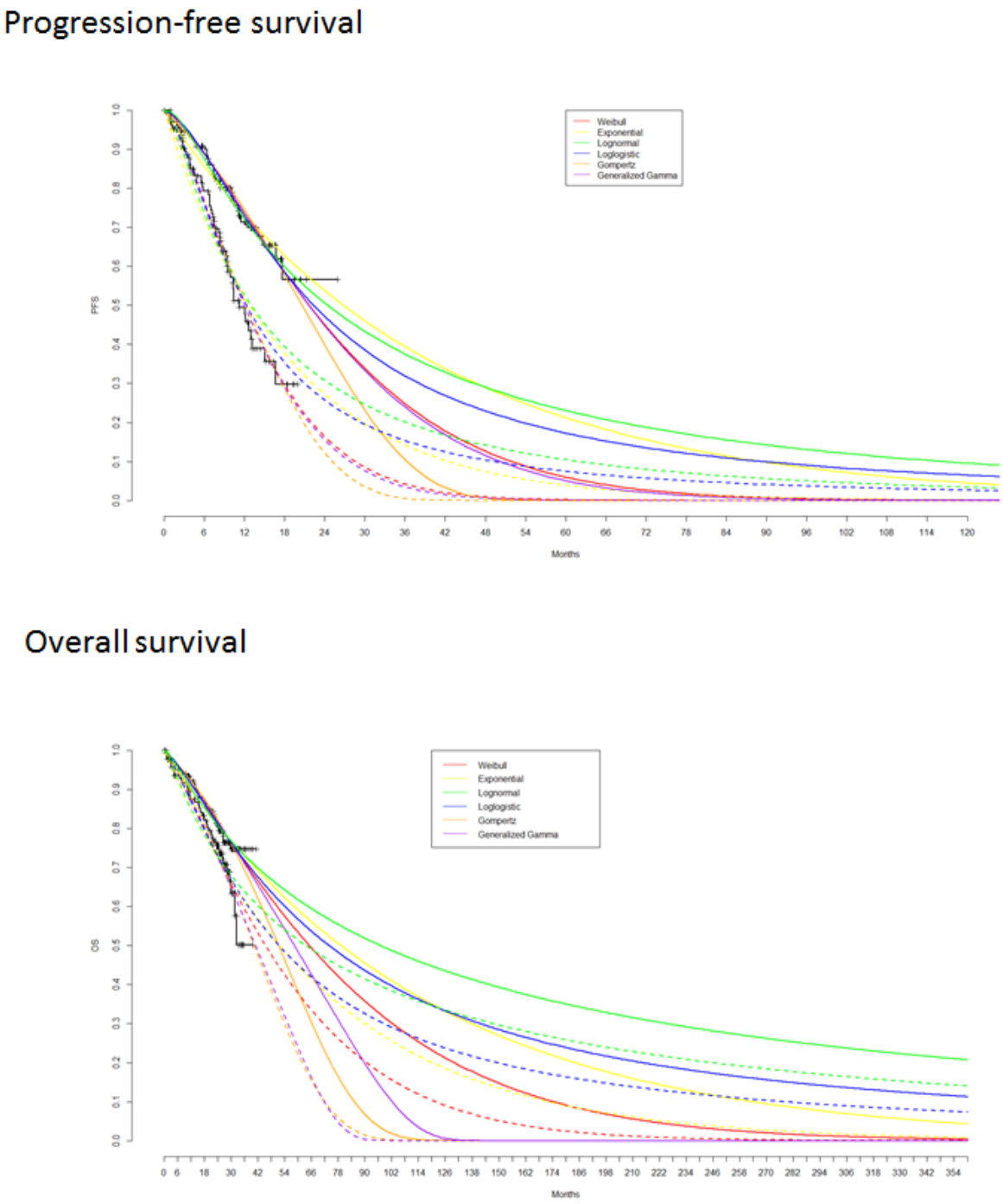


Note: Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ENDEAVOR subgroup of patients with one prior therapy and no prior bortezomib (Section 2.1).

Cd, carfilzomib/dexamethasone; KM, Kaplan-Meier; OS, overall survival; PFS, progression-free survival; Vd, bortezomib/dexamethasone

The standard parametric model fits on the weighted data are provided in Figure 2, and model fit statistics (AIC and BIC) are provided in Table 15.

Figure 2. PFS and OS parametric curves fitted to the joint parametric regression model weighted data (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



Note: Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ENDEAVOR subgroup of patients with one prior therapy and no prior bortezomib (Section 2.1). Solid lines are for Cd and dotted lines are for Vd.

OS, overall survival; PFS, progression-free survival

Table 15. Model fit statistics (AIC and BIC) for the parametric PFS and OS curves fitted to the joint parametric regression model weighted data (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	Model	AIC	BIC
PFS	Weibull	781.4131	792.1748
	Exponential	790.8056	797.9801
	Gompertz	785.0662	795.8279
	Generalized Gamma	783.3869	797.7359
	Lognormal	783.1394	793.9011
	Loglogistic	789.1732	799.9349
OS	Weibull	725.4559	736.2176
	Exponential	726.2330	733.4075
	Gompertz	723.4136	734.1754
	Generalized Gamma	725.8774	740.2264
	Lognormal	727.3010	738.0627
	Loglogistic	732.9990	743.7608

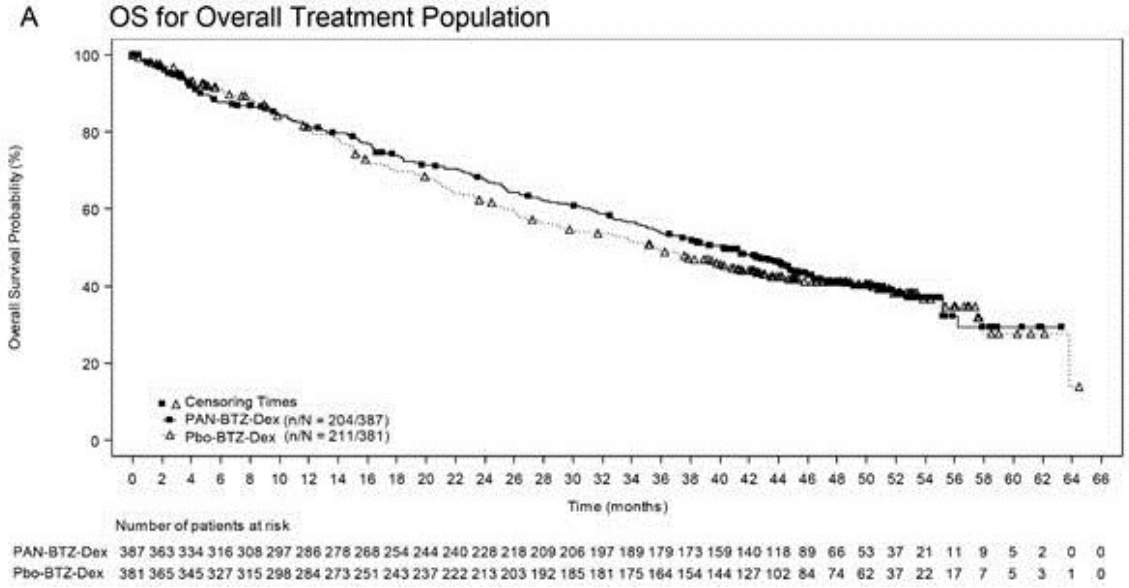
AIC, Akaike information criterion; BIC, Bayesian information criterion; OS, overall survival; PFS, progression-free survival

The AIC and BIC values were close across all curves for both PFS and OS. For PFS, the Weibull model was the best fitting model (both AIC and BIC). For OS, the Gompertz was the best fitting model (both AIC and BIC), though the Weibull model was not meaningfully different.

With respect to clinical plausibility, the log-logistic and log-normal curves predict a high proportion of Vd-treated patients to remain alive after 40 years and hence were considered inappropriate to model the survival of patients with relapsed and/or refractory multiple myeloma (R/RMM) receiving Vd as a second line (2L) therapy. The exponential curve predicts a high proportion of Vd-treated patients to remain alive after 10 and 20 years, and hence was also considered inappropriate. The Gompertz and Gamma curves present a very pessimistic estimate of long-term survival with 0% Vd-treated patients predicted to be alive after 10 years. This is inconsistent with long-term observational data from the HMRN registry which demonstrated that in 2L patients not exposed to prior bortezomib (likely to be treated with bortezomib in clinical practice), around 5% were still alive at 10 years,[9] despite being older and less fit than patients enrolled in ENDEAVOR, and therefore having a poorer prognosis. The Gompertz and Gamma curves were therefore considered inappropriate to model the survival of patients with R/RMM receiving Vd as a 2L therapy.

As discussed in our original submission, the Weibull distribution has been demonstrated to be an appropriate distribution for modelling survival outcomes in R/RMM based on curve fitting applied to the Haematological Malignancy Research Network (HRMN) long-term OS data. To further validate the choice of the Weibull distribution for OS extrapolations, we provided an analysis in response to ERG questions utilising long-term follow-up data from the recent PANORAMA-1 study in R/RMM (panobinostat/bortezomib/dexamethasone [PVd] versus Vd) (Figure 3).

Figure 3. OS Kaplan-Meier curves from PANORAMA-1

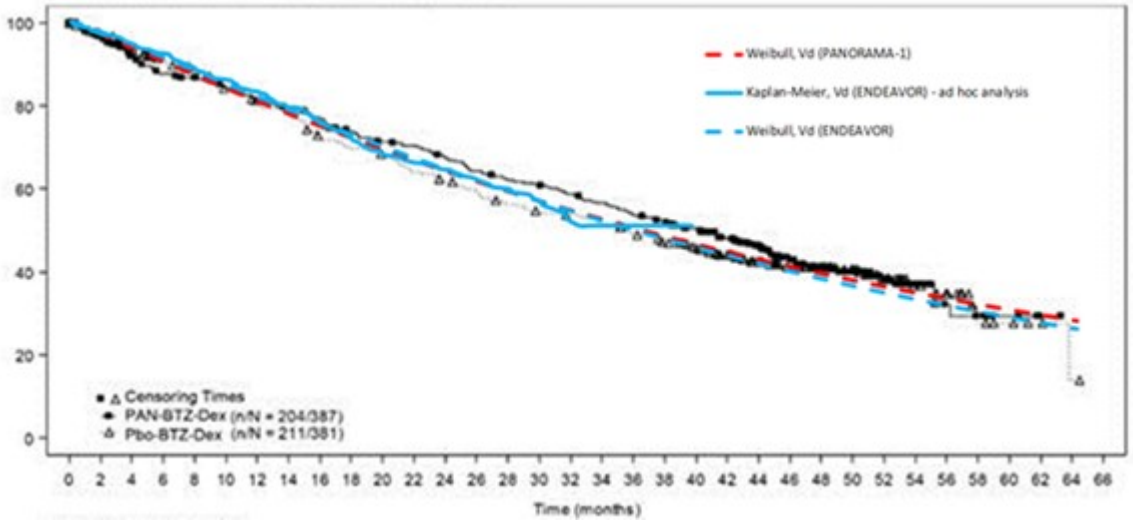


References: San-Miguel *et al.,* 2015[10]

OS, overall survival; PAN-BTZ-Dex, panobinostat/bortezomib/dexamethasone; Pbo-BTZ-Dex, placebo/bortezomib/dexamethasone

Specifically, the Kaplan-Meier OS curve for Vd was digitised, patient-level data were simulated such that they replicated the Kaplan-Meier curve, and a Weibull model was fitted to the simulated patient level data. In a next step, the fitted Weibull model was plotted on the Kaplan-Meier curve to visually assess its fit to the data (Figure 4). Since the OS curve for Vd patients in the ENDEAVOR study is very similar to the OS curve for Vd patients in the PANORAMA-1 study (up till the end of the follow-up time for ENDEAVOR), and because the Weibull distribution (both using the ENDEAVOR ITT population and the full PANORAMA-1 population) fitted the PANORAMA-1 OS data well, it was concluded that the Weibull distribution is likely the most appropriate choice for extrapolations for Vd and Cd, and in a broader context for treatments in R/RMM. Scenario analysis using alternative distributions for OS are presented in Appendix F.

Figure 4. OS curves for Vd in ENDEAVOR versus PANORAMA-1

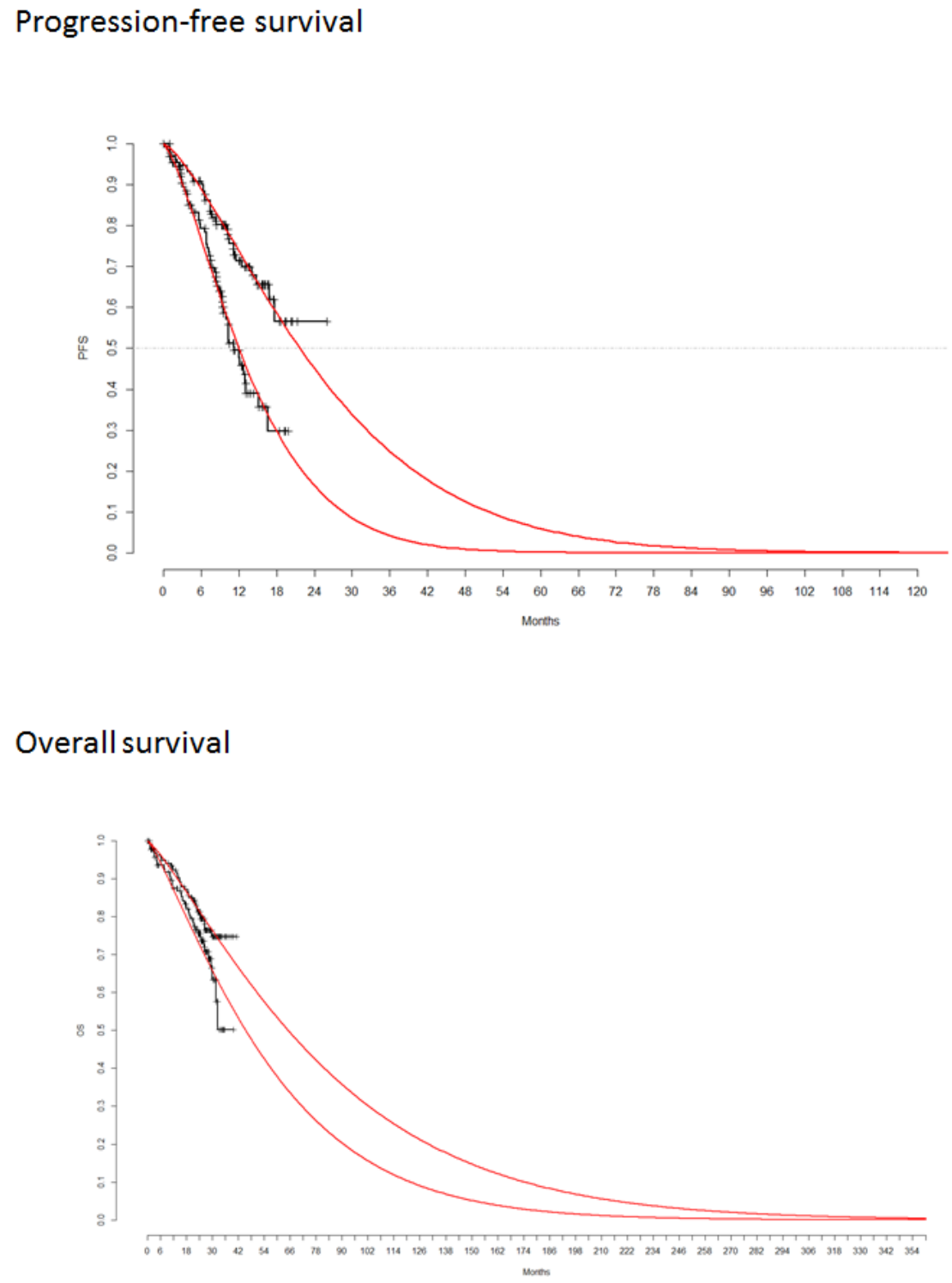


PAN-BTZ-Dex, panobinostat/bortezomib/dexamethasone; Pbo-BTZ-Dex, placebo/bortezomib/dexamethasone; Vd, bortezomib/dexamethasone

Based on the above, the Weibull model was considered the most appropriate model for OS given its generally good statistical fit and that it projected the most clinically plausible estimates of OS. The Weibull model was considered the most appropriate model for PFS given its generally good statistical fit and that it was considered to have a plausible extrapolation; in the NICE appraisal of PVd, the Weibull was considered the best fitting and most appropriate distribution for PFS for Vd based on almost complete follow-up data for PFS.[11] Scenario analysis using alternative distributions for PFS are presented in Appendix F.

A comparison of the IPW-based Weibull PFS and OS model projections with corresponding covariate-adjusted ENDEAVOR Kaplan-Meier curves is provided in Figure 5. This demonstrates that the IPW-based Weibull model provides an excellent fit to and closely matches the covariate-adjusted PFS and OS estimates from the ENDEAVOR trial data.

Figure 5. Comparison of IPW-based Weibull PFS and OS model projections with corresponding covariate-adjusted (weighted) ENDEAVOR Kaplan-Meier curves (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



Note: Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ENDEAVOR subgroup of patients with one prior therapy and no prior bortezomib (Section 2.1). Stabilised weights are estimates as: w[i] * N / sum(w[i]). The weighted Weibull model is based on weighted patient-level data with treatment indicator as the only covariate.
Note: upper curves are for Cd, lower curves are for Vd

The assumption of proportional hazards for the IPW-based Weibull model was assessed for the treatment indicator (given that treatment group was the only covariate included in the model) using Grambsch and Therneau tests for proportionality of hazards (p=0.582 for PFS, p=0.311 for OS) and visual assessment of Schoenfeld residual plots (Figure 6) and log-log plots (Figure 7). These assessments indicate that the proportional hazards assumptions for both PFS and OS based on the IPW-based Weibull model are satisfied.

Figure 6. Schoenfeld residuals plots – Cd vs Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

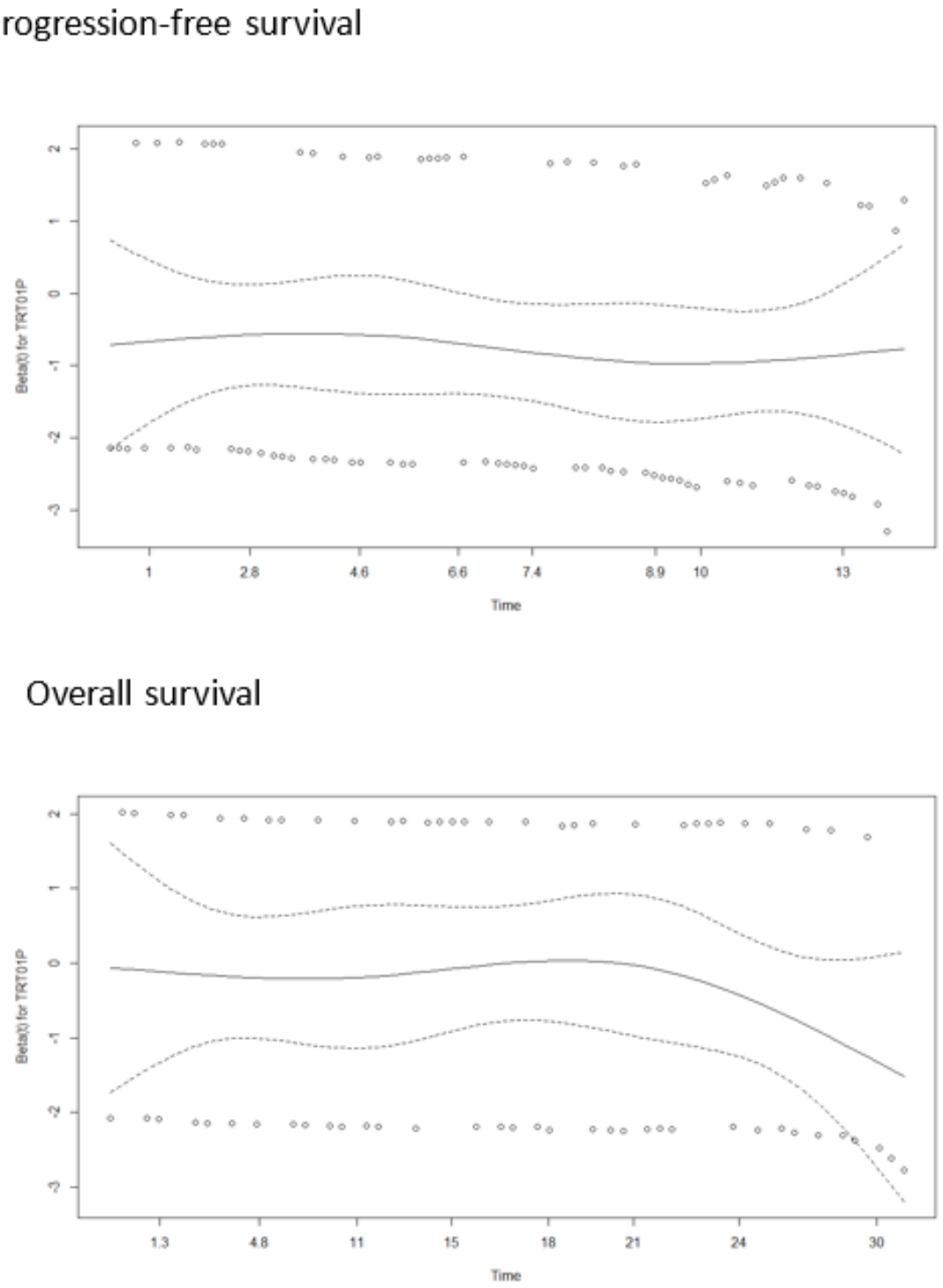
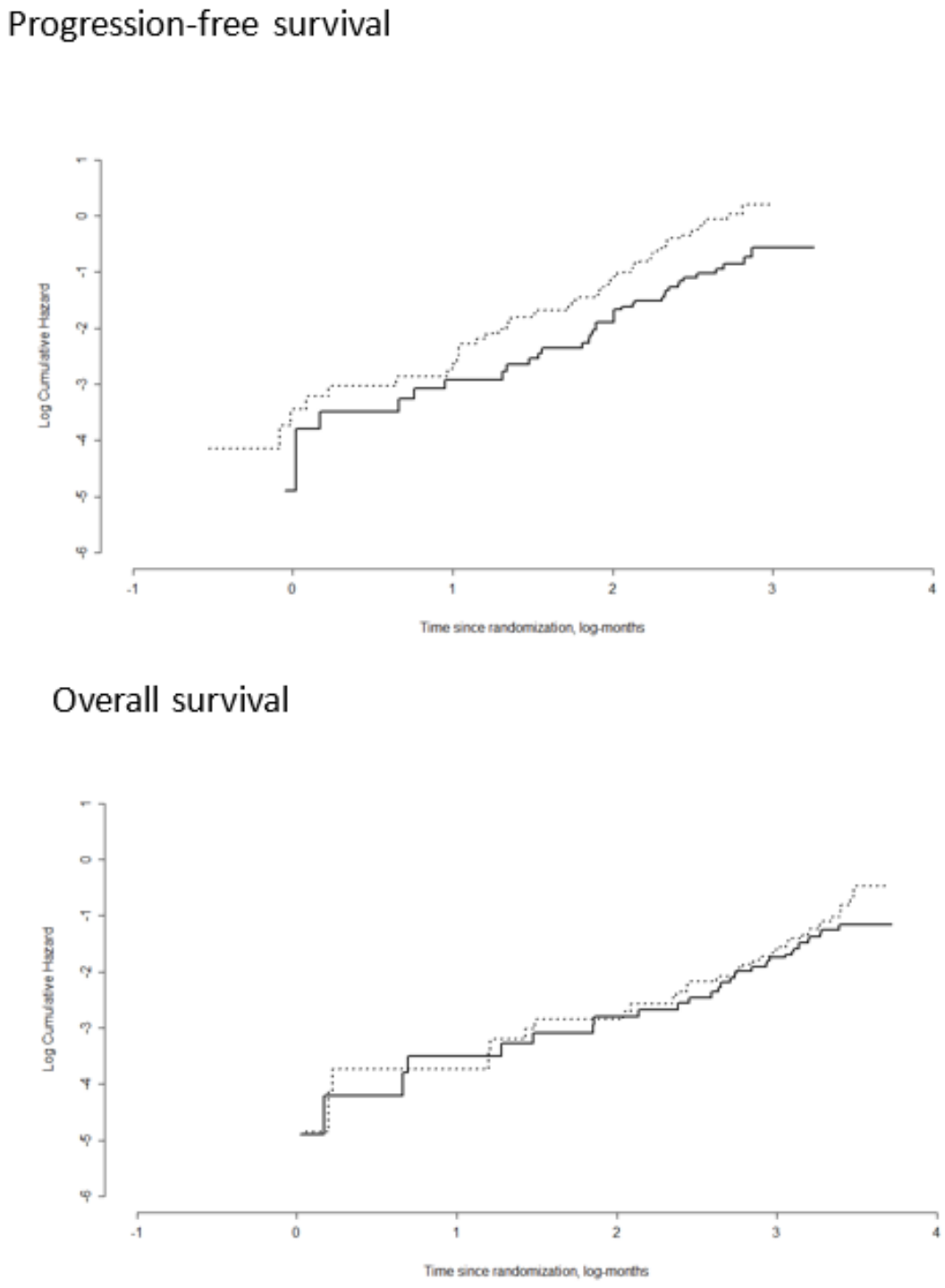


Figure 7. Log-log plots – Cd vs Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



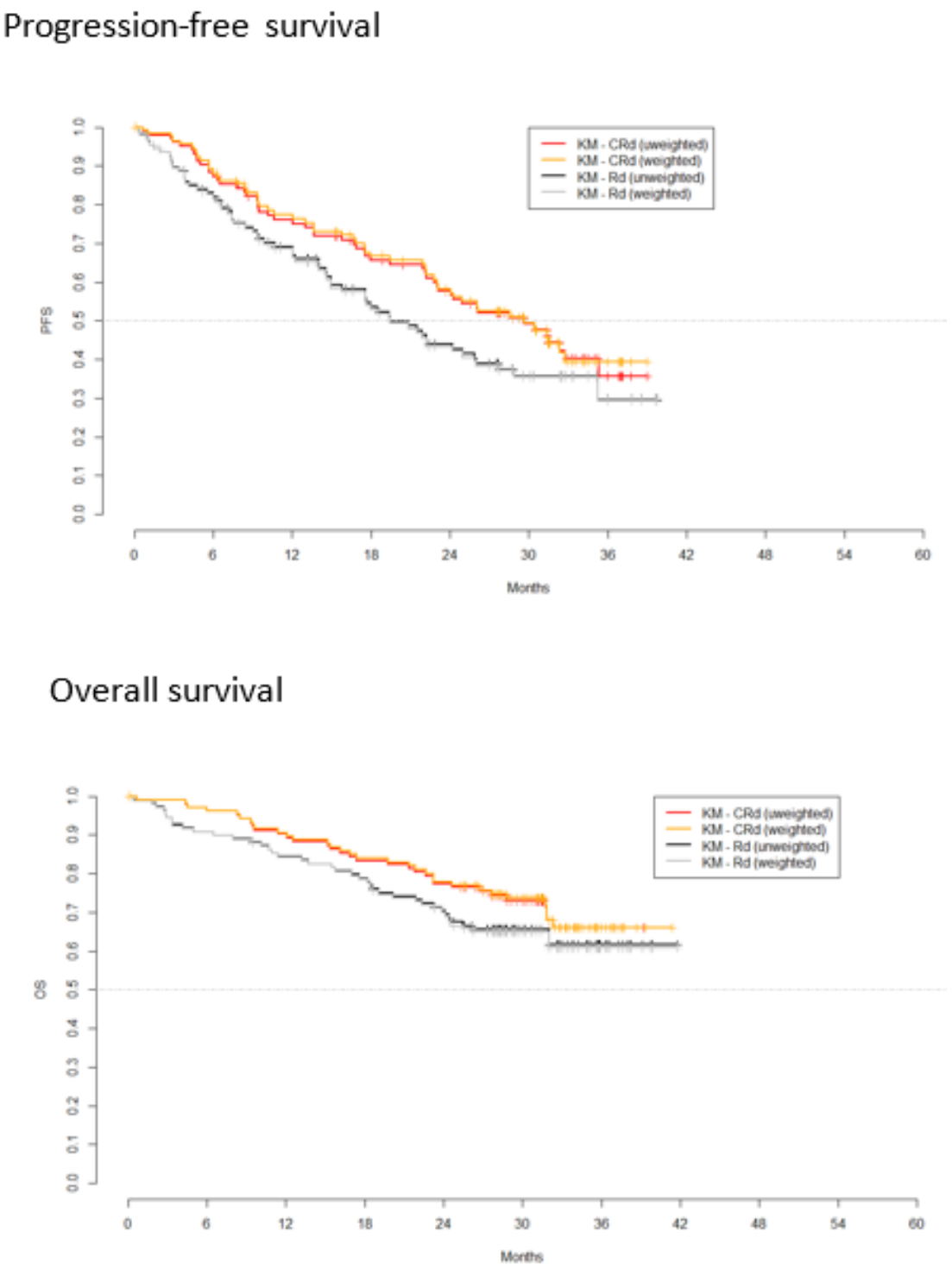
Results of the IPW analyses for CRd versus Rd (patients who have received two prior therapies and no prior lenalidomide)

PFS and OS curves from the fitted joint parametric regression model on the weighted patient-level data from ASPIRE in which treatment group was the only included covariate are provided in

Figure 8. For comparison, unweighted curves are also presented. The CRd:Rd HRs for PFS and OS from the weighted model were 0.698 and 0.743, respectively. For the unweighted model, the Cd:Vd HRs for PFS and OS were 0.726 and 0.769, respectively.

Notably, the weighted CRd:Rd HRs for PFS and OS (PFS: 0.698; OS: 0.743) are similar to the HRs estimated for the ASPIRE subgroup of interest in our original submission (PFS: 0.686; OS: 0.737), which supports the robustness of the treatment effect estimates used to inform the cost-effectiveness model in our original submission.

Figure 8. PFS and OS Kaplan-Meier curves from the joint parametric regression model (ASPIRE, patients with two prior therapies and no prior lenalidomide)



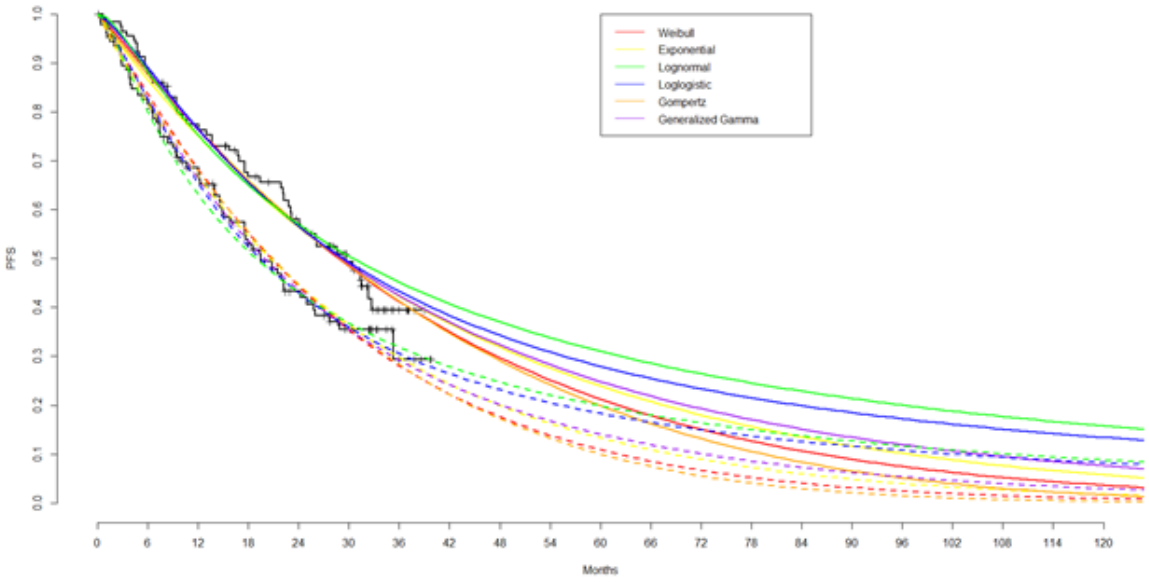
Note: Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ASPIRE subgroup of patients with two prior therapies and no prior lenalidomide (Section 2.1).

CRd, carfilzomib/lenalidomide/dexamethasone; KM, Kaplan-Meier; Rd, lenalidomide/dexamethasone

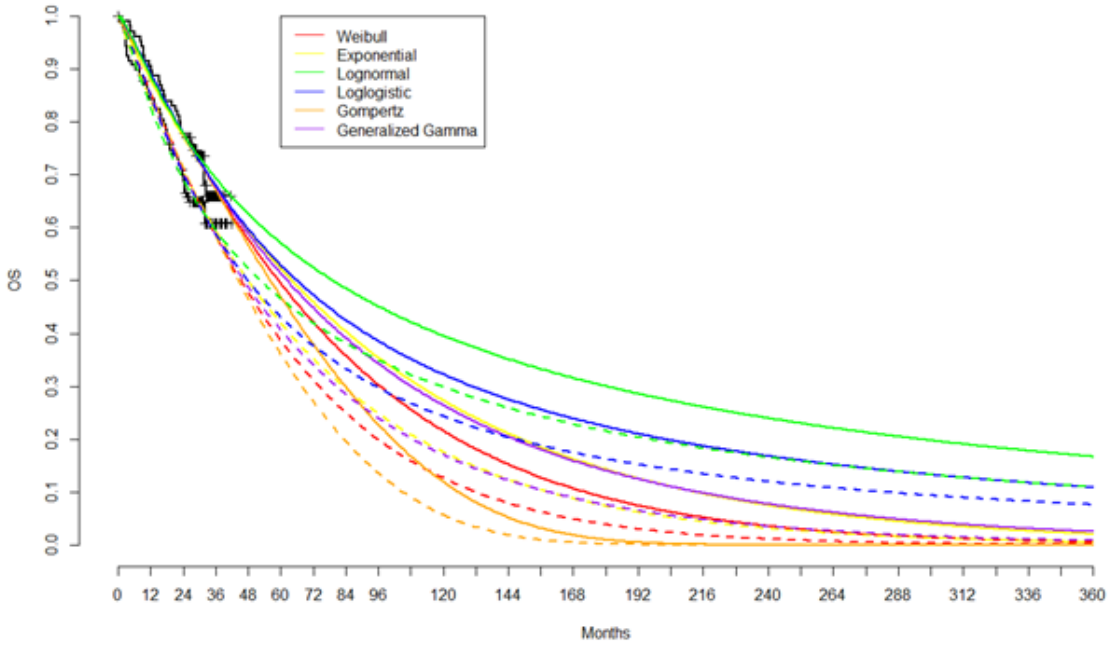
The standard parametric model fits on the weighted data are provided in Figure 9, and model fit statistics (AIC and BIC) are provided in Table 16.

Figure 9. PFS and OS parametric curves fitted to the joint parametric regression model weighted data (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Progression-free survival



Overall survival



Note: 'Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ASPIRE subgroup of patients with two prior therapies and no prior lenalidomide (Section 2.1).Solid lines are for CRd and dotted lines are for Rd.

Table 16. Model fit statistics (AIC and BIC) for the parametric PFS and OS curves fitted to the joint parametric regression model weighted data (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Model	AIC	BIC
PFS	Weibull	1049.215	1059.27
	Exponential	1048.412	1055.116
	Gompertz	1049.903	1059.958
	Generalized Gamma	1050.318	1063.726
	Lognormal	1048.775	1058.831
	Loglogistic	1051.575	1061.631
OS	Weibull	758.5665	768.6221
	Exponential	757.5528	764.2565
	Gompertz	758.9114	768.9670
	Generalized Gamma	760.4390	773.8464
	Lognormal	760.4414	770.4970
	Loglogistic	758.3595	768.415

AIC, Akaike information criterion; BIC, Bayesian information criterion; OS, overall survival; PFS, progression-free survival

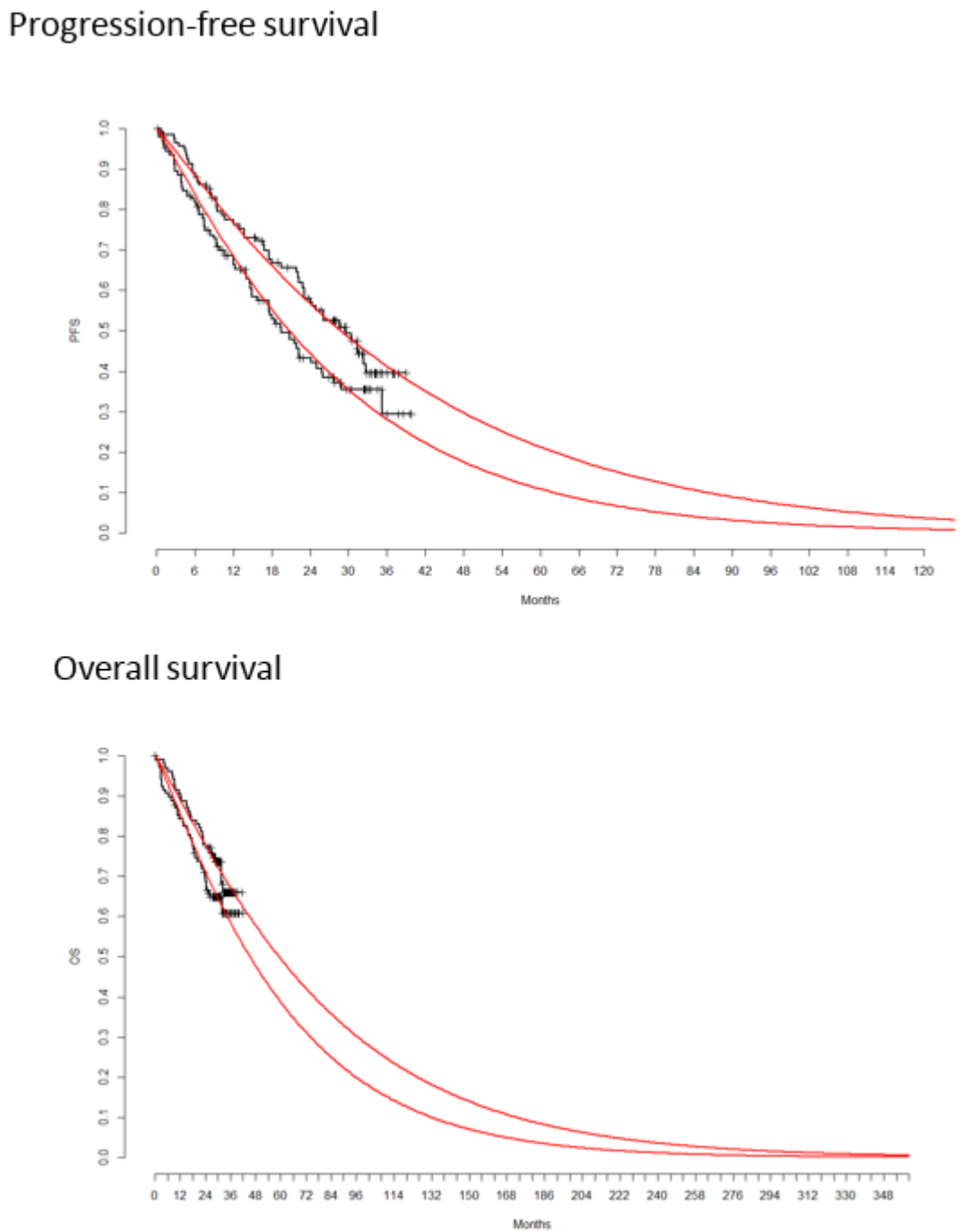
The AIC and BIC values were close across all curves for both PFS and OS. For both PFS and OS, the exponential model was the best fitting model (both AIC and BIC), though the Weibull model (considered the most plausible model for the earlier comparison of Cd versus Vd) was not meaningfully different.

As discussed earlier for the comparison for Cd versus Vd, the Weibull curve is the most appropriate shape for reflecting survival outcomes in R/RMM based on observational data from the HMRN registry and feedback from clinical experts.

Based on the above, the Weibull model was considered the most appropriate model for OS given its generally good statistical fit and that it projected the most clinically plausible estimates of OS. The Weibull model was considered the most appropriate model for PFS given its generally good statistical fit and that it was considered to have a plausible extrapolation; in the NICE appraisal of PVd, the Weibull was considered the best fitting and most appropriate distribution for PFS for Vd based on almost complete follow-up data for PFS.[11] Scenario analysis using alternative distributions for PFS and OS are presented in Appendix F.

A comparison of the Weibull PFS and OS model projections with corresponding covariate-adjusted ASPIRE Kaplan-Meier curves is provided in Figure 10, which demonstrates that the Weibull model provides an excellent fit to and closely matches the PFS and OS estimates from the ASPIRE trial data.

Figure 10. Comparison of IPW-based Weibull PFS and OS model projections with corresponding covariate-adjusted (weighted) ASPIRE Kaplan-Meier curves (ASPIRE, patients with two prior therapies and no prior lenalidomide)



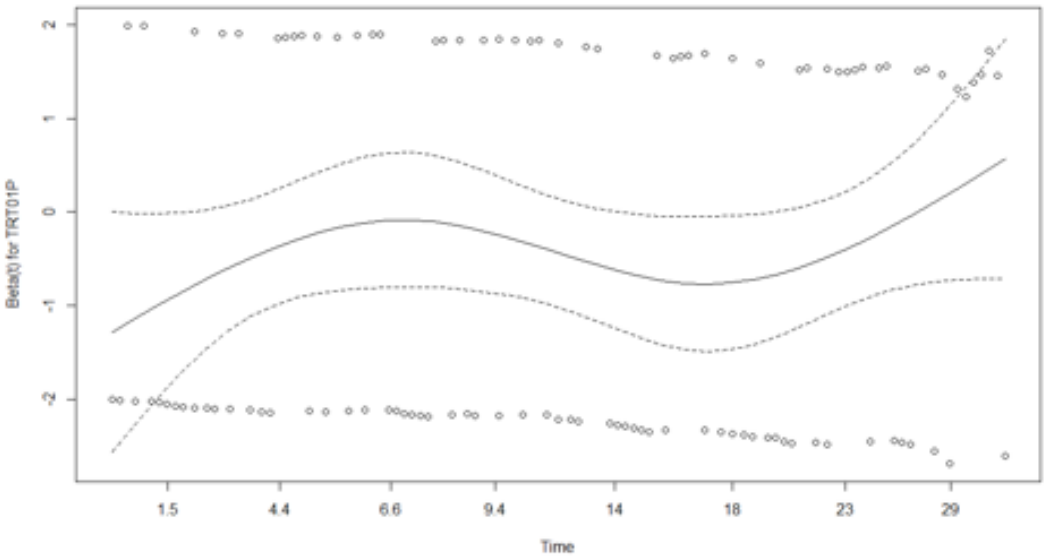
Note: 'Variables used for matching are the same as those selected based on the best fitting stepwise selection models for PFS and OS in the ASPIRE subgroup of patients with two prior therapies and no prior lenalidomide (Section 2.1). Stabilised weights are estimates as: w[i] * N / sum(w[i]). The weighted Weibull model is based on weighted patient-level data with treatment indicator as the only covariate.
Note: upper curves are for CRd, lower curves are for Rd

The assumption of proportional hazards for the IPW-based Weibull model was assessed for the treatment indicator (given that treatment group was the only covariate included in the

model) using Grambsch and Therneau tests for proportionality of hazards (p=0.365 for PFS, p=0.300 for OS) and visual assessment of Schoenfeld residual plots (Figure 11) and log-log plots (Figure 12). These assessments indicate that the proportional hazards assumptions for both PFS and OS based on the IPW-based Weibull model are satisfied.

Figure 11. Schoenfeld residuals plots – CRd vs Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Progression-free survival



Overall survival

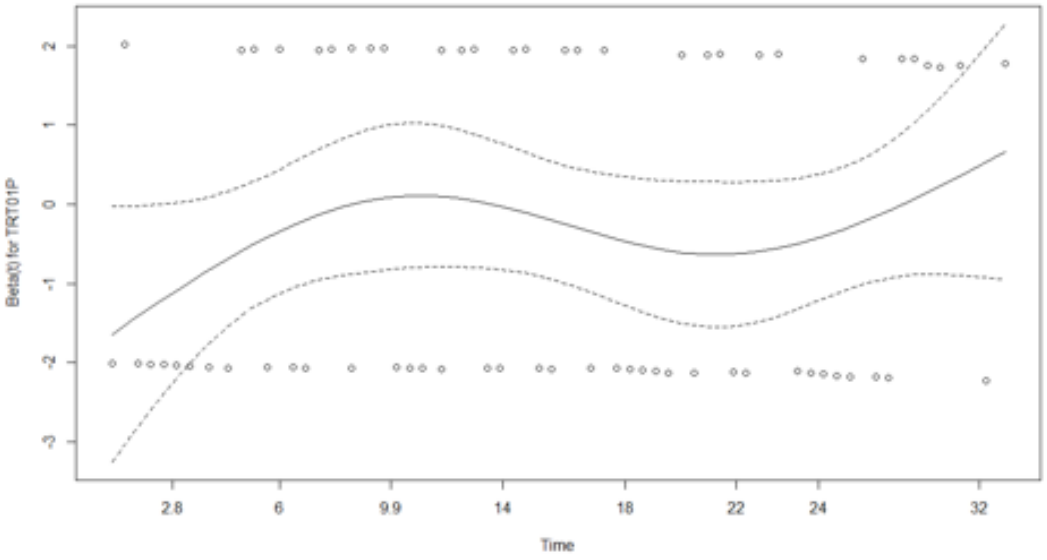
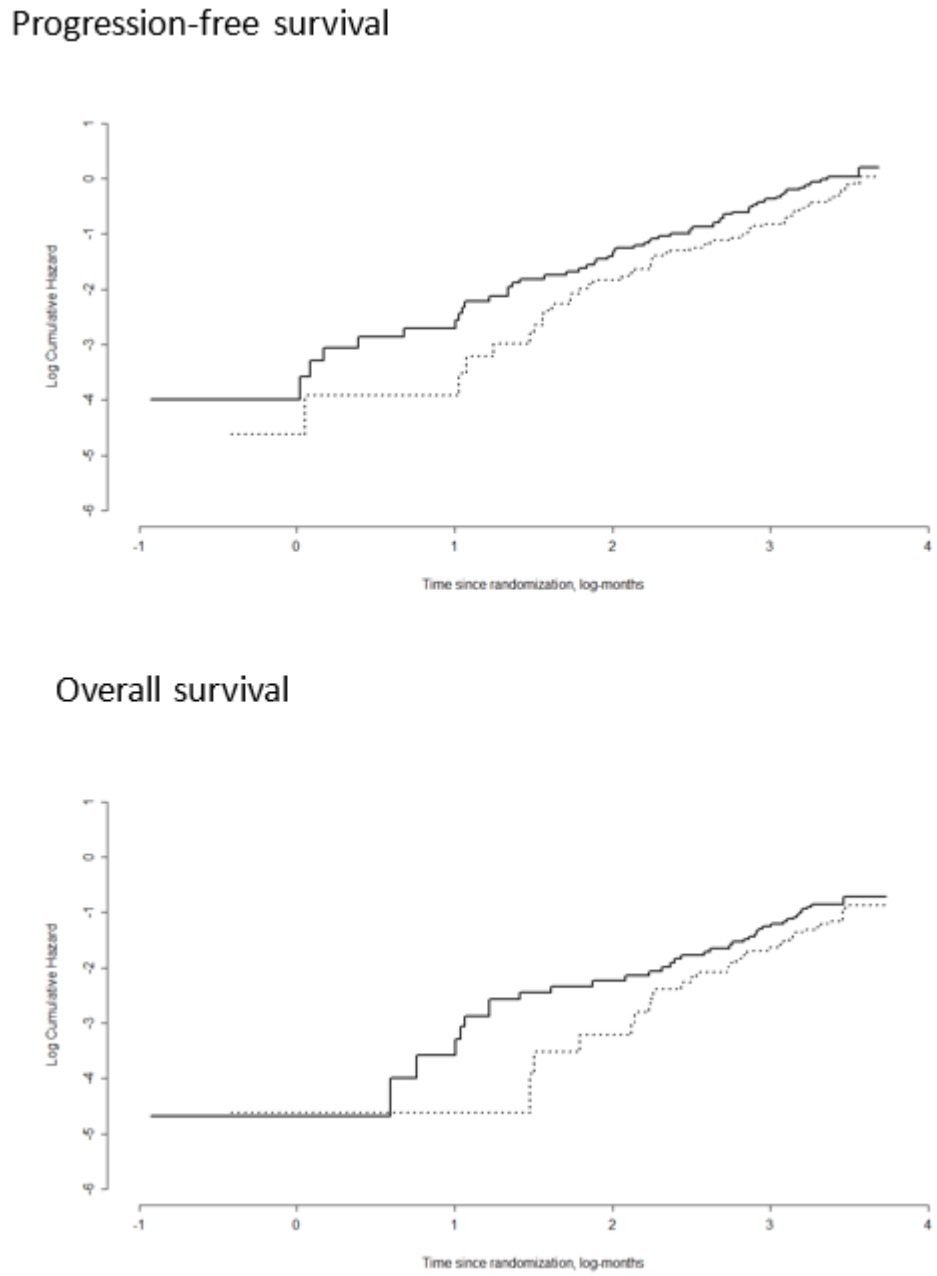


Figure 12. Log-log plots – Cd vs Vd (ASPIRE, patients with two prior therapies and no prior lenalidomide)



Conclusion

In order to address the Committee's concerns around separately modelling the different arms in the ENDEAVOR and ASPIRE trials, and to align with NICE DSU guidance, we have now fitted covariate-adjusted parametric models jointly to both treatment arms of ENDEAVOR and ASPIRE. As specifically requested by the Committee, we have explored of the effects of different extrapolation techniques (including exploring a weighted-adjusted covariate model) and fitting different parametric distributions, and have provided a comparison of the model predictions from the most appropriate extrapolation technique incorporating covariate

adjustment and most plausible parametric distributions to corresponding covariate-adjusted estimates from the trials.

Of the different techniques to incorporate covariate adjustment and different parametric distributions, The IPW-based Weibull model yielded the most clinically plausible projections of PFS and OS for both the ENDEAVOR and ASPIRE subgroups of interest. In addition, it provided a generally good statistical fit, an excellent fit to corresponding covariate-adjusted trial data, and satisfied proportional hazards assumptions for both PFS and OS in the ENDEAVOR and ASPIRE subgroups of interest. Notably, the IPW-weighted treatment effect estimates (Cd vs Vd and CRd vs Rd) for PFS and OS are similar to the treatment effects estimated for the ENDEAVOR and ASPIRE subgroups of interest in our original submission, which supports the robustness of the treatment effect estimates used to inform the cost-effectiveness model in our original submission.

In contrast to the IPW-Weibull based model, none of the parametric distributions using either the CGP or MoC methods resulted in clinically plausible projections of OS for either the ENDEAVOR or ASPIRE subgroups of interest.

Given the above, the IPW-based Weibull model was considered the most appropriate way in which to model survival (PFS and OS) for both ENDEAVOR and ASPIRE, with resulting ICERs presented in Table 17 and Table 18, and this has therefore been used in our revised cost-effectiveness analysis (Section 6).

Table 17. Revised cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib using IPW-based Weibull for OS and PFS (original base case ICER: £20,044)

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£93,769	4.23	2.94				
Cd	£117,660	5.74	4.09	£23,891	1.51	1.15	£20,766
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone							

Table 18. Revised cost-effectiveness analysis for CRd versus Rd in patients with two prior therapies and no prior lenalidomide using IPW-based Weibull for OS and PFS (original base case ICER: £33,467)

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Rd	£94,528	4.36	2.88				
CRd	£127,140	5.46	3.67	£32,612	1.10	0.79	£41,429
CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; Rd, lenalidomide/dexamethasone; QALYs, quality-adjusted life years							

3 Derived health state utility values

Context

The ACD states the Committee 'noted that the company had used a mixed method, using published utility values from Agthoven et al. (2004) and mapped utility values from the trials. The committee heard that the company had used the ERG's preferred approach in the sensitivity analysis. This derived utility values straight from trial data, using a mapping algorithm from Proskorovsky et al (2014). The committee noted that using values derived straight from trial data was more plausible and more closely followed the NICE reference case' (Section 4.14, page 13).

The Committee concluded that the utility estimates used in the base case analysis were not appropriate and that it would have preferred to see mapped utility from trial data.

Implementation

In our original submission, utilities were derived by applying the results of a mapping analysis to the health state utility values reported by van Aghtoven *et al.* 2004.[12] This approach was taken given that van Aghtoven *et al.* 2004 provides UK-specific EQ-5D utility values (contrary to mapping) derived directly from patients with MM that have been used in previous economic evaluations of treatments for MM,[13-15] and meets the NICE reference case (with respect to utilising EQ-5D data directly from patients).[16] To ensure consistency of the decision making process and to avoid limitations associated with mapping from the ASPIRE trial EORTC QLQ-C30 tool to EQ-5D, we considered that use of the van Agthoven baseline utilities were most appropriate for our original base case given the NICE reference case highlights the need for 'consistency across appraisals.'[16] In our response to clarification questions from the ERG, we conducted a scenario analysis using utility values mapped directly from European Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 (EORTC QLQ-C30) data collected in the ENDEAVOR (Cd vs. Vd) and ASPIRE (CRd vs. Rd) trials using the mapping algorithm from Proskorovsky *et al.,* 2014 (Table 19 and Table 20).

Table 19. Utility values mapped directly from EORTC QLQ-C30 data from ENDEAVOR for Cd versus Vd (patients with one prior therapy and no prior bortezomib)

Health state	Cd	Vd	Calculations/assumptions
Pre-progression (Cycles 1–2)	0.737	0.737	• Baseline utility in ENDEAVOR
Pre-progression (later cycles)	0.741	0.714	• Cd: baseline utility + average change over time (+ 0.004) • Vd: baseline utility minus the difference between Cd and Vd (– 0.027) • Utilities for off-treatment were assumed to be the same as on-treatment utilities because the impacts of adverse events associated with treatments were taken into account separately

Health state	Cd	Vd	Calculations/assumptions
Post-progression, subsequent treatment phase	0.638	0.638	• Pre-progression utility in later cycles for Vd (0.714) minus the disutility associated with progression following Vd (0.076)
Post-progression, BSC	0.638	0.638	• Assumed equal to the utility in post-progression for the subsequent treatment phase
BSC, best supportive care; Cd, carfilzomib/dexamethasone; Vd, bortezomib/dexamethasone			

Table 20. Utility values mapped directly from EORTC QLQ-C30 data from ENDEAVOR for Cd versus Vd (patients with one prior therapy and no prior bortezomib)

Health state	CRd	Rd	Calculations/assumptions
Pre-progression (Cycles 1–2)	0.690	0.690	• Baseline utility in ASPIRE patients with two prior treatments
Pre-progression (later cycles)	0.699	0.683	• CRd: baseline utility + average increase in utility from baseline (+ 0.00892) • Rd: baseline utility + average increase in utility from baseline (+ 0.00892) minus the utility difference between Rd and CRd (0.016) • Utilities for off-treatment were assumed to be the same to on-treatment utilities because the impact of adverse events associated with treatments were taken into account separately
Post-progression, subsequent treatment phase	0.637	0.637	• Pre-progression utility in later cycles for Rd (0.683) minus the disutility associated with progression following Vd (0.046)
Post-progression, BSC	0.637	0.637	• Assumed equal to the utility in post-progression for the subsequent treatment phase
BSC, best supportive care; CRd, carfilzomib/lenalidomide/dexamethasone; Rd, lenalidomide/dexamethasone; Vd, bortezomib/dexamethasone			

We have updated this analysis using data from the revised survival model approach described in Section 2 (i.e. the IPW-based Weibull PFS and OS model for both ENDEAVOR and ASPIRE). The cost-effectiveness results for Cd versus Vd and CRd versus Rd based on this updated survival model are provided in Table 21 and Table 22, respectively, using both the Committee's preferred approach (i.e. utility values mapped directly from trial data) and our original approach (based on mapping to HSUVs from Van Agthoven *et al.,* 2014) for comparison.

These show that the ICERs for both Cd versus Vd and CRd versus Rd slightly increase when utilities are mapped directly from the trial data. We acknowledge that it may be more appropriate to use utilities mapped directly from trial data, in line with the NICE reference case (with respect to deriving EQ-5D utility values directly from trial data)[16] and Committee's preferences, and these utilities have been used in our revised cost-effectiveness analysis (Section 6).

Table 21. Updated results with van Aghtoven *et al.* 2004[12] based utilities and directly mapped utilities and IPW-based Weibull for OS and PFS: Cd vs Vd (patients with one prior therapy and no prior bortezomib)

	Technology	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Original approach	Vd	£93,769	4.23	2.94				
	Cd	£117,660	5.74	4.09	£23,891	1.51	1.15	£20,766
Directly mapped utility from trial	Vd	£93,769	4.23	2.79				
	Cd	£117,660	5.74	3.88	£23,891	1.51	1.09	£22,009

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

Table 22. Updated results with van Aghtoven *et al.* 2004[12] based utilities and directly mapped utilities and IPW-based Weibull for OS and PFS: CRd vs Rd (patients with two prior therapies and no prior lenalidomide)

	Technology	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Original approach	Rd	£94,528	4.36	2.95				
	CRd	£127,140	5.46	3.77	£32,612	1.10	0.81	£40,198
Directly mapped utility from trial	Rd	£94,528	4.36	2.88				
	CRd	£127,140	5.46	3.67	£32,612	1.10	0.79	£41,429

CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Rd, lenalidomide/dexamethasone

Conclusion

We acknowledge that it could be considered a more appropriate option to use utilities mapped directly from trial data, in line with the NICE reference case and the Committee's stated preference. In order to address the Committee's concerns and stated preference, utility data mapped directly from the trials using the mapping algorithm from Proskorovsky *et al.,* 2014 have been used in our revised cost-effectiveness analysis (Section 6).

4 The length of treatment and dosing schedule with bortezomib

Context

The ACD states: The committee noted that there were discrepancies between the model and clinical practice in the dosing schedule and length of treatment for bortezomib. It noted that the marketing authorisation for bortezomib states that it can be given twice weekly for 8 cycles (21-day cycles equal to a total of 32 doses), whereas the model assumed bortezomib would be given twice weekly as an intravenous infusion until progression. The clinical experts clarified that in practice they prefer to give bortezomib once weekly and subcutaneously, because this is associated with fewer adverse reactions, and to give the full 32 doses. (ACD Section 4.13, pages 12 and 13).

The Committee concluded that the assumptions in the model for bortezomib did not accurately reflect its use in NHS clinical practice (ACD Section 4.13, page 13).

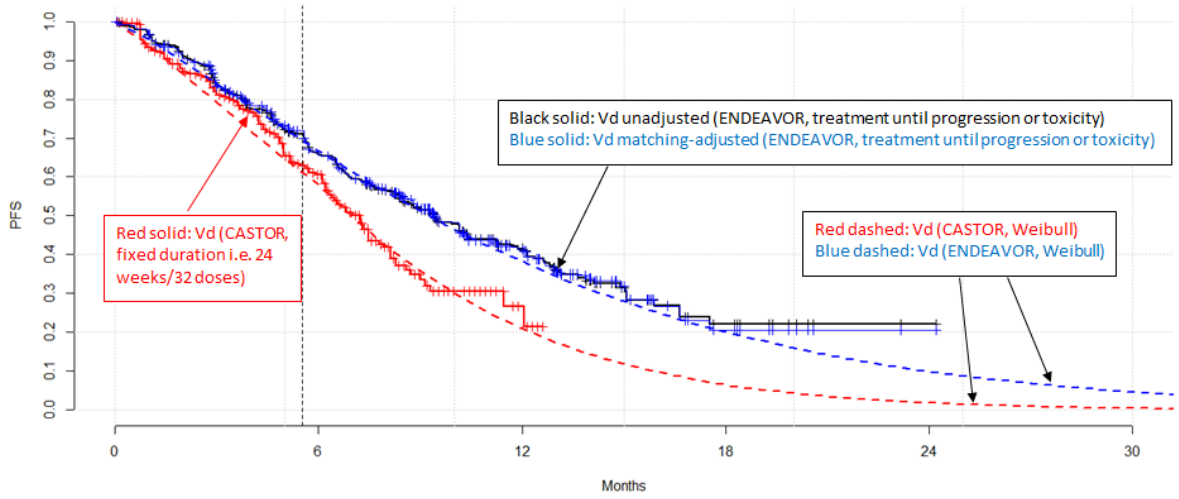
Implementation

We acknowledge that the dosing schedule, administration route, and length of treatment for bortezomib in ENDEAVOR does not fully reflect current clinical practice in England and Wales, and have explored these aspects of trial generalisability to address the Committee's concerns.

Length of treatment

In order to model the cost-effectiveness of Cd versus Vd given for different duration than that studied in the ENDEAVOR RCT, it is necessary to break the dose-efficacy link (as this results in a different cumulative dose of bortezomib) and make assumptions around how such differences might impact outcomes as well as costs. UK clinical experts consulted by Amgen considered it highly plausible that treatment duration limited to 8 cycles (32 doses) is likely to impact on outcomes (both PFS and OS) compared to treatment until progression as per the ENDEAVOR trial. We conducted an analysis comparing PFS and OS outcomes from the R/RMM CASTOR RCT (where treatment with Vd was given for 8 cycles [24 weeks, 32 doses] consistent with the marketing authorisation for bortezomib) and ENDEAVOR (where treatment with Vd was given until progression or unacceptable toxicity) using matching-adjusted indirect comparison (MAIC) methodology to minimise confounding associated with differences in patient characteristics across the studies. In summary, a piece-wise Weibull model was fitted on the matched ENDEAVOR Vd data and the approximated patient level CASTOR Vd data which allowed an assessment of the relative increase in the hazards due to stopping bortezomib after 24 weeks. Full details of the MAIC analysis are provided in Appendix D. Results from this analysis are provided below, and indicate that when bortezomib is given for 8 cycles (24 weeks, 32 doses), both PFS (Figure 13) and OS (Figure 14) outcomes are poorer than those observed in the ENDEAVOR RCT when treatment was continued to progression. The relative increase in the hazards when stopping bortezomib after 24 weeeks are 1.360 (PFS) and 1.349 (OS) and were applied to the Vd arm of ENDEAVOR after 24 weeks.

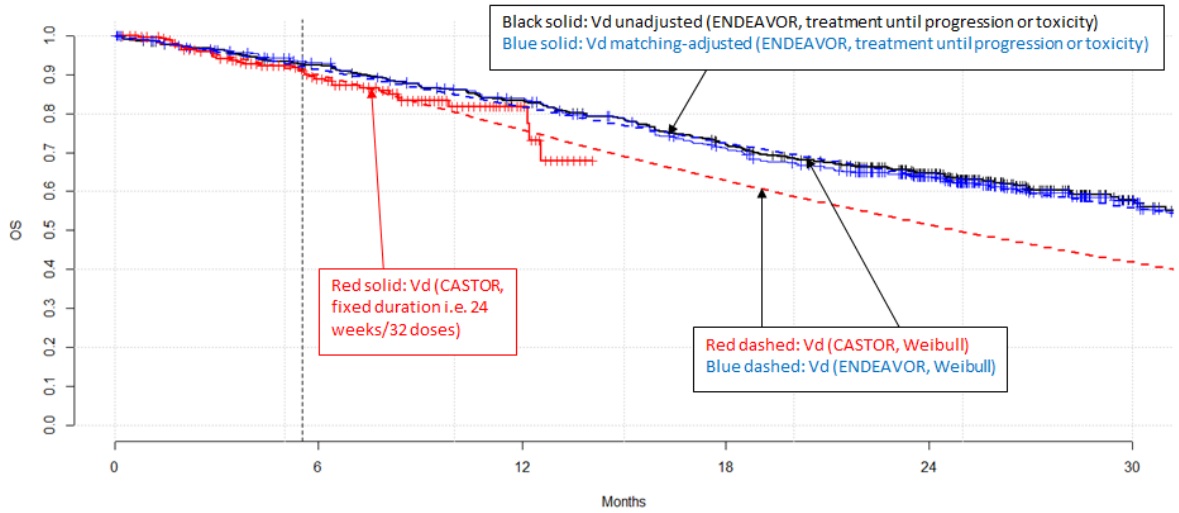
Figure 13. PFS curves from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity)



Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

MAIC, matching adjusted indirect comparison; Vd, bortezomib/dexamethasone

Figure 14. OS curves from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity)



Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

MAIC, matching adjusted indirect comparison; Vd, bortezomib/dexamethasone

Table 23. Progression-free survival and overall survival hazard ratios (first 24 weeks vs. after 24 weeks) from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity)

	PFS (Weibull model)	OS (2-arm Weibull model)
Piecewise matching-adjusted (first 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.344 (1.021 - 1. 769)	1.157 (0.679 – 1.972)
Piecewise matching-adjusted (after 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.828 (1.311 - 2.549)	1.561 (0.942 – 2.588)
Relative hazard increase when stopping bortezomib after 24 weeks, HR (95% CI) CASTOR:ENDEAVOR[a]	**1.360 (0.913 - 2.027)**	**1.349 (0.684 - 2.662)**

Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

[a] HR for first 8 cycles (24 weeks) versus subsequent cycles

CI, confidence interval; HR, hazard ratio; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone

Additional support for the assertion that prolonged treatment duration is associated with improved outcomes comes from an analysis of data from the phase 3 APEX RCT in R/RMM (bortezomib monotherapy [bortezomib dosing per license for eight cycles plus three 5-week maintenance cycles] vs. high-dose dexamethasone)[17] This showed that over 20% of patients who achieved at least a best response of CR in the bortezomib arm achieved that response in or after eight cycles of therapy, suggesting that prolonged bortezomib exposure improves the quality of response. Furthermore, an analysis of data from the VMP (bortezomib/melphalan/prednisone) arm of the phase 3 VISTA RCT in newly-diagnosed MM patients, which suggests that higher cumulative doses, indicative of extended treatment duration or dose intensity, may be associated with improved OS.[18]

Based on the above, we strongly believe that if costs of bortezomib are capped at 8 cycles per the duration of treatment outlined in the bortezomib SmPC, the efficacy of Vd in ENDEAVOR (given until progression or unacceptable toxicity) should be adjusted to appropriately take into account the impact of reduced treatment duration on outcomes seen in ENDEAVOR.

In the absence of a more robust alternative data source, the ENDEAVOR RCT is considered the most appropriate evidence to utilise for cost-effectiveness analyses. In order to model the cost-effectiveness of Cd versus Vd given for different duration than studied in the ENDEAVOR RCT, it is necessary to break the dose-efficacy link (as this results in a different cumulative dose of bortezomib) and adjust for efficacy as well as costs, given the strong evidence that shorter durations of treatment are associated with poorer outcomes. To explore the effect of

different bortezomib treatment durations on the cost-effectiveness of Cd versus Vd, we conducted an informative scenario analysis by adjusting PFS and OS for Vd based on the outcomes of the MAIC described above; we applied the matching-adjusted CASTOR:ENDEAVOR PFS and OS HRs reflecting the relative increase in the hazards when stopping bortezomib treatment after 24 weeks, and capped treatment costs at 8 cycles. However, as it is impossible to accurately quantify this impact, we believe it is more appropriate to base our cost-effectiveness analyses on the dose-efficacy relationship observed in a robust RCT setting.

Dosing schedule and method of administration

The dosing schedule does not impact costs nor outcomes as patients receive the same cumulative number of doses (32) for once- and twice-weekly dosing.

With respect to method of administration, patients could receive either intravenous (IV) or subcutaneous (SC) bortezomib in ENDEAVOR as both are licensed for use in Europe. As highlighted by the Committee, it is more common for bortezomib to be given SC in clinical practice in England and Wales due to its more favourable tolerability profile. Similarly in ENDEAVOR, SC dosing was by far the more common intended route of administration at randomisation (77% in the ITT population and 75.6% in the subgroup of interest), which is consistent with clinical practice. Furthermorethe method of administration does not impact costs as we assumed administration costs based on SC dosing for all patients receiving bortezomib to be consistent with clinical practice.

Based on the above, the Vd arm in ENDEAVOR can be considered to be broadly generalisable to clinical practice in England and Wales where a once-weekly dosing schedule and SC administration route is preferred.

Based on the revised survival model outlined in Section 2 and revised mapped utility values outlined in Section 3, the scenario analysis results capping bortezomib costs at 8 cycles and adjusting for efficacy as outlined above are provided in Table 24. Results of the updated cost-effectiveness analyses without the capping of costs or adjusting for efficacy are provided in Table 25 for comparison. These show that the ICERs are moderately increased when bortezomib costs are capped and efficacy accordingly adjusted (£27,387), or whether the costs and outcomes based on the duration of treatment with bortezomib in the ENDEAVOR RCT are used (£22,009).

Table 24. Updated cost effectiveness results capping the cost of bortezomib at 8 cycles and adjusting for efficacy (patients with one prior therapy and no prior exposure to bortezomib) – scenario analysis

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£73,114	3.41	2.25				
Cd	£117,660	5.74	3.88	£44,547	2.34	1.63	£27,397
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone							

Table 25. Updated cost effectiveness results without capping the cost of bortezomib at 8 cycles or adjusting for efficacy (patients with one prior therapy and no prior exposure to bortezomib)

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£93,769	4.23	2.79				
Cd	£117,660	5.74	3.88	£23,891	1.51	1.09	£22,009
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone							

Conclusion

In order to address the Committee's concerns around the generalisability of the dosing schedules, method of administration, and treatment duration with bortezomib in ENDEAVOR compared with current clinical practice in England and Wales, we explored these aspects of trial generalisability in further detail.

In the absence of a more robust alternative data source, the ENDEAVOR RCT is considered the most appropriate evidence to utilise for cost-effectiveness analyses. In order to model the cost-effectiveness of Cd versus Vd given for different duration than studied in the ENDEAVOR RCT, it is necessary to break the dose-efficacy link (as this results in a different cumulative dose of bortezomib) and adjust for efficacy as well as costs, given the strong evidence that shorter durations of treatment are associated with poorer outcomes. To explore the impact of different bortezomib treatment durations on the cost-effectiveness of Cd versus Vd, we conducted an informative scenario analysis adjusted for efficacy using data from a recently conducted RCT in a similar patient population where Vd was given for 8 cycles (32 doses). However, as it is impossible to accurately quantify the impact of different bortezomib treatment durations on outcomes, we believe it is more appropriate to base our cost-effectiveness analyses on the dose-efficacy relationship observed in a robust RCT setting (as in our original submission).

Differences in dosing schedule and administration route in clinical practice (driven by tolerability considerations), do not limit generalisability of ENDEAVOR given that: the same cumulative number of doses is given in clinical practice irrespective of whether bortezomib is dosed once-weekly (per clinical practice) or twice-weekly (per ENDEAVOR) e.g. 32 doses over 8 cycles of treatment, clinicians in ENDEAVOR could use an alternative once-weekly dosing schedule to manage toxicities, and the majority of bortezomib-treated patients in ENDEAVOR received SC dosing (> 75%) consistent with clinical practice.

Given the above, we have not incorporated any changes around bortezomib dosing schedule, method of administration, or duration of treatment in our revised cost-effectiveness analysis (Section 6).

5 Additional concerns

Context

The ACD states the Committee 'noted that the company did not include the patient access scheme for bortezomib for the comparison of carfilzomib and dexamethasone with bortezomib and dexamethasone. The committee agreed that the inclusion of the patient access scheme would decrease the cost of bortezomib and therefore increase the ICERs for carfilzomib' (Section 4.15, page 13).

Implementation

The bortezomib complex PAS for patients at first relapse (i.e. 2L) means that the manufacturer provides replacement stock or credit for patients who do not respond after four cycles of treatment with bortezomib.[19] Response is defined as a patient achieving at least a minimum response (a 25% or greater reduction in serum M-protein) within the first for cycles of treatment, and non-response is defined as having stable or progressive disease (i.e. less than a 25% improvement in serum M-protein) within the first four cycles of treatment.[19]

We agree with the Committee's conclusion that the incorporation of this PAS into the cost-effectiveness analysis for Cd versus Vd will to result in an increased ICER. However, accurately translating this PAS into an appropriate model input that reflects the uptake and utilisation of this PAS in current clinical practice is highly challenging for the following reasons:

- To the best of our knowledge, there is no routine monitoring of the enforcement and uptake of this PAS in routine clinical practice
- As highlighted by the Committee, the dosing schedule for bortezomib in routine clinical practice may differ to the marketing authorisation (8 21-day cycles, twice weekly, 32 doses)[20] as '*clinical experts clarified that in practice they prefer to give bortezomib <u>once weekly</u> and subcutaneously, because this is associated with fewer adverse reactions, and to give the full 32 doses.'* (ACD Section 4.13, pages 12 and 13). It is unclear whether such deviation from the marketing authorisation would preclude clinicians from utilising the bortezomib PAS.
- The NICE TA guidance for bortezomib (TA129) suggests that patients who do not respond after four cycles of treatment with bortezomib should stop treatment.[21] Our cost-effectiveness analysis is based on the ENDEAVOR RCT, where there was no such arbitrary stopping rule. Patients in the Vd arm of ENDEAVOR who didn't respond after four cycles of treatment may have continued to receive treatment with Vd, which may have had an impact on outcomes observed in the trial.

In the recent NICE part-review of TA171 (lenalidomide for the treatment of patients who have received one prior therapy with bortezomib), the cost saving to the NHS in bortezomib drug acquisition costs as a result of the bortezomib PAS was estimated to be between 8.3% and 15%.[22] Based on the manufacturer's expectation and the preference of the Committee in the NICE part-review of TA171,[19,22] we have provided an updated scenario analysis assuming a 15% cost saving in bortezomib drug acquisition costs. However for the reason outlined above we feel this represent an overestimation of the real impact of this complex PAS. Based on the revised survival model outlined in Section 2 and revised mapped utility values outlined in

Section 3, the revised cost-effectiveness results for Cd versus Vd with an assumed 15% reduction in bortezomib costs are provided in Table 26. Results of the revised analyses without the 15% reduction in bortezomib costs are provided in Table 27 for comparison. Unsurprisingly, the inclusion of a 15% reduction in bortezomib drug acquisition costs slightly increases the ICER (£26,306 vs. £22,009).

Table 26. Updated cost-effectiveness results with inclusion of 15% reduction in the cost of bortezomib (patients with one prior therapy and no prior exposure to bortezomib) – scenario analysis

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£89,105	4.23	2.79				
Cd	£117,660	5.74	3.88	£28,555	1.51	1.09	£26,306
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone							

Table 27. Updated cost-effectiveness results without inclusion of 15% reduction in the cost of bortezomib (patients with one prior therapy and no prior exposure to bortezomib)

	Total costs (£)	Total LYG	Total QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£93,769	4.23	2.79				
Cd	£117,660	5.74	3.88	£23,891	1.51	1.09	£22,009
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone							

Given the substantial uncertainty around the uptake and utilisation of the bortezomib PAS, and the difficulties in accurately translating it to an appropriate model input that reflects current clinical practice, we do not consider it appropriate to incorporate the bortezomib PAS into our revised cost-effectiveness analyses (Section 6). Doing so would rely on assuming an arbitrary cost saving that is not informed by robust empirical evidence.

Conclusion

We acknowledge the Committee's concern around the incorporation of the bortezomib complex PAS into the cost-effectiveness analysis for Cd versus Vd. However, translating the PAS into an appropriate model input that accurately reflects its uptake and utilisation in current clinical practice is highly challenging. We explored the potential impact of the bortezomib PAS in an informative scenario analysis assuming that the PAS translates to a 15% saving in bortezomib drug acquisition costs based on the manufacturer's expectation and the assumption preferred by the Committee in the recent NICE part-review of TA171. In this very conservative scenario analysis, the ICER for Cd versus Vd remained below £30,000/QALY gained. Given the substantial uncertainty around the uptake and utilisation of the bortezomib PAS, and the difficulties in accurately translating it to an appropriate model input that reflects current clinical practice, we do not consider it appropriate to incorporate the bortezomib PAS into our revised cost-effectiveness analysis (Section 6). Doing so would rely on assuming an arbitrary cost saving that is not informed by robust empirical evidence.

6 Revised cost-effectiveness analysis and justification

A summary of our revised cost-effectiveness analysis is provided in Table 28, along with a justification for any changes to (or decision not to change) our original approach.

Table 28. Summary of approach in the revised cost-effectiveness analyses and justification

Area of uncertainty	Item	Original approach	Revised approach and justification
The survival model to estimate long-term effects	Plausibility of trial covariate-adjusted efficacy estimates	Adjustment of all covariates considered to be prognostic of outcomes in MM by UK clinical experts	Use of stepwise variable selection model to identify relevant covariates for adjustment in the parametric survival models: • The stepwise selection method yielded the best statistical fit out of all of the alternative methods assessed for both PFS and OS for both comparisons of Cd versus Vd and CRd versus Rd.
	Plausibility of proportional hazards assumptions	Assumption of proportional hazards for PFS and OS within the subgroups of interest	Assumption of proportional hazards for PFS and OS was assessed and satisfied in the context of the revised model approach (IPW-based Weibull)

Area of uncertainty	Item	Original approach	Revised approach and justification
	Modelling of survival while ensuring balance on prognostic factors	Use of two separate regression models (for both arms of the trials) to extrapolate the effects over the full model time horizon, with the model to extrapolate the carfilzomib arms based on the post-hoc subgroup post-hoc estimated hazard ratios	<u>Use of an IPW-based Weibull model to jointly model both treatment arms in the ENDEAVOR and ASPIRE studies for PFS and OS:</u> • A jointly fitted model for both treatment arms in the studies is aligned with NICE DSU guidance. We explored several different modelling approaches (IPW, CGP, and MoC) and parametric distributions, and concluded that: • Of the different techniques to incorporate covariate adjustment and different parametric distributions, the IPW-based Weibull model yielded the most clinically plausible projections of PFS and OS for both the ENDEAVOR and ASPIRE subgroups of interest. In addition, it provided a generally good statistical fit, an excellent fit to corresponding covariate-adjusted trial data, and satisfied proportional hazards assumptions for both PFS and OS in the ENDEAVOR and ASPIRE subgroups of interest. • In contrast to the IPW-Weibull based model, none of the parametric distributions using either the CGP or MoC methods resulted in clinically plausible projections of OS for either the ENDEAVOR or ASPIRE subgroups of interest. • Given the above, the IPW-based Weibull model was considered the most appropriate way in which to model survival (PFS and OS) for both ENDEAVOR and ASPIRE
Derived health state utility values	-	Based on a mapping analysis to the health state utility values reported by van Aghtoven *et al.* 2004.	<u>Use of utilities mapped directly from trial data using the mapping algorithm from Proskorovsky *et al.,* 2014</u> • This approach is in line with the NICE reference case and Committee's stated preferences
The length of treatment and dosing schedule of bortezomib	Length of bortezomib treatment	Per the ENDEAVOR study	<u>Per the ENDEAVOR study (i.e. no change)</u> • In the absence of a more robust alternative data source, the ENDEAVOR RCT is considered the most appropriate evidence to utilise for cost-effectiveness analyses. • In order to model the cost-effectiveness of Cd versus Vd given for different duration than studied in the ENDEAVOR RCT, it is necessary to break the dose-efficacy link (as this

Area of uncertainty	Item	Original approach	Revised approach and justification
(*note: applies to Cd versus Vd comparison only*)			results in a different cumulative dose of bortezomib) and adjust for efficacy as well as costs, given the strong evidence that shorter durations of treatment are associated with poorer outcomes. • To explore the impact of different bortezomib treatment durations on the cost-effectiveness of Cd versus Vd, we conducted an informative scenario analysis adjusted for efficacy of Vd using data from a recently conducted RCT in a similar patient population where Vd was given for 8 cycles (32 doses). However, as it is challenging to accurately quantify this impact, we believe it is more appropriate to base our cost-effectiveness analyses on the dose-efficacy relationship observed in a robust RCT setting (as in our original submission).
	Dosing schedule and administration of bortezomib	Per the ENDEAVOR study	<u>Per the ENDEAVOR study (i.e. no change)</u> • Differences in dosing schedule and administration route in clinical practice (driven by tolerability considerations) do not limit generalisability of ENDEAVOR given that ▪ The same cumulative number of doses is given in clinical practice irrespective of whether bortezomib is dosed once-weekly (per clinical practice) or twice-weekly (per ENDEAVOR) e.g. 32 doses over 8 cycles of treatment ▪ Clinicians in ENDEAVOR could use an alternative once-weekly dosing schedule to manage toxicities ▪ The majority of bortezomib-treated patients in ENDEAVOR received SC dosing (> 75%) consistent with clinical practice.
Cd, carfilzomib/dexamethasone; CGP, corrected group prognosis; CRd, carfilzomib/lenalidomide/dexamethasone; IPW, inverse probability weighted; MoC, mean of covariates; NICE, National Institute for Health and Care Excellence; OS, overall survival; PAS, patient access scheme; PFS, progression-free survival; RCT, randomised controlled trial; Rd, lenalidomide/dexamethasone; SC, subcutaneous; Vd, bortezomib/dexamethasone			

We have updated the cost-effectiveness analyses based on the amendments described above. A summary of the impact of these different amendments on the cost-effectiveness results compared with the previous results considered by the Committee, and final revised cost-effectiveness estimates are provided in Table 29 for the Cd versus Vd comparison and Table 30 for the CRd versus Rd comparison.

Table 29. Revised cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib

	Cd	Vd	Incremental value
Original base case			
Total costs	£121,891	£95,213	£26,678
Total LYG	6.05	4.26	1.79
QALYs	4.28	2.95	1.33
ICER	-	-	£20,044
Using the updated survival model			
Total costs	£117,660	£93,769	£23,891
Total LYG	5.74	4.23	1.51
QALYs	4.09	2.94	1.15
ICER (compared to original base case)	-	-	£20,766
Using directly mapped utilities			
Total costs	£117,660	£93,769	£23,891
Total LYG	5.74	4.23	1.51
QALYs	3.88	2.79	1.09
ICER (compared to original base case)	-	-	£21,137
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£22,009**

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

Table 30. Revised cost-effectiveness results for CRd versus Rd in patients with two prior therapies and no prior lenalidomide

	CRd	Rd	Incremental value
Original base case			
Total costs	£128,654	£95,420	£33,234
Total LYG	6.31	4.93	1.37
QALYs	4.32	3.33	0.99
ICER	-	-	£33,467
Using the updated survival model			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.77	2.95	0.79
ICER (compared to original base case)	-	-	£40,198
Using directly mapped utilities			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.67	2.88	0.79
ICER (compared to original base case)	-	-	£34,404
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£41,429**

CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; Rd, lenalidomide/dexamethasone; QALYs, quality-adjusted life years

We have shown in this response that by applying appropriate alternative approaches to selection of covariates and modelling survival in line with DSU guidance (key areas of uncertainty for the Committee, the revised cost-effectiveness analyses predict outcomes similar to our original approach. Notably, the difference in ICERs for the comparison of Cd versus Vd in patients who have received one prior therapy and no prior bortezomib was marginal (£20,766/QALY gained compared with £20,044/QALY gained).

We also implemented into our revised cost-effectiveness analyses the utility data mapped directly from the trials per the Committee's specific preference, and have provided informative scenario analyses for the comparison of Cd versus Vd around:

- The duration of treatment with bortezomib given in accordance with clinical practice (i.e. 8 cycles; 32 doses), and appropriately adjusted for efficacy using data from a recently conducted RCT in a similar patient population (Section 4). This resulted in an ICER of £27,397/QALY gained. Given that this necessitates breaking the dose-efficacy relationship observed in a robust RCT setting and that it is impossible to accurately quantify the impact of different treatment durations on outcomes in the absence of an alternative robust data source, we did not believe it appropriate to incorporate this into the revised analyses

- The potential cost saving to the NHS resulting from the availability of the complex bortezomib PAS, which resulted in an ICER of £26,306/QALY gained (Section 5). Given the substantial uncertainty around the utilisation and uptake of this PAS, we did not believe it appropriate to incorporate this into the revised analyses

The final revised ICERs for CRd versus Rd in patients who have received two prior therapies and have not received prior lenalidomide are above the typical willingness to pay threshold of £30,000/QALY gained (revised ICER £41,429/QALY gained). As highlighted in our original submission, the decision problem in the current economic evaluation, assessing the cost-effectiveness of introducing carfilzomib as a treatment option, should not be penalised by the additional costs originating from the prolongation of a background therapy that has been considered cost effective in England and Wales (Rd). This approach is well accepted, including by the NICE DSU and ERGs in other therapeutic areas such as nephrology.[23,24] Additionally, the inconsistency of the standard approach of assessing the cost-effectiveness of combination therapies is also evident in some counterintuitive results. Usually the ICER is lower when the price of the comparator drug is higher; however, in the case of combination therapies the ICER is lower when the price of the comparator drug is lower, because the costs associated with the additional time on treatment with the comparator drug is lower. Consequently, removing the additional Rd costs in the CRd arm when calculating the ICER seems appropriate and provides useful insight for the overall assessment of the cost-effectiveness of carfilzomib in combination with lenalidomide and dexamethasone. Consequently, we urge the Committee to take this into account and to consider scenarios when the additional costs of lenalidomide have been discounted, resulting in an ICER of £36,455 when using the revised cost-effectiveness analysis described in Table 30, and £29,671 using our original base case (Table 31).

Table 31. Impact of discounting additional cost of lenalidomide on cost-effectiveness analyses of CRd versus Rd using original base case and final revised approach

	Additional cost of lenalidomide included (Base case)			**Additional cost of lenalidomide excluded**		
Original base case						
	CRd	Rd	Incremental values	CRd	Rd	Incremental values
Total costs	£128,654	£95,420	£33,234	£124,886	£95,420	£29,466
Total QALYs	4.32	3.33	0.99	4.32	3.33	0.99
ICER	-	-	£33,467			£29,671
Final revised cost-effectiveness analyses						
	CRd	Rd	Incremental values	CRd	Rd	Incremental values
Total costs	£127,140	£94,528	£32,612	£123,225	£94,528	£28,697
Total QALYs	3.67	2.88	0.79	3.67	2.88	0.79
ICER	-	-	£41,429			£36,455

CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; LYG, life years gained; Rd, lenalidomide/dexamethasone; QALYs, quality-adjusted life years

7 References

1. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. *Br J Cancer* 2003;89:781-6.
2. Judd CM, McClelland GH, Ryan CS. Data Analysis: A Model Comparison Approach, Second Edition: Taylor & Francis.2011.
3. National Institute for Health and Care Excellence (NICE). Carfilzomib for treated multiple myeloma in people who have received at least one prior therapy (ID 934). Evidence Review Group report. Available from: https://www.nice.org.uk/guidance/GID-TA10005/documents/committee-papers (Accessed November 2016).
4. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis: Springer.2001.
5. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16:385-95.
6. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition: Springer New York.2009.
7. Ghali WA, Quan H, Brant R *et al.* Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models. *JAMA* 2001;286:1494-7.
8. National Institute for Health and Care Excellence (NICE). Multiple myeloma - lenalidomide (post bortezomib) (part rev TA171) [ID667]. Evidence review group report. Available from https://www.nice.org.uk/guidance/GID-TAG452/documents/multiple-myeloma-lenalidomide-post-bortezomib-part-rev-ta171-evaluation-report2 (Accessed November 2016).
9. Amgen data on file. Haematological Malignancy Research Network (HMRN) clinical management and outcomes in multiple myeloma. 2016.
10. San-Miguel JF, Hungria VTM, Yoon S, Beksac M, M.A. D. Final analysis of overall survival from the phase 3 PANORAMA-1 trial of panobinostat plus bortezomib and dexamethasone versus placebo plus bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma. *Blood* 2015;126:Abstr 3026.
11. National Institute for Health and Care Excellence (NICE). Panobinostat for multiple myeloma after at least 2 previous treatments (TA380). 2016. Available from: http://www.nice.org.uk/guidance/ta380 (Accessed May 2016).
12. Van Agthoven M, Segeren CM, Buijt I *et al.* A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study. *Eur J Cancer* 2004;40:1159-69.
13. National institute for Health and Care Excellence (NICE). Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy (TA171). 2009. Available from: https://www.nice.org.uk/guidance/ta171. (Accessed November 2015).
14. Bacelar M, Durand A, Cooper C *et al.* The clinical and cost effectiveness of lenalidomide for people who have received at least one prior therapy with bortezomib (partial review of TA171). Single Technology Appraisal NIHR HTA Programme (13/07/01), Matrix and Peninsula Technology Assessment Group, 2014. Available from: http://www.nice.org.uk/guidance/indevelopment/gid-tag452/documents. (Accessed November 2014).
15. Novartis. Panobinostat for treating multiple myeloma in people who have received at least one prior therapy. ID663 Single technology appraisal (STA) manufacturer's submission. 2015. Available from: http://www.nice.org.uk/guidance/indevelopment/gid-tag477/documents (Accessed November 2015).

16. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. 2013. Available from: http://www.nice.org.uk/article/pmg9/chapter/Foreword (Accessed May 2016).
17. Richardson P, Sonneveld P, Schuster MW *et al.* Extended follow-up of a phase 3 trial in relapsed multiple myeloma: final time-to-event results of the APEX trial. *Blood* 2007;110:3557-3560.
18. Mateos M. Management of treatment-related adverse events in patients with multiple myeloma. *Cancer Treat Rev* 2010;36 Suppl 2:S24-32.
19. Department of Health (DoH). Summary of Velcade response scheme. 2007. Available from http://www.nice.org.uk/guidance/ta129/documents/department-of-health-summary-of-responder-scheme2 (Accessed 25 November 2016).
20. Janssen Cilag Ltd. Velcade (bortezomib) 3.5 mg powder for solution for injection. Summary of Product Characteristics. 2013. Available from: http://www.medicines.org.uk/emc/medicine/17109 (Accessed October 2015).
21. National Institute for Health and Care Excellence (NICE). Bortezomib monotherapy for relapsed multiple myeloma (TA129). 2007. Available from: http://www.nice.org.uk/TA129 (Accessed September 2015).
22. National Institute for Health and Care Excellence (NICE). Lenalidomide for the treatment of multiple myeloma in people who have received one prior therapy with bortezomib (part rev TA171) [ID667]. Appraisal Consultation 3. Committee Papers. Available from: https://www.nice.org.uk/guidance/GID-TAG452/documents/committee-papers (Accessed 25 November 2016).
23. Davis S. Assessing technologies that are not cost-effective at a zero price. 2014. Available from: http://www.nicedsu.org.uk/Not-costeffective-at-0(2973292).htm (Accessed December 2015).
24. National Institute for Health and Care Excellence (NICE). Cinacalcet for the treatment of secondary hyperparathyroidism in patients with end-stage renal disease on maintenance dialysis therapy Technology appraisal guidance (TA117). 2007. Available from: http://www.nice.org.uk/guidance/ta117 (Accessed February 2016).
25. Signorovitch JE, Sikirica V, Erder MH *et al.* Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012;15:940-7.
26. Signorovitch JE, Wu EQ, Yu AP *et al.* Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics* 2010;28:935-45.
27. Palumbo A, Chanan-Khan A, Weisel K *et al.* Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. *N Engl J Med* 2016;375:754-66.
28. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2012;12:9.

8 Appendices

8.1 Appendix A: List of covariates retained with each covariate selection method

Table 32 to Table 35 show which variables were retained within different covariate-selection models (analysis approaches 2 to 5 in Table 5) together with associated hazard ratios.

Table 32. Variables retained within different covariate-selection models and associated hazard ratios: Cd vs Vd PFS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Covariate	Stepwise variable selection model HR	All clinician-identified covariates model HR	LASSO variable selection model HR	Cox model using LASSO variables HR
Treatment (Cd vs Vd)	0.408	0.362	0.397	0.366
Prior lenalidomide	-	0.937	0.959	0.957
Prior stem cell transplantation	-	0.626	0.682	0.623
Age (≥65 vs <65)	-	1.209	1.152	1.226
ECOG status (1-2 vs 0)	-	1.310	1.213	1.279
Creatinine clearance (≥50 - <80 vs other)	-	0.793	-	-
Creatinine clearance (>80)	-	0.670	0.847	0.823
Time from diagnosis	0.990	0.991	0.992	0.991
Time from last relapse	-	1.016	1.009	1.016
ISS stage (II-III vs I)	2.455	2.440	2.162	2.289
B2-microglobulin (≥3.5 vs <3.5 mg/L)	-	0.880	-	-
Refractory to last prior treatment	-	1.303	1.249	1.281
AIC	**883.822**	**893.576**	**-**	**890.208**

AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; Vd, bortezomib/dexamethasone

Table 33. Variables retained within different covariate-selection models and associated hazard ratios: Cd vs Vd OS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Covariate	Stepwise variable selection model HR	All clinician-identified covariates model HR	LASSO variable selection model HR	Cox model using LASSO variables HR
Treatment (Cd vs Vd)	0.631	0.618	0.661	0.621
Prior lenalidomide	2.034	1.943	1.615	1.953

Prior stem cell transplantation	-	0.701	0.801	0.723
Age (≥65 vs <65)	-	0.996	-	-
ECOG status (1-2 vs 0)	2.154	2.218	2.008	2.170
Creatinine clearance (≥50 - <80 vs other)	-	0.783	-	-
Creatinine clearance (>80)	0.496	0.386	0.520	0.473
Time from diagnosis	0.986	0.987	0.989	0.987
Time from last relapse	-	0.995	0.998	0.996
ISS stage (II-III vs I)	-	1.356	1.352	1.370
B2-microglobulin (≥3.5 vs <3.5 mg/L)	2.926	2.186	2.081	2.263
Refractory to last prior treatment	2.010	2.0944	1.862	2.078
AIC	**640.675**	**648.1524**	**-**	**644.6873**

AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; OS, overall survival; Vd, bortezomib/dexamethasone

Table 34. Variables retained within different covariate-selection models and associated hazard ratios: CRd vs Rd PFS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Covariate	Stepwise variable selection model HR	All clinician-identified covariates model HR	LASSO variable selection model HR	Cox model using LASSO variables HR
Treatment (CRd vs Rd)	0.699	0.708	0.776	0.707
Prior bortezomib	1.448	1.479	1.153	1.470
Prior stem cell transplantation	-	1.025	-	-
Age (≥65 vs <65)	-	1.171	1.042	1.216
ECOG status (1-2 vs 0)	1.578	1.566	1.352	1.570
Creatinine clearance (≥50 - <80 vs other)	-	0.599	0.975	0.768
Creatinine clearance (>80 vs other)	-	0.681	-	-
Time from diagnosis	0.994	0.993	0.996	0.993
Time from last relapse	-	1.002	-	-
ISS stage (II-III vs I)	4.138	4.034	1.965	2.408
B2-microglobulin (≥3.5 vs <3.5 mg/L)	0.523	0.536	-	-
Refractory to last prior treatment	-	1.036	-	-
AIC	**1076.913**	**1087.011**	**-**	**1079.794**

AIC Akaike information criterion; CRd, carfilzomib/lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; Rd, lenalidomide/dexamethasone

Table 35. Variables retained within different covariate-selection models and associated hazard ratios: CRd vs Rd OS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Covariate	Stepwise variable selection model HR	All clinician-identified covariates model HR	LASSO variable selection model HR	Cox model using LASSO variables HR
Treatment (CRd vs Rd)	0.730	0.783	0.801	0.783
Prior bortezomib	-	1.311	1.200	1.306
Prior stem cell transplantation	0.569	0.622	0.651	0.621
Age (≥65 vs <65)	-	1.256	1.190	1.256
ECOG status (1-2 vs 0)	-	1.015	-	-
Creatinine clearance (≥50 - <80 vs other)	0.296	0.282	0.367	0.282
Creatinine clearance (>80 vs other)	0.298	0.302	0.403	0.302
Time from diagnosis	0.993	0.992	0.994	0.992
Time from last relapse	-	0.991	0.994	0.991
ISS stage (II-III vs I)	10.111	10.521	7.680	10.543
B2-microglobulin (≥3.5 vs <3.5 mg/L)	0.382	0.373	0.503	0.373
Refractory to last prior treatment	-	1.309	1.205	1.313
AIC	**675.733**	**683.207**	**-**	**681.210**

AIC Akaike information criterion; CRd, carfilzomib/lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; OS, overall survival; Rd, lenalidomide/dexamethasone

8.2 Appendix B: Grambsch and Therneau tests for proportionality of hazards (previously provided to the ERG)

We provided Grambsch and Therneau tests for proportionality of hazards in our response to ERG clarification questions. These were based on the full trial population model with treatment by subgroup interactions and adjusted for all clinician-identified prognostic factors. These are shown below in Table 36 (Cd vs Vd) and Table 37 (CRd vs Rd) in the subgroups of relevance.

Table 36. Previously provided Grambsch and Therneau tests for proportionality of hazards: Cd vs Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	P-value from Grambsch and Therneau test	
Covariate	**PFS**	**OS**
Treatment (Cd vs Vd)	0.729	0.637
Treatment x Prior BTZ	0.551	0.407
Treatment x Prior lines of therapy	0.111	0.536
Prior lines of therapy (2+ vs 1)	0.346	0.851
Prior stem cell transplantation	0.013	0.610
Prior bortezomib	0.463	0.673
Prior lenalidomide	0.693	0.477
Age (≥65 vs <65)	0.370	0.802
ECOG (1-2 vs 0)	0.011	0.770
Creatinine clearance (>=50 - <80 vs other)	0.936	0.595
Creatinine clearance (>=80 vs other)	0.579	0.771
Time from diagnosis	0.308	0.433
Time from last relapse	0.394	0.025
ISS at randomization (II-III vs I)	0.703	0.282
β2-microglobulin (>=3.5 vs <3.5 mg/L)	0.004	0.546
Refractory to last prior treatment	0.033	0.053
Time from diagnosis	0.729	0.637

BTZ, bortezomib; Cd, carfilzomib/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone

Table 37. Previously provided Grambsch and Therneau tests for proportionality of hazards: CRd vs Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

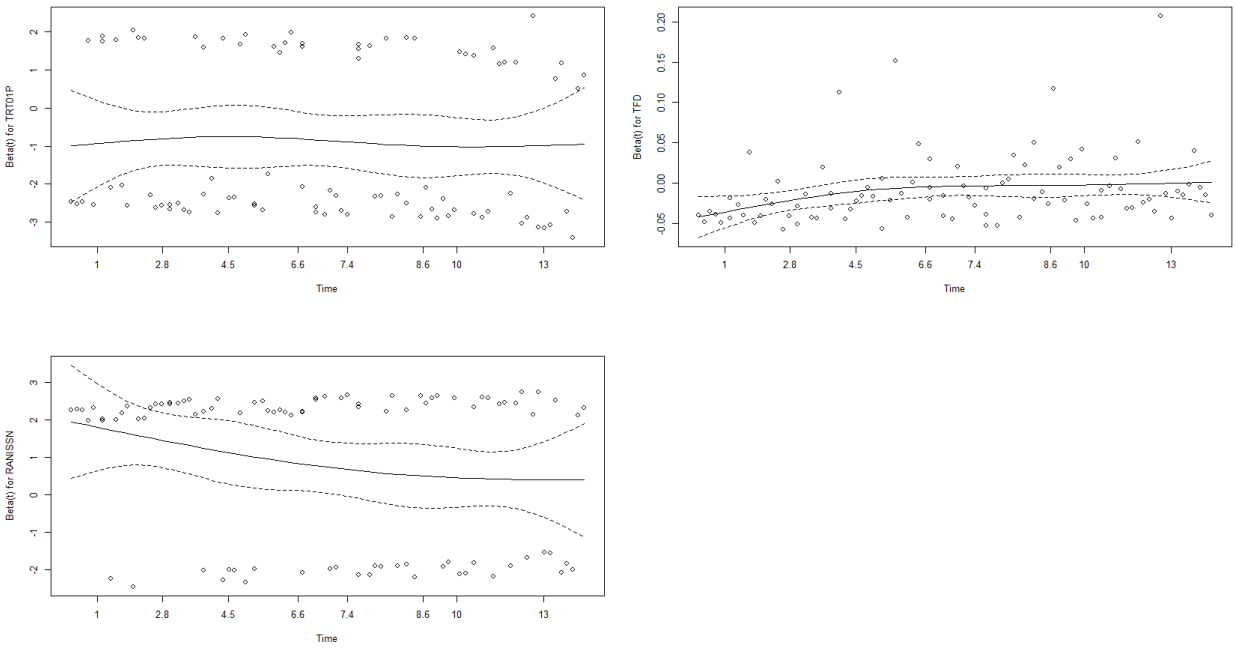
	P-value from Grambsch and Therneau test	
Covariate	**PFS**	**OS**
Treatment (CRd vs Rd)	0.036	0.219
Treatment x 2 prior LOT	0.715	0.869
Prior lenalidomide	0.869	0.646
Treatment x prior lenalidomide	0.164	0.653
1 prior LOT	0.185	0.071
2 prior LOT	0.952	0.639
Treatment x 1 prior LOT	0.621	0.432
Age (≥65 vs <65)	0.013	0.038
ECOG (1-2 vs 0)	0.018	0.000

Creatinine clearance (>=50mL/min - <80 mL/min vs other)	0.569	0.512
Creatinine clearance (>=80 vs other)	0.163	0.728
Time from diagnosis	0.068	0.901
Time from last relapse	0.540	0.187
ISS at randomization (II-III vs I)	0.177	0.095
Prior bortezomib	0.474	0.203
Prior stem cell transplantation	0.230	0.309
β2-microglobulin (>=3.5 vs <3.5 mg/L)	0.744	0.601
Refractory to lenalidomide	0.334	0.760
CRd, carfilzomib/lenalidomide/dexamethasone; ECOG, Eastern Cooperative Oncology Group; ISS, International Staging System; LOT, line of therapy; PFS, progression-free survival; OS, overall survival; Rd, lenalidomide/dexamethasone		

8.3 Appendix C: Shoenfeld residual plots for covariate-adjusted PFS and OS models

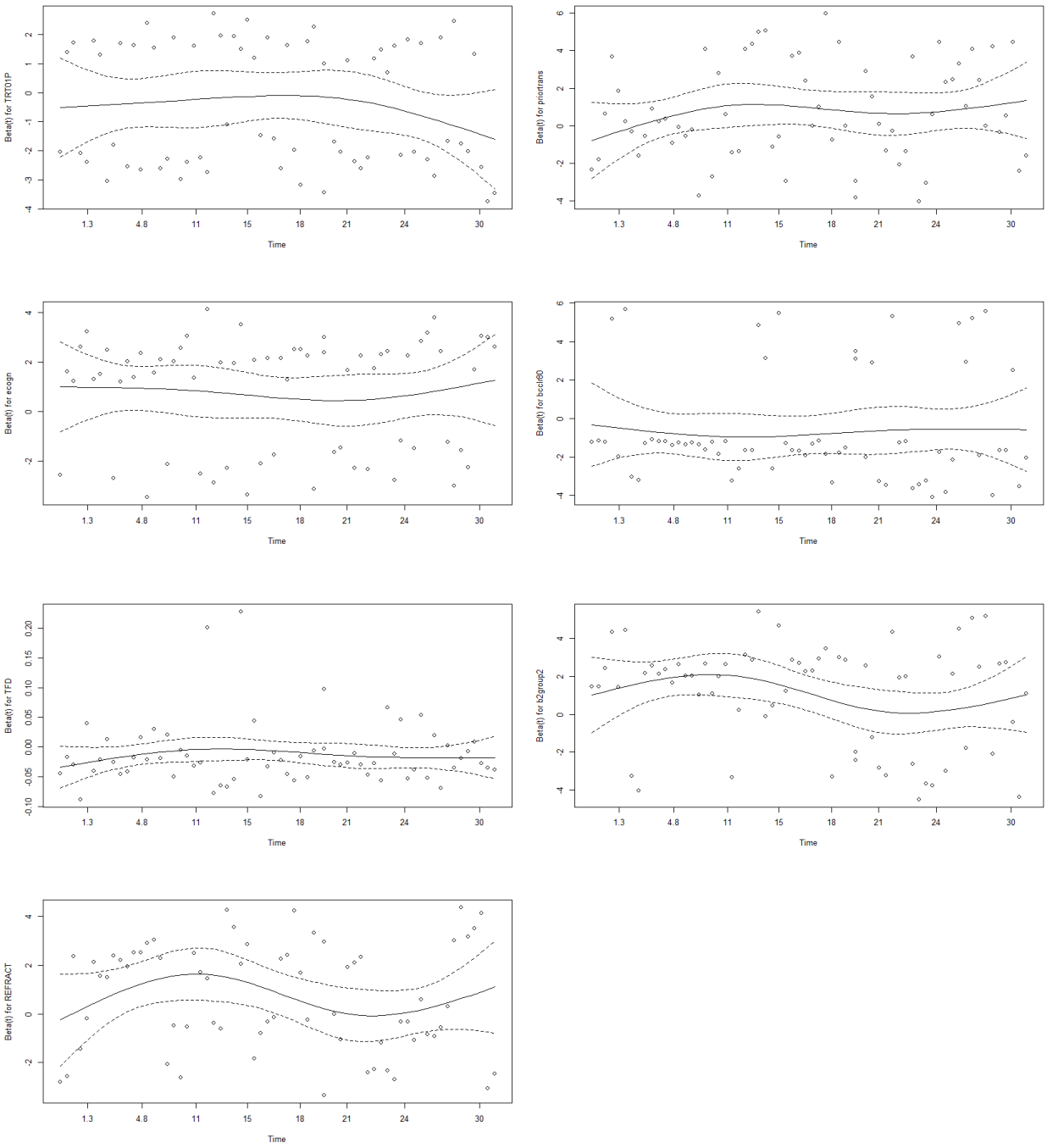
Schoenfeld residual plots for the covariate-adjusted PFS and OS models (stepwise variable selection models) are shown in Figure 15 to Figure 18. These relate to the analyses presented in Table 8 and Table 9.

Figure 15. Schoenfeld residuals plot – Cd vs Vd PFS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



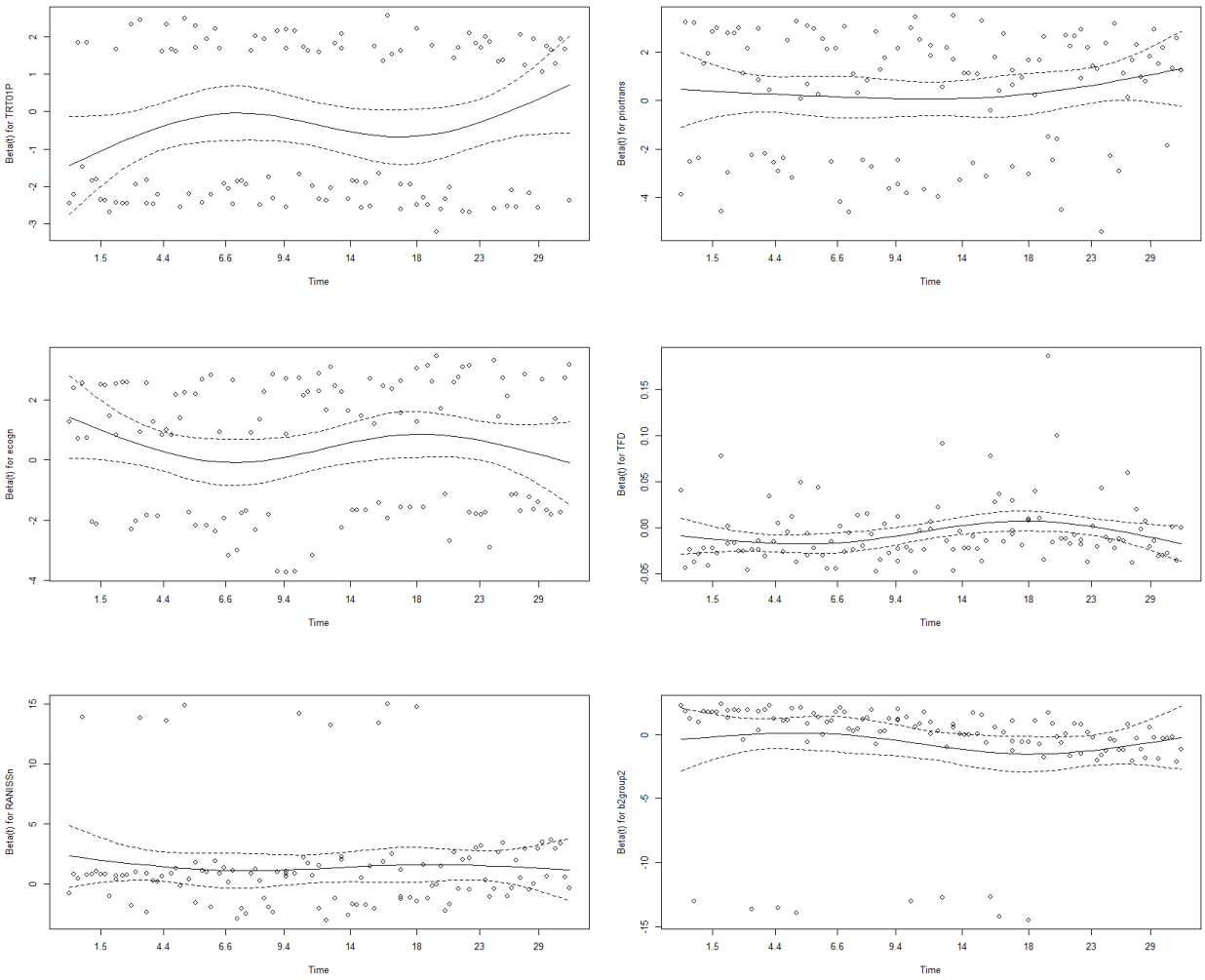
Cd, carfilzomib/dexamethasone; PFS, progression-free survival RANISSN, ISS stage at baseline (II-III vs I); TFD, time from diagnosis; TRT01P, treatment indicator (Cd vs Vd); Vd, bortezomib/dexamethasone.
Note: stepwise variable selection model

Figure 16. Schoenfeld residuals plot – Cd vs Vd OS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



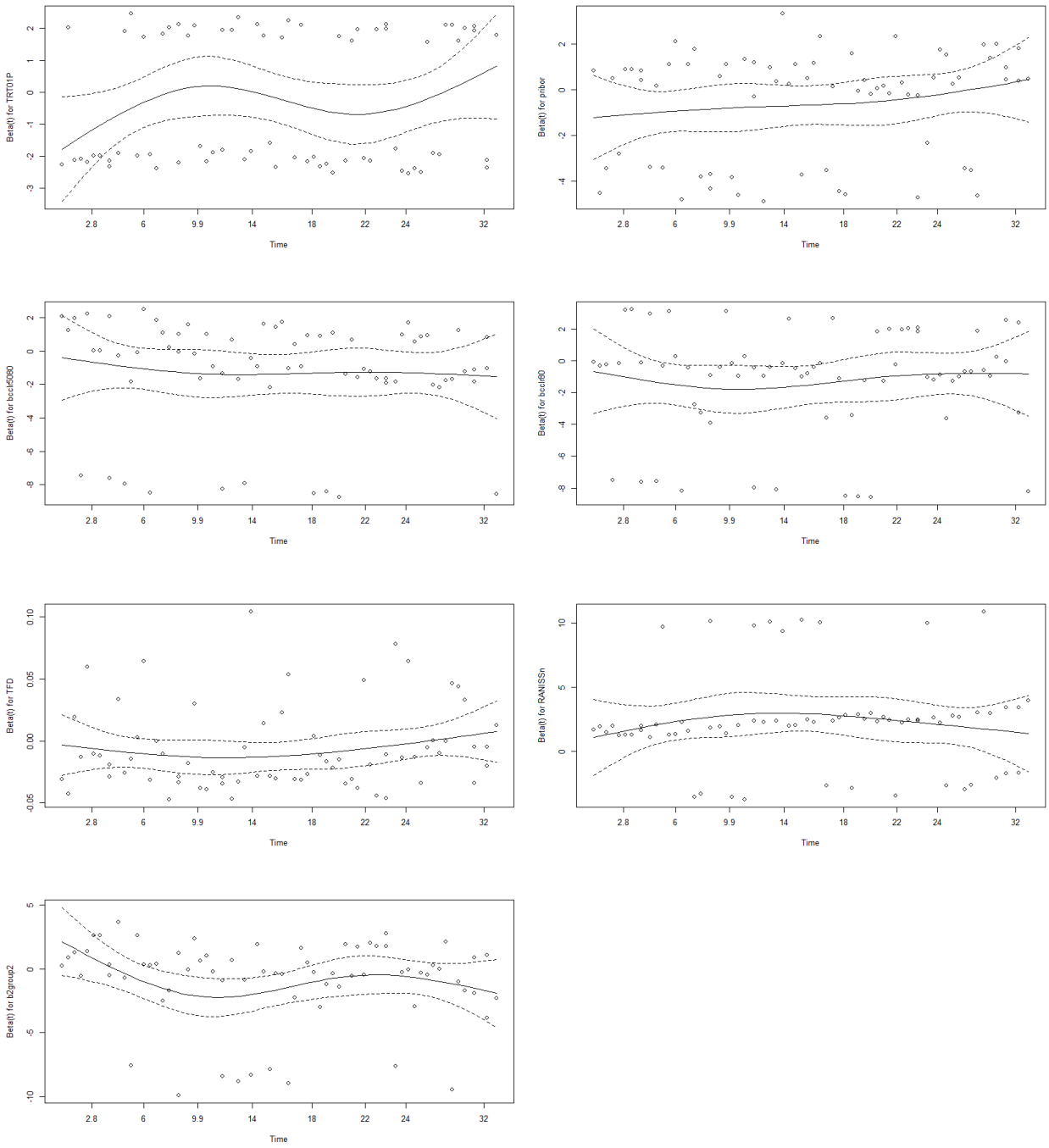
bcclr80, baseline creatinine clearance (≥80 mL/min vs. other); b2group2, β2-microglobulin (≥3.5mg/L vs other); Cd, carfilzomib/dexamethasone; ecogn, ECOG at baseline (1-2 vs 0); OS, overall survival; priortrans; prior stem cell transplantation (yes vs no); REFRACT, refractory to last prior treatment (yes vs no); TFD, time from diagnosis; TRT01P, treatment indicator (Cd vs Vd), Vd, bortezomib/dexamethasone.
Note: stepwise variable selection model

Figure 17. Schoenfeld residuals plot – CRd vs Rd PFS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)



b2group2, β2-microglobulin (≥3.5mg/L vs other); CRd, carfilzomib//lenalidomide/dexamethasone; ecogn, ECOG at baseline (1-2 vs 0); priortrans, prior stem cell transplantation (yes vs no); RANISSn, ISS stage at baseline (II-III vs I); Rd, lenalidomide/dexamethasone; TFD, time from diagnosis; TRT01P, treatment indicator (CRd vs Rd);
Note: stepwise variable selection model

Figure 18. Schoenfeld residuals plot – CRd vs Rd OS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)



b2group2, β2-microglobulin (≥3.5mg/L vs other); bcclr80, baseline creatinine clearance (≥80 mL/min vs. other); CRd, carfilzomib//lenalidomide/dexamethasone; pribor, prior bortezomib (yes vs no); RANISSn, ISS stage at baseline (II-III vs I); Rd, lenalidomide/dexamethasone; TFD, time from diagnosis; TRT01P, treatment indicator (CRd vs Rd),
Note: stepwise variable selection model

8.4 Appendix D: Results of the CGP and MoC survival models

Corrected group prognosis (CGP) method

The CGP approach was assessed as follows:

- Taking the covariates that were selected in the best-fitting stepwise variable selection models for PFS and OS (Section 2.1) and using these to generate a range of standard parametric survival models where we jointly fitted the two treatment arms
- Assessing the fit and clinical plausibility of the parametric survival models based on the AIC/BIC values and visual assessment.

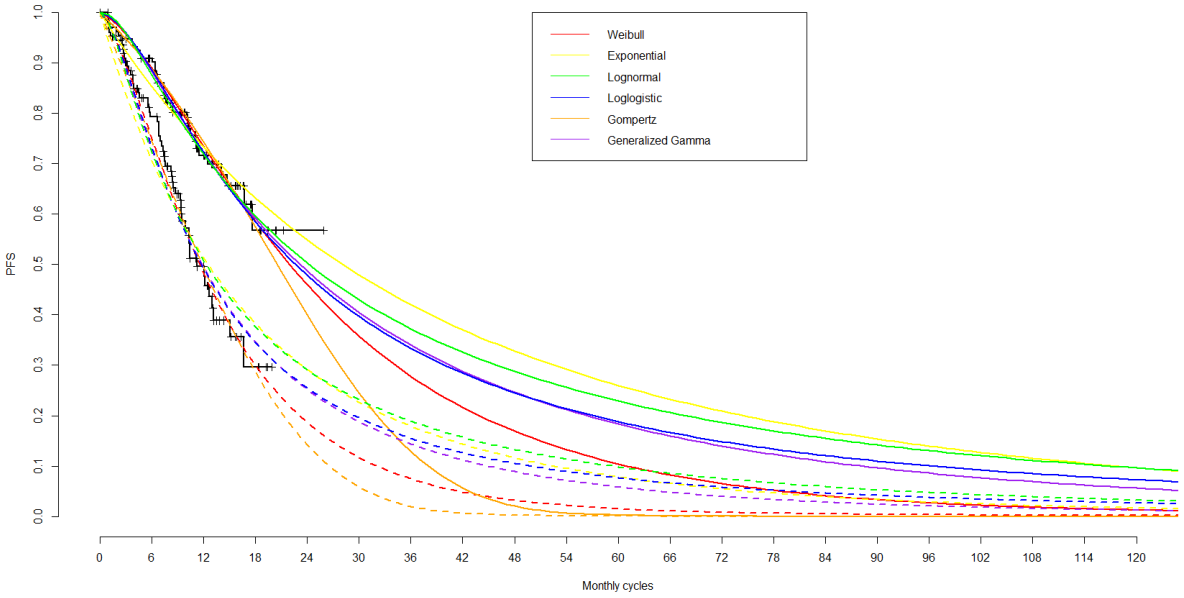
The results of these analyses are described below.

Cd versus Vd (patients who have received one prior therapy and no prior bortezomib)

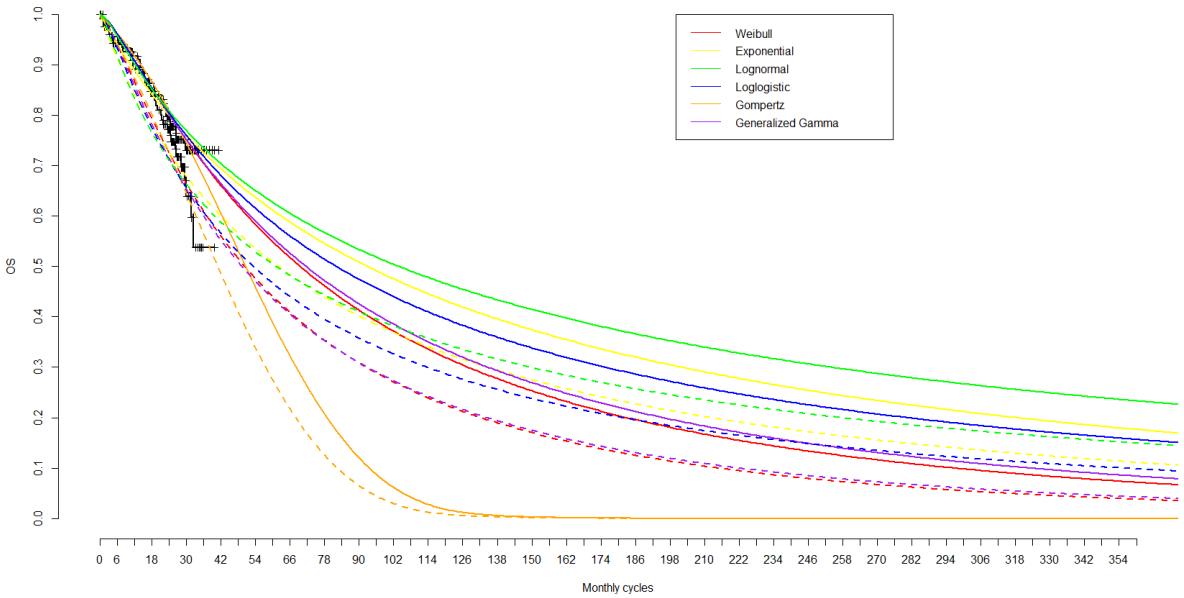
PFS and OS curves from the parametric models fitted jointly to the two ENDEAVOR treatment arms using the CGP approach are provided in Figure 19 and model fit statistics (AIC and BIC) are provided in Table 38.

Figure 19. PFS and OS parametric curves jointly fitted to the two treatment arms using the CGP method (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

PFS



OS



Note: Covariates included are those identified by the best-fitting stepwise selection model (Section 2.1).
Note: solid lines are for Cd and dotted lines are for Vd
Cd, carfilzomib/dexamethasone; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone.

Table 38. Model fit statistics (AIC and BIC) for the parametric PFS and OS curves jointly fitted to the two treatment arms using the CGP method (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	Model	AIC	BIC
PFS	Weibull	747.2038	765.0456
	Exponential	758.9913	773.2646
	Gompertz	751.1742	769.0159
	Generalized Gamma	746.6939	768.1040
	Lognormal	744.9386	762.7803
	Loglogistic	745.4302	763.2719
OS	Weibull	687.9296	720.0447
	Exponential	691.0898	719.6366
	Gompertz	684.3932	716.5083
	Generalized Gamma	689.8528	725.5363
	Lognormal	687.7871	719.9022
	Loglogistic	692.1623	724.2774

AIC, Akaike information criterion; BIC, Bayesian information criterion; OS, overall survival; PFS, progression-free survival

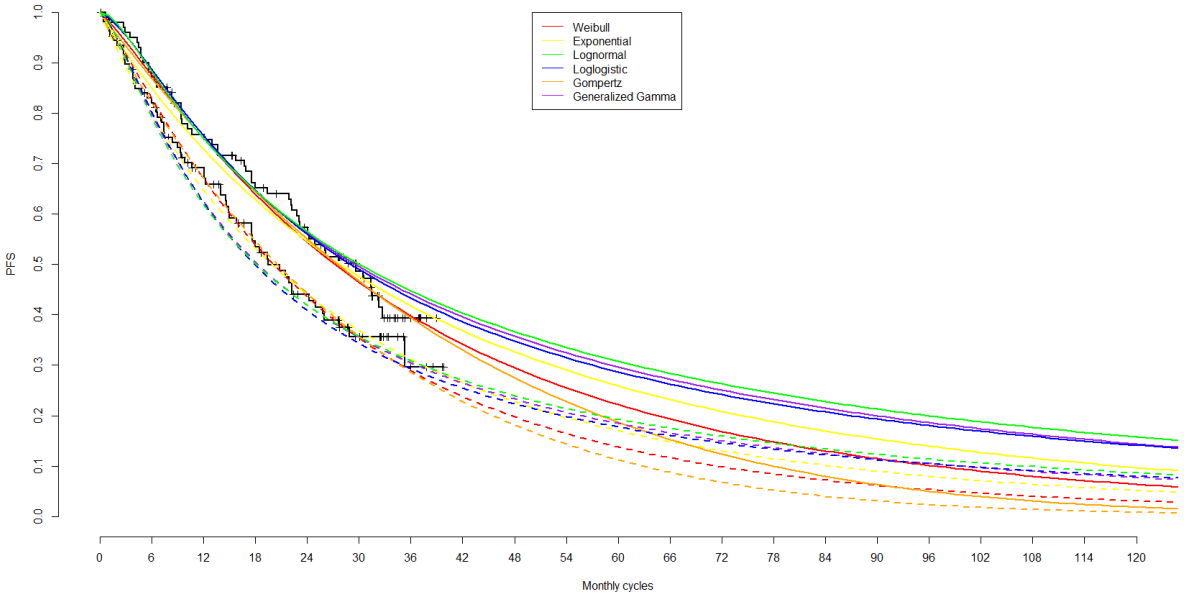
For PFS, the log-logistic, lognormal, exponential and gamma models resulted in very long survival predictions for both arms which are considered implausible. Among the Weibull and Gompertz, the Weibull model had a lower AIC/BIC and was also the best fitting distribution in the PANORAMA-1 trial where almost the complete PFS profile was observed for patients receiving Vd.
For OS, none of the models were considered plausible. The Gompertz model severely underestimated OS whereas the other models overestimated OS. Among those that overestimated OS, the Weibull model resulted in the most conservative projections.

CRd versus Rd (patients who have received two prior therapies and no prior lenalidomide)

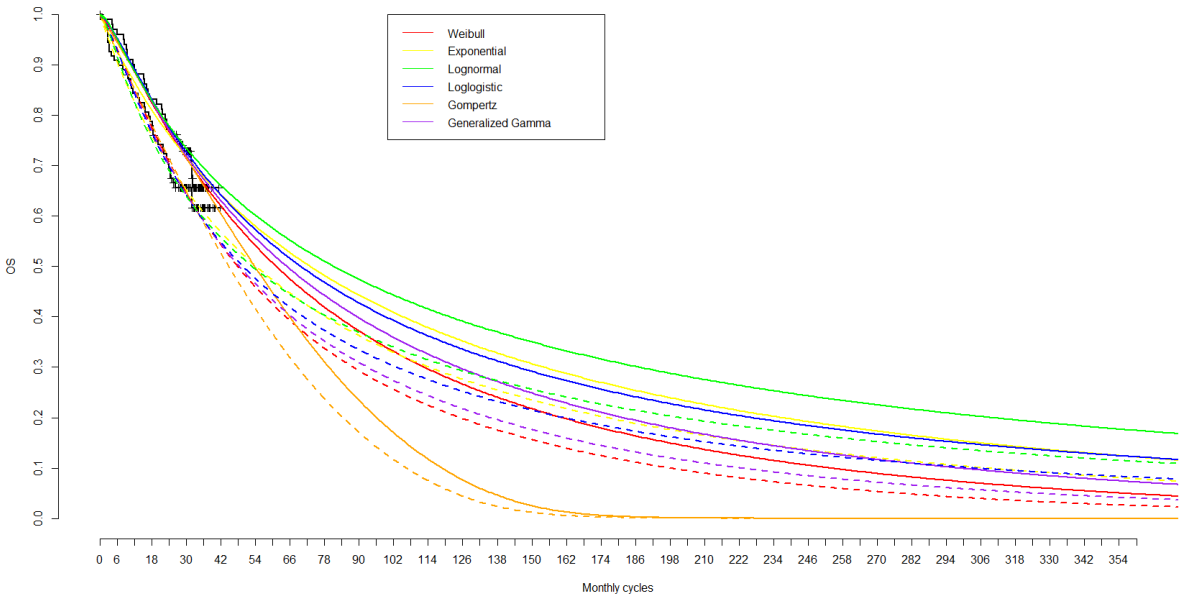
PFS and OS curves from the parametric models fitted jointly to the two ASPIRE treatment arms using the CGP approach are provided in Figure 20 and model fit statistics (AIC and BIC) are provided in Table 39.

Figure 20. PFS and OS parametric curves jointly fitted to the two treatment arms using the CGP method (ASPIRE, patients with one prior therapy and no prior bortezomib)

PFS



OS



Note: Covariates included are those identified by the best-fitting stepwise selection model (Section 2.1).
Note: solid lines are for CRd and dotted lines are for Rd
CRd, carfilzomib/lenalidomide/dexamethasone; PFS, progression-free survival; OS, overall survival; Rd, lenalidomide/dexamethasone.

Table 39. Model fit statistics (AIC and BIC) for the parametric PFS and OS curves jointly fitted to the two treatment arms using the CGP method (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Model	AIC	BIC
PFS	Weibull	1019.109	1045.924
	Exponential	1020.327	1043.79
	Gompertz	1020.229	1047.044
	Generalized Gamma	1015.458	1045.625
	Lognormal	1013.306	1040.121
	Loglogistic	1013.603	1040.418
OS	Weibull	718.366	748.533
	Exponential	720.050	746.865
	Gompertz	718.856	749.022
	Generalized Gamma	720.080	753.598
	Lognormal	718.450	748.617
	Loglogistic	720.599	750.766

AIC, Akaike information criterion; BIC, Bayesian information criterion; OS, overall survival; PFS, progression-free survival

Similar to ENDEAVOR, for PFS, the log-logistic, lognormal, exponential and gamma models resulted in very long survival predictions which are considered implausible. Among the Weibull and Gompertz models, the Weibull model had lower AIC/BIC values.

For OS, the Gompertz model resulted in similar OS estimates as the IPW Gompertz method (Section 2.3), while other models resulted in implausibly long predictions.

Mean of covariates (MoC) method

The MoC approach was assessed as follows:

- Taking the covariates that were selected in the best-fitting stepwise variable selection models for PFS and OS (Section 2.1) and using these to generate a range of standard parametric survival models where we jointly fitted the two treatment arms
- Assessing the fit and clinical plausibility of the parametric survival models based on the AIC/BIC values and visual assessment.

The results of these analyses are described below.

Cd versus Vd (patients who have received one prior therapy and no prior bortezomib)

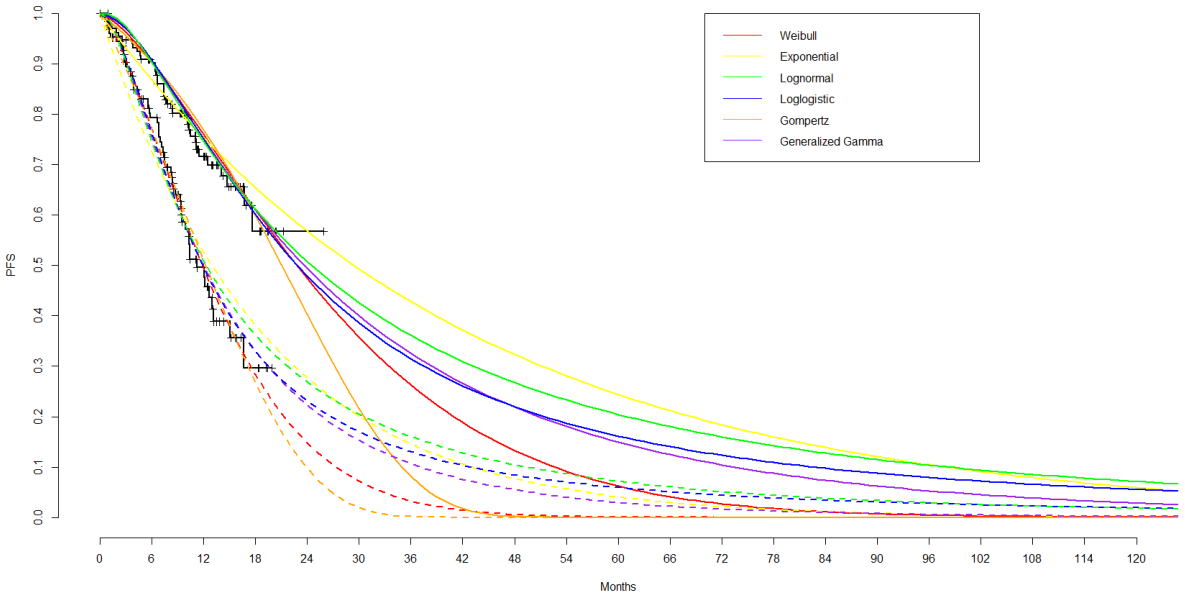
Since the same survival models are used for the MoC method as for the CGP method, model fit statistics (AIC and BIC) are not presented here. PFS and OS curves from the parametric models fitted jointly to the two ENDEAVOR treatment arms using the MoC approach are provided in Figure 21.

For PFS, the MoC method yielded reasonable estimates for the Weibull model. The Gompertz model underestimated PFS and the other models overestimated PFS, particularly for the Cd arm.

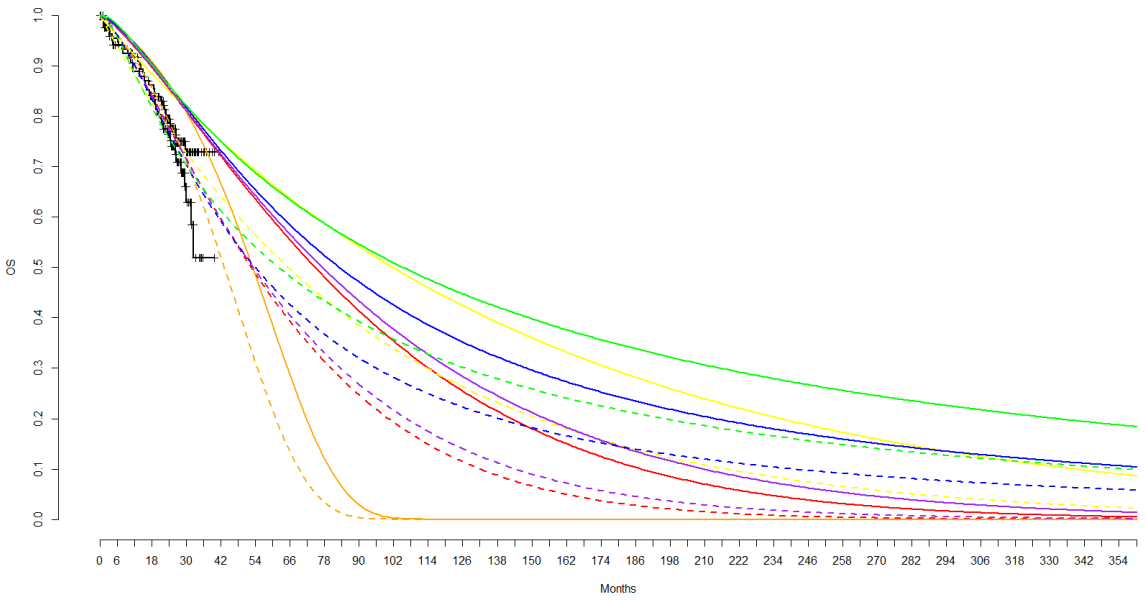
For OS the MoC method overestimated survival for all models, particularly for the Cd arm. This is likely due to the inherent limitations of the MoC approach (see Table 14) which lead to skewed survival estimates.

Figure 21. PFS and OS parametric curves jointly fitted to the two treatment arms using the MoC method (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

PFS



OS



Note: Covariates included are those identified by the best-fitting stepwise selection model (Section 2.1).
Note: solid lines are for Cd and dotted lines are for Vd
Cd, carfilzomib/dexamethasone; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone.

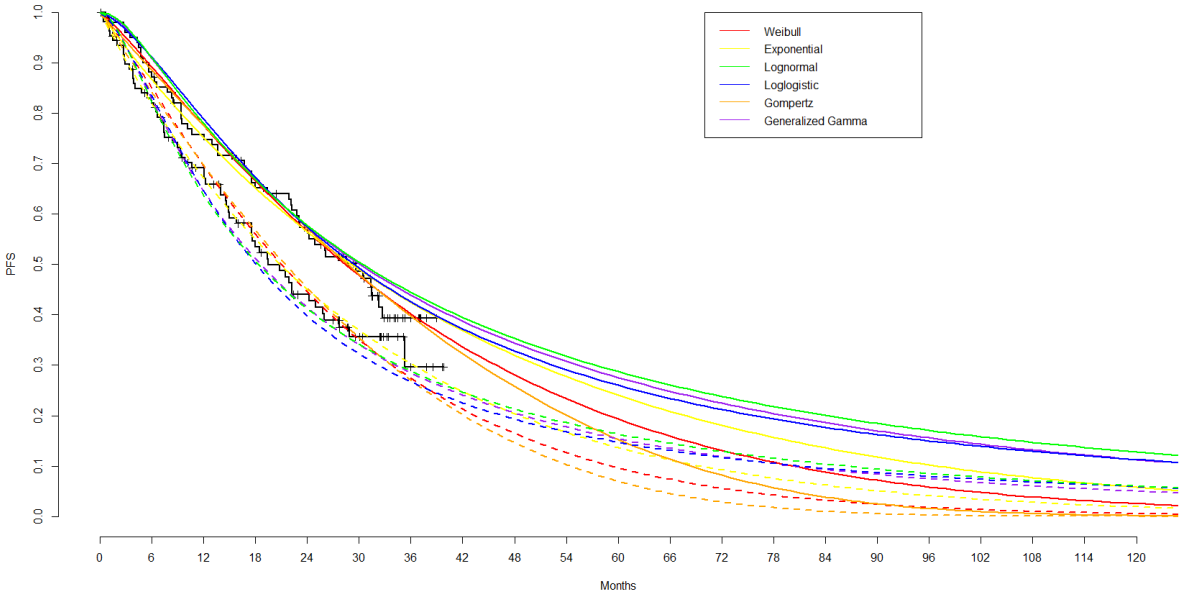
CRd versus Rd (patients who have received two prior therapies and no prior lenalidomide)

Since the same survival models are used for the MoC method as for the CGP method, model fit statistics (AIC and BIC) are not presented here. PFS and OS curves from the parametric models fitted jointly to the two ASPIRE treatment arms using the MoC approach are provided in Figure 22.

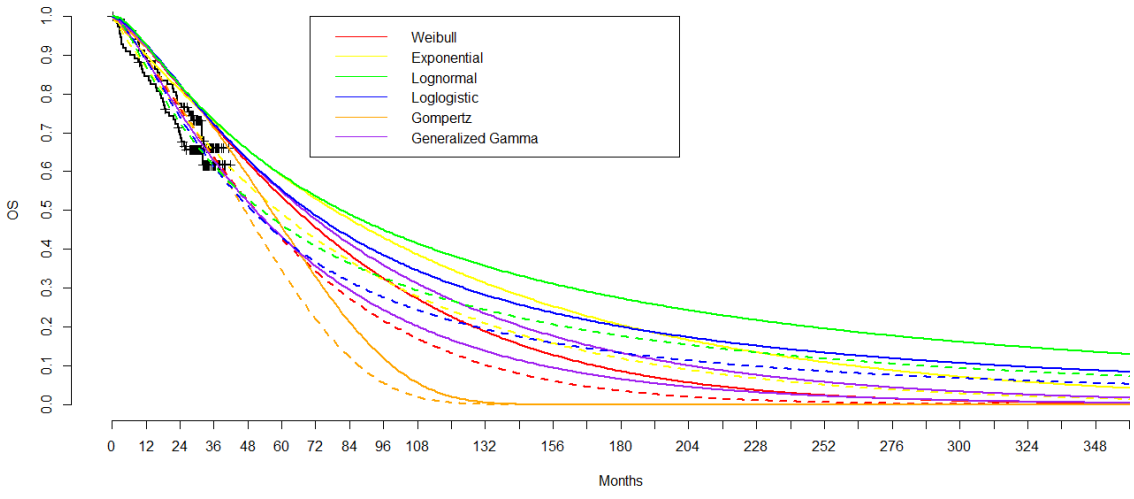
The MoC method yielded reasonable PFS estimates for the Weibull and Gompertz models, with the Weibull model having lower AIC/BIC values. For OS, the MoC method appeared to overestimate survival for both arms in all models.

Figure 22. PFS and OS parametric curves jointly fitted to the two treatment arms using the MoC method (ASPIRE, patients with two prior therapies and no prior lenalidomide)

PFS



OS



Note: Covariates included are those identified by the best-fitting stepwise selection model (Section 2.1).
Note: solid lines are for CRd and dotted lines are for Rd
CRd, carfilzomib/lenalidomide/dexamethasone; PFS, progression-free survival; OS, overall survival; Rd, lenalidomide/dexamethasone.

8.5 Appendix E: Matching-adjusted indirect comparison (Vd in ENDEAVOR and CASTOR) additional information

Matching-adjusted indirect comparison introduction

Matching-adjusted indirect comparison (MAIC) is a relatively new technique for making comparisons between treatments using a combination of individual patient level data for one treatment and aggregated data for another.[25,26] This method can overcome some of the limitations of indirect comparative effectiveness analysis that use aggregated data alone (e.g. network meta-analyses), including differences in baseline characteristics. For studies to be compatible they should be broadly similar in trial design, In addition, the patient populations should have sufficient overlap in characteristics that influence disease progression, e.g. a trial of patients with treatment-naïve MM would not be compatible with a trial of patients with R/RMM.

The key assumptions involved in an MAIC are that:

- the studies are generally compatible for comparison: that is, differences in outcomes can be attributed to differences in the population and the treatments received
- differences in design or background factors have little or negligible impact on outcomes, or can be taken into account in the analyses
- matching the mean values of the characteristics is sufficient. This implies, for instance, that the correlation between characteristics and the distribution around the mean is the same in the index and comparator trials
- all potential confounders are available and have been included in the matching
- the effects derived from the MAIC apply uniformly to the overall patient population once adjusted for available baseline characteristics.

The MAIC approach has been used in previous NICE technology appraisals to date in MM. The method has been generally accepted, although considered equivalent to observational evidence, and dependent on the details of the methods and the number of variables needed for matching.

The CASTOR study

Like ENDEAVOR, CASTOR is an open-label RCT in patients with relapsed and/or refractory MM (R/RMM).[27] Patients in CASTOR were randomised to receive either daratumumab/bortezomib/dexamethasone (DVd) or bortezomib/dexamethasone alone. In the Vd arm, patients received subcutaneous bortezomib (1.3 mg/m^2) on Days 1, 4, 8, and 11 for 8 x 3-week cycles i.e. 24 weeks duration of treatment and 32 doses. This is consistent with the marketing authorisation for bortezomib and clinical practice in England and Wales where patients receive a maximum of 32 doses of treatment. A comparison of study methodology is provided in Table 40.

Table 40. Comparison of ENDEAVOR and CASTOR study methodology

Item	ENDEAVOR	CASTOR
Blinding	Open-label (randomisation 2012-2014), 27 countries	Open-label (randomisation 2014-2015), 15 countries
Cycle length	Cd: 28 days Vd: 21 days	Cycles 1-8: 21 days Cycles 9+: 28 days
Dosing	Cd arm Carfilzomib: 20 mg/m^2 IV, Cycle 1: (Days 1-2); 56 mg/m^2 IV, Cycle 1 (Days 8, 9, 15, 16), Cycle 2+ (Days 1, 2, 8, 9, 15, 16) Dexamethasone: 20 mg PO/IV , Cycles 1 + (Days 1, 2, 8, 9, 15, 16, 22, 23) Vd arm Bortezomib: 1.3 mg/m^2 IV/SC, Cycles 1+ (Days 1, 4, 8, 11) Dexamethasone: 20 mg PO/IV , Cycles 1 + (Days 1, 2, 8, 9, 15, 16, 22, 23)	DVd arm Daratumumab: 16 mg/kg IV, Cycles 1-3 (weekly), Cycles 4-8 (every 3 weeks), Cycle 9+ (every 4 weeks) Bortezomib: 1.3 mg/m^2 SC, Cycles 1-8 (Days 1,4,8,11) Dexamethasone: 20 mg PO/IV, Cycles 1-8 (Days 1,2,4,5,8,9,11,12) Vd arm Bortezomib: 1.3 mg/m^2 SC, Cycles 1-8 (Days 1,4,8,11) Dexamethasone: 20 mg PO/IV, Cycles 1-8 (Days 1,2,4,5,8,9,11,12)
Stratification factors	• Number of prior therapies (1 vs. 2-3) • ISS at screening (I vs. II or III) • Prior proteasome inhibitor (yes vs. no) • Planned administration route of bortezomib at randomisation (SC vs. IV)	• Number of prior therapies(1 vs. 2-3 vs. 4+) • ISS at screening (1 vs. 2 vs. 3) • Prior bortezomib (yes vs. no)
Disease assessments	Based on IMWG criteria	Based on IMWG criteria
Key eligibility criteria	Inclusion: • 1-3 prior lines of therapy • Achieved at least a PR to ≥ 1 prior therapy • ECOG PS 0-2 Exclusion: • Did not response (< PR) to prior bortezomib or carfilzomib • Discontinuation of prior bortezomib or carfilzomib due to AEs • ≥ 6 months treatment-free interval since last proteasome inhibitor • Grade ≥2 PN or neuropathic pain • ANC < 10^9/L	Inclusion: • ≥1 prior line of therapy • Achieved at least a PR to ≥ 1 prior therapy Exclusion: • Refractory to proteasome inhibitors • Discontinuation of prior bortezomib or carfilzomib due to AEs • Grade ≥2 PN or neuropathic pain • ANC ≤ 10^9/L • Hemoglobin ≤ 7.5g/dL • Platelets ≤ 75*10^9/L

Item	ENDEAVOR	CASTOR
	• Platelets < 50*10^9/L • LVEF < 40% • CrCl < 15mL/min • MI/hearth failure within 4 months before randomisation	• AST ≥ 2.5 times ULN • Bilirubin ≥ 1.5 times ULN • CrCl < 20mL/min per 1.73m^2 BSA
AE, adverse event; ANC, absolute neutrophil count; Cd, carfilzomib/dexamethasone; CrCl, creatinine clearance; DVd, daratumumab/bortezomib/dexamethasone; ECOG PS, Eastern Cooperative Oncology Group performance status; IMWG, International Myeloma Working Group; ISS, international staging system; IV, intravenous; LVEF, left ventricular ejection fraction; PN, peripheral neuropathy; PO, oral; PR, partial response; SC, subcutaneous; ULN, upper limit of normal; Vd, bortezomib/dexamethasone		

Matching-adjusted indirect comparison methods

The MAIC was conducted to compare changes in PFS and OS for Vd in the CASTOR study (8 cycles, 24 weeks, 32 doses) and the ENDEAVOR study (treatment until progression or unacceptable toxicity). We digitised PFS and OS curves for Vd from the CASTOR trial and generated patient-level data,[28] and then applied the MAIC methodology described by Signorovitch *et al.,*[25,26] to adjust the Vd arm in ENDEAVOR to the Vd arm in CASTOR. The variables used for matching were all those considered by clinicians to be prognostic for outcomes (i.e. PFS and OS) in MM (Section 2.1) and reported for both the ENDEAVOR and CASTOR studies, with the exception of cytogenetic risk status as there was a substantial proportion of patients with of missing/unknown data in ENDEAVOR (15.5% of the ITT population).

The full list of variables used for matching is as follows:

- age
- ISS stage
- time since diagnosis
- creatinine clearance
- number of prior therapies
- prior SCT
- prior bortezomib exposure
- prior IMiD exposure (used as a proxy for prior lenalidomide exposure identified by the clinicians)
- refractory to last prior treatment

After adjusting for these baseline characteristics, we ran a jointly fitted Weibull model (consistent with the best fitting parametric distribution in the cost-effectiveness analyses; Section 2.3) and estimated piecewise time-dependent HRs using the approximated patient-level data for Vd in CASTOR and the matched (weighted) patient-level data for Vd in ENDEAVOR. The cut-off point for the HRs was defined as 24 weeks, which is equivalent to the 8 cycles treatment duration recommended in the bortezomib marketing authorisation. We then adjusted the post-24 week HRs to obtain HRs that reflect the likely effect of prolonged treatment with Vd beyond 8 cycles on PFS and OS.

MAIC results

The matching was successful, and yielded an effective ENDEAVOR Vd sample size of 335.5 patients (72% of ITT population). A summary of baseline characteristics in the Vd arms of CASTOR and ENDEAVOR before and after adjusting baseline characteristics is provided in Table 41.

Table 41. Matched and unmatched baseline characteristics in the MAIC (ENDEAVOR Vd vs. CASTOR Vd)

Characteristic	ENDEAVOR Vd N = 425[a]	ENDEAVOR Vd matched to CASTOR N = 335.5	CASTOR Vd N = 247
Age ≥ 65 years, %[b]	55.3	49.4	49.4
Age ≥ 75 years, %[b]	13.6	14.2	14.2
ISS II at baseline, %	31.5	40.5	40.5
ISS III at baseline, %	22.1	20.6	20.6
Time since diagnosis ≥ 3.72 years, %[c]	47.3	50	50
Creatinine clearance at baseline > 60 mL/min, %	64.9	66.0	66.0
2 prior lines of therapy, %	30.4	30.0	30.0
≥ 3 prior lines of therapy, %	17.4	24.3	24.3
Prior SCT, %	60.2	60.3	60.3
Prior bortezomib, %	51.8	66.4	66.4
Prior IMiD, %	74.4	80.2	80.2
Refractory to last therapy, %	37.6	34.4	34.4

Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

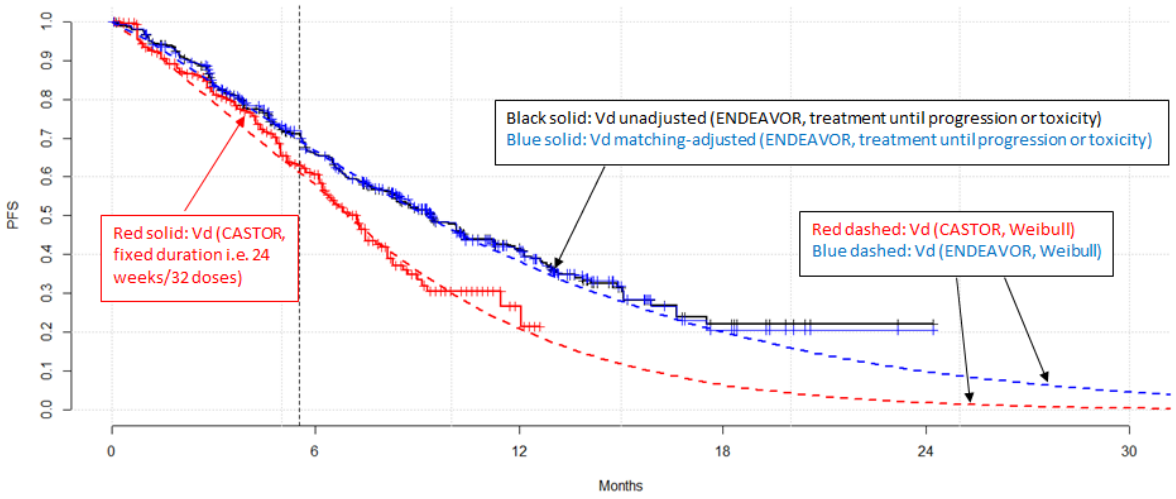
[a] Of the 645 patients in the Vd arm of the ENDEAVOR ITT population, 39 were excluded from the analysis for not meeting CASTOR eligibility criteria and 1 was excluded for not having a value for time since diagnosis
[b] Age cut-offs reported for CASTOR
[c] Median time since diagnosis reported for CASTOR

IMiD, immunomodulatory drug; ISS, International Staging System; MAIC, matching-adjusted indirect comparison; SCT, stem cell transplant; Vd, bortezomib/dexamethasone

PFS and OS results from the MAIC analysis are provided below, and indicate that when bortezomib is given for 8 cycles (24 weeks, 32 doses), both PFS (Figure 23) and OS (Figure 24) outcomes are poorer than those observed in the ENDEAVOR RCT after 8 cycles of

treatment. The matching-adjusted HRs for Vd in CASTOR compared with Vd in ENDEAVOR reflecting the residual effect of Vd treatment post-8 cycles in ENDEAVOR were 1.360 (PFS) and 1.349 (OS).

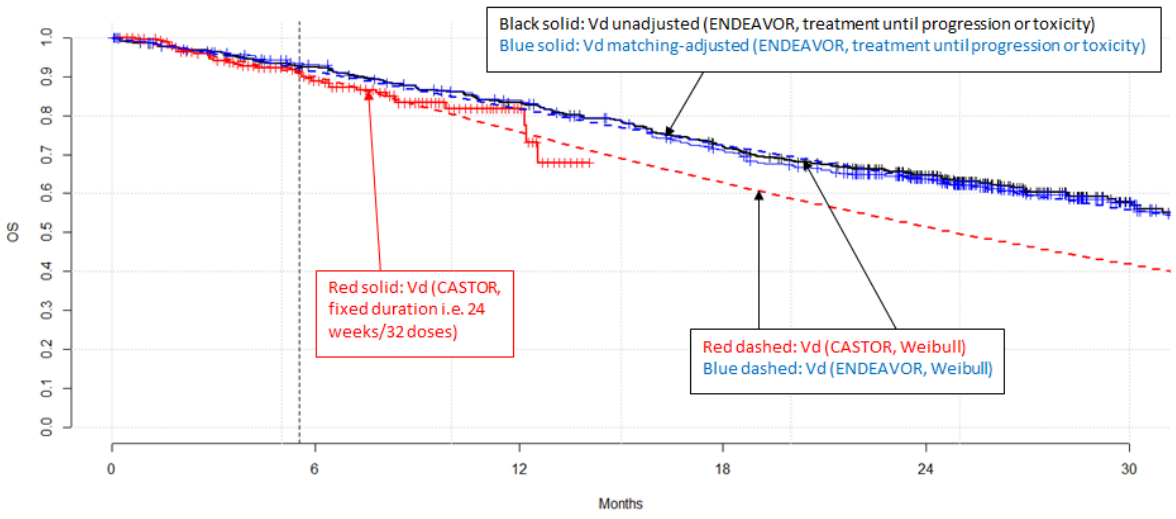
Figure 23. PFS curves from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity)



Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

MAIC, matching adjusted indirect comparison; Vd, bortezomib/dexamethasone

Figure 24. OS curves from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) versus ENDEAVOR (treatment until progression or unacceptable toxicity)



Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic

risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

MAIC, matching adjusted indirect comparison; Vd, bortezomib/dexamethasone

Table 42. PFS and OS hazard ratios (first 24 weeks vs. after 24 weeks) from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity)

	PFS (Weibull model)	**OS (2-arm Weibull model)**
Piecewise matching-adjusted (first 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.344 (1.021 - 1. 769)	1.157 (0.679 – 1.972)
Piecewise matching-adjusted (after 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.828 (1.311 - 2.549)	1.561 (0.942 – 2.588)
Matching-adjusted (residual effect of post-24 weeks treatment in ENDEAVOR), HR (95% CI) CASTOR:ENDEAVOR[a]	**1.360 (0.913 - 2.027)**	**1.349 (0.684 - 2.662)**

Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure, baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment

[a] HR for first 8 cycles (24 weeks) versus subsequent cycles

CI, confidence interval; HR, hazard ratio; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone

8.6 Appendix F: Sensitivity, scenario and probabilistic analyses based on the revised cost-effectiveness analyses

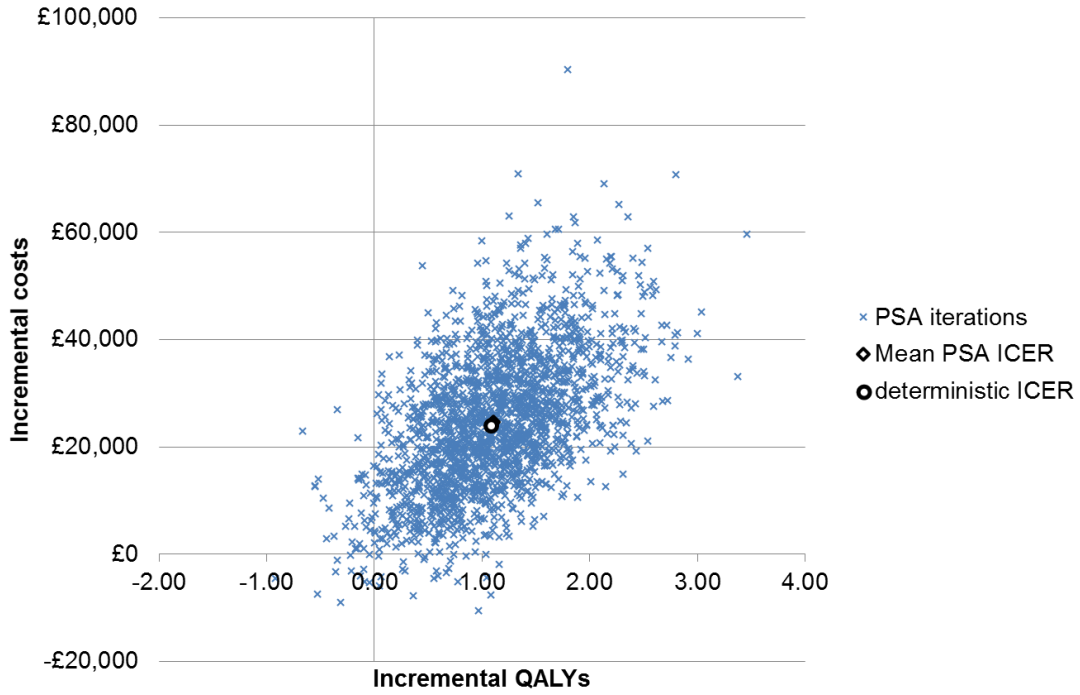
Probabilistic sensitivity analysis results for Cd versus Vd in patients who have received one prior therapy and have not received prior bortezomib using the final revised cost-effectiveness analysis (incorporating IPW-based survival models and revised mapped utility values) is presented in Table 43. A scatter plot is presented in Figure 25 and the cost-effectiveness acceptability curve in Figure 26.

Table 43. Cd versus Vd – Probabilistic ICER – revised cost-effectiveness analysis

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) incremental (QALYs)
Vd	£93,812	2.83			
Cd	£118,391	3.93	£24,579	1.10	£22,326

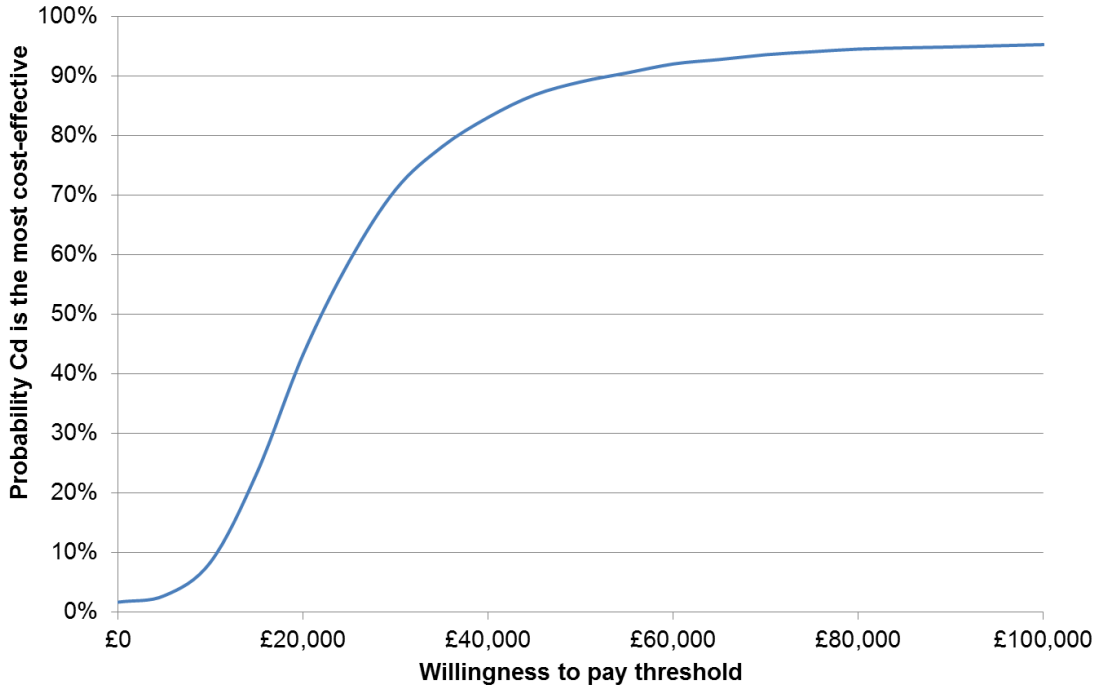
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year; Vd, bortezomib/dexamethasone

Figure 25. Scatter plot of incremental cost and QALYs: Cd versus Vd – revised cost-effectiveness analysis



Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; Vd, bortezomib/dexamethasone

Figure 26. Cost-effectiveness acceptability curve: Cd versus Vd – revised cost-effectiveness analysis



Probability Cd is cost-effective at		
£20,000/QALY	**£30,000/QALY**	**£50,000/QALY**
43.35%	71.00%	89.05%

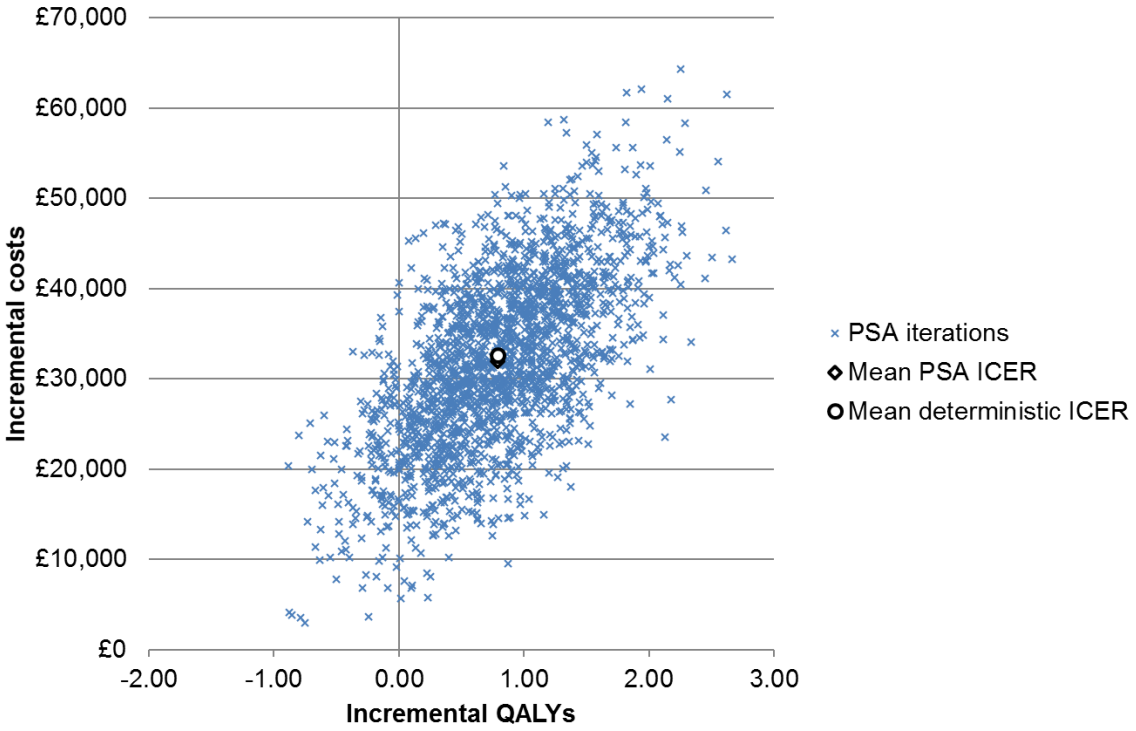
Cd, carfilzomib/dexamethasone; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone

Probabilistic sensitivity analysis results for CRd versus Rd in patients who have received two prior therapies and have not received prior lenalidomide using the revised cost-effectiveness analysis (incorporating IPW-based survival models and revised mapped utility values) is presented in Table 44. The corresponding scatter plot is presented in Figure 27 and the cost-effectiveness acceptability curve in Figure 28.

Table 44. CRd versus Rd – Probabilistic ICER – revised cost-effectiveness analysis

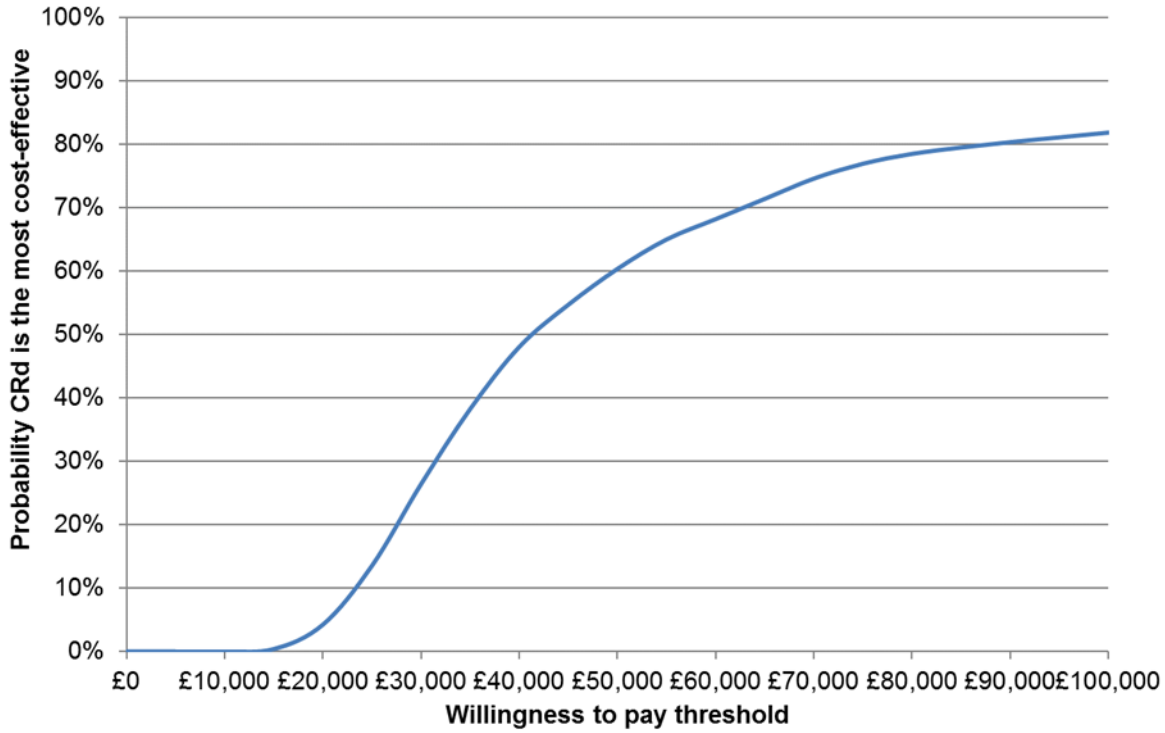
Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) incremental (QALYs)
Rd	£93,505	2.91			
CRd	£125,454	3.70	£31,949	0.79	£40,457
CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year; Rd, lenalidomide/dexamethasone					

Figure 27. Scatter plot of incremental cost and QALYs: CRd versus Rd – revised cost-effectiveness analysis



CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; Rd, lenalidomide/dexamethasone

Figure 28. Cost-effectiveness acceptability curve: CRd versus Rd – revised cost-effectiveness analysis



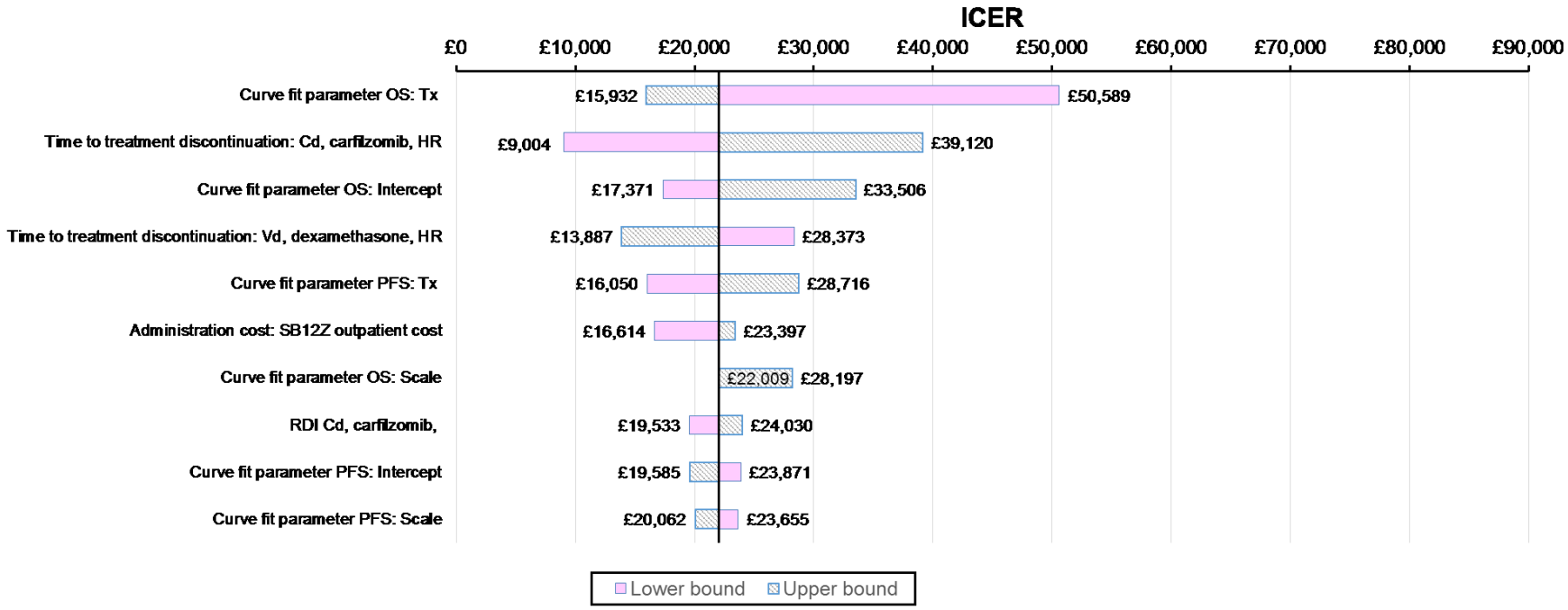
Probability CRd is cost-effective at		
£20,000/QALY	**£30,000/QALY**	**£50,000/QALY**
4.25%	26.45%	60.40%

CRd, carfilzomib/lenalidomide/dexamethasone; QALY, quality-adjusted life year; Rd, lenalidomide/dexamethasone

Deterministic sensitivity analyses for Cd versus Vd in patients who have received one prior therapy and have not received prior bortezomib using the revised cost-effectiveness analysis are presented in Figure 29.

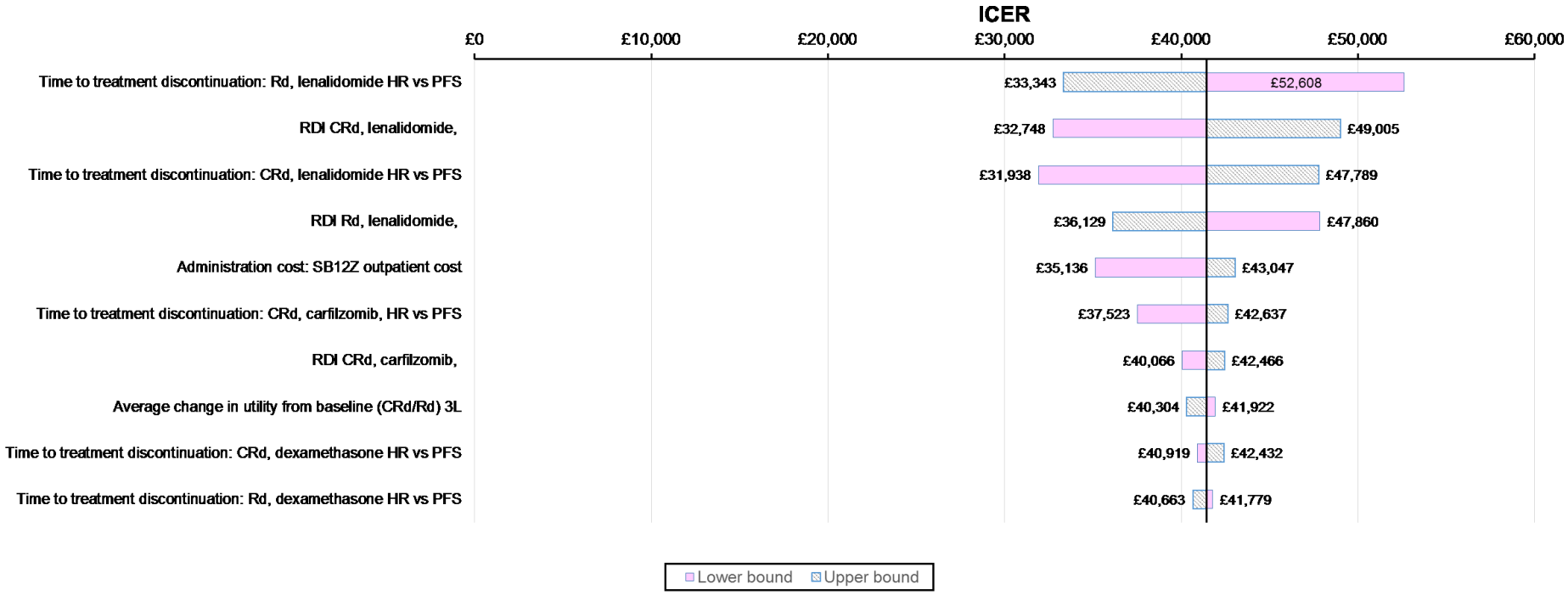
Deterministic sensitivity analyses for CRd versus Rd in patients who have received two prior therapy and have not received prior lenalidomide using the revised cost-effectiveness analysis are presented in Figure 30.

Figure 29. Tornado diagram Cd versus Vd using revised cost-effectiveness analysis



Cd, carfilzomib/dexamethasone; HR, hazard ratio ICER, incremental cost-effectiveness ratio; OS, overall survival, PFS, progression-free survival; RDI, relative dose intensity; Tx, treatment; Vd, bortezomib/dexamethasone.

Figure 30. Tornado diagram CRd versus Rd using revised cost-effectiveness analysis



3L, third line; CRd, carfilzomib/lenalidomide/dexamethasone; HR, hazard ratio ICER, incremental cost-effectiveness ratio; OS, overall survival, PFS, progression-free survival; Rd, lenalidomide/dexamethasone; RDI, relative dose intensity; Tx, treatment;

The impact on revised cost-effectiveness analysis of using alternative parametric functions for both PFS and OS is presented for the Cd versus Vd and CRd versus Rd comparisons in Table 45 and Table 46 respectively.

Table 45. Impact of using different parametric function for PFS and OS on the revised cost-effectiveness analysis of Cd versus Vd

Scenario	PFS	OS	ICER
Alternative cost-effectiveness analysis	*IPW Weibull*	*IPW Weibull*	*£22,009*
1	IPW Exponential	IPW Weibull	£23,740
2	IPW Gompertz	IPW Weibull	£22,721
3	IPW Gamma	IPW Weibull	£22,021
4	IPW Lognormal	IPW Weibull	£22,203
5	IPW Loglogistic	IPW Weibull	£21,254
6	IPW Weibull	IPW Exponential	£18,798
7	IPW Weibull	IPW Gompertz	£30,697
8	IPW Weibull	IPW Gamma	£24,933
9	IPW Weibull	IPW Lognormal	£23,606
10	IPW Weibull	IPW Loglogistic	£23,003
ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; PFS, progression-free survival; OS, overall survival.			

Table 46. Impact of using different parametric function for PFS and OS on the revised cost-effectiveness analysis of CRd versus Rd

Scenario #	PFS	OS	ICER
Alternative cost-effectiveness analysis	*IPW Weibull*	*IPW Weibull*	*£41,429*
1	IPW Exponential	IPW Weibull	£40,163
2	IPW Gompertz	IPW Weibull	£42,250
3	IPW Gamma	IPW Weibull	£40,554
4	IPW Lognormal	IPW Weibull	£40,470
5	IPW Loglogistic	IPW Weibull	£41,198
6	IPW Weibull	IPW Exponential	£37,072
7	IPW Weibull	IPW Gompertz	£52,439
8	IPW Weibull	IPW Gamma	£37,331
9	IPW Weibull	IPW Lognormal	£34,208
10	IPW Weibull	IPW Loglogistic	£39,268
ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; PFS, progression-free survival; OS, overall survival.			

Professor Eugene Milne
Chair, Appraisal Committee C
National Institute for Health and Care Excellence
Level 1A, City Tower
Piccadilly Plaza
Manchester
M1 4BT

Wednesday 30 November 2016

Dear Prof Milne

Myeloma UK response to NICE Appraisal Consultation Document (ACD) on carfilzomib

Myeloma UK welcomes the opportunity to comment on the NICE ACD on carfilzomib (Kyprolis®) in combination with lenalidomide and dexamethasone or dexamethasone alone.

We have a good working relationship with NICE and have absolute confidence in its appraisal methodology and processes. We also understand the difficulties faced by the committee in approving new medicines, particularly in the face of uncertainty. However, we are obviously very disappointed at the decision reached by the NICE appraisal committee on carfilzomib.

As myeloma is a complex and individual cancer, clinicians need a range of treatments available at every stage of the disease, to ensure that they are able to treat their patients optimally. The negative decision means that relapsed myeloma patients will face a further delay in accessing carfilzomib, a very effective treatment option, on the NHS.

The ACD highlights the appraisal committee's clear acceptance of the clinical case and need for both carfilzomib combinations. In particular, recognising the survival benefits, the clinical and patient need for carfilzomib and the quality-of-life benefits of the treatment to patients. NICE therefore has all the data available to them demonstrating why patients and their carers will benefit from accessing carfilzomib as part of their treatment pathway and agree that a "compelling" clinical case has been made for approval.

Analysing the Committee's concerns around cost-effectiveness highlighted in the ACD, we are cautiously optimistic that the challenges with the health economic modelling and uncertainty will be overcome by further clarification and dialogue. We therefore urge NICE and Amgen to collaborate to find a solution that benefits everyone and provides vital access to a new and innovative treatment for myeloma patients on the NHS.

We look forward to working with NICE to find a solution for myeloma patients and carers. Please do not hesitate to contact me if we can provide any further information to support the appraisal.

Yours sincerely

Eric Low OBE
Chief Executive

Myeloma UK
22 Logie Mill, Beaverbank Business Park, Edinburgh, EH7 4HG
Company No.190563 Charity No. SC 026116 Myeloma UK is an Investor in People
Myeloma Infoline 0800 980 3332

Janssen's Response to the Appraisal Consultation Document (ACD)

Carfilzomib for previously treated multiple myeloma [ID934]

Janssen is pleased to have the opportunity to comment on the above ACD for carfilzomib.

Sections 4.2 to 4.5

Janssen would like to highlight that in Sections 4.2 to 4.5 (Decision problem and treatment pathway), no mention is made of the 1st line treatment of patients eligible for a stem cell transplant. As noted in the Final scope for this appraisal:

"NICE technology appraisal guidance 311 recommends bortezomib as an option, in combination with dexamethasone or with dexamethasone and thalidomide, for the induction treatment of adults with untreated multiple myeloma who are eligible for high-dose chemotherapy with haematopoietic stem cell transplantation."

Furthermore, the exclusion of this aspect of the patient journey results in inconsistency between Myeloma appraisals.

Section 4.5

Within the ACD, Section 4.5, page 8, it is stated "carfilzomib with dexamethasone would only replace bortezomib with dexamethasone at second line if people had not had bortezomib therapy at first line (and instead had thalidomide therapy at first line, as the most commonly used regimen; see section 4.4).

Janssen would like to highlight that with the inclusion of patients eligible for stem cell transplant, bortezomib is the most commonly used 1st line treatment.

Single Technology Appraisal (STA)

Carfilzomib for treated multiple myeloma [ID934]

Comments on Appraisal consultation document

• Has all of the relevant evidence been taken into account?

Yes

• Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?

Yes

• Are the provisional recommendations sound and a suitable basis for guidance to the NHS?

Provisional recommendations would limit therapy options for patients with relapsed myeloma. Addition of an irreversible reasonably well tolerated proteasome inhibitor, also shown in UK based MUK5 trial (ASH 2015 Abstract No 1840) is likely to improve long-term outcomes for patients. Myeloma is genomically unstable with clonal tiding and additional mutations at relapse (Smith et al British Journal of Haematology, 2015, 171, 881–883). Outcomes for patients with high-risk disease (up to 30% of patients) at relapse are poor with significant management challenges. Carfilzomib in combination with lenalidomide and dexamethasone significantly improves clinical outcomes for high-risk patients in the ASPIRE trial (Blood 2016 128:1174-1180). The current standard of care Lenalidomide and dexamethasone is clearly suboptimal for this group of patients.

 Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination versus any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?

Nil

Comments on the ACD Received from the Public through the NICE Website

Name	[redacted]
Role	NHS Professional
Other role	Consultant haematologist
Organisation	Mid Yorkshire NHS Trust/University of Leeds
Location	England
Conflict	n/a
Notes	
Comments on individual sections of the ACD:	
Section 1 (Appraisal Committee's preliminary recommendations)	The ACD should be considered carefully as the clinical data is very impressive and providing Amgen can resolve the economic modelling this technology would help bridge a significant healthcare need in this group of patients who have limited prognosis and QOL on the limited alternative therapies at present available in this setting. It would appear that assumptions about the dosage have been derived by the company from the time on treatment rather than true numbers of cycles which is clearly an error. I am sure you will recieve many comments re this ACD as Carfilzomib clearly can provide a significant improvement in our present treatment armoury.
Section 2 (The technology)	
Section 3 (The manufacturer's submission)	
Section 4 (Consideration of the evidence)	
Section 5 (Implementation)	
Section 6 (Related NICE guidance)	
Section 7 (Proposed date of review of guidance)	

Name	[redacted]
Role	NHS Professional
Other role	
Organisation	
Location	England
Conflict	Yes
Notes	I am a director of [redacted]. I am [redacted] for the myeloma 11+ trial which has carfilizomib in one of the experimental arms. The trial has received funding from Amgen/Onyx and I personally have attended advisory boards for Amgen and have received honoraria for speaking at symposia. I was also on the safety monitoring committee for the Endeavour study.
Comments on individual sections of the ACD:	
Section 1 (Appraisal Committee's	As the CI of the myeloma XI+ trial I would like to firstly say that

preliminary recommendations)	Amgen/Onyx have been hugely supportive of clinical trials in the U.K. This has allowed many physicians to gain first hand experience of this powerful and effective proteosome inhibitor. The endeavour and aspire studies show considerable efficacy over the current standards of care. In addition it is not associated with a significant risk of neuropathy which is a difficult and frequently disabling irreversible side-effect of both velcade and thalidomide. Also the current nice guideance on single agent velcade is woefully out of date and there is an onus on Nice to provide better guideance on front and second line therapy. The FAD on second line single agent velcade is out of date as most patients now receive velcade as front line therapy and NICE and NHS England advice on second line therapy are currently at odds. The myeloma community would welcome the availability of carfilizomib in combination with either dexamethasone or dexamethasone/revlimid.
Section 2 (The technology)	
Section 3 (The manufacturer's submission)	
Section 4 (Consideration of the evidence)	
Section 5 (Implementation)	
Section 6 (Related NICE guidance)	
Section 7 (Proposed date of review of guidance)	

Carfilzomib for previously treated multiple myeloma

ERG's review of the company's comments on ACD

This report was commissioned by the NIHR HTA Programme as project number 15/121/18

BMJ Technology Assessment Group

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In response to the Appraisal Consultation Document (ACD) developed by the Committee, the company submitted updated analyses for the two main proposed positions for carfilzomib as a treatment for previously treated multiple myeloma. That is, as a second-line treatment in combination with dexamethasone, in patients who have not previously received bortezomib (2LnV), and as a third-line treatment in combination with lenalidomide and dexamethasone (3LnRC) in patients who have not previously received lenalidomide or carfilzomib.

As part of the comments on the ACD, the company provided revised models for the key subgroups of interest in the ENDEAVOR and ASPIRE models. The secondary positioning in the ASPIRE model has not been revised. The company also provided a range of updated analyses for the estimation of a covariate-adjusted treatment effect as well as a range of parametric regression models incorporating covariate adjustment.

The following sections outline where the company have attempted to address the key concerns of the Committee and the ERG's critique of the methods chosen by the company to address those concerns. The company's revised base case is presented in Section 5 and the ERG's revised base case is given in Section 6.

1. Survival analysis to estimate long-term effects

1.1. Plausibility of trial covariate-adjusted efficacy estimates

The ACD states that the Committee requested additional evidence on plausible efficacy estimates for all comparators, taking into account the covariates used in the adjustment of these estimates. In response to the Committee's request, the company firstly reiterated that the original method for variable selection, being informed by clinical expertise without any automated selection method applied, was preferable as it ensured that current knowledge and opinion regarding prognostic factors was taken into account. The ERG agrees that clinical expert opinion should be sought to identify those variables that are prognostic of the outcomes, but consideration of the uncertainty surrounding the estimated coefficients for each covariate and the potential impact that any highly uncertain coefficients could have on the adjusted treatment effect needs to be taken into account. It is therefore essential to apply suitable statistical methods to inform the appropriate variable selection and hence minimise potential bias in the estimate of the treatment effect.

The company provided efficacy results for progression-free survival (PFS) and overall survival (OS) using the following five methods of variable selection (including the company's original method):

1. All clinician-identified covariates fitted on the full trial population (Original model);
2. Stepwise variable selection fitted within subgroups;
3. All clinician-identified covariates fitted within subgroups;
4. Least absolute shrinkage and selection (LASSO) fitted within subgroups;
5. Cox PH model fitted within subgroups using LASSO variables from the previous method.

Although the company provided the results of the ERG's suggested method for variable selection, the company chose to apply the stepwise method for use in the updated cost-effectiveness model. The ERG are concerned at the lack of clarity in the regression results presented and find it difficult to fully assess the best model fit. The company did not provide confidence intervals around the coefficients for any other models except for the chosen stepwise model and did not provide AIC or BIC for the LASSO method, which prevents a comparison of the goodness-of-fit being made across all the models. Given the uncertainty this causes, the ERG considers the LASSO method to be the most reliable as it reduces the effect that large coefficients for variables with a large confidence intervals have on the adjusted treatment effect. This method also resulted in the smallest treatment effect and hence, considering the uncertainty surrounding the analysis, this method can at least be considered as the most conservative estimate, and therefore might be regarded as the most suitable in the absence of strong evidence for an alternative.

1.2. Proportional hazards assumptions

In response to the Committee's concerns about the violation of PH in some of the covariates included in the company's original regression model, the company provided an assessment of PH in the updated analysis for the stepwise regression models using Grambsch and Therneau tests. These tests showed that there were still violations of the PH assumption, for PFS in the two key subgroups, so the company used a piece-wise Cox model to allow a time-varying HR for covariates where PH was violated. This showed that the treatment effect was fairly robust to this violation. However, the ERG notes that a piece-wise model was not tested for OS, despite a very low p-value in the PH test for the B2-microglobulin covariate in the ENDEAVOR model population. Although it may not be rejected at a 5% threshold, this covariate shows some evidence that the PH assumption could be unsuitable and therefore the ERG

considers this an uncertainty that could have been assessed further. The coefficient of this covariate was fairly large (2.926) and so uncertainty in the PH assumption could have an important impact on the treatment effect estimate.

1.3. Survival modelling and extrapolation

The Committee had concerns around the company's survival modelling approach, which used the HR from a covariate-adjusted Cox PH model using the full trial population, and applied it to a parametric survival curve fitted to the subgroups of interest. The Committee's key concerns were in the assumption that the treatment effect is assumed to be constant over the entire time horizon, and the lack of consideration given to alternative parametric survival curves to assess the plausibility and the impact of the extrapolation on the results. In addition to this, the Committee wanted to see the impact of using different combinations of covariates used to adjust the treatment effects in the subgroups of interest.

In response to the Committee's requests, the company produced jointly fitted parametric regression models using the Weibull, exponential, Gompertz, generalized Gamma, lognormal and loglogistic distributions for comparison. The company assessed model fit using the Akaike information criteria (AIC) and Bayesian information criteria (BIC) as well as assessing the plausibility of the extrapolated projection of the resulting curves. The Weibull was considered to be the most appropriate for both PFS and OS for the key subgroups in the ENDEAVOR and ASPIRE trials. The company also incorporated covariate adjustment into these models and considered three different approaches for this: the inverse probability weighting (IPW); corrected group prognosis (CGP); and the mean of covariates (MoC) methods. The company concluded that the CGP and MoC methods did not result in clinically plausible projections of OS for either the ENDEAVOR or ASPIRE subgroups of interest and therefore chose to use the IPW method. The ERG notes that the analyses provided by the company are based on the IPW method only and so it could make no assessment of the impact of using different methods could be made. However, the ERG notes the discussion provided in Table 14 of the company's response to the Appraisal Consultation Document (ACD) and considers the IPW method to be reasonable. The ERG, has a concern, however, that these analyses use the stepwise variable selection, which may not be the most appropriate, and consideration of the other methods could have been explored further.

The ERG disagrees that the chosen Weibull parametric survival curve for OS in the ENDEAVOR model is the best fit after assessing the AIC and BIC. These statistics suggest that the best fitting curve is the Gompertz, which has the lowest AIC and second lowest BIC, while the Weibull curve has the second lowest AIC and third lowest BIC. In the ERG's opinion, the visual fit of the Gompertz curve is also

more plausible than that of the Weibull curve. This can be seen when compared to the long term survival estimates from the bortezomib group of the PANORAMA-1 trial given in Figure 3, which shows the survival at 5 years to be around 30%. This appears to be closer to the Gompertz curve shown in Figure 2 than the company's Weibull curve shown in Figure 1.

Figure 1. Company's Weibull OS curves for the ENDEAVOR (Company's revised economic model in response to the ACD)

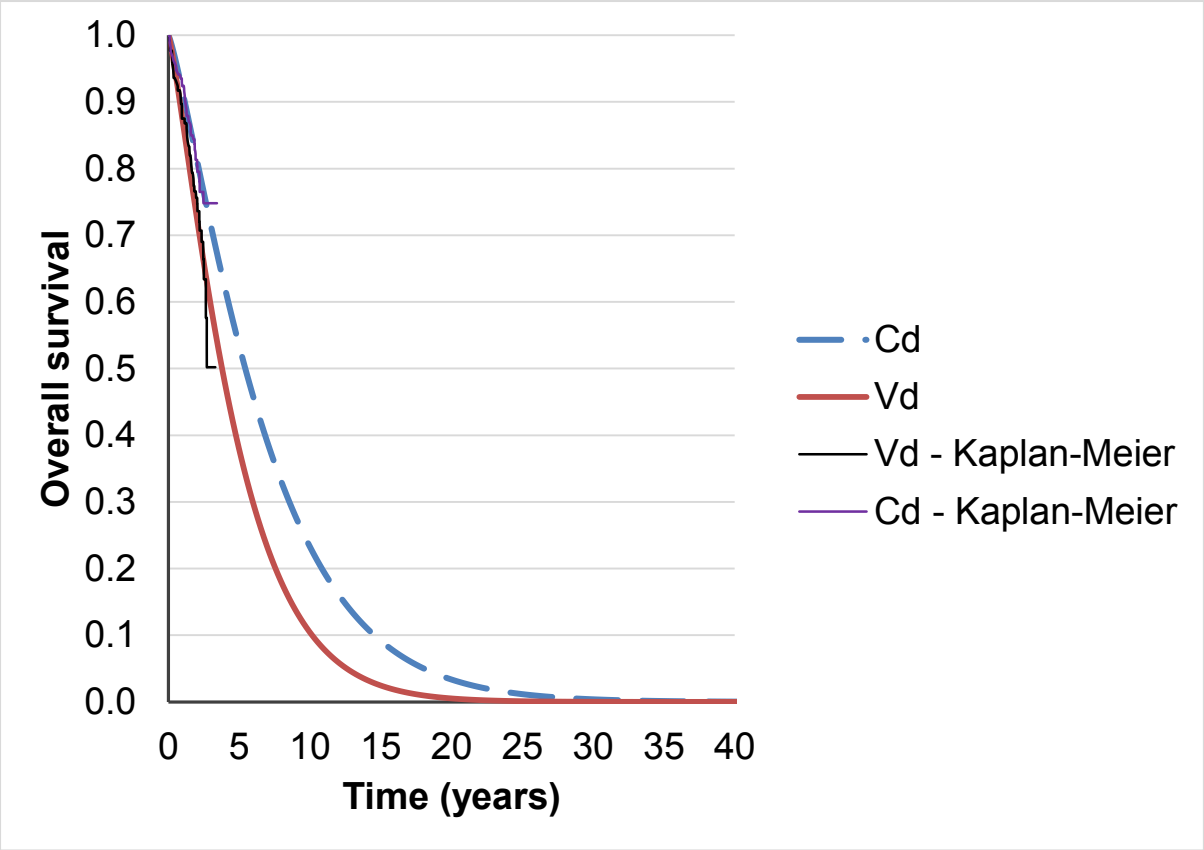


Figure 2. ERG's suggested Gompertz OS curves for the ENDEAVOR model (Company's revised economic model in response to the ACD)

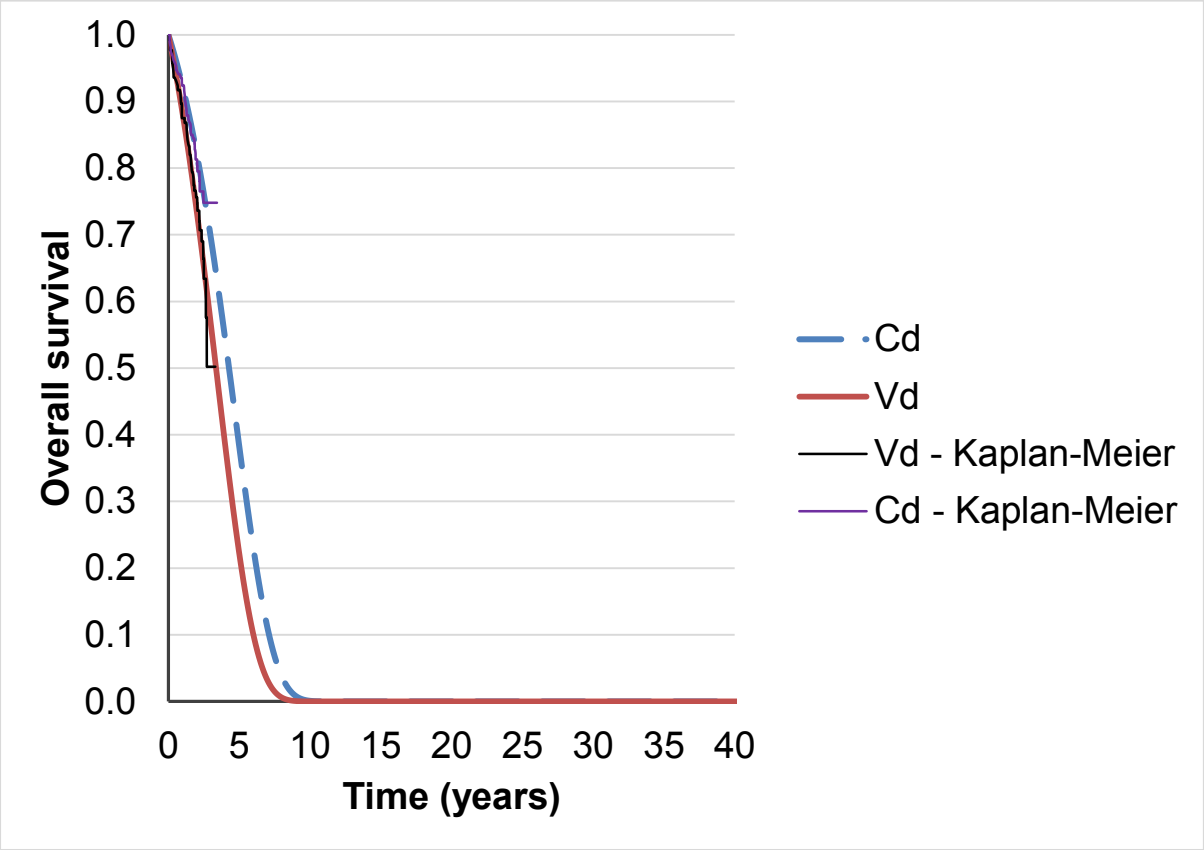
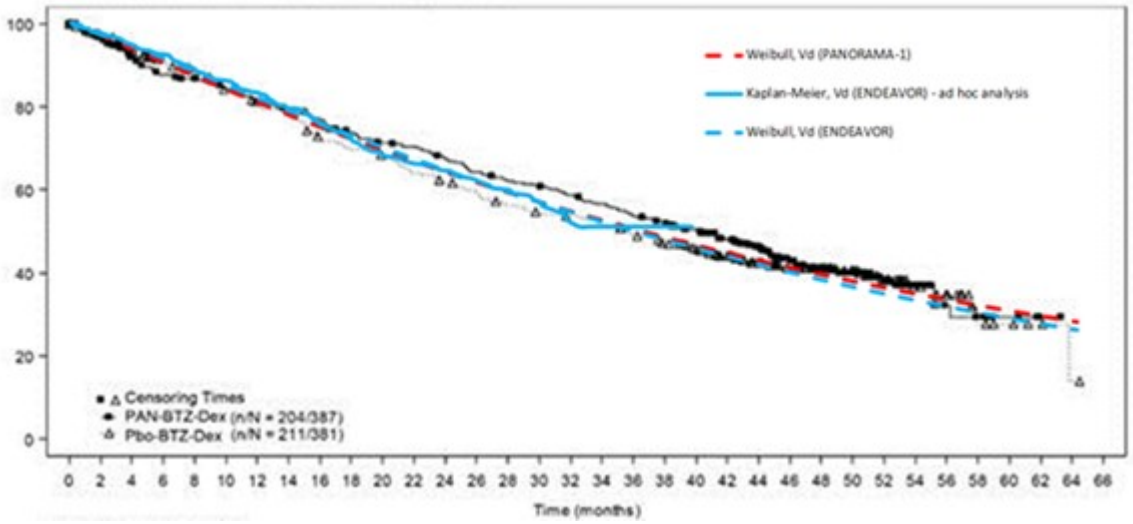


Figure 3. OS curves for Vd in ENDEAVOR versus PANORAMA-1 (Company's response to the ACD, page 34, Figure 3)



For the ASPIRE model, the company also chose to use the Weibull curve for PFS and OS despite the exponential curve showing the lowest AIC and BIC for both PFS and OS. Given that the AIC and BIC statistics are similar, this approach is reasonable if the extrapolation provided is more plausible. However, the ERG considers the Gompertz curve to have a more plausible extrapolation for OS in the ASPIRE model, and this corroborates with previously sought expert opinion regarding OS extrapolation for this population.

A key concern the ERG has for the ASPIRE model is that the log-log plots for both PFS and OS appear to indicate that a PH assumption is not valid and that the HR actually diminishes over time. This means that the difference in OS will be overestimated in a PH based analysis as time moves on and this would cause the ICER to be overestimated. The loglogistic model would be an option to introduce a diminishing HR, however, the extrapolation of the curves is implausibly long and unlikely to reflect long term OS. The results for this analysis should be interpreted with caution.

2. Derived health state utility values

The Committee considered the health state utility values (HSUVs) based on the company's mapping analysis using the EORTC scores from the ASPIRE trial to be more plausible than the approach taken in the company's original submission, which used a mixed approach to estimating HSUVs. The company agreed to this change, which was in line with the ERG's original base case, and it has been incorporated into the company's revised base case results shown in Section 5.

3. The length of treatment and dosing schedule with bortezomib

The company's original submission applied the costs of bortezomib to treatment progression to reflect what occurred in the trials. However, the marketing authorisation (MA) for bortezomib states that it should be stopped after 8 cycles (24 weeks) of treatment. The costs in the company's model are therefore inflated compared to the MA and current UK practice. To address the Committee's concerns about this, the company conducted a matched-adjusted indirect comparison (MAIC) of the bortezomib arm of ENDEAVOR in comparison to the bortezomib arm of the CASTOR trial, which stopped treatment after 8 cycles as per the MA. This MAIC aimed to provide an estimate of the relative effect beyond 8 cycles for the discontinued treatment compared to treatment until progression. The company considered this to be an important adjustment if the costing approach was adjusted to reflect UK practice.

Individual patient data (IPD) from ENDEAVOR was adjusted to aggregate data from CASTOR to minimise confounding associated with differences in patient characteristics across the studies. The IPD from ENDEAVOR was adjusted for all variables reported in both the ENDEAVOR and CASTOR trials which were considered to be prognostic factors for PFS and OS by clinicians. The only variable which was captured in both trials that was not adjusted for was cytogenetic status. This was due to missing data in ENDEAVOR (15.5% of the ITT population). The company did not give any further details to support the choice of variables to adjust for, or provide any evidence that any residual differences were accounted for. It is highly unlikely that all prognostic variables would have been measured and so the results are likely to be affected by residual bias due to unobserved prognostic variables or undocumented effect modifiers.

The ERG agrees with the company's approach to adjust for all possible prognostic factors, however, the justification for the choice of variables was limited and residual bias was not accounted for. Also, according to the ERG's clinical experts, cytogenetic status is an important prognostic factors in MM. Hence, cytogenetic status should have been adjusted for despite the resulting reduction in sample size.

Using the aggregate data for bortezomib in CASTOR and the matched (weighted) IPD for bortezomib in ENDEAVOR, the company estimated HRs for the first 8 cycles and for post-8 cycles. The company then calculated HRs for first 8 cycles versus subsequent cycles to obtain HRs that reflect the likely effect of prolonged treatment with bortezomib beyond 8 cycles on PFS and OS.

The ERG notes that despite adjusting for all available prognostic factors the HRs for both PFS and OS for the first 8 cycles of treatment show longer survival in ENDEAVOR than in CASTOR with PFS being statistically significantly different between the bortezomib groups in the two trials (Table 1). It is the ERG's view that the differences observed in the first 8 cycles indicate unobserved prognostic or treatment modifying effects that have not been accounted for. The ERG also notes that the relative increase in OS and PFS HR when stopping bortezomib after 8 cycles is not statistically significant.

Hence it is likely the adjustments proposed by the company for the post-8 cycle period are unreliable.

Table 1. Progression-free survival and overall survival hazard ratios (first 24 weeks vs. after 24 weeks) from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity) Reproduced from the company ACD response (Table 23)

	PFS (Weibull model)	**OS (2-arm Weibull model)**
Piecewise matching-adjusted (first 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.344 (1.021 - 1. 769)	1.157 (0.679 – 1.972)
Piecewise matching-adjusted (after 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.828 (1.311 - 2.549)	1.561 (0.942 – 2.588)
Relative hazard increase when stopping bortezomib after 24 weeks, HR (95% CI) CASTOR:ENDEAVOR[a]	**1.360 (0.913 - 2.027)**	**1.349 (0.684 - 2.662)**

Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts (Section 2.1) reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR)

[a] HR for first 8 cycles (24 weeks) versus subsequent cycles

CI, confidence interval; HR, hazard ratio; PFS, progression-free survival; OS, overall survival; Vd, bortezomib/dexamethasone

The ERG agrees in theory that the impact of treating with bortezomib until progression would need to be adjusted for if the costing is adjusted to only consider 8 cycles of treatment, otherwise the outcomes for the bortezomib group may be overestimated and the relative treatment effect of carfilzomib versus bortezomib underestimated. However, the ERG is concerned with the MAIC approach taken by the company in order to estimate this adjustment in treatment effect. The ERG considers the results of the company's MAIC to be unreliable. Therefore, the ERG deems it more appropriate to assume no reduction in treatment effect for the 8 cycle restriction, to act as an upper bound for the ICER. This assumption was considered by the ERG's clinical experts to be a reasonable one.

4. Additional concerns

The Committee were concerned that the patient access scheme (PAS) for bortezomib had not been considered in the model, which meant that the company's model overestimated the cost of bortezomib and so the ICER would increase if the PAS was applied.

As the bortezomib PAS was complex, the company attempted to estimate a simple reduction in the cost of bortezomib when it was applied in TA171. This reduction in cost was estimated to be between 8.3% and 15% and the company provided the results of a scenario analysis using a 15% reduction in the cost of bortezomib in Section 5 of their response to the ACD. The ERG consider this to be a fairly reasonable approximation, but given that the proportion of patients who do not respond to treatment within four cycles is at least 18% based on the proportion of patients who did not progress at any time, it may be an underestimation of the discount (ENDEAVOR CSR, Page 120, Table 23). Excluding those patients who were unable to be evaluated, the proportion of patients who did not respond is approximately 20%. Assuming that these patients incur no cost and the remaining 80% incur the cost of the full 8 cycles, then a 20% discount may be a closer approximation. The ERG produced the results of this analysis in Section 6. A further consideration is that those who progressed in the trial may have progressed after four cycles, and so in practice, these patients would not receive further treatment and the cost of the four cycles would not incur a cost.

5. Company's revised cost-effectiveness analyses and justification

In response to the ACD, the company revised its cost-effectiveness model by making two changes to both the ENDEAVOR and the ASPIRE models. The first was to update the survival curves to the joint parametric covariate-adjusted regression models, and the second was to change the HSUVs to those used in the ERG's original base case. The results of the two models are given in Table 2 and Table 3, respectively, and these show the impact on the ICER that each change has on the original base case as well as the overall ICER when all changes are applied.

Table 2. Revised cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib (Company's response to the ACD, Table 29)

Original base case	**Cd**	**Vd**	**Incremental value**
Total costs	£121,891	£95,213	£26,678
Total LYG	6.05	4.26	1.79
QALYs	4.28	2.95	1.33
ICER	-	-	£20,044
Using the updated survival model			
Total costs	£117,660	£93,769	£23,891
Total LYG	5.74	4.23	1.51
QALYs	4.09	2.94	1.15
ICER (compared to original base case)	-	-	£20,766
Using directly mapped utilities			
Total costs	£117,660	£93,769	£23,891

Original base case	Cd	Vd	Incremental value
Total LYG	5.74	4.23	1.51
QALYs	3.88	2.79	1.09
ICER (compared to original base case)	-	-	£21,137
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£22,009**

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

Table 3. Revised cost-effectiveness results for CRd versus Rd in patients with two prior therapies and no prior lenalidomide (Company's response to the ACD, Table 30)

	CRd	Rd	Incremental value
Total costs	£128,654	£95,420	£33,234
Total LYG	6.31	4.93	1.37
QALYs	4.32	3.33	0.99
ICER	-	-	£33,467
Using the updated survival model			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.77	2.95	0.79
ICER (compared to original base case)	-	-	£40,198
Using directly mapped utilities			
Total costs	£127,140	£94,528	£32,612
Total LYG	5.46	4.36	1.10
QALYs	3.67	2.88	0.79
ICER (compared to original base case)	-	-	£34,404
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£41,429**

Abbreviations used in the table: CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; Rd, lenalidomide/dexamethasone; QALYs, quality-adjusted life years

6. ERG revised cost-effectiveness analysis and justification

The ERG considered the company's updated survival model for OS to produce implausible extrapolations as well as not being the best statistical fit. The ERG therefore chose to alter the

distribution to ones that appeared more plausible and better fitting statistically. The ERG's chosen parametric curve was the Gompertz distribution for both the ENDEAVOR and ASPIRE models. The ERG also considered the company's model to lack relevance to clinical practice given that the MA for bortezomib had not been applied to the model. The ERG has concerns around the appropriateness of the adjustments the company has proposed in its MAIC to account for the impact a reduced number of cycles of treatment with bortezomib has on PFS and OS. In addition, the results of the MAIC demonstrate no statistically significant difference in OS. The ERG, therefore, considers a conservative approach to be no change in PFS or OS as the company did not present robust evidence for the adjustments. This assumption was considered reasonable by the ERG's clinical experts. The company's estimation of the potential bortezomib discount resulting from the complex PAS was deemed to be fairly reasonable and so the ERG applied this to their base case also. However, the ERG considers that this discount may be an underestimate as discussed previously in Section 4.

The results of the respective models are given in Table 4 and Table 5, each showing the result of the cumulative changes to the model as well as the ICER when each change is applied separately to the original base case.

6.1. Base case

Table 4. ERG revised base case ICER (ENDEAVOR)

Results per patient	Cd	Vd	Incremental value
Company's revised base case			
Total costs (£)	£117,660	£93,769	£23,891
QALYs	3.88	2.79	1.09
ICER			£22,009
Using the Gompertz distribution for OS			
Total costs (£)	£108,436	£90,814	£17,622
QALYs	2.70	2.13	0.57
ICER (compared with base case)			£30,697
ICER with all changes incorporated			£30,697
Restricting the cost of bortezomib to 8 treatment cycles			
Total costs (£)	£108,436	£73,789	£34,647
QALYs	2.70	2.13	0.57
ICER (compared with base case)			£37,694
ICER with all changes incorporated			£60,357
Using a discount of 15% for the cost of bortezomib			
Total costs (£)	£108,436	£71,512	£36,924
QALYs	2.70	2.13	0.57

Results per patient	Cd	Vd	Incremental value
ICER (compared with base case)			£26,306
ICER with all changes incorporated			£64,325
ERG preferred base case ICER			**£64,325**

Abbreviation used in the table: Cd, carfilzomib and dexamethasone; EQ-5D, European Quality of Life 5 dimensions; HSUV, healthy state utility value; ICER, incremental cost effectiveness ratio; MA, market authorisation; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life-years; Vd, bortezomib and dexamethasone.

Table 5. ERG revised base case ICER (ASPIRE)

Results per patient	CRd	Rd	Incremental value
Company's revised base case			
Total costs (£)	£127,140	£94,528	£32,612
QALYs	3.67	2.88	0.79
ICER			£41,429
Using the Gompertz distribution for OS			
Total costs (£)	£122,944	£92,263	£30,681
QALYs	3.15	2.56	0.59
ICER (compared with base case)			£52,439
ICER with all changes incorporated			£52,439
ERG preferred base case ICER			**£52,439**

Abbreviation used in the table: CRd, carfilzomib, lenalidomide and dexamethasone; EQ-5D, European Quality of Life 5 dimensions; HSUV, healthy state utility value; ICER, incremental cost effectiveness ratio; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life-years; Rd, lenalidomide and dexamethasone.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal (STA)

Carfilzomib for treating multiple myeloma in people who have received at least one prior therapy

Response to additional queries raised by the ERG on Amgen's response to the ACD

Prepared by:

Date: February 2017

File name	Version	Contains confidential information	Date
	1.0 (without PAS)	No	7 February 2017

Contents

List of tables

List of figures

1 Additional queries raised by the ERG

1.1 Context

Following our response to the NICE Appraisal Consultation Document (ACD), the Evidence Review Group (ERG) requested further information relating to analyses conducted by Amgen in response to the following Committee request in Section 4.12 of the ACD (analyses presented in Section 2 of the ACD response):

'Plausible efficacy estimates for all comparisons, adjusted by covariates, including a treatment effect, and to explore the plausibility of different combinations of covariates on the efficacy estimates. The covariates to adjust the model, presented with a rationale for why they had been chosen.'

Specifically, the ERG requested:

'In your ACD response you provided different methods for selecting the covariates and provided analyses–incorporating these with another statistical method; but you have not provided confidence intervals for individual variables or statistical fit data for the LASSO model in appendix A in the response to the ACD. Neither have you explored the uncertainty and impact on the ICER of incorporating other covariate selections, particularly the LASSO method.

Please can we request additional information relating to the other covariate selection methods and data around the uncertainty in the ICERs by providing an additional analysis in which you fit adjusted parametric survival curves (standard distributions as previously) using the LASSO method, and incorporate the resulting curves in to your economic model. If possible, can these be fit independently and dependently and to explore any potential areas of non-proportionality. This would help inform the ERG review of your ACD response analysis and best inform committee members before the next appraisal committee meeting on 15 February 2017.'

Additional analyses to address these ERG requests for additional information have been conducted, as described below in Section 1.2.

1.2 Implementation

As highlighted in email correspondence with NICE relating to the above ERG request, to the best of our knowledge there are no statistical packages available that would enable an analysis involving fitting adjusted parametric survival curves using the least absolute shrinkage and selection operator (LASSO) variable selection method, and consequently such analyses were considered unfeasible. The ERG subsequently acknowledged the difficulties in applying the LASSO method in this way, and NICE accepted this rationale for why such an analysis is not technically feasible. In addition, the LASSO R statistical package used to run the LASSO variable selection model does not include functionality to estimate statistical fit values (e.g. AIC and BIC), and we are consequently unable to provide these. A summary of the additional analyses included in this response is provided in Table 1.

Table 1. Summary of additional analyses provided in this response to the ERG – both Cd versus Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib) and CRd versus Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Additional analyses exploring the uncertainty associated with the choice of covariate selection method to inform the survival model and resulting ICERs	• Confidence intervals for the PFS and OS hazard ratios for each individual variable retained within the different covariate selection models included in our response to the ACD • Summary of resulting ICERs across standard parametric distributions when the different covariate selection methods are used to inform the IPW-based survival model: ▪ Subset of UK clinician-identified prognostic covariates selected using the stepwise selection method (per the revised base case cost-effectiveness analyses) ▪ All UK clinician-identified prognostic covariates ▪ Subset of UK clinician-identified prognostic covariates selected using the LASSO method
Additional analyses exploring the uncertainty associated with method of survival curve fitting (joint vs. independent) and the non-proportionality of hazards for the revised base case cost-effectiveness analyses presented our response to the ACD (IPW-based Weibull model with stepwise covariate selection)	• Comparison of PFS and OS curves derived from: ▪ IPW-based Weibull model jointly fitted to both study arms with covariates included in the model identified using the stepwise selection method (per the revised base case cost-effectiveness analyses) ▪ IPW-based Weibull model independently fitted to the Cd and Vd arms with covariates included in the model identified using the stepwise selection method • Summary of resulting ICERs from the jointly vs. independently fitted models
Additional analyses repeating the above joint versus independent survival curve fitting for the other covariate selection methods and standard parametric distributions	Summary of resulting ICERs from the jointly (per the revised base case cost-effectiveness analyses) vs. independently fitted models when the other covariate selection methods and other standard parametric distributions are used to inform the IPW-based survival model: • All UK clinician-identified prognostic covariates • Subset of UK clinician-identified prognostic covariates selected using the LASSO method
ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; ERG, evidence review group' ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival.	

It should be noted that all of these additional analyses incorporate the revised utility values (derived by mapping utilities directly from the trial data using the Proskorovsky *et al.,* 2014 mapping algorithm[1]) included in the revised base case cost-effectiveness analyses presented in our ACD response (see Section 3 of the ACD response for details). Furthermore, all cost-effectiveness results are inclusive of the simple patient access scheme (PAS) discount for carfilzomib.

1.3 Additional analyses requested by the ERG – Cd versus Vd (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

1.3.1 Additional analyses exploring uncertainty with the choice of covariate selection method to inform the survival model

Confidence intervals for the progression-free survival (PFS) and overall survival (OS) hazard ratios for each individual variable retained within the different covariate selection models described in our ACD response are provided in Table 2 (PFS) and Table 3 (OS).

Table 2. Variables retained within the different covariate selection models and associated hazard ratios: Cd vs. Vd PFS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Covariate	Stepwise variable selection model, HR (95% CI)	All UK clinician-identified covariates model, HR (95% CI)	LASSO variable selection model, HR (95% CI)	Cox model using LASSO variables, HR (95% CI)
Treatment (Cd vs. Vd)	0.408 (0.267, 0.624)	0.362 (0.231, 0.567)	0.397 (0.272, 0.692)	0.366 (0.235, 0.572)
Prior SCT (yes vs. no)[a]	-	0.937 (0.542, 1.618)	0.959 (0.611, 1.354)	0.957 (0.565, 1.621)
Prior lenalidomide (yes vs. no)[a]	-	0.627[b] (0.371, 1.059)	0.682 (0.414, 1.000)	0.623 (0.369, 1.051)
Age (≥ 65 vs. < 65 years)	-	1.209 (0.738, 1.979)	1.152 (0.800, 1.781)	1.226 (0.752, 1.996)
ECOG PS (1-2 vs. 0)	-	1.310 (0.843,2.036)	1.213 (0.944, 1.768)	1.279 (0.829, 1.974)
Creatinine clearance (≥50 to <80 mL/min vs. other)	-	0.793 (0.430, 1.462)	-	-
Creatinine clearance (>80 mL/min vs. other)	-	0.670 (0.334, 1.344)	0.847 (0.554, 1.064)	0.823 (0.509, 1.332)
Time from diagnosis	0.990 (0.983, 0.997)	0.991 (0.982, 0.999)	0.992 (0.983,1.000)	0.991 (0.983, 0.999)
Time from last relapse	-	1.016 (0.984, 1.049)	1.009 (0.984, 1.038)	1.016 (0.985, 1.047)
ISS stage (II-III vs. I)	2.455 (1.590, 3.792)	2.44 (1.299, 4.581)	2.162 (1.017, 4.244)	2.289 (1.437, 3.646)
β_2-microglobulin (≥3.5 vs. <3.5 mg/L)	-	0.88 (0.463, 1.671)	-	-
Refractory to last prior treatment (yes vs. no)	-	1.303 (0.799, 2.124)	1.249 (0.967, 1.991)	1.281 (0.789, 2.078)
AIC	**883.822**	**893.576**	**-[c]**	**890.208**

Note: PFS data are from the prespecified interim analysis (10 November 2014 data cut-off date). 95% CIs were estimated by bootstrapping. If the variable was shrunk to zero for a bootstrap sample, the HR for that sample was 1.

[a] The rows for prior lenalidomide and prior SCT were incorrectly labelled the wrong way round in Appendix A (Table 32) of the ACD response.
[b] Incorrectly stated to be 0.626 in Appendix A (Table 32) of the ACD response.
[c] The LASSO R statistical package used to run this model did not include functionality to estimate statistical fit values.

ACD, appraisal consultation document; AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; HR, hazard ratio; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

Table 3. Variables retained within the different covariate selection models and associated hazard ratios: Cd vs. Vd OS model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Covariate	Stepwise variable selection model, HR (95% CI)	All UK clinician-identified covariates model, HR (95% CI)	LASSO variable selection model, HR (95% CI)	Cox model using LASSO variables, HR (95% CI)
Treatment (Cd vs. Vd)	0.631 (0.384, 1.039)	0.618 (0.373, 1.024)	0.661 (0.388, 1.000)	0.621 (0.377, 1.023)
Prior SCT (yes vs. no)[a]	2.034 (1.125, 3.678)	1.943 (1.034, 3.651)	1.615 (1.000, 3.108)	1.953 (1.067, 3.575)
Prior lenalidomide (yes vs. no)[a]	-	0.701 (0.386, 1.273)	0.801 (0.433, 1.193)	0.723 (0.402, 1.300)
Age (≥ 65 vs. < 65 years)	-	0.996 (0.574, 1.731)	-	-
ECOG PS (1-2 vs. 0)	2.154 (1.265, 3.665)	2.218 (1.290, 3.813)	2.008 (1.000, 3.654)	2.170 (1.270, 3.707)
Creatinine clearance (≥50 to <80 mL/min vs. other)	-	0.783 (0.410, 1.495)	-	-
Creatinine clearance (>80 mL/min vs. other)	0.496 (0.264, 0.931)	0.386 (0.163, 0.916)	0.520 (0.264, 0.943)	0.473 (0.246, 0.909)
Time from diagnosis	0.986 (0.976, 0.997)	0.987 (0.976, 0.997)	0.989 (0.974, 1.000)	0.987 (0.976, 0.997)
Time from last relapse	-	0.995 (0.952, 1.041)	0.998 (0.958, 1.018)	0.996 (0.953, 1.042)
ISS stage (II-III vs. I)	-	1.356 (0.595, 3.093)	1.352 (0.832, 2.964)	1.370 (0.601, 3.123)
β_2-microglobulin (≥3.5 vs. <3.5 mg/L)	2.926 (1.634, 5.238)	2.186 (0.927, 5.154)	2.081 (1.000, 4.531)	2.263 (0.970, 5.277)
Refractory to last prior treatment (yes vs. no)	2.010	2.094	1.862	2.078

	(1.158, 3.487)	(1.189, 3.688)	(1.000, 3.332)	(1.182, 3.654)
AIC	**640.675**	**648.1524**	**-[b]**	**644.6873**

Note: OS data are from the EMA ad hoc-analysis (3 March 2016 data cut-off date). 95% CIs were estimated by bootstrapping. If the variable was shrunk to zero for a bootstrap sample, the HR for that sample was 1.

[a] The rows for prior lenalidomide and prior SCT were incorrectly labelled the wrong way round in Appendix A (Table 33) of the ACD response.
[b] The LASSO R statistical package used to run this model did not include functionality to estimate statistical fit values.

ACD, appraisal consultation document; AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; EMA, European Medicines Agency; HR, hazard ratio; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; OS, overall survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

We explored the uncertainty associated with method of covariate selection and impact on the incremental cost-effectiveness ratio (ICER) for Cd versus Vd by incorporating other covariate selection methods in the IPW-based Weibull model as per the revised base case analysis presented in our ACD response (Table 4). The ICERs for Cd versus Vd are highly consistent irrespective of what covariate selection method is used:

- Stepwise selection method as per our revised base case analysis (£22,009 per quality-adjusted life year [QALY] gained)
- All UK clinician-identified prognostic covariates (£22,441 per QALY gained)
- LASSO variable selection method (£22,481 per QALY gained)

Table 4. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection for inclusion in the IPW-based Weibull survival model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variable selection model**
Cd versus Vd PFS and OS hazard ratios (95% CIs)[a]			
PFS[b]	0.408 (0.267, 0.624)	0.362 (0.231, 0.567)	0.397 (0.272, 0.692)
OS[b]	0.631 (0.384, 1.039)	0.618 (0.373, 1.024)	0.661 (0.388, 1.000)
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – joint curve fitting (as per the revised base case cost-effectiveness analysis)[c]			
IPW-based Weibull model	**£22,009 (revised base case)[d]**	£22,441	£22,481

[a] HRs were derived from a covariate-adjusted analysis within the relevant subgroup (patients with one prior therapy and no prior bortezomib).
[b] PFS data are from the prespecified interim analysis (10 November 2014 data cut-off date). OS data were from the EMA ad-hoc analysis (3 March 2016 data cut-off date).
[c] ICERs were derived using an IPW-based model with a Weibull distribution for both PFS and OS as per our revised base case analysis.
[d] Revised base case ICER submitted in response to the ACD.

ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; EMA, European Medicines Agency; ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; ITT, intent to treat; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.

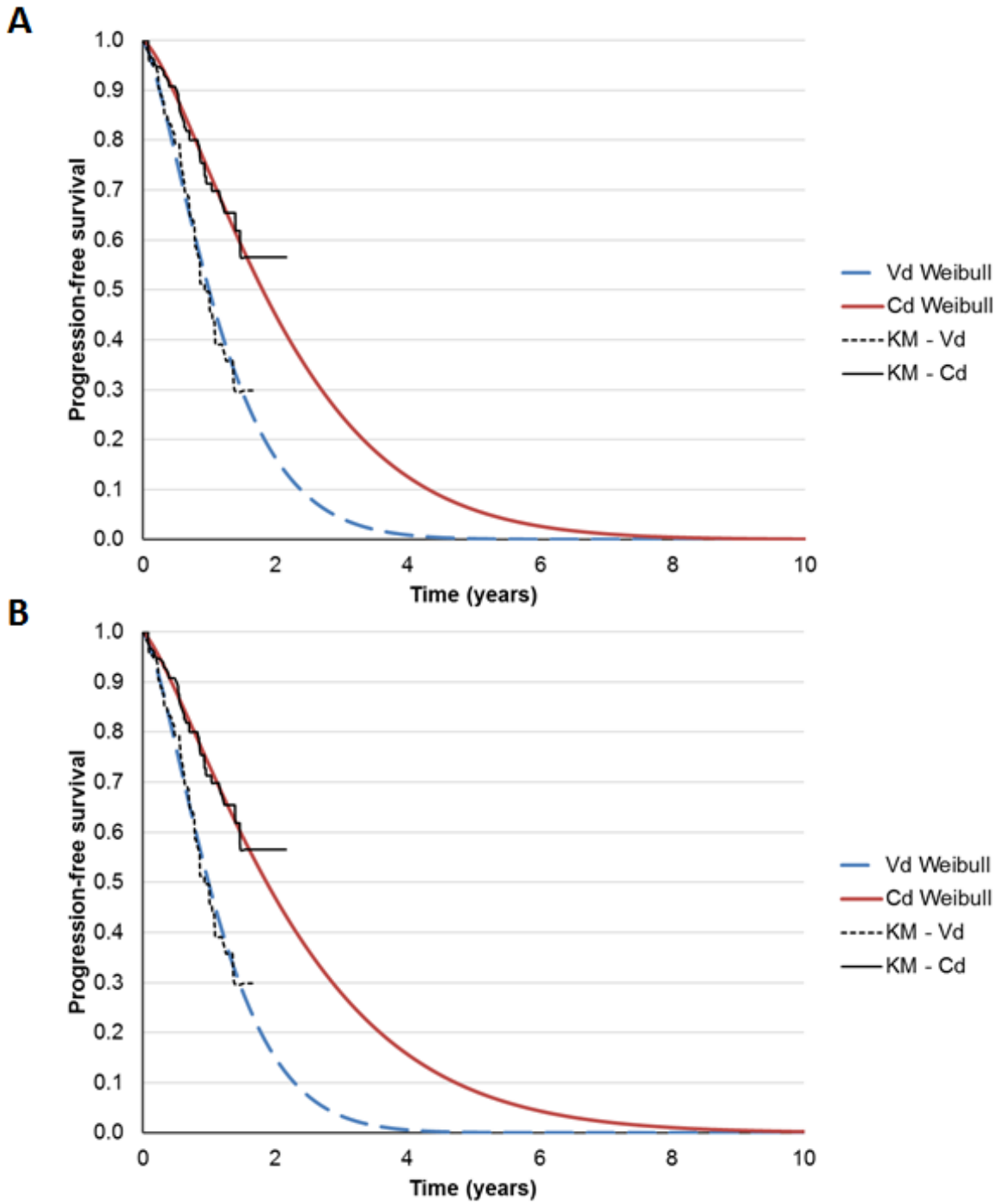
Further, the ICERs are similarly highly consistent (also within a margin of £500/QALY gained) for the different covariate selection methods across all other standard parametric distributions for OS (Appendix A). It is also pertinent to note that all of the covariate selection methods result in similar ICERs to our original base case survival model approach (Cox proportional hazards Weibull model), which yields an ICER of £21,137 per QALY gained.

These results suggest that the choice of method for selecting covariates for inclusion in the IPW-based survival model has a negligible impact on the cost effectiveness of Cd versus Vd, and underlines the robustness of the cost-effectiveness results from both the original base case survival model reported in the initial company submission and revised base case survival model reported in our ACD response.

1.3.2 Additional analyses exploring the uncertainty associated with method of curve fitting (joint vs. independent) and the non-proportionality of hazards in the IPW-based model (stepwise selection covariates) used in our revised base case cost-effectiveness analyses

To further explore the uncertainty associated with the method of curve fitting and potential non-proportionality of hazards in the IPW-based survival model, and impact on the ICER for Cd versus Vd, we applied joint (as per our revised base case analysis) and independent curve fits for Cd and Vd within the cost-effectiveness model. A comparison of jointly and independently fitted PFS and OS curves for the IPW-based Weibull model (using the covariates identified through the stepwise selection method as per our revised base case analysis) are provided below in Figure 1 (PFS) and Figure 2 (OS).

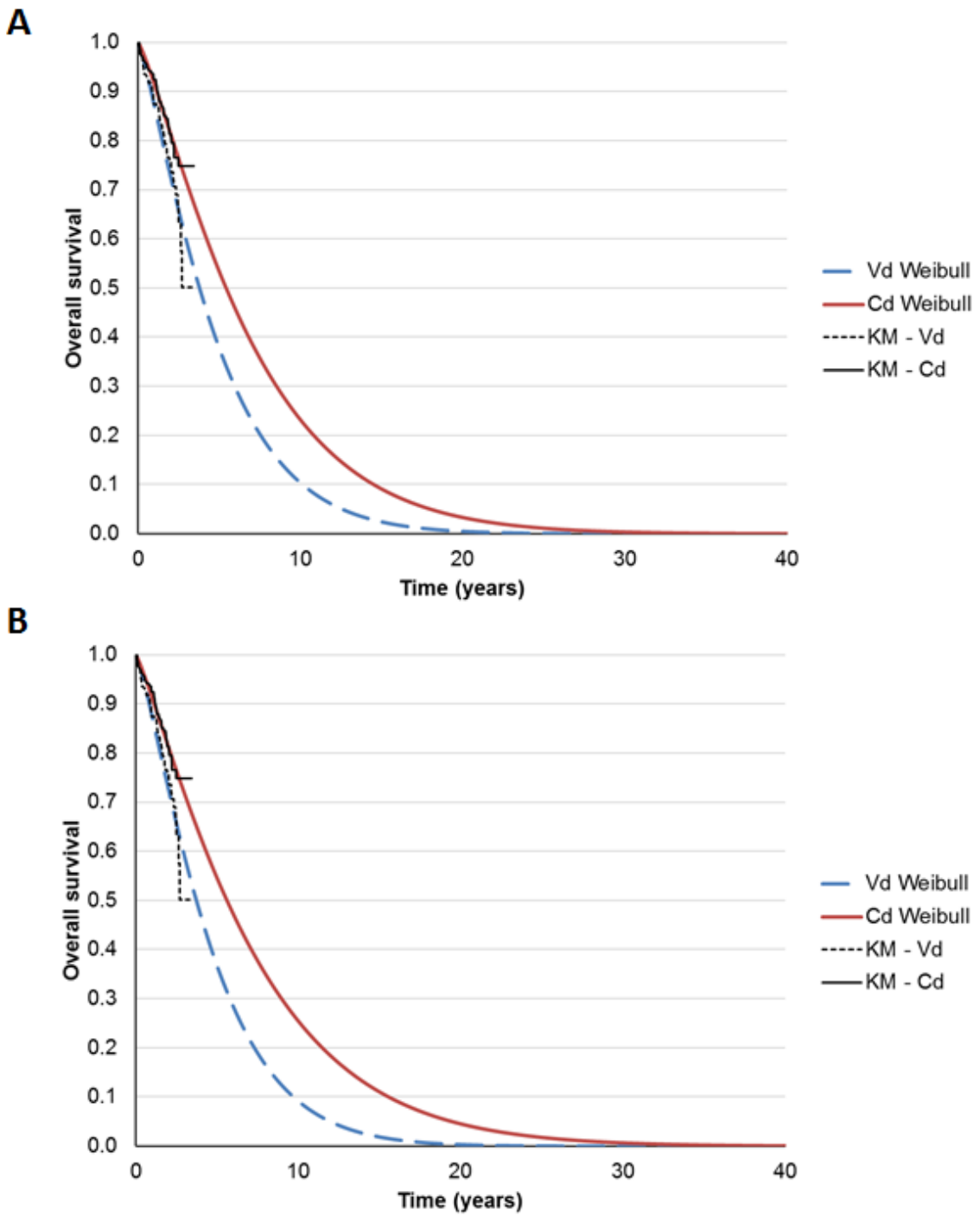
Figure 1. PFS Kaplan-Meier curves and IPW-based Weibull model (stepwise selection covariates) curves fitted (A) jointly and (B) independently to the Cd and Vd arms (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



Note: PFS data are from the second interim analysis (10 November 2014 data cut-off date). Covariates included in the IPW-based Weibull model jointly fitted to both study arms within the relevant subgroup were identified through a stepwise selection procedure within the relevant subgroup: treatment (Cd vs. Vd), time from diagnosis (continuous variable), and ISS stage (II-III vs. I).

Cd, carfilzomib/dexamethasone; IPW, inverse probability weighted; ISS, International Staging System; KM, Kaplan-Meier; PFS, progression-free survival; Vd, bortezomib/dexamethasone.

Figure 2. OS Kaplan-Meier curves and IPW-based Weibull model (stepwise selection covariates) curves fitted (A) jointly and (B) independently to the Cd and Vd arms (ENDEAVOR, patients with one prior therapy and no prior bortezomib)



Note: OS data are from the EMA ad hoc-analysis (3 March 2016 data cut-off date). Covariates included in the IPW-based Weibull model jointly fitted to both study arms within the relevant subgroup were identified through a stepwise selection procedure within the relevant subgroup: treatment (Cd. vs Vd), prior SCT (yes vs. no), ECOG PS (1 to 2 vs. 0), creatinine clearance (≥80 mL/min vs. other), time from diagnosis (continuous variable), β2-microglobulin (≥3.5 vs. <3.5 mg/L), and refractory to last prior treatment (yes vs. no).

Cd, carfilzomib/dexamethasone; ECOG PS, Eastern Cooperative Oncology Group performance status; IPW, inverse probability weighted; KM, Kaplan-Meier; OS, overall survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

The resulting ICERs are generally consistent but marginally more favourable using the independently fitted curves (£18,775 per QALY gained) than the jointly fitted curves in which proportional hazards is assumed (£22,009 per QALY gained) as per our revised base case analysis.

These results suggest that the method of curve fitting and potential for non-proportionality of hazards has a negligible impact on the ICER for Cd versus Vd. This underlines the robustness of the cost-effectiveness results from the revised base case analysis reported in our ACD response.

1.3.3 Additional analyses exploring the uncertainty associated with method of curve fitting (joint vs. independent) for the other covariate selection methods and standard parametric distributions

A summary of the ICERs resulting from the joint (as per our revised base case analysis) versus independent curve fitting using the different covariate selection methods to select covariates for inclusion in the IPW-based Weibull model is provided in Table 5. Similar to previous results, the ICERs are generally consistent using both jointly and independently fitted curves (marginally more favourable using the independently fitted curves) irrespective of the method used to select covariates for inclusion in the model. When the other standard parametric distributions are used to extrapolate OS, the ICERs are also consistently more favourable when curves are independently fitted across the different methods for covariate selection (Appendix A).

Table 5. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection and curve fitting for the IPW-based Weibull survival model (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variable selection model**
Cd versus Vd PFS and OS hazard ratios (95% CIs)[a]			
PFS[b]	0.408 (0.267, 0.624)	0.362 (0.231, 0.567)	0.397 (0.272, 0.692)
OS[b]	0.631 (0.384, 1.039)	0.618 (0.373, 1.024)	0.661 (0.388, 1.000)
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based Weibull survival model – (as per revised the base case cost-effectiveness analysis)[c]			
PFS and OS curves jointly fitted	**£22,009 (revised base case)**[d]	£22,441	£22,481
PFS and OS curves independently fitted	£18,775	£20,506	£19,903

[a] HRs were derived from a covariate-adjusted analysis within the relevant subgroup (patients with one prior therapy and no prior bortezomib).
[b] PFS data are from the prespecified interim analysis (10 November 2014 data cut-off date). OS data were from the EMA ad-hoc analysis (3 March 2016 data cut-off date).
[c] ICERs were derived using an IPW-based model with a Weibull distribution for both PFS and OS as per our revised base case analysis.
[d] Revised base case ICER submitted in response to the ACD.

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; EMA, European Medicines Agency; ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.			

1.3.4 Conclusion

A summary of cost-effectiveness results exploring the impact of method of covariate selection (stepwise selection vs. LASSO vs. all UK clinician identified prognostic covariates) and curve fitting (joint vs. independent) on the revised base case analysis presented in our response to the ACD is provided in Table 6.

These analyses show that the impact of both covariate selection method and curve fitting method on the cost-effectiveness of Cd versus Vd is negligible. This underlines the robustness of our revised base case analysis and provides additional certainty that Cd represents a cost-effective use of NHS resources.

Table 6. Summary of ICERs for Cd versus Vd from the revised base case analysis and key additional analyses presented in this response (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Curve fitting method	Covariate selection method	ICER (£/QALY gained)
Revised base case cost-effectiveness analysis presented in our response to the ACD		
Joint	Stepwise	22,009
Key additional analyses presented in this response		
Joint	LASSO	22,481
Joint	All UK clinician identified prognostic covariates	22,441
Independent	Stepwise	18,775
Independent	LASSO	19,903
Independent	All UK clinician identified prognostic covariates	20,506
Note: An IPW-based survival model with a Weibull distribution for both PFS and OS was used for all analyses. ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.		

1.4 Additional analyses requested by the ERG - CRd versus Rd (ASPIRE, patients with two prior therapies and no prior lenalidomide)

1.4.1 Additional analyses exploring uncertainty with the choice of covariate selection method to inform the survival model

Confidence intervals for the PFS and OS hazard ratios for each individual variable retained within the different covariate selection models described in our response to the ACD are provided in Table 7 (PFS) and Table 8 (OS).

Table 7. Variables retained within different covariate selection models and associated hazard ratios: CRd vs. Rd PFS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Covariate	Stepwise variable selection model, HR (95% CI)	All UK clinician-identified covariates model, HR (95% CI)	LASSO variable selection model, HR (95% CI)	Cox model using LASSO variables, HR (95% CI)
Treatment (CRd vs. Rd)	0.699 (0.481, 1.015)	0.708 (0.478, 1.047)	0.776 (0.532, 1.000)	0.707 (0.482, 1.038)
Prior SCT (yes vs. no)[a]	1.448 (0.932, 2.250)	1.479 (0.947, 2.309)	1.153 (0.990, 2.004)	1.470 (0.946, 2.284)
Prior bortezomib (yes vs. no)[a]	-	1.025 (0.655, 1.604)	-	-
Age (≥ 65 vs. < 65 years)	-	1.171 (0.768, 1.785)	1.042 (0.845, 1.626)	1.216 (0.805, 1.838)
ECOG PS (1-2 vs. 0)	1.578 (1.065, 2.339)	1.566 (1.047, 2.341)	1.352 (1.000, 2.091)	1.570 (1.059, 2.328)
Creatinine clearance (≥50 to <80 mL/min vs. other)	-	0.599 (0.267, 1.341)	0.975 (0.753, 1.561)	0.768 (0.507, 1.165)
Creatinine clearance (>80 mL/min vs. other)	-	0.681 (0.295, 1.572)	-	-
Time from diagnosis	0.994 (0.988, 0.999)	0.993 (0.988, 0.999)	0.996 (0.988, 1.000)	0.993 (0.988, 0.999)
Time from last relapse	-	1.002 (0.974, 1.030)	-	-
ISS stage (II-III vs. I)	4.138 (1.964, 8.721)	4.034 (1.898, 8.572)	1.965 (1.282, 8.245)	2.408 (1.581, 3.669)
β_2-microglobulin (≥3.5 vs. <3.5 mg/L)	0.523 (0.257, 1.063)	0.536 (0.253, 1.135)	-	-
Refractory to last prior treatment (yes vs. no)	-	1.036 (0.661, 1.624)	-	-
AIC	**1076.913**	**1087.011**	**-[b]**	**1079.794**

Note: PFS data are from the prespecified interim analysis (16 June 2014 data cut-off date). 95% CIs were estimated by bootstrapping. If the variable was shrunk to zero for a bootstrap sample, the HR for that sample was 1.

[a] The rows for prior bortezomib and prior SCT were incorrectly labelled the wrong way round in Appendix A (Table 34) of the ACD response.
[b] The LASSO R statistical package used to run this model did not include functionality to estimate statistical fit values.

ACD, appraisal consultation document; AIC Akaike information criterion; CRd, carfilzomib/lenalidomide/dexamethasone; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; HR, hazard ratio; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; PFS, progression-free survival; Rd, lenalidomide/dexamethasone; SCT, stem cell transplant

Table 8. Variables retained within different covariate selection models and associated hazard ratios: CRd vs. Rd OS model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Covariate	Stepwise variable selection model, HR (95% CI)	All UK clinician-identified covariates model, HR (95% CI)	LASSO variable selection model, HR (95% CI)	Cox model using LASSO variables, HR (95% CI)
Treatment (CRd vs. Rd)	0.730 (0.454, 1.174)	0.783 (0.471, 1.299)	0.801 (0.503, 1.079)	0.783 (0.471, 1.299)
Prior SCT (yes vs. no)[a]	-	1.311 (0.735, 2.336)	1.200 (0.770, 2.243)	1.306 (0.746, 2.285)
Prior bortezomib (yes vs. no)[a]	0.569 (0.333, 0.973)	0.622 (0.356, 1.089)	0.651 (0.366, 1.000)	0.621 (0.356, 1.085)
Age (≥ 65 vs. < 65 years)	-	1.256 (0.745, 2.119)	1.190 (0.820, 2.096)	1.256 (0.745, 2.119)
ECOG PS (1-2 vs. 0)	-	1.015 (0.601, 1.713)	-	-
Creatinine clearance (≥50 to <80 mL/min vs. other)	0.296 (0.142, 0.616)	0.282 (0.133, 0.599)	0.367 (0.165, 1.000)	0.282 (0.133, 0.599)
Creatinine clearance (>80 mL/min vs. other)	0.298 (0.138, 0.645)	0.302 (0.137, 0.669)	0.403 (0.180, 1.000)	0.302 (0.137, 0.669)
Time from diagnosis	0.993 (0.986, 1.000)	0.992 (0.985, 1.000)	0.994 (0.984, 1.000)	0.992 (0.985, 1.000)
Time from last relapse	-	0.991 (0.954, 1.029)	0.994 (0.949, 1.009)	0.991 (0.954, 1.029)
ISS stage (II-III vs. I)	10.111 (4.263, 23.986)	10.521 (4.368, 25.343)	7.680 (2.162, 22.729)	10.543 4.390, 25.317)
β_2-microglobulin (≥3.5 vs. <3.5 mg/L)	0.382 (0.176, 0.829)	0.373 (0.170, 0.815)	0.503 (0.186, 1.000)	0.373 (0.171, 0.815)
Refractory to last prior treatment (yes vs. no)	-	1.309 (0.744, 2.301)	1.205 (0.811, 2.141)	1.313 (0.755, 2.281)
AIC	**675.733**	**683.207**	**-[b]**	**681.21**

Note: OS data are from the prespecified interim analysis (16 June 2014 data cut-off date). 95% CIs were estimated by bootstrapping. If the variable was shrunk to zero for a bootstrap sample, the HR for that sample was 1.

[a] The rows for prior bortezomib and prior SCT were incorrectly labelled the wrong way round in Appendix A (Table 35) of the ACD response.
[b] The LASSO R statistical package used to run this model did not include functionality to estimate statistical fit values.

ACD, appraisal consultation document; AIC Akaike information criterion; CRd, carfilzomib/lenalidomide/dexamethasone; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; EMA, European Medicines Agency; HR, hazard ratio; ISS, International Staging System; LASSO, least absolute shrinkage and selection operator; OS, overall survival; Rd, lenalidomide/dexamethasone; SCT, stem cell transplant.

We explored the uncertainty associated with method of covariate selection and impact on the ICER for CRd versus Rd by incorporating other covariate selection methods in the IPW-based Weibull model as per the revised base case analysis presented in our ACD response (Table 9).

The stepwise selection method yielded more favourable ICERs (£41,429 per QALY gained) than the all UK clinician-identified prognostic covariates method (£54,095 per QALY gained) or the LASSO variable selection method (£55,247 per QALY gained).

Table 9. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection for inclusion in the IPW-based Weibull survival model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
CRd versus Rd PFS and OS hazard ratios (95% CIs)[a]			
PFS[b]	0.699 (0.481, 1.02)	0.708 (0.478, 1.047)	0.776 (0.532, 1.000)
OS[b]	0.730 (0.454, 1.174)	0.783 (0.471, 1.299)	0.801(0.503, 1.079)
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – joint curve fitting (as per the revised base case cost-effectiveness analysis)[c]			
IPW-based Weibull model	**£41,429 (revised base case)[d]**	£54,095	£55,247

[a] HRs were derived from a covariate-adjusted analysis within the relevant subgroup (patients with two prior therapies and no prior lenalidomide).
[b] PFS and OS data are from the prespecified interim analysis (16 June 2014 data cut-off date).
[c] ICERs were derived using an IPW-based model with a Weibull distribution for both PFS and OS, jointly fitted to both study arms within the relevant subgroup.
[d] Revised base case ICER submitted in response to the ACD.

ACD, appraisal consultation document; CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Rd, lenalidomide/dexamethasone.

The ICERs were also more favourable for the stepwise selection method than the other covariate selection methods for the other standard OS parametric distributions (Appendix A).

These results suggest that the method of covariate selection has an impact on the cost effectiveness of CRd versus Rd. As highlighted in our ACD response, the stepwise selection method was considered the most appropriate method to use in our revised base case analysis, given that:

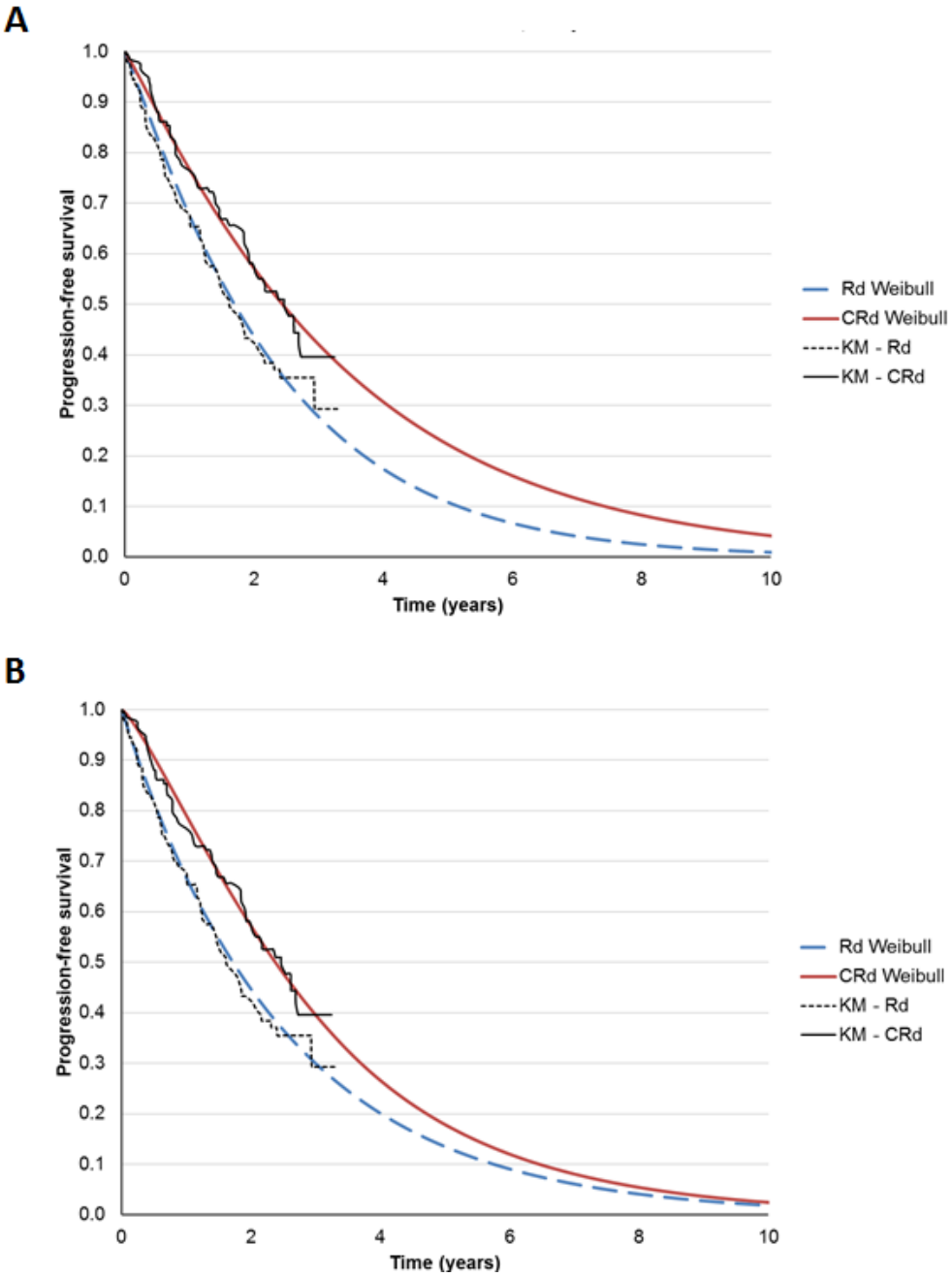
- It is more parsimonious than a model with all clinician-identified covariates (e.g. 6 covariates selected versus 10 for OS).
- It is less prone to bias than the LASSO variable selection model (LASSO trades off unbiasness with variance).[2]
- It has the best statistical fit i.e. AIC value. As noted previously, the LASSO R statistical package used to run the LASSO variable selection model does not include functionality to estimate statistical fit values so a direct comparison with this model cannot be made. However, the Cox model fitted using LASSO selected covariates had a higher AIC value than the Cox model using covariates selected with the stepwise selection model.

1.4.2 Additional analyses exploring the uncertainty associated with method of curve fitting (joint vs. independent) and non-proportionality of hazards in the IPW-based model (stepwise selection covariates) used in our revised base case cost-effectiveness analysis

To further explore the uncertainty associated with the method of curve fitting and potential non-proportionality of hazards in the IPW-based survival model, and impact on the ICER, we applied both joint (as per our revised base case analysis) and independent curve fits for CRd and Rd within the cost-effectiveness model. To ensure that all extrapolations of PFS and OS were clinically plausible, it was assumed that the survival hazard for the modelled CRd arm should not exceed the survival hazard for the modelled Rd arm. This approach is similar to a constraint commonly applied to OS (and applied within cost-effectiveness models included with our original submission) whereby the survival hazard of the modelled treatment arm should not exceed the survival hazard of an age-and gender-matched general population.

A comparison of jointly and independently fitted PFS and OS curves for the IPW-based Weibull model including the covariates identified through the stepwise selection method (as per our revised base case analysis) are provided below in Figure 3 (PFS) and Figure 4 (OS).

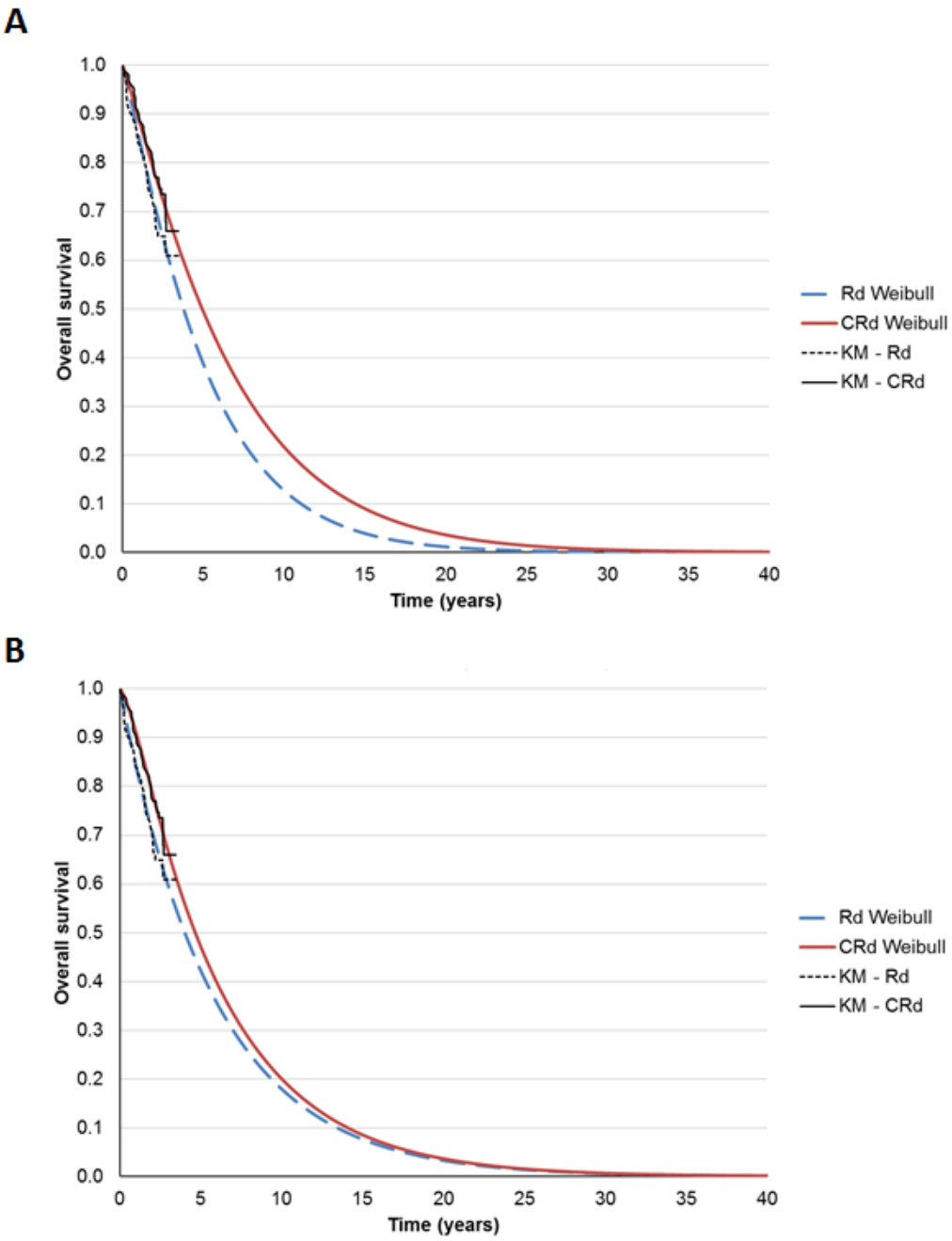
Figure 3. PFS Kaplan-Meier curves and IPW-based Weibull model (stepwise selection covariates) curves fitted (A) jointly and (B) independently to the CRd and Rd arms (ASPIRE, patients with two prior therapies and no prior lenalidomide)



Note: PFS data are from the prespecified interim analysis (16 June 2014 data cut-off date). Covariates included in the IPW-based Weibull model jointly fitted to both study arms within the relevant subgroup were identified through a stepwise selection procedure within the relevant subgroup: treatment (CRd vs. Rd), prior SCT (yes vs. no), ECOG PS (1-2 vs. 0), time from diagnosis (continuous variable), ISS stage (II-III vs. I), and β2-microglobulin (≥3.5 vs. <3.5 mg/L).

CRd, carfilzomib/lenalidomide; dexamethasone; ECOG PS, Eastern Cooperative Oncology Group performance status; IPW, inverse probability weighted; ISS, International Staging System; KM, Kaplan-Meier; PFS, progression-free survival; Rd, lenalidomide/dexamethasone; SCT, stem cell transplant.

Figure 4. OS Kaplan-Meier curves and IPW-based Weibull model (stepwise selection covariates) curves fitted (A) jointly and (B) independently to the CRd and Rd arms (ASPIRE, patients with two prior therapies and no prior lenalidomide)



Note: OS data are from the prespecified interim analysis (16 June 2014 data cut-off date). Covariates included in the IPW-based Weibull model jointly fitted to both study arms within the relevant subgroup were identified through a stepwise selection procedure within the relevant subgroup: treatment (CRd vs. Rd), prior bortezomib (yes vs. no), creatinine clearance (≥50 to <80 mL/min vs. other), creatinine clearance (>80 mL/min vs. other), time from diagnosis (continuous variable), ISS stage (II-III vs. I), and β2-microglobulin (≥3.5 vs. <3.5 mg/L).

CRd, carfilzomib/lenalidomide; dexamethasone; ECOG PS, Eastern Cooperative Oncology Group performance status; IPW, inverse probability weighted; ISS, International Staging System; KM, Kaplan-Meier; OS, overall survival; Rd, lenalidomide/dexamethasone; SCT, stem cell transplant.

The resulting ICERs are more favourable using the jointly fitted curves in which proportional hazards is assumed as per our revised base case analysis (£41,429 per QALY gained) than when using independently fitted curves (£95,876 per QALY gained).

These results suggest that the method of covariate selection has an impact on the cost effectiveness of CRd versus Rd. This is likely a result of the prespecified stopping rule for carfilzomib in ASPIRE where treatment with carfilzomib was discontinued after 18 cycles (61% of patients in the CRd arm remained on treatment in Cycle 18),[3] in contrast with the marketing authorisation (MA) for carfilzomib which permits treatment until progression or unacceptable toxicity.[4]

1.4.3 Additional analyses exploring the uncertainty associated with method of curve fitting (joint vs. independent) for the other covariate selection methods and standard parametric distributions

A summary of the ICERs resulting from the joint (as per our revised base case analysis) versus independent curve fitting using the different covariate selection methods to select covariates for inclusion in the IPW-based Weibull model is provided in Table 10. Similar to previous results, the ICERs are more favourable when curves are jointly fitted than when independently fitted across the different methods for covariate selection. When the other standard parametric distributions are used to extrapolate OS, the ICERs are also generally more favourable when curves are jointly fitted (Appendix A).

Table 10. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection and curve fitting for the IPW-based Weibull survival model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
CRd versus Rd PFS and OS hazard ratios (95% CIs)[a]			
PFS[b]	0.699 (0.481, 1.02)	0.708 (0.478, 1.047)	0.776 (0.532, 1.000)
OS[b]	0.730 (0.454, 1.174)	0.783 (0.471, 1.299)	0.801(0.503, 1.079)
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based Weibull survival model (as per the revised base case cost-effectiveness analysis)[c]			
PFS and OS curves jointly fitted	**£41,429 (revised base case)[d]**	£54,095	£55,247
PFS and OS curves independently fitted	£95,876	£108,049	£108,864

[a] HRs were derived from a covariate-adjusted analysis within the relevant subgroup (patients with two prior therapies and no prior lenalidomide).
[b] PFS and OS data are from the prespecified interim analysis (16 June 2014 data cut-off date).
[c] ICERs were derived using an IPW-based model with a Weibull distribution for both PFS and OS, jointly fitted to both study arms within the relevant subgroup.
[d] Revised base case ICER submitted in response to the ACD.

ACD, appraisal consultation document; CRd, carfilzomib/lenalidomide/dexamethasone; ICER, incremental cost-effectiveness ratio; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Rd, lenalidomide/dexamethasone.

1.4.4 Conclusion

A summary of cost-effectiveness results exploring the impact of method of covariate selection (stepwise selection vs. LASSO vs. all UK clinician identified prognostic covariates) and curve fitting (joint vs. independent) on the revised base case analysis presented in our response to the ACD is provided in Table 11.

These analyses show that both covariate selection method and curve fitting method have an impact on the cost effectiveness of CRd versus Rd. As outlined earlier in this response and in our response to the ACD, the stepwise selection method used to inform our revised base case analysis is considered more appropriate than the other methods for covariate selection. In addition, the less favourable ICERs resulting from independent curve fitting are likely a result of the arbitrary stopping rule for carfilzomib in ASPIRE (prespecified in the study protocol) where treatment with carfilzomib was discontinued after 18 cycles, in contrast with the MA for carfilzomib which permits treatment until progression or unacceptable toxicity.

Table 11. Summary of ICERs for CRd versus Rd from the revised base case cost-effectiveness analysis and key additional analyses presented in this response (ASPIRE, patients with two prior therapies and no prior lenalidomide)

Curve fitting method	Covariate selection method	ICER (£/QALY gained)
Revised base case cost-effectiveness analyses presented in our response to the ACD		
Joint	Stepwise	41,429
Key additional analyses presented in this response		
Joint	LASSO	54,095
Joint	All UK clinician identified prognostic covariates	55,247
Independent	Stepwise	95,876
Independent	LASSO	108,049
Independent	All UK clinician identified prognostic covariates	108,864

Note: An IPW-based survival model with a Weibull distribution for both PFS and OS was used for all analyses.

ACD, appraisal consultation document; CRd, carfilzomib/lenalidomidedexamethasone; ICER, incremental cost-effectiveness ratio; Rd, lenalidomide/dexamethasone; QALY, quality-adjusted life year.

2 References

1. Proskorovsky I, Lewis P, Williams CD *et al.* Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health Qual Life Outcomes* 2014;12:35.
2. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. *Br J Cancer* 2003;89:781-6.
3. Onyx. PX-171-009-A1 (ASPIRE) Clinical Study Report. 2014. (Data cutoff date 16-June 2014).
4. Amgen Europe. Kyprolis (carfilzomib) 10 mg, 30 mg and 60 mg powder for solution for infusion. Summary of Product Characteristics. 2016. Available from: http://www.medicines.org.uk/emc/medicine/31222 (Accessed February 2017).

Appendix A: Additional results for alternative parametric distributions

Table 12. Summary of ICERs for Cd versus Vd when the different methods are used to select covariates and fit curves for the IPW-based survival model across different parametric distributions (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model - joint curve fitting			
IPW-based Weibull model (PFS and OS)	**£22,009 (revised base case)[a]**	£22,441	£22,481
IPW-based Gompertz model (OS)[b]	£30,697	£31,172	£31,065
IPW-based exponential model (OS)[b]	£18,798	£19,130	£19,202
IPW-based gamma model (OS)[b]	£24,933	£25,438	£25,441
IPW-based log-logistic model (OS)[b]	£23,003	£23,282	£23,359
IPW-based log-normal model (OS)[b]	£23,606	£23,604	£23,767
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – independent curve fitting			
IPW-based Weibull model (PFS and OS)	£18,775	£20,506	£19,903
IPW-based Gompertz model (OS)[b]	£19,645	£21,906	£21,529
IPW-based exponential model (OS)[b]	£18,590	£19,076	£19,153
IPW-based gamma model (OS)[b]	£16,485	£17,200	£17,390
IPW-based log-logistic model (OS)[b]	£19,263	£20,764	£20,094
IPW-based log-normal model (OS)[b]	£20,600	£21,394	£21,003

[a] Revised base case ICER submitted in response to the ACD.
[b] The Weibull distribution was still used for PFS.

ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.

Table 13. Summary of ICERs for CRd versus Rd when the different methods are used to select covariates and fit curves for the IPW-based survival model across different parametric distributions (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model - joint curve fitting			
IPW-based Weibull model (PFS and OS)	**£41,429 (revised base case)[a]**	£55,247	£54,095
IPW-based Gompertz model (OS)[b]	£52,439	£73,931	£72,066
IPW-based exponential model (OS)[b]	£37,072	£47,962	£47,102
IPW-based gamma model (OS)[b]	£37,331	£48,786	£48,038
IPW-based log-logistic model (OS)[b]	£39,268	£49,372	£48,665
IPW-based log-normal model (OS)[b]	£34,208	£39,648	£39,383
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – independent curve fitting			
IPW-based Weibull model (PFS and OS)	£95,876	£108,864	£108,049
IPW-based Gompertz model (OS)[b]	£80,282	£87,548	£87,376
IPW-based exponential model (OS)[b]	£44,763	£58,163	£57,705
IPW-based gamma model (OS)[b]	£80,787	£86,554	£86,157
IPW-based log-logistic model (OS)[b]	£55,674	£42,778	£43,281
IPW-based log-normal model (OS)[b]	£60,588	£51,583	£52,230

[a] Revised base case ICER submitted in response to the ACD.
[b] The Weibull distribution was still used for PFS.

ACD, appraisal consultation document; CRd, carfilzomib/lenalidomide/dexamethasone; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Rd, lenalidomide/dexamethasone.

Carfilzomib for previously treated multiple myeloma

ERG's review of the company's additional analysis following the ACD

February 2017

This report was commissioned by the NIHR HTA Programme as project number 15/121/18

BMJ Technology Assessment Group

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Following the Company's response to the ACD, new information was requested from the company, including the following analyses:

1. Applying the different covariate adjustment methods (in particular the LASSO) that the company used in the Cox proportional hazards (PH) models, to fit adjusted parametric curves using all the standard distributions that have previously been assessed.

2. Producing both jointly fitted and independently fitted adjusted models for each adjustment method.

3. Producing the results of the economic models using all the survival models outlined above.

The Company highlighted that it would not be feasible to implement the LASSO method to fit adjusted parametric curves due to the limitations of the available software. The Company did, however, include an analysis that used the included covariates from the LASSO adjusted Cox PH model, but this does not incorporate the shrinkage factor, which is the key reason why the ERG favoured the LASSO method. The ERG accepted that this was not as easy to implement as anticipated.

The ERG's assessment of the additional analyses provided by the Company are discussed in the sections below. As some issues apply to both the ENDEAVOR-based model and the ASPIRE-based model, they will be discussed together to avoid repetition. Where there are key differences that require separate discussion, these are clearly separated into subsections.

1 CHOICE OF COVARIATE SELECTION

In the response to the request for further analyses, the Company reproduced the results of the covariate adjusted Cox proportional hazards models, including the confidence intervals for the coefficients of all the included covariates in each method, as these were not presented in the Company's response to the Appraisal Consultation Document (ACD) initially. However, the Company did not provide the equivalent results for the parametric model fits, so the ERG was unable to assess the appropriateness of the parametric models selected by the Company.

In Table 4 and Table 10 of the Company's response, the Company presents the hazard ratios (HRs) and the resulting ICERs for each of the covariate selection methods (except the LASSO) using the IPW-based Weibull curves for both progression-free survival (PFS) and overall survival (OS), for the

ENDEAVOR-based model and the ASPIRE-based model, respectively. However, the ERG notes that the HRs presented are those from the Cox PH models produced in the Company's initial response to the ACD, and not from the Weibull models as defined in the Company's revised economic models. The ERG recalculated the HRs using the data in the economic models and these are presented in Table 1 and Table 2, respectively, along with the resulting ICERs produced by the Company. The ERG would also like to highlight that the analysis using the LASSO variables does not represent the full LASSO method as it does not incorporate the shrinkage factor but instead merely uses the variables that were selected using this method when applied in the Cox PH model. The ICERs for the Company's ENDEAVOR-based model appear to be robust to using alternative methods to the stepwise method used in the Company's base case, due to an increase in costs for the carfilzomib group as a result of improved PFS, and an increase in QALYs as a result of the increased OS. However, the ICERs for the ASPIRE-based model appear to increase due to the reduced treatment effects.

Table 1. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection for inclusion in the IPW-based Weibull survival model for the ENDEAVOR-based model. (Adapted from the Company's response, Page 9, Table 4)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variables model**
ERG's recalculated HRs based on Weibull model used in Company's revised model			
PFS	0.441	0.407	0.406
OS	0.645	0.622	0.621
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – joint curve fitting (as per the revised base case cost-effectiveness analysis)			
IPW-based Weibull model	**£22,009 (revised base case)**	£22,441	£22,481

Abbreviations in Table: ERG, evidence review group; HR, hazard ratio; ICER, incremental cost effectiveness ratio; IPW, inverse probability weighting; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality adjusted life-year.

Table 2. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection for inclusion in the IPW-based Weibull survival model for the ASPIRE based analysis. (Adapted from the Company's response, Page 17, Table 9)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variables model**
ERG's recalculated HRs based on Weibull model used in Company's revised model			
PFS	0.676	0.717	0.695
OS	0.741	0.818	0.816

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variables model
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based Weibull survival model (as per the revised base case cost-effectiveness analysis)			
PFS and OS curves jointly fitted	**£41,429 (revised base case)**	£54,095	£55,247

Abbreviations in Table: ERG, evidence review group; HR, hazard ratio; ICER, incremental cost effectiveness ratio; IPW, inverse probability weighting; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality adjusted life-year.

2 COMPARISON BETWEEN JOINTLY FITTED CURVES AND INDEPENDENTLY FITTED CURVES

The Company provided the results of the economic models using the different covariate adjustment methods and different distributions, both fitted as a joint model and as independently fitted model for each treatment group. These results are given in Appendix A of the Company's response and replicated in Section 3 of this document. The Company's discussion around the independent and jointly fitted models were based around the Weibull distribution as per the Company's base case analyses. However, the following discussion from the ERG gives a more general perspective on the use of independent and jointly fitted models in the ENDEAVOR-based and ASPIRE-based models, respectively.

2.1 ENDEAVOR

The ERG notes that, for the independently fitted models, the ICERs were lower compared to the jointly fitted models for all distributions tested and for all methods of covariate selection tested (See Section 3 for ICERs). The ERG considers the independent models in the ENDEAVOR based analysis to produce implausible extrapolations because they produce curves that show diverging hazards, i.e. the treatment effect continues to increase over time, which contradicts the ERG's, and the Company's assessment, that the hazards can be considered proportional. Given that the Company and the ERG considered a PH assumption to hold, a jointly fitted model was considered to be more appropriate by the ERG for the ENDEAVOR based model. Therefore, the ERG have not revised the ERG's base case, which used the jointly fitted Gompertz model for OS, and the jointly fitted Weibull for PFS, as per the Company's base case. The results of the ERG's base case following the ACD are reported in Section 4. The Company's ICERs for the jointly fitted and independently fitted Weibull model are given in Table 3 for comparison, and the ICERs for all the distributions tested are presented in Section 3.

Table 3. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection and curve fitting for the IPW-based Weibull survival model for the ENDEAVOR based analysis. (Adapted from the Company's response, Page 13, Table 5)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variable selection model**
ERG's recalculated HRs based on Weibull model used in Company's revised model			
PFS	0.441	0.407	0.406
OS	0.645	0.622	0.621
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based Weibull survival model – (as per revised the base case cost-effectiveness analysis)[c]			
PFS and OS curves jointly fitted	**£22,009 (revised base case)**	£22,441	£22,481
PFS and OS curves independently fitted	£18,775	£20,506	£19,903

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
Abbreviations in Table: ERG, evidence review group; HR, hazard ratio; ICER, incremental cost effectiveness ratio; IPW, inverse probability weighting; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality adjusted life-year.			

2.2 ASPIRE

In contrast to the ENDEAVOR-based model, the ERG considered the independently fitted models for the ASPIRE-based model to be more plausible than the jointly fitted models. This was because of a PH assumption being inappropriate on inspection of log-log plots (Company's response to the ACD, Page 46, Figure 12) as stated in the ERG's review of the Company's initial response to the ACD. The log-log plots appeared to show converging hazards, and so imposing a PH assumption is likely to overestimate the survival benefit for the carfilzomib group and therefore underestimate the ICERs. The independently fitted models allowed for a diminishing treatment benefit to reflect the converging hazards, and so this approach was considered to be more appropriate by the ERG. The ERG, therefore, revised its base case to incorporate the independently fitted curves, and the results are given in Section 4. The ERG notes that the ICERs increased significantly by using the independently fitted curves compared to the jointly fitted curves, reflecting the impact of the overestimated survival, and therefore the high degree of uncertainty in the Company's base case analysis. The resulting ICERs for the jointly fitted and independently fitted Weibull model are given in Table 4 for comparison. The ICERs for all the distributions tested are given in Section 3.

Table 4. Summary of PFS and OS HRs and ICERs resulting from different methods of covariate selection and curve fitting for the IPW-based Weibull survival model (ASPIRE, patients with two prior therapies and no prior lenalidomide)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
ERG's recalculated HRs based on Weibull model used in Company's revised model			
PFS	0.676	0.717	0.695
OS	0.741	0.818	0.816
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based Weibull survival model (as per the revised base case cost-effectiveness analysis)[c]			
PFS and OS curves jointly fitted	**£41,429 (revised base case)**	£54,095	£55,247
PFS and OS curves independently fitted	£95,876	£108,049	£108,864

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
Abbreviations in Table: ERG, evidence review group; HR, hazard ratio; ICER, incremental cost effectiveness ratio; IPW, inverse probability weighting; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality adjusted life-year.			

3 COMPANY'S RESULTS

The results of the Company's analyses are given in Table 5 and Table 6 for the ENDEAVOR-based model and the ASPIRE-based model, respectively. This includes scenarios for each combination of parametric curve and covariate adjustment method, and for curves fitted both jointly and independently.

Table 5. Summary of ICERs for Cd versus Vd when the different methods are used to select covariates and fit curves for the IPW-based survival model across different parametric distributions for the ENDEAVOR-based model (Appendix A of the Company's response, page 24, Table 12)

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model - joint curve fitting			
IPW-based Weibull model (PFS and OS)	**£22,009 (revised base case)**	£22,441	£22,481
IPW-based Gompertz model (OS)	£30,697	£31,172	£31,065
IPW-based exponential model (OS)	£18,798	£19,130	£19,202
IPW-based gamma model (OS)	£24,933	£25,438	£25,441
IPW-based log-logistic model (OS)	£23,003	£23,282	£23,359
IPW-based log-normal model (OS)	£23,606	£23,604	£23,767
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – independent curve fitting			
IPW-based Weibull model (PFS and OS)	£18,775	£20,506	£19,903
IPW-based Gompertz model (OS)	£19,645	£21,906	£21,529

	Stepwise variable selection model	All UK clinician-identified covariates model	LASSO variable selection model
IPW-based exponential model (OS)	£18,590	£19,076	£19,153
IPW-based gamma model (OS)	£16,485	£17,200	£17,390
IPW-based log-logistic model (OS)	£19,263	£20,764	£20,094
IPW-based log-normal model (OS)	£20,600	£21,394	£21,003

Abbreviations in table: ACD, appraisal consultation document; Cd, carfilzomib/dexamethasone; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.

Table 6. Summary of ICERs for CRd versus Rd when the different methods are used to select covariates and fit curves for the IPW-based survival model across different parametric distributions for the ASPIRE-based model (Appendix A of the Company's response, page 25, Table 13)

	Stepwise variable selection model	**All UK clinician-identified covariates model**	**LASSO variable selection model**
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model - joint curve fitting			
IPW-based Weibull model (PFS and OS)	**£41,429 (revised base case)**	£55,247	£54,095
IPW-based Gompertz model (OS)	£52,439	£73,931	£72,066
IPW-based exponential model (OS)	£37,072	£47,962	£47,102
IPW-based gamma model (OS)	£37,331	£48,786	£48,038
IPW-based log-logistic model (OS)	£39,268	£49,372	£48,665
IPW-based log-normal model (OS)	£34,208	£39,648	£39,383
ICER (£/QALY gained) when analysis is used to select covariates for inclusion in the IPW-based survival model – independent curve fitting			
IPW-based Weibull model (PFS and OS)	£95,876	£108,864	£108,049
IPW-based Gompertz model (OS)	£80,282	£87,548	£87,376
IPW-based exponential model (OS)	£44,763	£58,163	£57,705
IPW-based gamma model (OS)	£80,787	£86,554	£86,157
IPW-based log-logistic model (OS)	£55,674	£42,778	£43,281
IPW-based log-normal model (OS)	£60,588	£51,583	£52,230
Abbreviations in table: ACD, appraisal consultation document; CRd, carfilzomib/lenalidomide/dexamethasone; IPW, inverse probability weighted; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life year; Rd, lenalidomide/dexamethasone.			

4 ERG REVISED BASE CASES

4.1 ENDEAVOR

After reviewing the Company's response to the request for further analyses, the ERG considered the ERG's base case for the ENDEAVOR-based model following the ACD to be appropriate, and so no revision of this has been made. This base case, reported in Table 7, incorporated three changes from the Company's base case:

- Using the Company's jointly fitted Gompertz parametric curves for OS;
- Restricting the costs of bortezomib to 8 cycles to reflect UK practice; and,
- Applying the approximation that the Company provided for the effective discount expected to be incurred by the bortezomib patient access scheme (PAS).

Given the Company's concern about breaking the dose-efficacy link for bortezomib, the ERG would like to reiterate that it considers the matched adjusted indirect comparison (MAIC), to adjust the efficacy of bortezomib beyond 8 cycles, to be unreliable. This is because there is at least one prognostic indicator (cytogenetic status) that the Company was unable to adjust for, and the Company has made no reference to any unobserved prognostic indicators that are present but unaccounted for, which is an inherent problem in any unanchored MAIC. However, in order to give the Committee an insight into a diminished treatment effect, the ERG has presented a series of analyses that reduce the benefit for PFS and OS, beyond 8 treatment cycles in the bortezomib group, by increments of 10% (from 0% [ERG base case] to 40%). The results of these are given in Table 8. In addition to this, based on the Company's estimated HRs derived from the MAIC, suggesting a reduced benefit of 36.0% for PFS and 34.9% for OS, the resulting ICER was £44, 842.

Table 7. ERG base case following the ACD (ENDEAVOR)

Results per patient	Cd	Vd	Incremental value
Company's revised base case			
Total costs (£)	£117,660	£93,769	£23,891
QALYs	3.88	2.79	1.09
ICER			£22,009
Using the Gompertz distribution for OS			
Total costs (£)	£108,436	£90,814	£17,622
QALYs	2.70	2.13	0.57
ICER (compared with base case)			£30,697
ICER with all changes incorporated			£30,697
Restricting the cost of bortezomib to 8 treatment cycles			

Results per patient	Cd	Vd	Incremental value
Total costs (£)	£108,436	£73,789	£34,647
QALYs	2.70	2.13	0.57
ICER (compared with base case)			£37,694
ICER with all changes incorporated			£60,357
Using a discount of 15% for the cost of bortezomib			
Total costs (£)	£108,436	£71,512	£36,924
QALYs	2.70	2.13	0.57
ICER (compared with base case)			£26,306
ICER with all changes incorporated			£64,325
ERG preferred base case ICER			**£64,325**

Abbreviation used in the table: Cd, carfilzomib and dexamethasone; EQ-5D, European Quality of Life 5 dimensions; HSUV, healthy state utility value; ICER, incremental cost effectiveness ratio; MA, market authorisation; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life-years; Vd, bortezomib and dexamethasone.

Table 8. Scenario analysis of PFS and OS efficacy reduction for bortezomib after 8 treatment cycles.

Increase in HR	ICER
10%	£55,952
20%	£50,410
30%	£46,499
40%	£43,615

Abbreviations in table: ICER, incremental cost effectiveness ratio.

4.2 ASPIRE

The ERG revised its base case for the ASPIRE-based model following the Company's response to the request for further analysis, to incorporate independently fitted parametric curves. The ERG consider these to be more appropriate given the lack of PH for both PFS and OS, however, the Company did not provided goodness-of-fit statistics to assess the different distributions. For this reason, the ERG chose to use the Weibull for both PFS and OS, as the Company focussed their analysis around the Weibull throughout the Company's response. The results of the revised base case are given in Table 9.

Table 9. ERG revised base case following the ACD (ASPIRE)

Results per patient	CRd	Rd	Incremental value
Company's revised base case			
Total costs (£)	£127,140	£94,528	£32,612

Results per patient	CRd	Rd	Incremental value
QALYs	3.67	2.88	0.79
ICER			£41,429
Using the independently fitted Weibull curves for PFS and OS			
Total costs (£)	£130,946	£94,455	£36,491
QALYs	3.58	3.20	0.38
ICER (compared with base case)			£95,876
ICER with all changes incorporated			£95,876
ERG preferred base case ICER			**£95,876**

Abbreviation used in the table: CRd, carfilzomib, lenalidomide and dexamethasone; EQ-5D, European Quality of Life 5 dimensions; HSUV, healthy state utility value; ICER, incremental cost effectiveness ratio; OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life-years; Rd, lenalidomide and dexamethasone.

5 CONCLUSION

The Company's revised analyses highlight the high degree of uncertainty in the Company's base case analyses for both the ENDEAVOR-based model and the ASPIRE-based model. The ERG considers the survival curves used in the Company's base case analyses to have implausible extrapolations that have not been justified, and considers the ERG base case analyses to be a more appropriate reflection of the data available. This is particularly the case for the ASPIRE-based model where the treatment effect appears to diminish over time. Not accounting for this overestimates the benefits of carfilzomib and therefore underestimates the ICER.

Another serious issue in the Company's ENDEAVOR-based model is that the costs of bortezomib do not reflect UK clinical practice, in which the treatment schedule would be restricted to 8 cycles. This underestimates the costs of bortezomib, and given the lack of evidence showing a benefit of extended treatment beyond 8 cycles, this is likely to cause an underestimation of the ICER. Further to this, the Company's base case does not account for a potentially substantial reduction in the acquisition cost of bortezomib resulting from the PAS, which further underestimates the ICER.

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Single Technology Appraisal (STA)

Carfilzomib for treating multiple myeloma in people who have received at least one prior therapy [ID934]

Additional evidence for consideration by the Committee

Prepared by:

March 2017

File name	Version	Contains confidential information	Date
	1.0	Yes AIC and CIC redacted	27 March 2017

Contents

List of tables

List of figures

1 Executive Summary

We welcome the opportunity for further consideration of this appraisal following suspension of the final appraisal determination.

Following the second Appraisal Committee (AC) meeting, the most important residual uncertainty for the Committee related to the modelling of overall survival (OS), in particular the choice of the most appropriate parametric function for extrapolation. The incremental cost-effectiveness ratios (ICERs) for carfilzomib/dexamethasone (Cd) versus bortezomib/dexamethasone (Vd) as per the Committee's preferred modelling assumptions were £28,797*[*] per quality-adjusted life year (QALY) gained based on a Weibull distribution (Amgen base case extrapolation) and £44,842* per QALY gained based on a Gompertz distribution (the Evidence Review Group's [ERG's] preferred extrapolation). The Committee could not choose the most appropriate function, though recognised that the Gompertz extrapolation was very conservative. In addition, the ERG's choice of the Gompertz function for their preferred scenario was communicated to Amgen just three days in advance of the second AC meeting, leaving us with no opportunity to comprehensively challenge this ERG preference.

We now present compelling new evidence which removes this cloud of uncertainty and supports our base case extrapolation using the Weibull distribution. Further, like NICE, we have always strongly believed in the need to demonstrate value and took the important step of offering a [redacted] patientaccess scheme [redacted]% confidential PAS discount) from the outset. We are confident that our response will allow the Committee to make a positive recommendation on the use of Cd in the subgroup of patients who have received one prior therapy and no prior bortezomib.

New and more mature OS data from ENDEAVOR have demonstrated a statistically significant survival benefit for Cd over Vd. The ICER using the Weibull function remains extremely stable when the new data are incorporated into the cost-effectiveness analysis.

Since the last AC meeting, new and more mature OS data from a second prespecified interim analysis of OS in ENDEAVOR have become available. These new data have demonstrated a statistically significant improvement for Cd compared with bortezomib/dexamethasone (Vd) in the intent-to-treat (ITT) population (hazard ratio [HR] 0.79, 95% confidence interval [95% CI] 0.65, 0.96; 1-sided p = 0.0100), and with a covariate-adjusted HR for the subgroup of patients who have received one prior therapy and no prior bortezomib of 0.62 (95% CI 0.40, 0.95). These more mature data are very consistent with the previous OS data considered by the Committee at the last AC meeting.

[*] ICERs include an adjustment for bortezomib treatment duration with regards both costs and efficacy (maximum 8 cycles with efficacy adjustment using the ENDEAVOR:CASTOR matching-adjusted indirect comparison) as per the Committee's preferred modelling assumption. The Weibull ICER previously considered by the Committee at the last AC meeting (£26,306 per QALY gained) did not include any adjustment for bortezomib treatment duration.

The Weibull parametric function continues to provide a highly plausible fit to the new OS data, and the resulting ICER using the new OS data remains extremely stable at £27,629 per QALY gained (slightly lower than the equivalent ICER of £28,797 per QALY gained using the previous OS data) lending greater credence to our base case modelling approach. In contrast, the ICER does not remain stable when the Gompertz function is used, and markedly decreases to £39,052 per QALY gained (from £44,482 per QALY gained, the equivalent ICER using the previous OS data).

Published long-term OS data for bortezomib-treated R/RMM patients from Orlowski *et al.* provide a robust external clinical validation of our modelling approach, and endorse the Weibull function as the most clinically plausible parametric function for ENDEAVOR.

Long-term OS data (median 8.6 years follow-up) for bortezomib-treated relapsed and/or refractory multiple myeloma (R/RMM) patients from a randomised controlled trial (RCT) reported by Orlowski *et al.* provides a rich source of external evidence to assess the clinical plausibility of the Weibull and Gompertz extrapolations. Dosing and treatment duration with bortezomib were consistent with that in ENDEAVOR, and patients had not received prior bortezomib consistent with the subgroup of interest for this appraisal (i.e. patients who have received one prior therapy and no prior bortezomib).

When the Gompertz function is used to extrapolate OS for the ENDEAVOR Vd arm, the curve falls below the Orlowski *et al.* curve after approximately 7 years, and predicts that just 3.2% of patients will be alive at 9 years, which is substantially less than the 13.4% observed in the Orlowski *et al*. study. This is highly implausible, especially since the survival rates from the Orlowski *et al.* study are likely to represent an already conservative proxy for Vd in the ENDEAVOR subgroup of interest. In contrast, the Weibull extrapolated ENDEAVOR Vd curve is highly plausible with a very similar shape to the Orlowski *et al.* Kaplan-Meier curve, and predicts a similar proportion of patients remaining alive at 9 years (15.7% vs. 13.4%).

In addition to using long-term published data on bortezomib to validate the plausibility of the Weibull OS extrapolation, we sought feedback from 13 consultant haematologists practicing in England on the clinical plausibility of the Weibull versus Gompertz OS extrapolations. Eleven supported the Weibull curve as being the most clinically plausible, and none the Gompertz curve. This provides an overwhelming external clinical validation of our approach and use of the Weibull function.

Unlike the Gompertz function which results in highly conservative and clinically implausible long-term OS projections, the Weibull function results in clinically plausible model curves and long-term OS projections, and is the most appropriate choice for extrapolation of OS from ENDEAVOR.

Our final base case ICER incorporating all of the Committee's preferred modelling assumptions is £27,629, demonstrating that Cd is a cost-effective use of NHS resources for the treatment of patients who have received one prior therapy and no prior bortezomib.

All of the cost-effectiveness analyses reported in this document incorporate the following preferred modelling assumptions considered appropriate and accepted as suitable for decision-making by the Committee:

- Inverse probability weighted (IPW)-based survival model jointly fitted to both ENDEAVOR study arms
  - Including UK clinical expert-identified prognostic covariates retained in a stepwise selection procedure (Appendix A)
  - Assuming proportional hazards
- Utility values mapped from ENDEAVOR using Proskorovsky *et al.,* 2014[1]
- Bortezomib complex PAS assumed to translate to 15% cost saving for the NHS
- Bortezomib treatment capped at a maximum of 24 weeks (8 Cycles) with efficacy adjusted using the ENDEAVOR:CASTOR matching-adjusted indirect comparison (MAIC) (Appendix B)

The final base case cost-effectiveness results are presented in Table 1-1, with a final base case ICER of £27,629 per QALY gained.

Table 1-1 Final base case cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib

	Total costs (£)	Total LYGs	Total QALYs	Inc. costs (£)	Inc. LYGs	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£69,626	3.34	2.20				
Cd	£118,077	5.87	3.96	£48,451	2.54	1.75	£27,629
Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYGs, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone.							

We believe that the robust new evidence presented in this document (i.e. new and more mature ENDEAVOR OS data and published long-term OS data for bortezomib-treated patients) clearly demonstrates that the Weibull function is appropriate for modelling OS and the Gompertz function is not. The final base case ICER is £27,629, which is very consistent with the equivalent based on the previous OS data. The ICER remains firmly below £30,000 per QALY demonstrating that Cd represents a cost-effective use of NHS resources for the treatment of patients who have received one prior therapy and no prior bortezomib.

2 New ENDEAVOR overall survival data

2.1 Context

We highlighted in Section 4.13 of the company submission that the second planned interim analysis of OS for ENDEAVOR was anticipated to be conducted in Q4 2016, with the final planned analysis anticipated to be conducted in Q4 2018.

On 28 February 2017, Amgen released results from the second interim analysis which demonstrated a statistically significant survival benefit for Cd over Vd.[2-4] ENDEAVOR is the first and only head-to-head study comparing proteasome inhibitors to demonstrate statistically significant improved OS. Given the statistically significant survival benefit observed for Cd, there will be no subsequent formal analyses of OS in ENDEAVOR; these results should therefore be considered final. Follow-up for safety in ENDEAVOR is planned to continue until July 2017 as part of an FDA-mandated safety follow-up.

Results from this analysis, and its impact on the cost-effectiveness analysis, are reported below. These new OS data provide supportive evidence that the Weibull parametric function is the most appropriate for extrapolating long-term OS.

2.2 Summary of new ENDEAVOR overall survival results

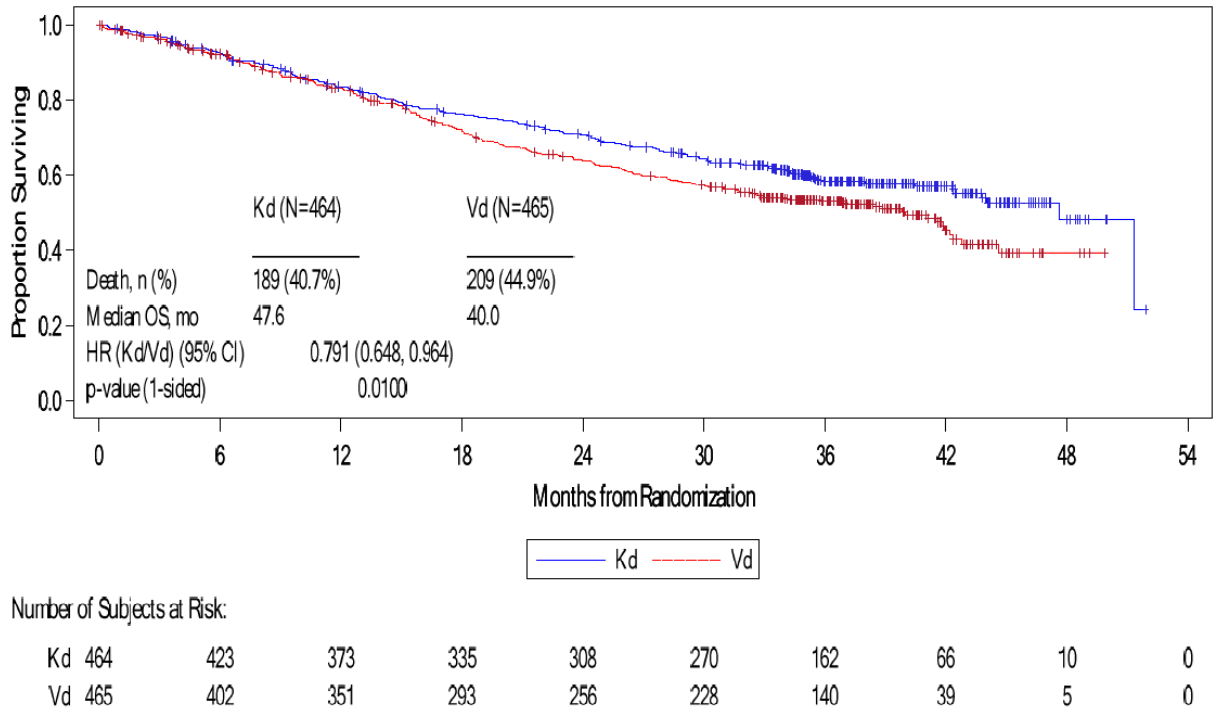
ENDEAVOR ITT population

A total of 398 (42.8%) deaths had occurred at the time of data cut-off for the second interim analysis (3 January 2017); 189 (40.7%) in the Cd arm and 209 (44.9%) in the Vd arm.[2] This represents 80% of the final target number of 496 OS events. A statistically significant OS improvement was observed for Cd compared with Vd (HR 0.79, 95% CI 0.65, 0.96; 1-sided p = 0.0100), with a p-value lower than the prespecified boundary for this second interim analysis of OS (0.0123). This HR is similar to the HR of 0.81 from the previous European Medicines Agency (EMA) ad-hoc analysis (3 March 2016 data cut-off date) reported in Section 4.6.1 of the company submission.

Median OS was 47.6 months (95% CI 42.5, not estimable [NE]) in the Cd arm and 40.0 months (95% CI 32.6, 42.3) in the Vd arm, representing a 7.6-month improvement for Cd over Vd. Median follow-up was 37.5 months and 36.9 months in the Cd and Vd arms, respectively.

A Kaplan-Meier plot of OS from the second interim analysis is provided in Figure 2-1. Consistent with the first prespecified interim analysis and EMA ad-hoc analysis, once the curves separate they remain separated for the remainder of follow-up.

Figure 2-1 Kaplan–Meier plot of OS – prespecified second interim analysis (ENDEAVOR, ITT population)



References: Dimopoulos *et al.,* 2017[2]

Note: New ENDEAVOR OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; Kd, carfilzomib/dexamethasone; OS, overall survival; Vd, bortezomib/dexamethasone.

ENDEAVOR subgroup of patients who have received one prior therapy and no prior bortezomib

The Committee considered the subgroup of patients who have received one prior therapy and no prior bortezomib to be relevant for decision-making as it reflects the anticipated place in the treatment pathway for Cd in England and Wales.

A Kaplan-Meier plot of OS for this subgroup based on the new ENDEAVOR OS data is provided in Figure 2-2, which shows that once the curves separate they remain separated for the duration of follow-up.

Figure 2-2 Kaplan–Meier plot of OS – prespecified second interim analysis (ENDEAVOR, patients with one prior therapy and no prior bortezomib)

Note: New ENDEAVOR OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date). Vertical bars represent censoring.

Cd, carfilzomib/dexamethasone; OS, overall survival; Vd, bortezomib/dexamethasone.

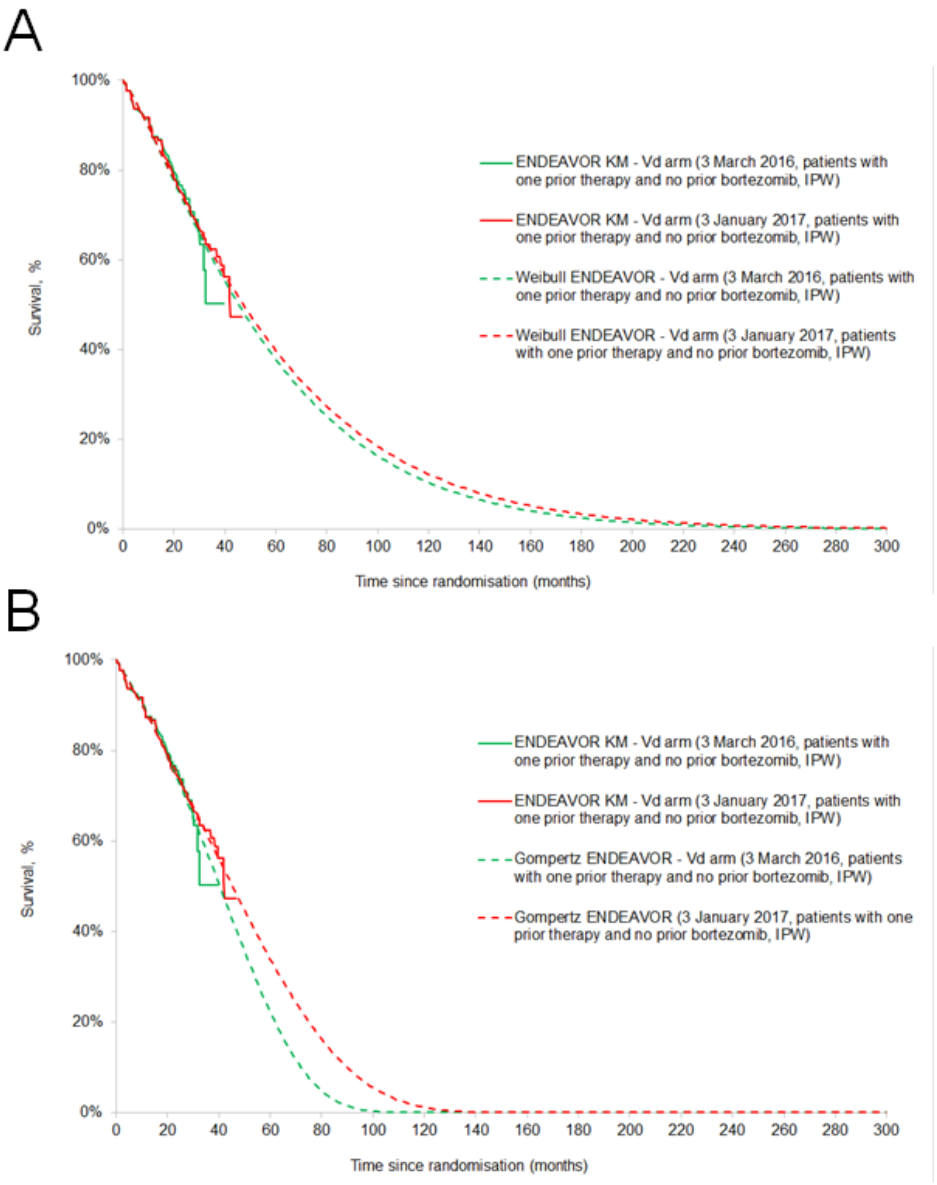
The unadjusted OS HR for Cd versus Vd based on the new ENDEAVOR OS data is XXX (95% CI [redacted], [redacted]) which is highly consistent with the unadjusted HR of 0.74 (95% CI 0.46, 1.20) based on the previous OS data as reported in our response to clarification questions from the ERG. The Committee considered the covariate-adjusted treatment effect for OS for this subgroup to be relevant and our revised approach (using the subset of UK clinical expert-identified prognostic covariates retained in a stepwise selection procedure; Appendix A) to be reasonable. The resulting covariate-adjusted OS HR estimated for Cd versus Vd based on the new ENDEAVOR OS data is 0.62 (95% CI 0.40, 0.95) which is highly consistent with the HR based on the previous OS data reported in our response to the appraisal consultation document (ACD) (HR 0.63; 95% CI 0.38, 1.04), with narrower 95% CIs that provide more certainty around the point estimate.

In short, Kaplan-Meier Cd and Vd curves based on the new ENDEAVOR OS data separate and remain separated for the duration of follow-up, and both unadjusted and covariate-adjusted treatment effect estimates are highly similar to those previously considered by the Committee based on the previous OS data

2.3 *Impact of the new ENDEAVOR overall survival data on the cost-effectiveness analysis based on the Weibull and Gompertz functions*

A comparison of the Kaplan-Meier and modelled (Weibull and Gompertz) OS curves for the ENDEAVOR subgroup of interest based on the previous and new ENDEAVOR OS data is provided in Figure 2-3.

Figure 2-3 Comparison of Kaplan-Meier and modelled OS curves based on the previous vs. new OS data from ENDEAVOR using (A) a Weibull distribution and (B) a Gompertz distribution for the modelled curves (patients with one prior therapy and no prior bortezomib)



Note: Previous ENDEAVOR OS data are from the EMA ad-hoc analysis (3 March 2016 data cut-off date) and new ENDEAVOR OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

Cd, carfilzomib/dexamethasone; IPW, inverse probability weighted; KM, Kaplan-Meier; OS, overall survival.

This comparison demonstrates that the modelled curves using the Weibull function continue to provide a highly plausible fit to the new OS data, and the resulting ICER for the ENDEAVOR subgroup using the new OS data (Table 2-1) remains extremely stable and is well below the £30,000 threshold at £27,629 per QALY gained (a marginal decrease relative to the equivalent ICER of £28,797 per QALY gained using the previous OS data). A probabilistic sensitivity analysis scatter plot and cost-effectiveness acceptability curve based on the new OS data are provided in Appendix C. Of note, if assuming bortezomib treatment duration as per ENDEAVOR (i.e. until progression or unacceptable toxicity), the ICER would be lower at £26,762 (£26,306 with the previous OS data).

In contrast to the above, the ICER does not remain stable when the Gompertz function is used, and markedly decreases to £39,052 per QALY gained with the new ENDEAVOR OS data (from £44,482 per QALY gained, the equivalent ICER using the previous OS data) (Table 2-1). This suggests that there is an inherent instability with the Gompertz function introducing substantial uncertainty to the ICER estimates.

Table 2-1 Impact of incorporating the new ENDEAVOR OS data on the cost-effectiveness results for the Weibull versus Gompertz parametric functions (patients with one prior therapy and no prior bortezomib)

	Weibull	Gompertz
ICER using previous OS data, £/QALY gained[a]	£28,797	£44,842
ICER using new OS data, £/QALY gained[b]	£27,629	£39,052
Absolute difference, £	- £1,168	- £5,790
Relative difference, %	- 4.2%	- 14.8%

[a] Previous OS data from the EMA ad-hoc analysis (3 March 2016 data cut-off date).
[b] New OS data from the second prespecified interim analysis of OS in ENDEAVOR (3 January 2017 data cut-off date).

AC, appraisal committee; IPW, inverse probability weighted; OS, overall survival; QALY, quality-adjusted life year.

The new ENDEAVOR OS data showing a statistically significant survival benefit provides additional certainty about the clinical effectiveness of Cd over Vd. Unlike the Gompertz function, the ICER remains extremely stable when the Weibull function is used to extrapolate OS, lending greater credence to our base case modelling approach.

3 Clinical plausibility of the modelled Weibull and Gompertz curves using the new ENDEAVOR overall survival data

3.1 Context

There is little to justify the choice of either Weibull or Gompertz over the other based on visual inspection of modelled versus Kaplan-Meier curves and statistical goodness of fit alone (Appendix D). Consequently, the clinical plausibility of the long-term extrapolations should be a consideration of great importance and have significant weight in determining the selection of the most appropriate parametric function for extrapolation of OS from ENDEAVOR, as per NICE Decision Support Unit (DSU) TSD14 which states that:[5]

'However it is of even greater importance to justify the plausibility of the extrapolated portion of the survival model chosen, as this is likely to have a very large influence on the estimated mean survival. This is difficult, but may be achieved through the use of external data sources, biological plausibility, or clinical expert opinion.'

The ERG's preference for the Gompertz function does not adequately take into account the clinical plausibility of the long-term extrapolations. We would like to bring to the attention of the Committee compelling additional evidence from a rich external data source – published long-term OS data for bortezomib-treated R/RMM patients with a median of 8.6 years follow-up - which validates the Weibull function as the most clinically plausible for extrapolating OS from the ENDEAVOR study. Additionally, we sought feedback from 13 consultant haematologists practicing in England which provides an overwhelming external clinical validation of our approach and use of the Weibull function.

3.2 Clinical validation of the Weibull function using published long-term overall survival data from bortezomib-treated patients (Orlowski et al., 2016)

Overview of the study

Orlowski *et al.* conducted an open-label RCT comparing bortezomib monotherapy with bortezomib in combination with pegylated liposomal doxorubicin and recruited 646 patients with R/RMM between December 2004 and March 2006, of whom 322 were randomised to the bortezomib monotherapy arm.[6] Patients were required to have received ≥ 1 prior therapy, and consistent with the subgroup being considered by NICE for Cd, had not received prior bortezomib. In the bortezomib monotherapy arm patients received bortezomib at a dose, dosing schedule, and treatment duration consistent with patients in the Vd arm of ENDEAVOR (1.3 mg/m^2 on Days 1, 4, 8, and 11 of repeated 21-day cycles until progression or unacceptable toxicity). At the time of data cut-off for the most recent analysis (May 2014), 80% of patients in the bortezomib monotherapy arm had died and median follow-up was 8.6 years.[7]

To the best of our knowledge, Orlowski *et al.* have reported the longest follow-up for OS of bortezomib-treated R/RMM patients to date. Given this, and that important aspects of study design and patient populations are consistent with the ENDEAVOR and the subgroup of interest (i.e. bortezomib dosing and treatment duration, no prior bortezomib exposure), we consider this study to represent a rich external data source from which clinical plausibility of the OS extrapolation for Vd can be meaningfully assessed. We acknowledge there are some differences between the Orlowski *et al.* study compared with ENDEAVOR. However, these are likely to mean that long-term OS data from the Orlowski *et al.* study represent a conservative proxy for Vd in the ENDEAVOR subgroup of interest, as explained Table 3-1.

Table 3-1 Comparison of key characteristics of the Orlowski *et al.* study vs. ENDEAVOR

Trial characteristics	Discussion
Bortezomib monotherapy vs. combination with dexamethasone	• Bortezomib was administered as a monotherapy in the Orlowski *et al.* study.[6] In contrast, bortezomib was administered in combination with dexamethasone (Vd) in the ENDEAVOR study.[8] • Vd is recognised by the EMA,[9] and well accepted by clinicians, to have superior outcomes in R/RMM relative to bortezomib monotherapy. • The use of bortezomib monotherapy in the Orlowski *et al.* study therefore likely means that patients had a poorer prognosis than in ENDEAVOR where Vd was given.
Number of prior therapies	• Two-thirds (66%) of patients had received ≥ 2 prior therapies in the Orlowski *et al.* study[6] In contrast, all patients in the ENDEAVOR subgroup of interest (i.e. patients who have received one prior therapy and no prior bortezomib) received only one prior therapy. • The heavily pretreated patients in the Orlowski *et al.* study are therefore likely to have a poorer prognosis than those in the ENDEAVOR subgroup.
Timing of study and subsequent therapies	• Orlowski *et al.* enrolled patients over a decade ago (December 2004 to March 2006),[6] implying that most patients were diagnosed with MM in the early 2000s. In contrast, ENDEAVOR enrolled patients between June 2012 and June 2014.[8] • The prognosis for MM patients has improved over the last decade. Kumar *et al.* showed an improvement in OS for patients diagnosed in 2006 to 2010 vs. 2001 to 2005; the 5-year survival rate was 10 to 15% higher for the overall MM population.[10] Cancer Research UK data similarly demonstrate improving survival in recent years, with 5-year survival rates estimated to be 35.9% for patients diagnosed in 2005 to 2006 vs. 47.0% for patients diagnosed in 2010 to 2011.[11] The equivalent 10-year survival rates are 21.4% and 32.5%, respectively. • Although there are likely multiple reasons to explain this trend in improving survival, the availability of novel therapies is likely a substantial driver. Within the UK treatment pathway there have been more regimens available and approved by NICE and/or funded by the CDF than would have been available to patients relapsing on study treatment in the Orlowski *et al.* study where the most common

Trial characteristics	Discussion
	subsequent therapies were cyclophosphamide (51%), thalidomide (31%), and lenalidomide (21%).[7] • The timing of the Orlowski *et al.* study and the limited availability of subsequent therapies that improve survival is therefore likely to mean that the Orlowski *et al.* population had a poorer prognosis for OS than in ENDEAVOR.
CDF, Cancer Drugs Fund; EMA, European Medicines Agency; NICE, National Institute of Health and Care Excellence; MM, multiple myeloma; OS, overall survival; R/RMM, relapsed and/or refractory multiple myeloma; Vd, bortezomib/dexamethasone.	

Comparison of observed and modelled survival outcomes from ENDEAVOR (based on the new OS data) with long-term data published by Orlowski *et al.*

A comparison of the Kaplan-Meier OS curves from the bortezomib arms in the ENDEAVOR study and Orlowski *et al.* study is provided in Figure 3-1. The shapes of the ENDEAVOR and Orlowski *et al.* curves are highly similar, though curves for both the ENDEAVOR ITT population and subgroup of interest curves remain above the Orlowski *et al.* curve for the duration of follow-up. This is unsurprising given that the population enrolled in the Orlowski *et al.* study is likely to have had a poorer prognosis than the population in ENDEAVOR and the subgroup of interest, as discussed in Table 3-1.

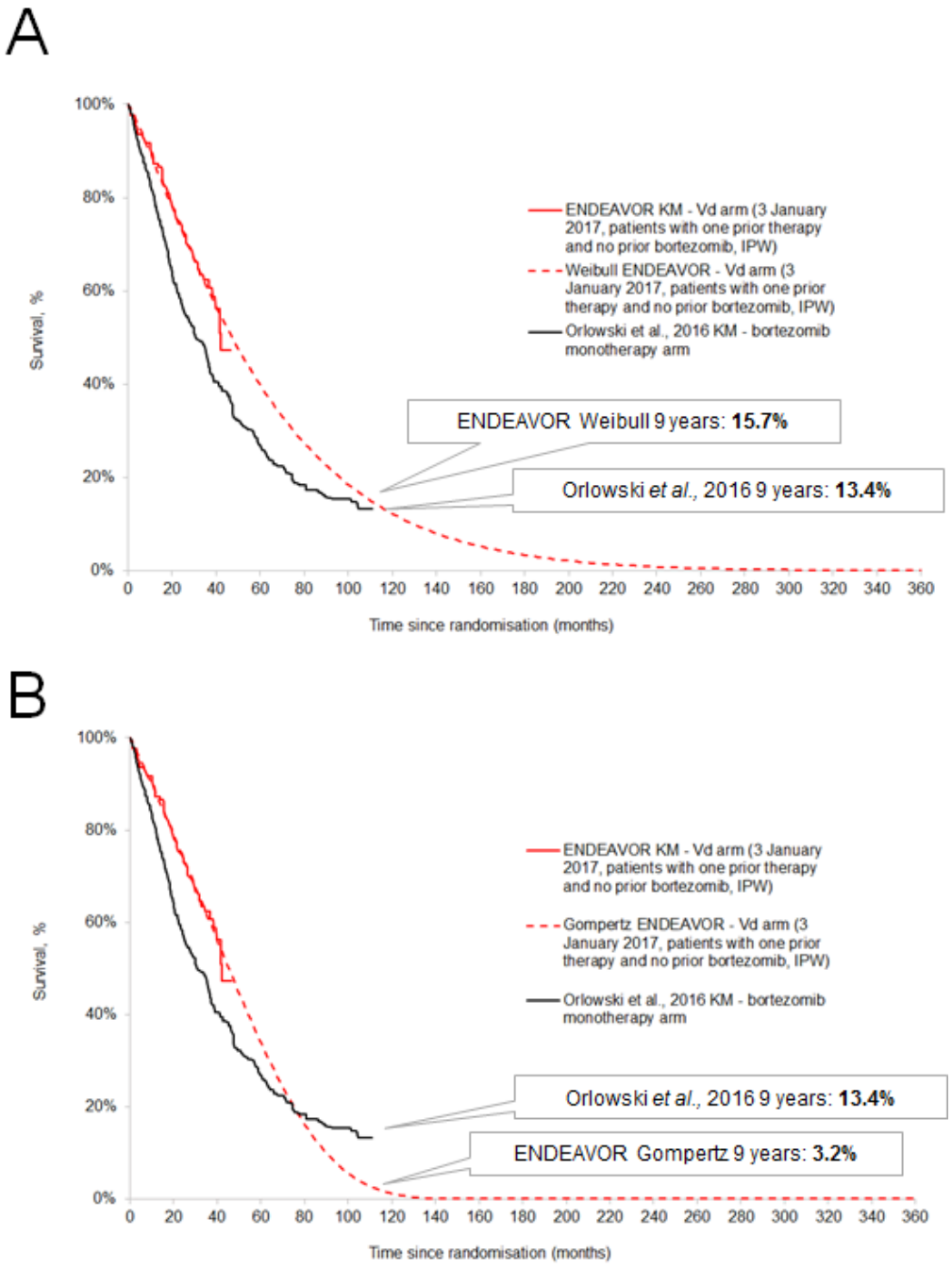
Figure 3-1 Comparison of Kaplan-Meier curves of OS from the ENDEAVOR study (ITT population and patients who have received one prior therapy and no prior bortezomib) and long-term data published by Orlowski *et al.*

Note: New ENDEAVOR OS data are from the second prespecified interim analysis of OS (3 January 2017 data cut-off date).

ITT, intent-to-treat; KM, Kaplan-Meier; OS, overall survival; Vd, bortezomib/dexamethasone.

A comparison of the modelled OS curves for the Vd arm in the ENDEAVOR subgroup of interest using both the Weibull and Gompertz functions versus the observed long-term data from the bortezomib monotherapy arm in the study by Orlowski *et al.* is provided in Figure 3-2.

Figure 3-2 Comparison of Kaplan-Meier curves of OS from Orlowski *et al.* and ENDEAVOR (patients who have received one prior therapy and no prior bortezomib) vs. modelled OS curves using (A) Weibull and (B) Gompertz parametric functions to extrapolate OS beyond the end of the trial



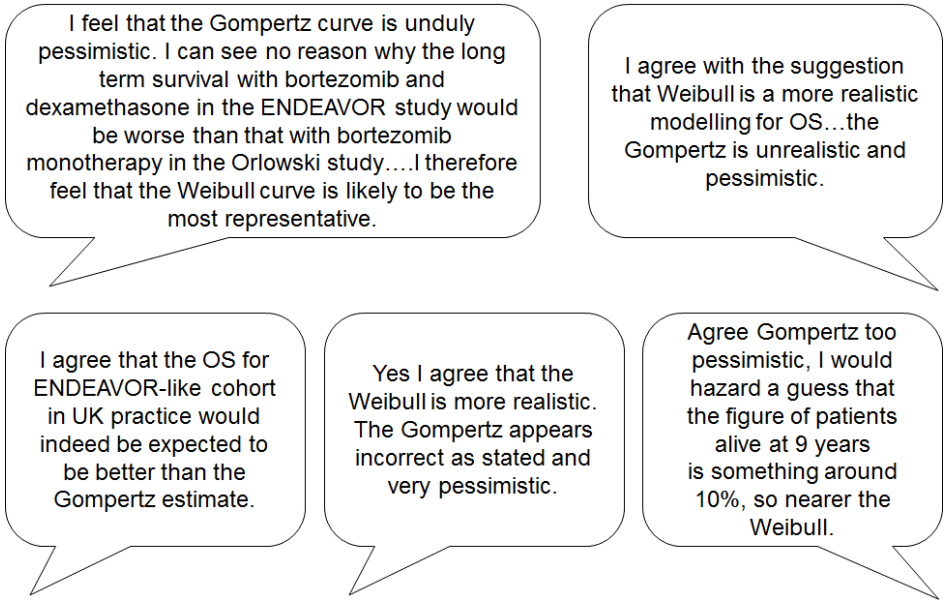
Note: New OS data are from the second prespecified interim analysis of OS in ENDEAVOR (3 January 2017 data cut-off date).

IPW, inverse probability weighted; KM, Kaplan-Meier; OS, overall survival; Vd, bortezomib/dexamethasone.

The Gompertz function does not provide a clinically plausible model of long-term OS for the ENDEAVOR subgroup of interest as the Gompertz curve falls below the Orlowski *et al.* curve after approximately 7 years, and predicts that just 3.2% of patients will be alive at 9 years which is substantially less than the 13.4% observed in the Orlowski *et al.* study. This is highly implausible, as for the reasons outlined above, the survival rates from Orlowski *et al.* are likely to represent an already conservative proxy for the ENDEAVOR subgroup of interest. In contrast, the shapes of the ENDEAVOR curves using the Weibull function and the long-term OS data from Orlowski *et al.* are very well matched. Survival rates at 9 years based on the ENDEAVOR Weibull curve (15.7%) and observed data from the Orlowski *et al.* study (13.4%) are also very similar. These comparisons show that the Weibull function provides a highly clinically plausible model of long-term OS for the ENDEAVOR subgroup of interest.

In order to further substantiate the case for Weibull over Gompertz, we sought feedback from 13 consultant haematologists practicing in England on the clinically plausibility of the Weibull versus the Gompertz OS extrapolations. Of the 12 who responded, 11 (92%) indicated that the Weibull curve was the most clinically plausible; the remaining respondent did not indicate a preference for either curve. Excerpts from these responses are provided in Figure 3-3 and the full set of responses are provided in Appendix E. This feedback provides an overwhelming external clinical validation of our approach and use of the Weibull function for extrapolation of OS from ENDEAVOR.

Figure 3-3 Excerpts from consultant haematologist responses on the clinical plausibility of the modelled Weibull and Gompertz curves for Vd in patients who have received one prior therapy and no prior bortezomib



OS, overall survival.

Indeed, NICE DSU TSD14 states that '*… For example, if a registry states that 5-year survival for a particular disease is 10%, parametric models that result in 0% survival at 5 years may not be appropriate, and neither may be those that estimate 40% survival at 5 years.*'[5] Therefore, we strongly believe that it is inappropriate to choose a parametric function (Gompertz) that predicts substantially fewer patients survive in the long-term than what is likely to be an already conservative proxy (the Orlowski *et al*. study). In contrast, the Weibull function

results in highly plausible projections that are clinically congruent with the data from the Orlowski *et al.* study endorsing Weibull as the most clinically plausible curve fit**.**

In conclusion, long-term OS data from the Orlowski *et al.* study (median 8.6 years follow-up), supplemented by overwhelmingly consistent feedback from consultant haematologists, provide a robust clinical validation of our OS modelling approach. The Gompertz function is not only very conservative but results in clinically implausible model curves and long-term OS projections, and is therefore an inappropriate choice for extrapolation of OS from ENDEAVOR. In contrast, the Weibull function results in clinically plausible model curves and long-term OS projections, and is therefore the most appropriate choice for extrapolation of OS from ENDEAVOR.

4 Discussion and conclusion

All of the cost-effectiveness analyses reported in this document incorporate the following preferred modelling assumptions considered appropriate and accepted as suitable for decision-making by the Committee:

- IPW-based survival model jointly fitted to both ENDEAVOR study arms
  - Including clinical expert-identified prognostic covariates retained in a stepwise selection procedure (Appendix A)
  - Assuming proportional hazards
- Utility values mapped from ENDEAVOR using Proskorovsky *et al.,* 2014[1]
- Bortezomib complex PAS assumed to translate to 15% cost saving for the NHS
- Bortezomib treatment capped at a maximum of 24 weeks (8 Cycles) with efficacy adjusted using the ENDEAVOR:CASTOR MAIC (Appendix B)

We have incorporated the new and more mature ENDEAVOR OS data into our final base case cost-effectiveness analysis, and have also provided compelling evidence to demonstrate that the Weibull function is the most appropriate for extrapolation of OS. The final base case ICER for Cd at £27,629 (Table 4-1) remains firmly below £30,000 per QALY, demonstrating that it represents a cost-effective use of NHS resources for the treatment of patients who have received one prior therapy and no prior bortezomib. This ICER may potentially be conservative given we have assumed that the bortezomib complex PAS translates into a saving of 15% in drug acquisition costs for the NHS. This does not take into account the costs of administering the complex PAS scheme and the likely negative impact on its uptake resulting from the high administrative burden.[12]

Table 4-1 Final base case cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib using the new ENDEAVOR OS data

	Total costs (£)	Total LYGs	Total QALYs	Inc. costs (£)	Inc. LYGs	Inc. QALYs	ICER (£) inc. (QALYs)
Vd	£69,626	3.34	2.20				
Cd	£118,077	5.87	3.96	£48,451	2.54	1.75	£27,629

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYGs, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone.

We are confident that the robust new evidence presented in this document removes the cloud of uncertainty around the most appropriate parametric function for extrapolation of OS from ENDEAVOR, supports our base case extrapolation using a Weibull distribution, and will allow the Committee to make a positive recommendation on the use of Cd in the subgroup of patients who have received one prior therapy and no prior bortezomib.

References

1. Proskorovsky I, Lewis P, Williams CD *et al.* Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health Qual Life Outcomes* 2014;12:35.
2. Dimopoulos M, Goldschmidt H, Niesvizky R *et al.* Overall Survival of Patients with Relapsed or Refractory Multiple Myeloma Treated with Carfilzomib and Dexamethasone versus Bortezomib and Dexamethasone in the Randomized Phase 3 ENDEAVOR Trial. 2017. Oral presentation at the 16th International Myeloma Working Group Workshop. 1-4 March 2017. New Dehli, India.
3. Amgen. Phase 3 Head-to-Head Trial Showed KYPROLIS® (Carfilzomib) Significantly Improved Overall Survival Compared To Velcade® (Bortezomib) In Relapsed Or Refractory Multiple Myeloma Patients. Press release. 2017. Available from: http://wwwext.amgen.com/media/news-releases/2017/02/phase-3-headtohead-trial-showed-kyprolis-carfilzomib-significantly-improved-overall-survival-compared-to-velcade-bortezomib-in-relapsed-or-refractory-multiple-myeloma-patients/ (Accessed March 2017)
4. Amgen. Amgen Presents Overall Survival Data From KYPROLIS® (carfilzomib) Phase 3 ENDEAVOR Trial At 16th International Myeloma Workshop. Press release. 2017. Available from: http://wwwext.amgen.com/media/news-releases/2017/03/amgen-presents-overall-survival-data-from-kyprolis-carfilzomib-phase-3-endeavor-trial-at-16th-international-myeloma-workshop/ (Accessed March 2017)
5. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from: http://www.nicedsu.org.uk/Survival-analysis-TSD(2892878).htm (Accessed March 2017).
6. Orlowski RZ, Nagler A, Sonneveld P *et al.* Randomized phase III study of pegylated liposomal doxorubicin plus bortezomib compared with bortezomib alone in relapsed or refractory multiple myeloma: combination therapy improves time to progression. *J Clin Oncol* 2007;25:3892-901.
7. 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;122:2050-6.
8. 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;17:27-38.
9. European Medicines Agency. Velcade EPAR public assessment report - variation. 2013. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000539/WC500161881.pdf (Accessed January 2017).
10. Kumar SK, Dispenzieri A, Lacy MQ *et al.* Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia* 2014;28:1122-8.
11. Cancer Research UK. Myeloma survival trends over time. 2014. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/survival#heading-Two (Accessed March 2017).
12. Williamson S, Thomson D. A report into the uptake of patient access schemes in the NHS. . *Clinical Pharmacist* 2010;2.

Appendix A – Update to the covariate stepwise selection model with new ENDEAVOR overall survival data

Updated results from the stepwise selection model (originally reported in our response to the ACD) with the new ENDEAVOR OS data, used to identify covariates for inclusion in covariate adjusted-treatment effect estimates and the IPW-based survival model, are provided in Table A-1.

Table A-1 Variables retained within the stepwise selection models and associated hazard ratios: Cd vs Vd OS model using the old versus new ENDEAVOR OS data (patients with one prior therapy and no prior bortezomib)

Covariate	Stepwise variable selection model based on previous ENDEAVOR OS data[a] HR (95% CI)	Updated stepwise variable selection model based on new ENDEAVOR OS data[b] HR (95% CI)
Treatment (Cd vs. Vd)	0.631 (0.384, 1.039)	0.615 (0.397, 0.953)
Prior SCT (yes vs. no)[c]	2.034 (1.125, 3.678)	1.890 (1.111, 3.215)
Prior lenalidomide (yes vs. no)[c]	-	0.652 (0.384, 1.108)
Age (≥ 65 vs. < 65 years)	-	-
ECOG PS (1-2 vs. 0)	2.154 (1.265, 3.665)	1.840 (1.166, 2.904)
Creatinine clearance (≥50 to <80 mL/min vs. other)	-	-
Creatinine clearance (>80 mL/min vs. other)	0.496 (0.264, 0.931)	0.605 (0.356, 1.028)
Time from diagnosis	0.986 (0.976, 0.997)	0.986 (0.977, 0.995)
Time from last relapse	-	-
ISS stage (II-III vs. I)	-	-
β_2-microglobulin (≥3.5 vs. <3.5 mg/L)	2.926 (1.634, 5.238)	3.040 (1.818, 5.083)
Refractory to last prior treatment (yes vs. no)	2.010 (1.158, 3.487)	1.943 (1.176, 3.208)
AIC	**640.675**	**824.582**

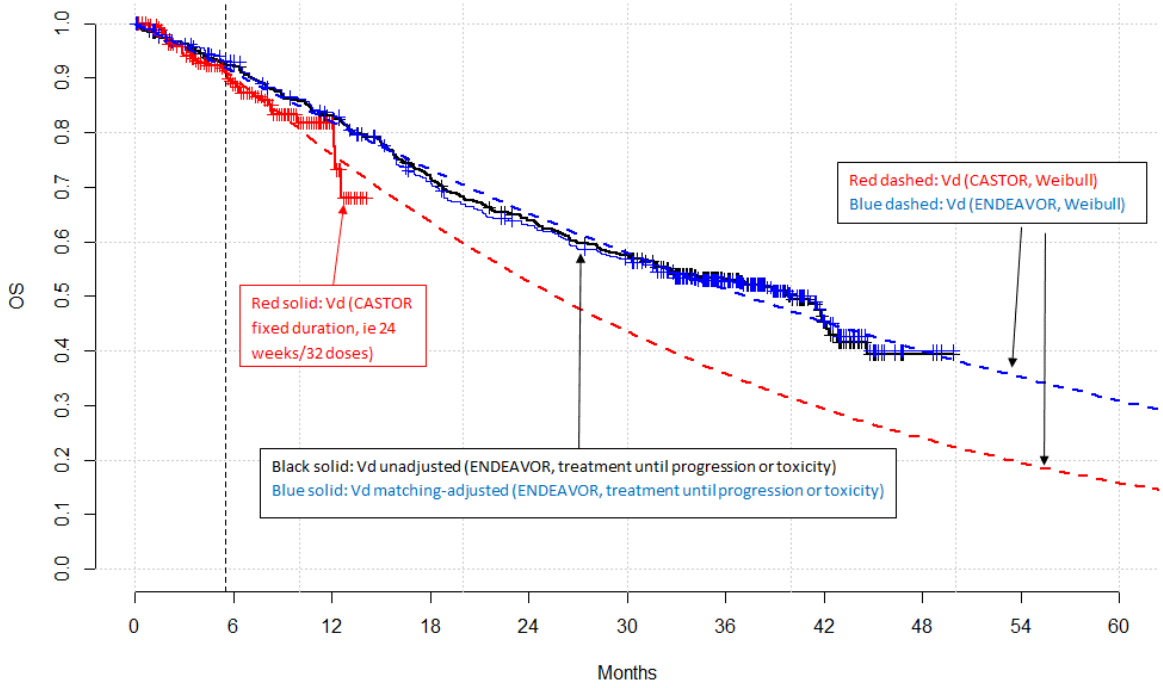
[a] Previous OS data are from the EMA ad-hoc analysis (3 March 2016 data cut-off date).
[b] New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).
[c] The rows for prior lenalidomide and prior SCT were incorrectly labelled the wrong way round in Appendix A (Table 33) of the ACD response.

ACD, appraisal consultation document; AIC Akaike information criterion; Cd, carfilzomib/dexamethasone; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EMA, European Medicines Agency; ISS, International Staging System; OS, overall survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

Appendix B – Update to the ENDEAVOR:CASTOR matching-adjusted indirect comparison with the new ENDEAVOR overall survival data

Updated results from the ENDEAVOR:CASTOR MAIC (originally reported in our response to the ACD) with the new ENDEAVOR OS data, used to estimate the impact of continued treatment with bortezomib beyond 8 cycles/24 weeks, are provided in Figure B-1 and Table B-1. The revised base case ICER with the updated MAIC results is £27,629 (Section 4), which is similar to the equivalent ICER when HRs from the previous MAIC are used instead (£28,677).

Figure B-1 Updated OS curves from the updated MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity) using the new ENDEAVOR OS data



Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR). However, the proportions of patients with high cytogenetic risk, defined similarly across studies, was similar in ENDEAVOR and CASTOR (24.3% and 21.3% in the Vd arms) suggesting that any residual bias is likely to be minimal. New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

CI, confidence interval; EMA, European Medicines Agency; IMiD, immunomodulatory drug; ISS, international staging system; MM, multiple myeloma; PFS, progression-free survival; OS, overall survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

Table B-1 Updated OS hazard ratios (first 24 weeks vs. after 24 weeks) from the MAIC analysis of Vd in CASTOR (8 cycles; 24 weeks) and ENDEAVOR (treatment until progression or unacceptable toxicity) using the new ENDEAVOR OS data

	Hazard ratio using previous ENDEAVOR OS data[a]	Hazard ratio using new ENDEAVOR OS data[b]
Piecewise matching-adjusted (first 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.157 (0.679, 1.972)	1.099 (0.648, 1.863)
Piecewise matching-adjusted (after 8 cycles/24 weeks), HR (95% CI) CASTOR:ENDEAVOR	1.561 (0.942, 2.588)	1.608 (0.973, 2.660)
Relative hazard increase when stopping bortezomib after 24 weeks, HR (95% CI) CASTOR:ENDEAVOR[c]	**1.349 (0.684, 2.662)**	**1.465 (0.748, 2.869)**

Note: MAIC matched for all covariates identified as prognostic factors for outcomes in MM by UK clinical experts reported for both the ENDEAVOR and CASTOR studies: age, ISS stage, number of prior therapies, prior bortezomib exposure, prior IMiD exposure (used as a proxy for prior lenalidomide), baseline creatinine clearance, time since diagnosis, prior SCT, and refractory to last prior treatment. An exception was cytogenetic risk status given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR). However, the proportions of patients with high cytogenetic risk, defined similarly across studies, was similar in ENDEAVOR and CASTOR (24.3% and 21.3% in the Vd arms) suggesting that any residual bias is likely to be minimal.

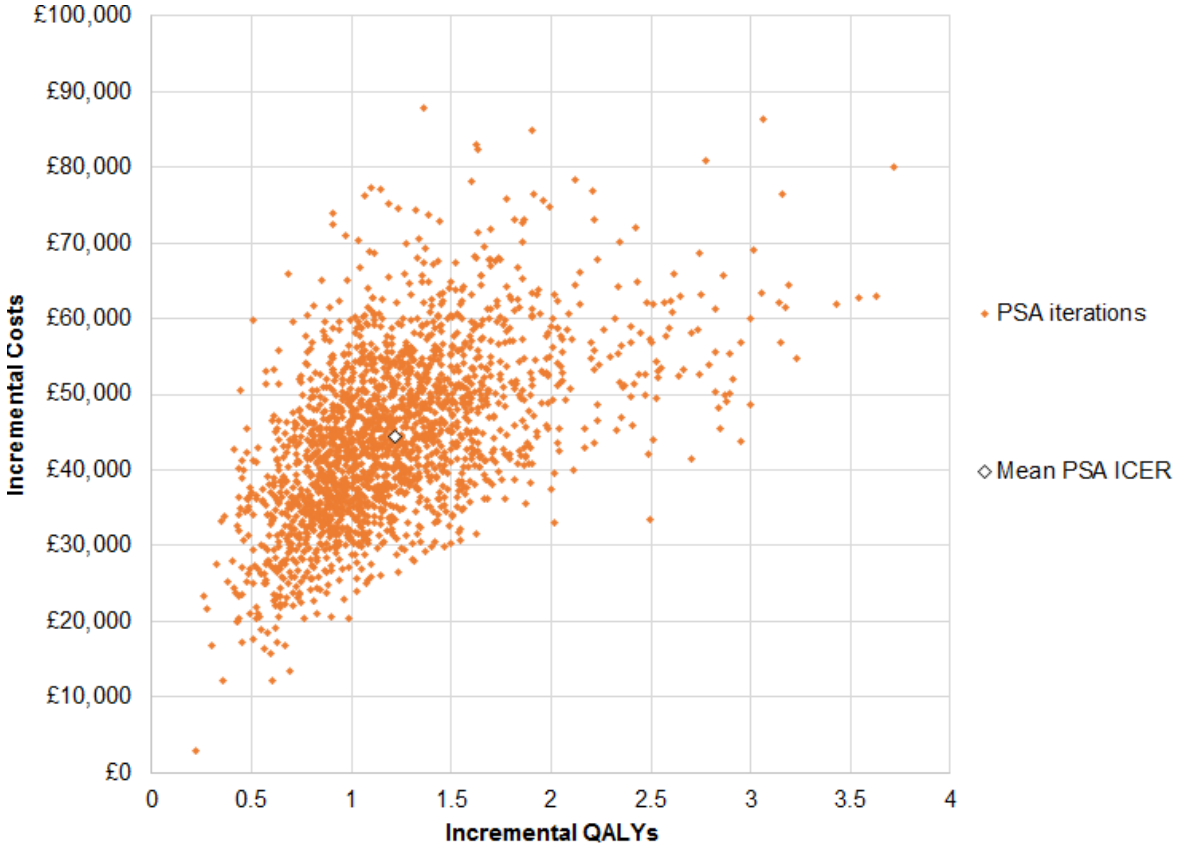
[a] Previous OS data are from the EMA ad-hoc analysis (3 March 2016 data cut-off date).
[b] New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).
[c] HR for first 8 cycles (24 weeks) versus subsequent cycles.

CI, confidence interval; EMA, European Medicines Agency; HR, hazard ratio; IMiD, immunomodulatory drug; ISS, international staging system; MM, multiple myeloma; OS, overall survival; SCT, stem cell transplant; Vd, bortezomib/dexamethasone.

Appendix C – Update to the probabilistic sensitivity analyses

A scatter plot of incremental costs and QALYs for Cd versus Vd based on of our final base case cost-effectiveness analysis using the new ENDEAVOR OS data (Section 4) is provided in Figure C-1.

Figure C-1 Scatter plot of incremental costs and QALYs: Cd versus Vd using the new ENDEAVOR OS data (patients with one prior therapy and no prior bortezomib)

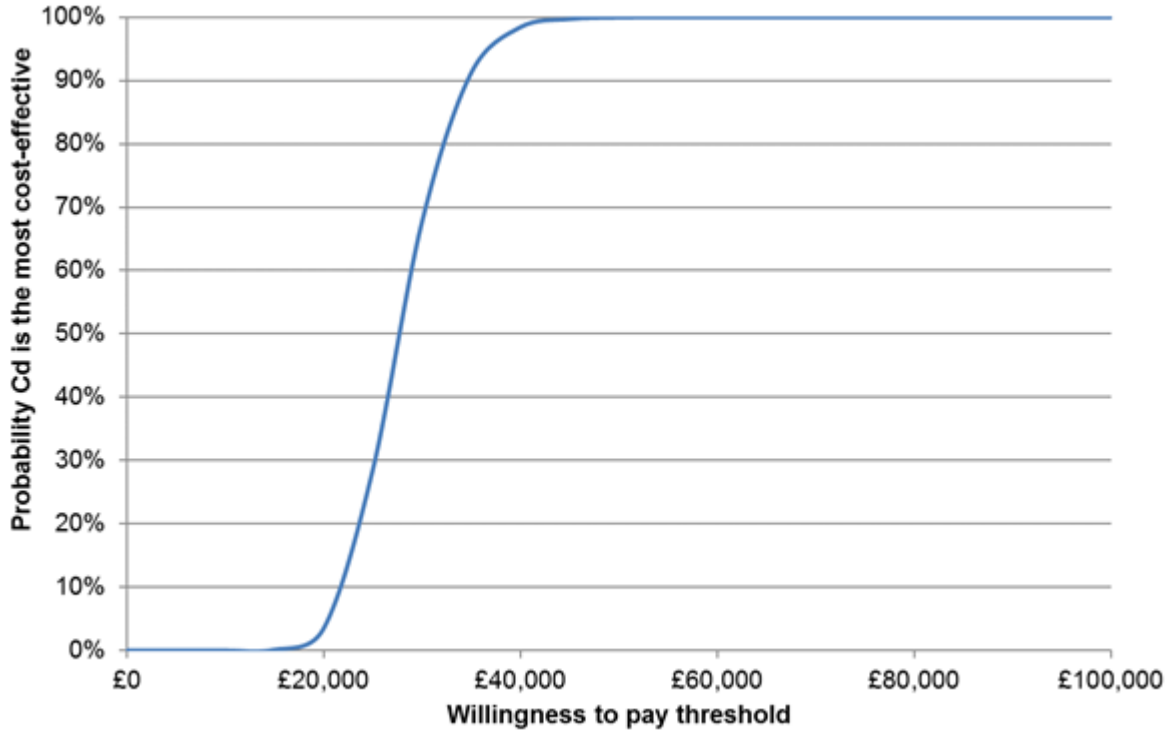


Note: New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; OS, overall survival; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life year; Vd, bortezomib/dexamethasone.

A cost-effectiveness acceptability curve for Cd versus Vd based on of our final base case cost-effectiveness analysis using the new ENDEAVOR OS data (Section 4) is provided in Figure C-2. This shows that there is a 67.6% chance of Cd being cost effective at a willingness-to-pay threshold of £30,000 per QALY gained (Figure C-2).

Figure C-2 Cost-effectiveness acceptability curve: Cd versus Vd using the new ENDEAVOR OS data (patients with one prior therapy and no prior bortezomib)



Note: The probability of Cd being cost effective at a willingness to pay threshold of £30,000 per QALY gained is 67.55%. New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

Cd, carfilzomib/dexamethasone; OS, overall survival; Vd, bortezomib/dexamethasone.

Appendix D – Modelled versus Kaplan-Meier overall survival curves and statistical fit values for the Weibull and Gompertz functions using the new ENDEAVOR overall survival data

Updated OS Weibull and Gompertz curve fits for Cd and Vd using the new ENDEAVOR OS data in the subgroup of patients who have received one prior therapy and no prior bortezomib are shown in Figure D-1. Based on visual inspection of the modelled curves versus the Kaplan-Meier curves, both the Weibull and Gompertz distributions fit the trial data well and there is little to justify the choice of one over the other. This was also the case with the previous ENDEAVOR OS data.

Figure D-1 Weibull and Gompertz modelled vs. Kaplan-Meier Vd OS curves based on the new ENDEAVOR OS data (patients who have received one prior therapy and have not received prior bortezomib)

Note: New ENDEAVOR OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

Cd, carfilzomib/dexamethasone; KM, Kaplan-Meier; OS, overall survival; Vd, bortezomib/dexamethasone.

The corresponding AIC and BIC values for the modelled Weibull and Gompertz curves shown above are provided in Table D-1. The AIC and BIC values are extremely similar for the Weibull and Gompertz curves suggesting there little to justify the choice of one over the other based on statistical fit of the two curves.

Table D-1 AIC and BIC for the Weibull and Gompertz OS curves jointly-fitted to both study arms based on the new ENDEAVOR OS data (patients who have received one prior therapy and no prior bortezomib)

	AIC	BIC
Weibull	940.741	951.502
Gompertz	939.598	950.360

Note: New OS data are from the second prespecified interim analysis (3 January 2017 data cut-off date).

AIC, Akaike information criterion; BIC, Bayesian information criterion; OS, overall survival.

Appendix E – Clinical expert feedback on the plausibility of the modelled Weibull and Gompertz overall survival curves

We contacted 13 consultant haematologists practicing in England to ask for their opinions on the clinical plausibility of the modelled Weibull and Gompertz OS curves for Vd in the ENDEAVOR subgroup of interest (i.e. patients who have received one prior therapy and no prior bortezomib). The vast majority of the respondents (11/12) suggested that the Weibull curve is the most clinically plausible, with many also highlighting that predictions of long-term survival from the Gompertz curve are very pessimistic. The responses are provided in full in Table E-1.

Table E-1 Responses from consultant haematologists practicing in England on the clinical plausibility of the modelled Weibull and Gompertz curves for Vd in patients who have received one prior therapy and no prior bortezomib

Response
"The Weibull is the most logical. Your points are clear and well made."
"Agree – Weibull OS extrapolation is the most realistic"
"Thank you for asking this question. The two curves are certainly interesting. I feel that the Gompertz curve is unduly pessimistic. I can see no reason why the long term survival with bortezomib and dexamethasone in the ENDEAVOR study would be worse than that with bortezomib monotherapy in the Orlowski study. As well as the points you make which I fully support, there is also data from Shaji Kumar demonstrating the improved OS over the decades and with the impact of newer therapies. Orlowski described outcomes for patients treated between 2004-2006, whereas ENDEAVOR was 2012-2014. This supports your statement regarding the improved range of subsequent line therapies resulting in improved outcomes. I therefore feel that the Weibull curve is likely to be the most representative."
"Agree Gompertz too pessimistic, I would hazard a guess that the figure of patients alive at 9 years is something around 10%, so nearer the Weibull."
"The Weibull extrapolation seems more realistic."
"I have looked at this and agree regarding the Weibull being the better estimate."
"I agree with the suggestion that Weibull is a more realistic modelling for OS. As you state the Gompertz is unrealistic and pessimistic and we would expect at least 25% of patients to be alive at 10 years – there are significant flaws with the Orlowski comparison for the reasons you state – I also note that there is a bias towards earlier stage patients in the Orlowski group. We now have greater availability of efficacious therapy beyond 3rd line and survival is significantly improving – we are seeing significant improvements in 10 year OS with more recent patient cohorts as a direct result of the introduction of newer therapies that include carfilzomib."
"I agree that the OS for ENDEAVOR-like cohort in UK practice would indeed be expected to be better than the Gompertz estimate, especially as most patients receiving VelDex 2nd-line would be able to receive Lenalidomide subsequently (as per NICE), as well as VelDexPanobinostat provided Vel-responsive/tolerant, and additionally Pomalidomide has now been relisted by NICE for subsequent treatment line (with OS benefit even after Len-refractoriness). ONS real-world figures for MM survival demonstrate improvement in recent years with access to newer agents, so Orlowski curve from mid-2000 would expect to be exceeded."

Response
“I agree with you in considering the Weibull extrapolation more accurate.” “It is true that we have more options available now but it is also true that part of the Orlowski patients received other new and effective treatments after they relapsed, especially in US. In other words part of the benefit of having new drugs available can be seen already in Orlowski’s curves. The curves are related to clinical trials and one argument could be that this is not representative of the general myeloma population who are generally less fit than the population in clinical trials. In this respect the drugs now available (Pomalidomide and Panobinostat) will help to improve the outcome in particular of the non-fit patients due to their good tolerability profile. The introduction of these new drugs should balance the fact that we are discussing on selected populations.”
“The latest data from CRUK (http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/survival#heading-Two) gives an age standardised 10 year survival (in England and Wales) for myeloma of 21.4% for patients diagnosed in 2005-6; estimates for those diagnosed 2010-11 are currently 32.5%. These data pre-date some of the additional lines of therapy that have since come along. Whilst they are survival from diagnosis, I would agree that the Gompertz extrapolation of 3.2% at 9 years from second line therapy is too pessimistic. Assuming that first line response is 24-30 months, and taking the 2005-6 estimates from CRUK, then that would imply a drop from 21.4% at 10 years to 3.2% at 11-12 years- clearly not likely. The Weibull extrapolation of 15.7% is a better fit, so yes, I agree with your interpretation.”
“Yes I agree that the Weibull is more realistic. The Gompertz appears incorrect as stated and very pessimistic. I do not understand the methodology but agree with all the points raised. You can also show Shaji Kumar survival curves from Mayo publications in Blood showing incremental improvement in OS and all points below are very valid.”
“Firstly, as a trialist and clinician, I have a very limited ability to criticise or critique either analysis. Secondly, the Orlowski study isn't a particular good comparator, as we don't use bortezomib monotherapy and Bortezomib/lipsomal Dox isn't used at all in the UK. But I'm sure you have considered this.”
Note: One of the consultant haematologists we contacted had not responded by the time of submission of this addendum. CDF, Cancer Drugs Fund; CRUK, Cancer Research UK; Dex, dexamethasone; NICE, National Institute of Health and Care Excellence; ONS, Office for National Statistics; OS, overall survival; Vel, Velcade (bortezomib)

Carfilzomib for previously treated multiple myeloma

ERG's review of the company's additional evidence

April 2017

TABLE OF CONTENTS

LIST OF TABLES

1. Changes made to the Company's updated model

The Company submitted an updated model, which incorporated the changes to their base case analysis that were requested by the Committee at the first Appraisal Committee Meeting (AMC). These changes were: capping bortezomib treatment at 8 treatment cycles; adjusting the efficacy of bortezomib beyond the 8th cycle to account for reduced treatment duration; and to apply the estimate of 15% for the bortezomib patient access scheme (PAS) discount. This is similar to the ERG's previous base case analysis, with the exception of the bortezomib efficacy adjustment and the choice of overall survival (OS) curve used.

The Company's updated model has based its estimates of efficacy on an updated set of data, with a cut-off date of 3rd January 2017 (previous model based on 3rd March 2016 cut-off). The updated analysis using the same stepwise variable selection process, showed an increased treatment effect in favour of carfilzomib, with a hazard ratio (HR) of 0.615 in comparison to the previously estimated HR of 0.631, The Company do not state which model is used to derive this HR, but the ERG note that 0.631 was a value previously presented as derived from a Cox proportional hazards (PH) model. The ERG would like to highlight that no other variable selection methods appear to have been tested on the updated data, which may have shown different results. The updated analysis using the stepwise approach included an additional covariate for prior lenalidomide, which has a large confidence interval around the coefficient. This shows the uncertainty in the estimated treatment effect and also demonstrates the instability in the selection method.

The ERG are concerned that the treatment effect may be overestimated, and would like to reiterate the variability of the HR estimated using the different selection methods on the Cox PH model in the Company's previous analyses. The least absolute shrinkage and selection (LASSO) method, which accounts for the uncertainty in the coefficient estimates, resulted in an OS HR of 0.661; higher than the stepwise method of 0.631. The OS HR based on the Weibull curve in the Company's updated model is 0.669, so a more reliable HR may be closer to the Company-reported unadjusted HR of [redacted]. The ERG used the unadjusted HR (based on the updated data cut) in a scenario analysis around the Company's base case analysis and the ERG's preferred base case analysis, and the results are presented in Section 3.2.1 and 3.4.1, respectively.

The updated OS data were also used to update the analysis used to estimate the bortezomib efficacy adjustment beyond 8 treatment cycles, which was done using a matched adjusted indirect comparison (MAIC) performed by the Company as a scenario analysis in the Company's response to the first

Appraisal Consultation Document (ACD). The updated analysis resulted in a greater reduction in efficacy for bortezomib following the discontinuation of treatment, but this effect was not significant and had a large confidence interval. The ERG still consider this analysis to be unreliable as it is an "unanchored" MAIC and so is likely to be confounded by unobserved (and so unadjusted for) prognostic indicators. In addition, the company has disclosed that their MAIC does not adjust for all observed prognostic indicators – specifically cytogenetic risk status – and the HR estimated could be very sensitive to its incorporation into the MAIC. The ERG does not consider that the proportion of patients at high risk being similar in ENDEAVOR and CASTOR to be suitable reassurance for not including cytogenetic risk status in the MAIC, in particular given the high proportion of patients with missing/unknown data in ENDEAVOR (13.2% and 29.6% of the Vd arms in ENDEAVOR and CASTOR). Therefore, the ERG considers assuming that there is no effect following treatment discontinuation as a more conservative estimate. A range of scenarios to show the increasing effect of a reduction in efficacy for bortezomib is explored in Section 3.4.4.

2. Plausibility of Weibull and Gompertz for OS

A key source of uncertainty in both the Company's and the ERG's previous analyses, was in the extrapolation of the OS curves. The Company presented an argument that the jointly fitted Weibull distribution resulted in the best fitting curve, which they validated using longer term survival data from the PANORAMA-1 trial, which had a follow up period of around 5 years.[(1)] The ERG took the same approach but found that the better fitting curve was the Gompertz distribution, from a statistical perspective and from validation using PANAORAMA-1 data.

The Company have now submitted evidence showing survival data from Orlowski *et al.*, with a median follow up of 8.6 years.[(2)] Orlowski *et al.* does provide robust long-term OS data for bortezomib-treated relapsed/refractory multiple myeloma (R/RMM) patients up to 7–8 years of follow-up and patient baseline characteristics reported in both ENDEAVOR and Orlowski *et al.* are similar. The studies are also similar in terms of bortezomib dosing schedule and the lack of prior bortezomib exposure (ENDEAVOR subgroup). However, there are also several important differences between the studies; bortezomib was administered as a monotherapy in the Orlowski *et al.* study but in combination with dexamethasone (Vd) in the ENDEAVOR study, and the median duration of treatment was substantially longer in ENDEAVOR compared to Orlowski *et al.* (188 days and 105 days, respectively). It is unclear if the studies were similar in terms of the proportion of patients from Western Europe and subsequent therapies received.

Overall, Orlowski *et al.* patients are likely to have a worse prognosis than patients in ENDEAVOR. Hence, the Kaplan-Meier (KM) curve for the Vd treated patients in ENDEAVOR would not be expected to cross the survival curve for bortezomib treated patients in Orlowski *et al*. The Company demonstrates that the ERG's originally preferred Gompertz distribution crosses the Orlowski *et al.* KM curve after approximately 7 years, and results in a much lower proportion of patients predicted to be alive at 9 years in comparison to the KM data (3.2% compared to 13.4%, respectively). The Company conclude that the Gompertz extrapolation is, therefore, clinically implausible, as the population in Orlowski *et al.* would be expected to have a poorer prognosis resulting in a survival curve that remains lower than that for the ENDEAVOR trial population. The Company's preferred Weibull curve generated a higher proportion of patients being alive at 9 years (15.7% compared to 13.4% for Orlowski *et al.*), which the Company considered more plausible.

Although the Weibull curve does appear to give a more plausible survival extrapolation compared with the Gompertz curve, this does not necessarily mean that it gives the best fit compared to other standard curves. Also, there is substantial uncertainty around the survival estimate at 9 years follow up for Orlowski *et al.,* as this is the tail of the KM data, which has fewer patients at risk. A more reliable estimate at 95 months (7.8 years) shows a survival of 15% in Orlowski *et al.* and around 24% using the Weibull extrapolation, which is a substantial difference.

Given the expected prognosis of the population in Orlowski *et al.* the ERG agrees with the Company that the Orlowski *et al.* study is a reasonable justification for using the Weibull curve for OS in preference to the Gompertz curve. Therefore, the ERG's updated base case analysis, presented in Section 3.3, incorporates this change. However, the ERG would like to highlight that no further assessment of alternative distributions was provided by the Company for the updated OS data, which may have resulted in a different distribution as the best fit. The ERG considers that, while the Gompertz distribution may produce an overly pessimistic curve, the Weibull may be producing an overly optimistic curve. A better fitting curve may lie somewhere between the two, albeit given the data in Orlowski *et al.,* probably closer to the Weibull. Given the uncertainty, the ERG have presented a scenario around the Company's base case and the ERG's preferred base case analysis using the Gompertz distribution for OS. The results of these are given in Section 3.2.2 and 3.4.2, respectively.

3. Results

3.1. Company's updated base case analysis

The Company's updated base case analysis results are given in Table 1, showing an incremental cost-effectiveness ratio (ICER) of £27,629 per QALY gained. The key components of the Company's base case are:

- Jointly fitted Weibull models for progression-free survival (PFS) and OS;
- Bortezomib treatment schedule capped at 8 treatment cycles;
- Bortezomib PAS estimated to be a 15% discount; and
- Bortezomib efficacy adjustment of 1.36 and 1.46 applied for PFS and OS, respectively. Note that in the Company's previous analysis, which was based on the 3rd March 2016 cut-off, the OS adjustment value was 1.35.

Table 1. Final base case cost-effectiveness results for Cd versus Vd in patients with one prior therapy and no prior bortezomib using the new ENDEAVOR OS data (Adapted from Company's additional evidence submission, Page 19, Table 4-1)

	Total costs (£)	Total LYGs	Total QALYs	Inc. costs (£)	Inc. LYGs	Inc. QALYs	ICER
Vd	£69,626	3.34	2.20	-	-	-	-
Cd	£118,077	5.87	3.96	£48,451	2.54	1.75	£27,629

Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYGs, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone.

3.2. ERG's scenario analyses around the Company's updated base case

3.2.1. Scenario analysis using unadjusted OS HR

A scenario using the unadjusted OS HR of [redacted], as reported by the Company, is presented in Table 2. This aims to assess the impact of the potentially underestimated HR derived using the stepwise variable selection, which includes some variables with a large confidence interval for the estimate of the coefficient. The LASSO method, which accounts for this uncertainty, showed an increased HR when applied to the Cox PH model. Therefore, a more reliable HR may lie between that used in the Company's model, 0.669, and the unadjusted value of [redacted].

Table 2. Scenario analysis using unadjusted OS HR

Unadjusted HR for OS	Cd	Vd	Incremental value
Total costs	£115,658	£69,626	£46,032
Total LYG	5.53	3.34	2.19
QALYs	3.74	2.20	1.53
ICER	-	-	£29,995

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

3.2.2. Scenario analysis using jointly fitted Gompertz for OS

As the ERG considered that the Weibull to be potentially over optimistic in its extrapolation, a scenario using the more pessimistic curve from the Gompertz distribution is presented in Table 3.

Table 3. Scenario analysis using jointly fitted Gompertz for OS

Unadjusted HR for OS	Cd	Vd	Incremental value
Total costs	£113,486	£69,466	£44,019
Total LYG	4.52	2.97	1.56
QALYs	3.10	1.97	1.13
ICER	-	-	£39,052

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

3.2.3. Scenario analysis using 20% estimate for bortezomib PAS

The ERG conducted a scenario analysis to test the impact of an increased PAS estimate for bortezomib on the results of the Company's base case analysis. The bortezomib PAS states that replacement stock or credit will be given to the healthcare provider for patients who do not achieve at least a minimal response, i.e. those who have stable disease or progression, within the first 4 cycles of treatment. The best overall response rates from the ENDEAVOR trial are given in Table 4, which provides a lower bound for this proportion of patients, as some patients best overall response may have been below that in the first 4 treatment cycles. The proportion of progressive and stable diseased patients is 18.1% based on this data, so the Company's estimate may be an underestimate. Also, 8.2% of patients were not evaluable and so this potentially underestimates the proportion further. The estimate used by the ERG for the scenario analysis was taken as 20% to account for the potential underestimation from the data in Table 4.[3] The results of the scenario analysis are given in Table 5.

Table 4. Best overall response rates for bortezomib (adapted from ENDEAVOR CSR, Page 120, Table 23)

Best overall response	Vd (N=465)
Stringent complete response	9 (1.9%)
Complete response	20 (4.3%)
Very good partial response	104 (22.4%)
Partial response	157 (33.8%)
Minimal response	53 (11.4%)
Stable disease	53 (11.4%)
Progressive disease	31 (6.7%)
Unable to evaluate	38 (8.2%)
Abbreviations in table: CSR, clinical study report; Vd, bortezomib and dexamethasone.	

Table 5. Scenario analysis using 20% estimate for bortezomib PAS

Unadjusted HR for OS	Cd	Vd	Incremental value
Total costs	£118,077	£68,867	£49,210
Total LYG	5.87	3.34	2.54
QALYs	3.96	2.20	1.75
ICER	-	-	£28,062
Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; PAS, patient access scheme; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone			

3.3. ERG's updated preferred base case analysis

The ERG agreed with the use of the Weibull for the OS extrapolation, so the only difference between the ERG's updated base case analysis and the Company's is the removal of the bortezomib efficacy adjustment, due to the potential unreliability of the analysis and the lack of significance in the resulting HR adjustment. The results of the ERG's updated base case are shown in Table 6.

Table 6. Updated ERG's base case analysis

Company's updated base case	Cd	Vd	Incremental value
Total costs	£118,077	£69,626	£48,451
Total LYG	5.87	3.34	2.54
QALYs	3.96	2.20	1.75
ICER			£27,629
Removing the bortezomib efficacy reduction following treatment discontinuation.			
Total costs	£118,077	£75,417	£42,660
Total LYG	5.87	4.42	1.45
QALYs	3.96	2.91	1.05
ICER (compared to original base case)	-	-	£40,744

Company's updated base case	Cd	Vd	Incremental value
Final revised cost-effectiveness analysis with both changes included	**-**	**-**	**£40,744**

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

3.4. Scenario analyses around the ERG's base case analysis

3.4.1. Scenario analysis using unadjusted OS HR

As per the scenario around the Company's base case analysis that uses the unadjusted OS HR of [redacted], Table 7 shows the results of the equivalent analysis performed around the ERG's preferred base case.

Table 7. Scenario analysis using unadjusted OS HR

Gompertz scenario	Cd	Vd	Incremental value
Total costs	£115,658	£75,417	£40,241
Total LYG	5.53	4.42	1.11
QALYs	3.74	2.91	0.83
ICER			£48,598

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

3.4.2. Scenario analysis using jointly fitted Gompertz for OS

As the ERG considered that the Weibull to be potentially over optimistic in its extrapolation, a scenario using the more pessimistic curve from the Gompertz distribution is presented in Table 8.

Table 8. Scenario analysis using jointly fitted Gompertz for OS

Gompertz scenario	Cd	Vd	Incremental value
Total costs	£113,486	£74,083	£39,403
Total LYG	4.52	3.68	0.85
QALYs	3.10	2.44	0.66
ICER			£59,764

Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone

3.4.3. Scenario analysis using 20% estimate for bortezomib PAS

A scenario using the PAS discount of 20%, as per the analysis performed around the Company's base case in Section 3.2, is presented in Table 9.

Table 9. Scenario analysis using 20% estimate for bortezomib PAS

Bortezomib PAS estimate of 20%	Cd	Vd	Incremental value
Total costs	£118,077	£74,658	£43,419
Total LYG	5.87	4.42	1.45
QALYs	3.96	2.91	1.05
ICER			£41,469
Abbreviations in table: Cd, carfilzomib/dexamethasone; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; PAS, patient access scheme; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone			

3.4.4. Scenario analyses for bortezomib efficacy adjustment

A range of scenarios for the efficacy adjustment for bortezomib beyond the treatment cap of 8 cycles, is given in Table 10. This shows the impact on the results of increasing the PFS and OS HRs applied to the bortezomib group beyond the 8th treatment cycle in increments of 0.1 up to the point at which the ICER becomes less than £30k per QALY. In this case, up to a HR of 1.3. The final rows of Table 10 show the results of the Company's base case, which is based on a PFS HR of 1.36 and an OS HR of 1.46.

Table 10. Scenario analyses for bortezomib efficacy adjustment

Bortezomib efficacy adjustment	Cd	Vd	Incremental value
HR = 1.0 (base case)			
Total costs	£118,077	£75,417	£42,660
Total LYG	5.87	4.42	1.45
QALYs	3.96	2.91	1.05
ICER	-	-	£40,744
HR = 1.1			
Total costs	£118,077	£74,235	£43,842
Total LYG	5.87	4.12	1.75
QALYs	3.96	2.72	1.24
ICER	-	-	£35,324
HR = 1.2			
Total costs	£118,077	£73,121	£44,956
Total LYG	5.87	3.87	2.00
QALYs	3.96	2.55	1.41
ICER	-	-	£31,922
HR = 1.3			
Total costs	£118,077	£72,065	£46,012
Total LYG	5.87	3.65	2.23
QALYs	3.96	2.40	1.55
ICER	-	-	£29,612
Company's base case HRs (PFS HR = 1.36, OS HR = 1.46)			

Total costs	£118,077	£69,626	£48,451
Total LYG	5.87	3.34	2.54
QALYs	3.96	2.20	1.75
ICER			£27,629
Abbreviations in table: Cd, carfilzomib/dexamethasone; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; inc., incremental; LYG, life years gained; QALYs, quality-adjusted life years; Vd, bortezomib/dexamethasone			

REFERENCES

1. San-Miguel JF, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, 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. Lancet Haematol. 2016;3(11):e506-e15.
2. Orlowski RZ, Nagler A, Sonneveld P, Blade J, Hajek R, Spencer A, 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):2050-6.
3. Amgen Europe. Clinical Study Report (CSR) for the ENDEAVOR trial. Amgen data on file. 2016.