Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy with bortezomib (part-review of TA171). Celgene's response to Technology Appraisal Committee's Critique of the Economic Case

A Critique of the Response Submitted by Celgene

June 2014

Matrix and Matrix Knowledge are trading names of TMKG Limited (registered in England and Wales under registration number 07722300) and its subsidiaries: Matrix Decisions Limited (registered in England and Wales under registration number 07610972); Matrix Insight Limited (registered in England and Wales under registration number 06000446); Matrix Evidence Limited (registered in England and Wales under registration number 07538753); Matrix Observations Limited (registered in England and Wales under registration number 05710927); and Matrix Knowledge Group International Inc. (registered in Maryland, USA under registration number D12395794).

Disclaimer In keeping with our values of integrity and excellence, Matrix has taken reasonable professional care in the preparation of this document. Although Matrix has made reasonable efforts, we cannot guarantee absolute accuracy or completeness of information/data submitted, nor do we accept responsibility for recommendations that may have been omitted due to particular or exceptional conditions and circumstances.

Confidentiality This document contains information which is proprietary to Matrix and may not be disclosed to third parties without prior agreement.

Except where permitted under the provisions of confidentiality above, this document may not be reproduced, retained or stored beyond the period of validity, or transmitted in whole, or in part, without Matrix's prior, written permission.

© TMKG Ltd, 2012

Any enquiries about this report should be directed to enquiries@matrixknowledge.com

Report commissioned by:

NHS R&D HTA Programme

On behalf of:

NICE

Produced by

Matrix
Peninsula Technology Assessment Group (PenTAG)

Although Matrix are primarily responsible for the work in this report, PenTAG retains responsibility for the standard of the report and the quality of the advice that it contains.

Authors

Mariana Bacelar, Economist, Matrix Knowledge
Adeline Durand, Senior Economist, Matrix Knowledge
Chris Hyde, Professor of Public Health and Clinical Epidemiology, PenTAG
Nick Latimer, Research Fellow in Health Economics, University of Sheffield

Acknowledgements

We particularly acknowledge the help of Claudius Rudin, Royal Devon and Exeter Hospital who advised us.

Declaration of competing interest of the authors

None.

Rider of responsibility for report

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

This report should be referenced as follows

Bacelar M, Durand A, Hyde C, Latimer N. The clinical and cost-effectiveness of lenalidomide for people who have received at least one prior therapy with bortezomib (partial review of TA171). Celgene's response to Technology Appraisal Committee's Critique of the Economic Case. A Critique of the Response Submitted by Celgene. Matrix and Peninsula Technology Assessment Group, 2014.

Contribution of authors

Mariana Bacelar: Contributed to project management, the critique of the manufacturer's submission, report writing and editing.

Adeline Durand: Contributed to the critique of the manufacturer's submission, report writing and editing.

Chris Hyde: Contributed to the critique of the manufacturer's submission and commented on drafts of the report.

Nick Latimer: Contributed to the critique of the manufacturer's submission.

Correspondence to

Matrix Knowledge
1st Floor | Kemp House | 152-160 City Road | London | EC1V 2NP
enquiries@matrixknowledge.com

Date completed

June 2014

Contents

Executive summary 9
1.0 Introduction 11
2.0 Lenalidomide effectiveness 12
2.1 The effectiveness of Len/Dex compared with Bort 12
Summary of previous submission and Committee's comment 12
Overview of method applied in the revised analysis 13
Critique of the method applied in the revised analysis 26
2.2 The effectiveness of Len/Dex compared with chemotherapy 45
Summary of previous submission and Committee's comment 45
Overview of method applied in the revised analysis 46
Critique of the method applied in the revised analysis 51
3.0 Model structure 54
4.0 Costs 56
Summary of previous submission and Committee's comment 56
Overview of method applied in the revised analysis 56
Critique of the method applied in the revised analysis 58
5.0 Use of utilities in the cost-effectiveness model 61
Summary of previous submission and Committee's comment 61
Overview of method applied in the revised analysis 61
Critique of the method applied in the revised analysis 62
6.0 Economic modelling results 64
Base case 64
Scenario analysis 69
Sensitivity analysis 74
7.0 Conclusion 75
Implications of the revised analysis 75
8.0 References 77

List of tables

Table 1. Potential prognostic factors identified by Celgene 18

Table 2.Stepwise regression results reported by Celgene..... 19

Table 3: Final models – full OS 20

Table 4: Final models – OS beyond week 100 (for piecewise, Kaplan–Meier and curve scenario analysis) 21

Table 5: Final models – PFS 21

Table 6: Final models – TTF 21

Table 7: Pooled MM-010 and MM-009 trial data – patient characteristics..... 22

Table 8: Patient characteristics from the literature identified – BOR retreatment..... 22

Table 9: OS HR estimates..... 25

Table 10: PFS HR estimates..... 25

Table 11.Stepwise regression results reported by Celgene in the Excel model..... 31

Table 12: Final models – full OS 35

Table 13: Final models – OS beyond week 100 (for piecewise, Kaplan–Meier and curve scenario analysis) 36

Table 14: Final models – PFS with values added by the ERG 36

Table 15: Final models – TTF with values added by the ERG..... 36

Table 16. PFS data in previous submission..... 41

Table 17. PFS data in revised submission (Gamma distribution) 42

Table 18: ERG OS HR estimates..... 43

Table 19: ERG PFS HR estimates..... 44

Table 20: Patient characteristics from the literature identified – chemotherapy 47

Table 21: OS HR estimates – chemotherapy agents compared to Len/Dex 48

Table 22: Key study & patient characteristics of non-RCTs identified for chemotherapy..... 49

Table 23: Key clinical and safety outcomes from non-RCTs identified for chemotherapy..... 50

Table 24. Patient characteristics in Petrucci (1989)..... 52

Table 25: OS HR estimates – chemotherapy compared to Len/Dex..... 52

Table 26. Bort SPC..... 57

Table 27. EQ-5D utilities derived by mapping EORTC QLQ-C30 scores as presented by the ERG in TA228 62

Table 28. EQ-5D utilities by health state – lower utility from fourth-line treatment scenario 62

Table 29. Len/Dex vs bortezomib retreatment – Celgene base case 65

Table 30. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS)..... 65

Table 31. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS) 65

Table 32: Summary results vs MP – Celgene base case 66

Table 33. Len/Dex vs MP - ERG corrections (HRs) 66

Table 34. Len/Dex vs MP - CGP method and ERG corrections (HRs) 66

Table 35. Len/Dex vs bortezomib retreatment – second-line 66

Table 36. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) -second-line 67

Table 37. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS) - second-line 67

Table 38. Len/Dex vs MP with ERG corrections (HRs) – second-line 67

Table 39. Len/Dex vs bortezomib retreatment – 19 cycles 68

Table 40. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) -19 and 8 cycles 68

Table 41. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS) - 19 and 8 cycles 68

Table 42. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS and CGP) 69

Table 43. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS and CGP) – second-line 69

Table 44. Len/Dex vs CY-DEX 70

Table 45. Use of alternative distributions in the economic model – all treatment lines included and Bort given until progression 71

Table 46. Use of alternative distributions in the economic model – second-line treatment and Bort given for 19 cycles 71

Table 47. Use of alternative sources for Bort effectiveness– all treatment lines included and Bort given until progression 72

Table 48. Len/Dex vs bortezomib retreatment with ERG corrections (HRs) with no Bort PAS 72

Table 49. Len/Dex vs bortezomib retreatment with ERG corrections (HRs) with no Bort PAS – second-line treatment 72

Table 50. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) using different utilities sources 73

List of figures

Figure 1. KM for OS for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for OS 15

Figure 2. KM for PFS for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for PFS 16

Figure 3. KM for TTF for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for TTF 17

Figure 4. Len/Dex base case OS, PFS and TTF models using all trial data adjusted to prognostic factors and second-line KM curves 20

Figure 5. Distribution of patients in the pooled MM-010 and MM-009 studies by myeloma duration . 24

Figure 6. OS, PFS and TTF adjusted (with MOC) Len/Dex curves and KM second-line data 35

Figure 7. Celgene calculations for the matrix of possible patients 38

Figure 8. ERG alternative calculation of weights 38

Figure 9. Myeloma categories 39

Figure 10. OS, PFS and TTF adjusted Bort curves 45

Executive summary

Text cited directly from Celgene's revised submission is presented in italic and cross referenced.

The original submission from Celgene considered the use of lenalidomide (Revlimid®) in combination with dexamethasone for adults with multiple myeloma (MM) for whom thalidomide is contradicted and whose disease has progressed after at least one prior treatment with bortezomib. The comparator used was bortezomib.

The ERG found several important logical errors in the economic model first submitted by Celgene[1]. On the initial request for clarification, we suggested that Celgene addressed some of these issues. However briefly after this, the ERG found other methodological errors in the model. In result of this, Celgene submitted an updated analysis[2], with the goal to address the problems identified in the original submission.

After considering the updated submission, the Committee concluded that the ICER resulting from the analysis was likely to be above £30,000 per QALY gained and that the economic analysis was based on several uncertainties related with the model structure and the data extrapolation process. The Committee therefore concluded that lenalidomide could not be recommended as a cost effective use of NHS resources for people with multiple myeloma for whom thalidomide treatment and stem cell transplant were not appropriate, and who had received 1 prior treatment with bortezomib.

The main uncertainties in Celgene' analysis were related to the following matters:

- The estimated **clinical effectiveness of lenalidomide compared with bortezomib.** Even though the evidence seems clear in establishing that lenalidomide is more effective than placebo and standard chemotherapy, it was unclear whether lenalidomide was clinically more effective than retreatment with bortezomib.

- The Committee pointed that there were two relevant **subgroups of patients**. Those for whom bortezomib retreatment would be appropriate, in which case the treatment options considered should include lenalidomide, bortezomib and standard chemotherapy and those for whom retreating with bortezomib would not be appropriate, in which case the treatment options should include lenalidomide and standard chemotherapy.

- The **model structure**. Third and fourth-treatment lines were modelled inconsistently. Also, in the bortezomib arm of the model, the approach taken to modelling second and third-line treatment outcomes reflected an unlikely scenario from a clinical point of view.

- Several issues related with the **data extrapolation process** were identified. More specifically, the estimated progression-free survival hazard ratio of 0.9 favouring bortezomib compared to lenalidomide was found to be clinical implausible. This also raised concerns related to the method used to carry adjustments in the survival models.

- Finally, the Committee noted several issues with **data inputs** related with the costs assumptions made in the economic model, such as patient access schemes being excluded

[1] Hereafter referred to as the "original submission"
[2] Hereafter referred to as the "updated submission"

from the base-case analysis. It also noted that utility values used in the analysis were based on some uncertainty.

Following the NICE appraisal consultation process in February 2014, Celgene have submitted additional evidence and a revised cost-effectiveness model.

The primary focus of this critique is on the additional evidence submitted by Celgene. However, some of the issues initially raised by the ERG have not been addressed hence we provide a brief summary on these throughout the report.

Given the nature of the STA process and considering this was a review of additional evidence, the ERG were bound by time constraints. Therefore we have focused on certain aspects of the revised analysis and only provide some insight on others. Specifically, the ERG could not review in depth excel formulae, particularly the ones related to subsequent treatment lines.

Having considered Celgene's revised submission the ERG view is that:

- The estimated **clinical effectiveness of lenalidomide compared with bortezomib** is still a main source of concern**.** The process to estimate the HRs is methodologically weak and is potentially biased.

- Analysis undertaken for the small **subgroup of patients** for whom the main current treatment option is standard chemotherapy rather than retreatment with bortezomib continue to suggest high ICERs.

- There are outstanding issues related with the **model structure**. After second-line of treatment, the manufacturer only consider the utility associated with the disease progression state and also the ERG question the value of including third and fourth-treatment lines, especially in the intervention arm of the model, as only cost data is available and the basket of drugs considered might not accurately reflect current clinical practice.

- There are still several issues related with the **data extrapolation process**. The adjustment processes undertaken to estimate survival models are fundamentally flawed and are potentially introducing a high level of noise to the analysis. PFS and TTF survival curves seem to flatten off too soon in the model. OS curves cross PFS and TTF curves after 10 years in the intervention arm of the model.

- Finally, the ERG are still concerned with the assumption of **keeping patients on bortezomib until disease progression** in the economic model. Literature and clinical opinion consistently report that bortezomib is given for a fixed period of time to MM patients. Furthermore, the duration of treatment is one of the key drivers of the final ICER. With regards to the utility values used in the analysis, even though Celgene tried to address some of the issues raised by the Committee, the values are still based on some uncertainty.

Overall there is a high level of uncertainty embedded in several steps of the analysis. It is extremely difficult to assess the extent to which the issues raised impact the final ICER. Therefore all ICERs here presented should be interpreted with extreme caution. The key determinants of the final ICERs seem to be the **duration of bortezomib treatment** and the **inclusion (or exclusion) of subsequent treatment lines**.

1.0 Introduction

The original submission from Celgene considered the use of lenalidomide (Revlimid®) in combination with dexamethasone[3] for adults with multiple myeloma (MM) for whom thalidomide is contradicted and whose disease has progressed after at least one prior treatment with bortezomib[4]. The comparator used was Bort.

After considering the previous submission, the Committee concluded that the ICER resulting from the analysis was not plausible based on the uncertainties around model parameters and assumptions.

Following the NICE appraisal consultation process in February 2014, Celgene have requested to submit additional analyses in relation to the appraisal of Len/Dex compared with Bort.

This document presents a critique of the additional evidence and the revised cost-effectiveness model submitted by Celgene in their revised analysis as a response of the concerns highlighted by the Committee.

The remainder of this report is structured as follows:

- Section 2 presents a review of the updated analysis regarding the effectiveness of Len/Dex
- Section 3 reviews the model structure
- Section 4 reviews the additional evidence relating to costs
- Section 5 reviews the utility values used in the revised model
- Section 6 refers to the results of the economic model
- Section 7 concludes.

Where relevant, the sections introduced above are organised as follows:

- Summary of previous submission and Committee's comment
- Overview of method applied in the revised analysis
- Critique of the method applied in the revised analysis.

[3] Referred to as "Len/Dex" in the remainder of this report.
[4] Referred to as "Bort" in the remainder of this report

2.0 Lenalidomide effectiveness

2.1 The effectiveness of Len/Dex compared with Bort

Summary of previous submission and Committee's comment

The main source for clinical effectiveness data in Celgene previous submission was the MM-010 trial, complemented with data from Taverna (2012) study. These data were used directly or indirectly to inform the calculation of transition probabilities within the model. The manufacturer decided to exclude data from MM-009 from the base case analysis. The Committee recognised that there was little difference in the survival data between MM-009 and MM-010 however, it noted that the identical trial designs meant that pooling data would provide more precise estimates and agreed that it was feasible and appropriate to pool the data (ACD, section 4.7).

In the original submission, an exponential piecewise model was used to fit the MM-010 OS data in order to extrapolate study results to a 25 year horizon. However, in their updated submission Celgene changed the distribution used to fit MM-010 OS data to a log-logistic. This decision was reported to be in response to the issues raised by the ERG upon review of the original submission, where it noted that overall survival (OS) curves (fitted with the exponential piecewise model) crossed progression-free survival (PFS) and time to treatment failure (TTF) curves in the model. The Committee heard from the manufacturer that when using the piecewise exponential curve approximately 3% of people survived after 25 years (that is, beyond an age of 88 years) in the model, whereas using the log-logistic curve considerably increased this to 11%. The Committee agreed that it was clinically implausible for 11% of people with multiple myeloma to live beyond the age of 88 years. The Committee concluded that the piecewise exponential curve was a better choice for extrapolation because it showed a better fit to the data and provided more realistic estimates of overall survival, however, the Committee were concerned about the validity of this modelling given the overlapping of the progression-free survival and overall survival curves. (ACD, section 4.6)

PFS, TTF and OS curves derived from MM-010 data were adjusted using the mean of covariates method, by which average values of covariates (like for example the beta-2 microglobulin count and presence of bone lesions for the baseline MM-10 population) were entered into a proportional hazards regression equation. The choice of relevant predictors of PFS, TTF and was not very transparent in the submission. The Committee agreed that several factors predicting survival were not adjusted for, and that it was not clear why the 3 factors used had been chosen. The Committee therefore concluded that the data had not been appropriately adjusted (ACD, section 4.8)

For the bortezomib arm of the model, the same transition probabilities from the intervention arm were used, only exponentiated to computed hazard ratios (HRs). To estimate HRs, Taverna (2012) data were used. Celgene compared the median survival estimates (OS and PFS) between studies of interest and MM-010 and derived a crude approximation of a HR for Len/Dex and each comparator. The Committee noted that the assumptions applied to allow hazard ratios to be calculated in this way only hold when exponential distributions are used (ACD, section 4.8).

The multivariate parametric models designed to predict PFS and OS were used to improve exchangeability between studies. This was done by adjusting the median survival estimates (PFS and OS) from MM-010 to reflect the characteristics of the population in the comparator study (e.g. Taverna, 2012). The Committee discussed how the hazard ratios comparing the effectiveness of Len/Dex with Bort were calculated. The ERG noted an inconstancy in the original HR calculation of

the PFS curve. While OS and TTF respective HRs were adjusted to reflect the population characteristics of the Taverna study, the same approach wasn't followed for PFS. The manufacturer acknowledged this as an oversight and adjusted the HR of Bort retreatment relative to Len/Dex in their updated submission. The PFS adjusted HR obtained was 0.9. The Committee noted that progression-free survival was longer with Bort retreatment, whereas overall survival was longer with Len/Dex. It heard from the clinical specialist that this was neither plausible nor seen in clinical practice (ACD, section 4.8)

The Committee concluded that the methods used to estimate the efficacy of Len/Dex compared with Bort were not robust and therefore the hazard ratios were highly uncertain. It therefore agreed that it would be appropriate and helpful to include in the analysis:

- A model that assumed Len/Dex and Bort to have equal effectiveness (that is, a hazard ratio of 1)

- A model that compared lenalidomide with placebo because it would be based on high-quality data from MM-009 and MM-010 (ACD, section 4.9).

The Committee considered the approach to modelling third and fourth-line treatment in the model. It noted that the manufacturer had modelled the costs and effects differently in the Len/Dex and the comparator arms. The Committee agreed it was important to consider treatments that would follow second-line treatment, and to include both their costs and effectiveness. However, it noted that costs and effectiveness should be modelled consistently in the Len/Dex and comparator arms. The Committee also agreed that it was not clinically plausible that, as an individual's disease progresses, the utility value remained constant, despite receiving third and fourth-line treatments, which should be associated with a progression-free state for some time. In addition, the Committee noted that, by including third and fourth-line treatments, Len/Dex treatment was included in the comparator arm, meaning that the model compared Len/Dex with Bort followed by Len/Dex. It therefore agreed that it would be appropriate to run the analysis:

- Excluding third and fourth-line treatments in the model.

- Including sequences of Len/Dex followed by Bort compared with Bort followed by Len/Dex (without asymmetry between the Len/Dex and comparator arms) (ACD, section 4.10).

Overview of method applied in the revised analysis

Celgene present additional evidence regarding an MTC undertaken for TA 171. It is reported that a new study (DOXIL) has been included in the MTC analysis since TA 171, which provided greater power to the analysis. The common comparator across studies was reported to be Dex, thus allowing a connected network of treatment comparisons between Bort and Len/Dex. Celgene report the HRs however the data presented are not used in the base case analysis (but instead are included in scenario analysis) as they compare the use of Len/Dex with Bort initial treatment (i.e. the patients included in the MTC analysis are Bort naïve).

In their revised analysis, Celgene have used MM-010 and MM-009 pooled data, along with the Taverna (2012) study to estimate the effectiveness of Len/Dex compared with Bort (note that 63% of patients receive dexamethasone in the comparator study). Survival analysis was re-run with the pooled dataset and parametric survival models were developed to estimate OS, PFS and TTF curves. Time to progression (TTP) was also included in the revised analysis.

Celgene describe their process for estimating OS, PFS and TTF curves through the following steps:

1. They started by running unadjusted models to determine which parametric survival models provided the best fit to the MM-010 and MM-009 pooled data.

2. Celgene then identified the potentially relevant covariates to test for prognostic significance in the survival models. The manufacturer claim to have identified the outcome predictive factors through publications, CSRs and clinical feedback.

3. Subsequently, Celgene report to have used stepwise regression analysis to determine which of the variables identified in step 2 were statistically significant in predicting survival outcomes.

4. Finally, covariate-adjusted parametric survival models were estimated to produce the final OS, PFS and TTF survival curves which were used to estimate the transition probabilities in the economic model.

We now report in turn the steps undertaken by Celgene to estimate OS, PFS and TTF.

1. Unadjusted models and selection of parametric model

Unadjusted pooled data from MM-010 and MM-009 was used to assess which parametric models provided the best fit to the data.

- Overall survival

A Gompertz distribution was used to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon.

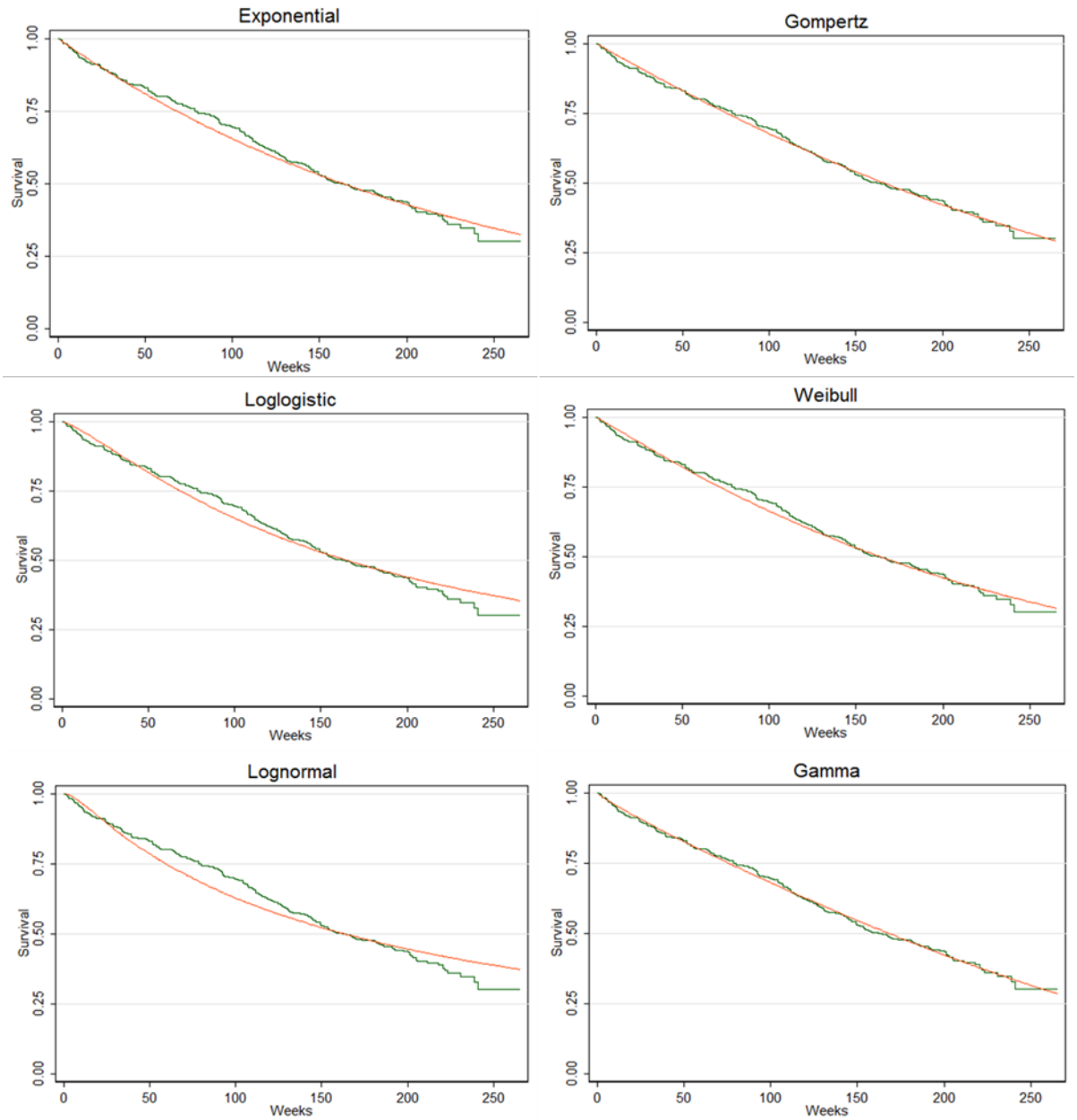
Celgene report undertaking visual inspections of the fitted curves and using Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) to assess the best model fit.

Celgene claim that based upon visual inspection of the KM for MM-010 and MM-009 pooled data the piecewise approach does not seem to be necessary. Furthermore it was mentioned that the piecewise exponential and the piecewise Gompertz models were incorporated as scenario analyses, where the parametric models were fitted to the survival data only after 100 weeks. Survival during the initial 100 weeks (i.e. preceding the parametric curve) were informed by KM data. Celgene say the KM data during the initial 100 weeks are based on a large number of observations and that there is a change in gradient of the KM after 100 weeks. Celgene conclude this provides a piecewise approach, with OS split into observed and extrapolated sections.

Celgene provide the KM curves along with the OS curves extrapolated up to the point where the KM curves end, with different distributions. These are shown in Figure 1. The unadjusted extrapolated curves for the piecewise exponential model were not provided.

.

Figure 1. KM for OS for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for OS



Source: Celgene revised submission, Appendix A

- Progression-free survival

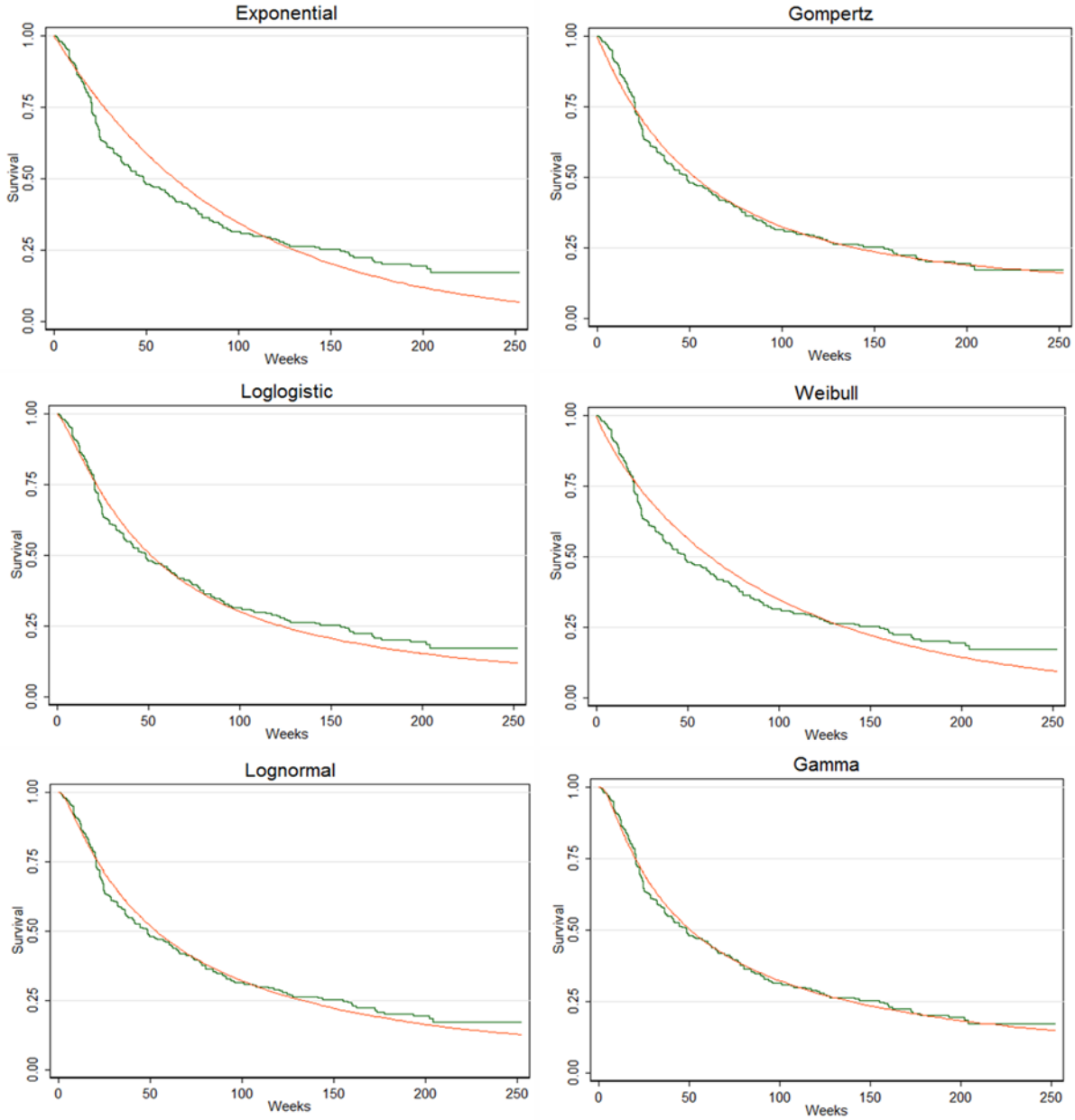
A Gamma distribution was used to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon.

Celgene report undertaking visual inspections of the fitted curves and using AIC and BIC to assess the best model fit. It is also mentioned that the exponential and the Weibull models were a poor fit

and that the Gompertz model appeared to visually fit the KM data well despite the relatively high AIC and BIC statistics.

The manufacturer provide the KM curves along with the PFS curves extrapolated up to the point where the KM curves end, with different distributions. These are shown in Figure 2.

Figure 2. KM for PFS for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for PFS



Source: Celgene revised submission, Appendix A

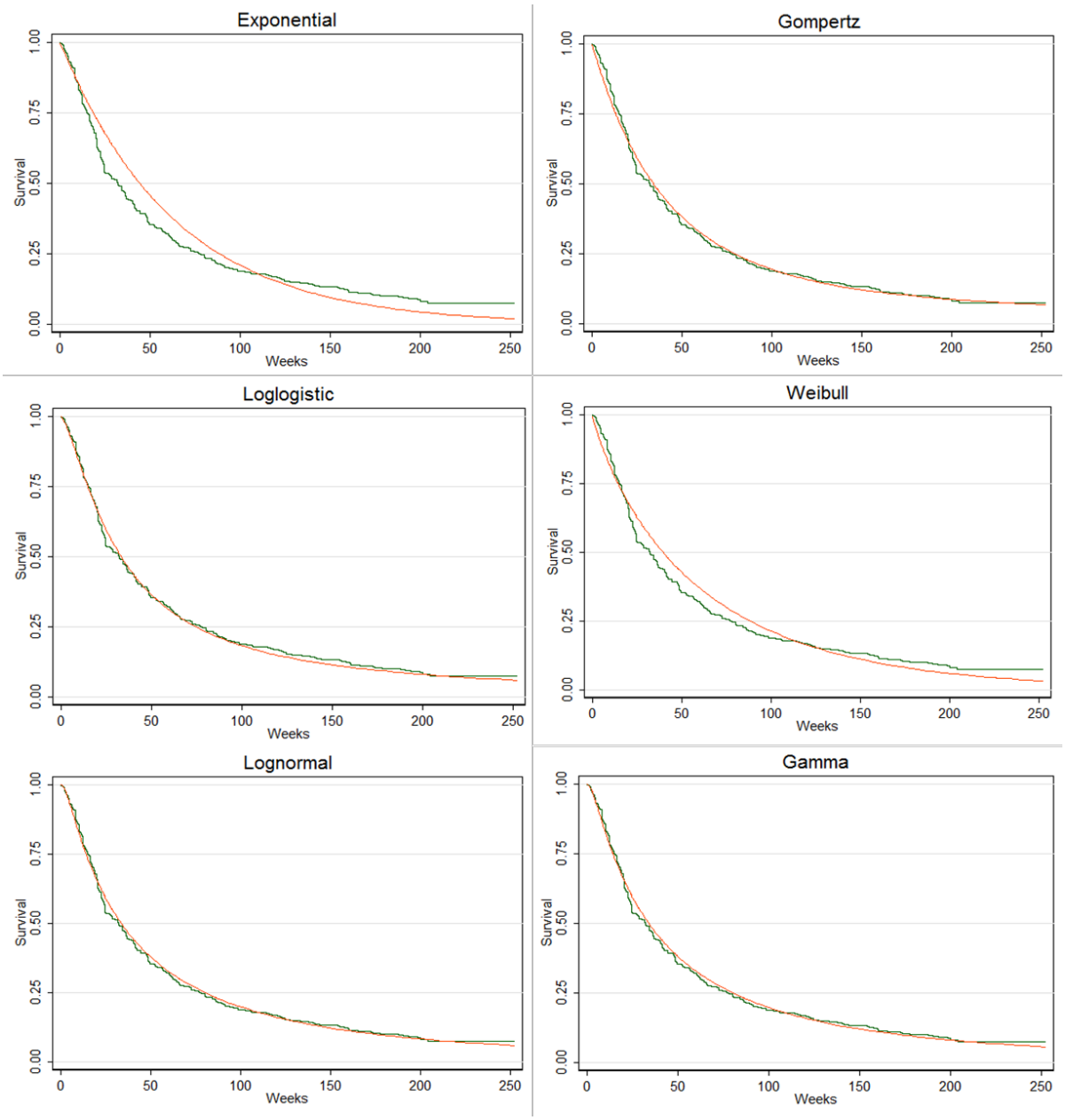
- Time to treatment failure

A Gamma distribution was used to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon.

Similarly to the PFS, it is mentioned that the exponential and the Weibull models were a poor fit and that the Gompertz model appeared to visually fit the KM data well despite the relatively high AIC and BIC statistics.

The manufacturer provide the KM curves along with the TTF curves extrapolated up to the point where the KM curves end, with different distributions. These are shown in Figure 3.

Figure 3. KM for TTF for the entire population in MM-010 and MM-009 pooled dataset and unadjusted survival curves for TTF



Source: Celgene revised submission, Appendix A

Celgene also presented data on time to progression (TTP) even though this is not used in the economic model.

2. Identification of relevant covariates

Celgene present a list of the variables considered as potentially relevant determinants of survival outcomes in Table 4 of their revised submission (page 17). It is stated that analysis presented in the CSRs for MM-010 and MM-009 along with other publications and clinical opinion were used to inform the initial identification of relevant variables. Table 1 replicates the list of potentially relevant covariates identified by the manufacturer.

Table 1. Potential prognostic factors identified by Celgene

Source for inclusion	***Variable***	***Definition***
MM-010 CSR[7]	*Prior thalidomide*	*Yes or no*
	Prior doxorubicin	*Yes or no*
	Prior bortezomib	*Yes or no*
	Worsening extramedullary plasmacytoma disease	*Yes or no*
	Baseline bone marrow cellularity-aspirate/biopsy	*Normal, hyperplasia, hypoplasia or missing*
MM-009 CSR[6]	*Prior melphalan*	*Yes or no*
	Baseline plasma cell %	*High or low*
Dimopoulos et al[16]	*Baseline beta-2 microglobulin*	*>2.5 or ≤2.5 mg/L*
	Duration of multiple myeloma	*Years*
	Lytic bone lesions present at baseline	*Yes or no*
	Prior high-dose therapy or stem cell transplant	*Yes or no*
	Prior dexamethasone	*Yes or no*
	Baseline ISS score	*1, 2 or 3*
	Baseline plasma cell %	*High or low*
Stadtmauer et al[17]	*Number of prior anti-myeloma therapies*	*1 or 2+* *Note: required to produce estimates for the population of interested (1 prior therapy)*
	Age	*Years*
Clinical advice	*Sex*	*Male or female*
Key: *CSR, clinical study report; ISS, International Staging System.*		

Source: Revised submission, Table 4.

3. Covariates used in the economic model

Celgene report undertaking stepwise regression analysis on the full ITT population data to assess which of the identified covariates were statistically significant predictors of OS, PFS and TTF.

Celgene present the covariates found to be statistically significant predictors of survival outcomes in Appendix C of their revised submission. The manufacturer decided to omit the p-values for the non-significant variables (>0.05).

The variables and p-values reported by the manufacturer are presented in Table 2.

Celgene explain that covariates would be included in the economic model if they were found statistically significant in predicting *OS and one or both of the other outcomes* (i.e. PFS and TTF).

The manufacturer mention that *OS was the preferred outcome as it is the most important clinical outcome and the driver of the cost-effectiveness results*. Celgene go on to explain that this approach prevented the *inclusion of variables that statistically appear to affect only one variable, avoiding clinical inconstancies whereby only one of two related variable is adjusted.*

Celgene identified **myeloma duration, beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease** as being the only statistically significant prognostic factors. Celgene also claim that *the proportion of patients who had received one or two or more therapies was forcibly kept in the stepwise process, as this covariate is required to present outcomes for a second-line population.*

Table 2.Stepwise regression results reported by Celgene

Parameter	P-values		
	OS - Gompertz	PFS - Gamma	TTF - Gamma
Patient age			
BL ECOG = 0			
BL ECOG = 1			
BL ECOG = 2			
Myeloma duration (yrs)	0.008		
Sex = Male			
Beta-2 mic >2.5mg/l	0.000	0.001	0.001
2+ prior therapies	0.003	0.026	0.004
Prior HDT/SCT			
BL plasma = high			
BL plasma = low			
BL bone lesions			
Earlier DEX			
Prior doxorubicin		0.001	
Prior melphalan			0.034
Prior thalidomide			
Prior bortezomib			
Worsening Ex PD	0.004	0.001	0.007
Myeloma stage 1			
Myeloma stage 2			
Myeloma stage 3	0.028	0.027	
BMCA biop, normal			
Hyperplasia	0.022		0.010
Hypoplasia			

Source: Adapted from Celgene revised submission, Appendix C

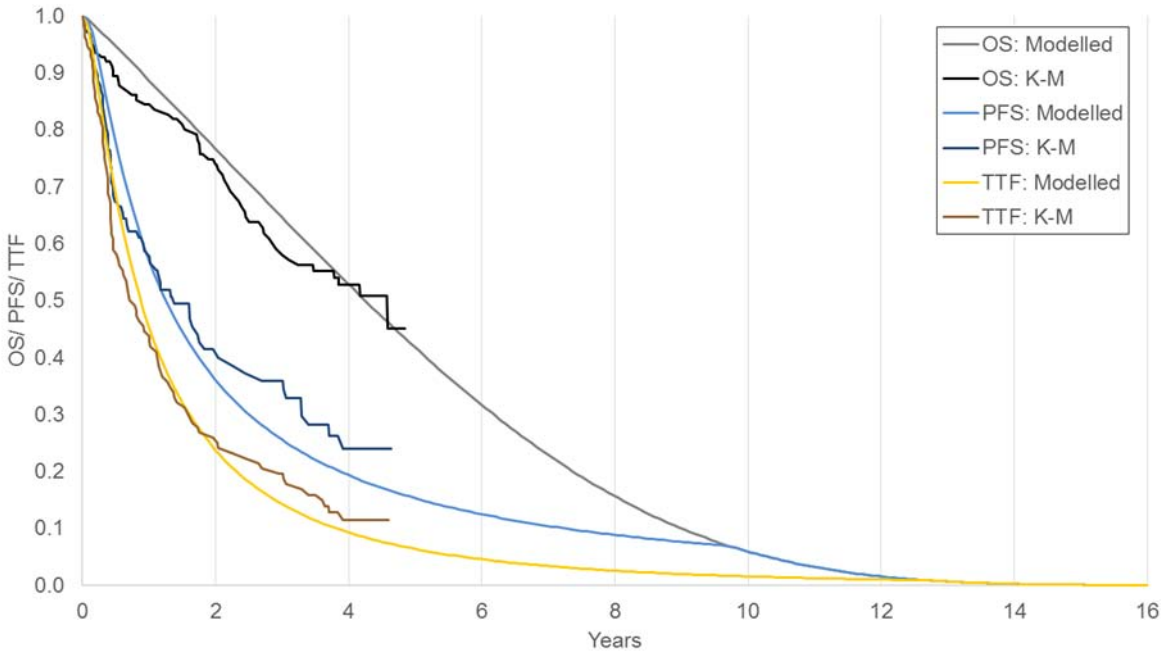
4. Adjusted survival models

The covariates selected to adjust the survival models (described in the previous subsection) were used in regression analysis to estimate the adjusted OS, PFS and TTF curves. This was done using two alternative methodologies, the mean of covariates (MOC) method which was used in the base case analysis and the corrected group prognosis (CGP) method, presented as scenario analysis.

The Len/Dex curves were reported by Celgene and are provided in Figure 4. KM curves for the second-line MM-010 and MM-009 pooled data are presented, as well as the OS curve fitted with the Gompertz distribution and the PFS and TTF curves, both fitted with the Gamma distribution.

Table 3 to Table 6 present the regression coefficients for the different distributions used to fit the OS, PFS and TTF curves respectively.

Figure 4. Len/Dex base case OS, PFS and TTF models using all trial data adjusted to prognostic factors and second-line KM curves



Source: Celgene revised submission, Figure 2

Table 3: Final models – full OS

Parameter	***Coefficients***			
	Exponential	***Gamma***	***Gompertz***	***Weibull***
Myeloma duration (years)	*-0.081*	*0.063*	*-0.086*	*-0.084*
Beta-2 microglobulin >2.5mg/L	*0.745*	*-0.609*	*0.791*	*0.767*
2+ prior therapies	*0.583*	*-0.517*	*0.610*	*0.598*
Worsening Ex PD	*0.902*	*-0.722*	*0.966*	*0.943*
Intercept	*-6.119*	*6.148*	*-6.445*	*-6.711*
Gamma			*0.003*	
Rho				*1.113*
Sigma		*0.651*		
Kappa		*1.658*		
Key: *Ex PD, extramedullary plasmacytoma disease; OS overall survival.*				

Source: Celgene revised submission, Table 5

Table 4: Final models – OS beyond week 100 (for piecewise, Kaplan–Meier and curve scenario analysis)

Parameter	***Coefficients***			
	Exponential	***Gamma***	***Gompertz***	***Weibull***
Myeloma duration (years)	*-0.099*	*0.099*	*-0.099*	*-0.100*
Beta-2 microglobulin >2.5mg/L	*0.726*	*-0.700*	*0.740*	*0.738*
2+ prior therapies	*0.789*	*-0.755*	*0.794*	*0.794*
Worsening Ex PD	*1.011*	*-0.997*	*1.014*	*1.015*
Intercept	*-5.949*	*5.834*	*-6.055*	*-6.260*
Gamma			*0.002*	
Rho				*1.066*
Sigma		*1.010*		
Kappa		*0.862*		
Key: *Ex PD, extramedullary plasmacytoma disease; OS overall survival.*				

Source: Celgene revised submission, Table 6

Table 5: Final models – PFS

Parameter	***Coefficients***			
	Gamma	***Gompertz***	***Log-logistic***	***Log-normal***
Myeloma duration (years)	*0.060*	*-0.051*	*0.056*	*0.058*
Beta-2 microglobulin >2.5mg/L	*-0.647*	*0.524*	*-0.679*	*-0.653*
2+ prior therapies	*-0.405*	*0.476*	*-0.510*	*-0.452*
Worsening Ex PD	*-0.880*	*1.041*	*-0.911*	*-0.951*
Intercept	*4.349*	*-4.716*	*4.569*	*4.523*
Gamma			*0.743*	
Rho				
Sigma	*1.308*			*1.277*
Kappa	*-0.236*			
Key: *Ex PD, extramedullary plasmacytoma disease; PFS, progression-free survival.*				

Source: Celgene revised submission, Table 7

Table 6: Final models – TTF

Parameter	***Coefficients***			
	Gamma	***Gompertz***	***Log-logistic***	***Log-normal***
Myeloma duration (years)	*0.043*	*-0.029*	*0.041*	*0.047*
Beta-2 microglobulin >2.5mg/L	*-0.618*	*0.533*	*-0.616*	*-0.603*
2+ prior therapies	*-0.409*	*0.353*	*-0.390*	*-0.403*
Worsening Ex PD	*-0.665*	*0.727*	*-0.564*	*-0.614*
Intercept	*4.161*	*-4.319*	*4.055*	*4.032*
Gamma			*0.704*	
Rho				
Sigma	*1.216*			*1.237*
Kappa	*0.172*			

Key: *Ex PD, extramedullary plasmacytoma disease; TTF, time to treatment failure.*

Source: Celgene revised submission, Table 8

- Mean of covariates method

PFS, TTF and OS curves were adjusted using the MOC method, in which average values of the covariates (myeloma duration, beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease) are entered into a proportional hazards regression equation. The mean values observed in the MM studies are reported in Table 7.

Table 7: Pooled MM-010 and MM-009 trial data – patient characteristics

Variable	***ITT population***	***Second-line population***	***Third-line population***
1 prior therapy (vs 2+ prior therapies)	*64.9%*	*100%*	*0.0%*
Myeloma duration	*3.9 years*	*2.7 years*	*4.6 years*
Beta-2 microglobulin level >2.5 mg/L	*70.8%*	*63.7%*	*74.7%*
Worsening extramedullary plasmacytoma disease	*3.4%*	*5.6%*	*2.2%*
Key: *ITT, intention-to-treat.*			

Source: Celgene submission, Table 9

To adjust the survival curves derived from MM-010 and MM-009 pooled data to the comparator studies using the MOC method, Celgene used the baseline patients' characteristics in the comparator studies (reported in Table 8) together with the respective coefficients (reported in Table 3 to Table 6) to obtain an estimate of the Len/Dex results if the MM-010 and MM-009 populations had the same characteristics as the populations in the comparator studies, for each relevant covariate.

Table 8: Patient characteristics from the literature identified – BOR retreatment

Source	***Characteristic***	***Value***
Taverna et al[1]	*Myeloma duration*	*Median = 2.8 years*
Hrusovsky et al[2]	*Myeloma duration*	*Mean = 4.3 years*
Sood et al[19]	*Myeloma duration*	*Median = 3.7 years*
Petrucci et al[12]	*Myeloma duration* *1 prior therapy (vs 2+ prior therapies)*	*Median = 4.5 years* *11.5%*
Dispenzieri et al[14]	*Beta-2 microglobulin level >2.5 mg/L*	*100.0%*
MM-009 and MM-010 ITT population[6, 7]	*Myeloma duration* *1 prior therapy (vs 2+ prior therapies)* *Beta-2 microglobulin level >2.5 mg/L* *Worsening extramedullary plasmacytoma disease*	*Median = 3.9 years* *35.1%* *70.8%* *3.4%*
Key: *ITT, intention-to-treat.*		

Source: Celgene submission, Table 10

- Corrected group prognosis method

Due to the ERG concerns related to the use of the mean of covariates method, Celgene have undertaken scenario analysis using the CGP method.

Celgene explain that with the CGP method each relevant clinical outcome is predicted for each possible patient. The average survival is then calculated by weighting each possible outcome by the likelihood of the clinical outcome occurring. Celgene provide the following illustrative example:

For example, 70.8% of pooled MM-010 and MM-009 patients had a beta-2 microglobulin level of >2.5mg/L. …The corrected group prognosis approach would instead estimate survival for a patient who presents with beta-2 microglobulin >2.5mg/L, and every possible combination of other covariates, and then do the same for a patient who presents with beta-2 microglobulin level of ≤2.5mg/L. The resulting array of survival estimates are then weighted by the likelihood of each type of patient presenting and summed to give a corrected group prognosis survival estimate.

However, while the MOC method only requires the average covariate value observed in the study population, the CGP method requires patient-level data. As the Taverna (2012) retrospective study did not provide such data Celgene made the following assumptions in order to adjust the survival curves to the bortezomib population:

1. The likelihood of a patient presenting in one particular clinical diagnosis has no impact on the likelihood of the patient falling into any other possible clinical combination of outcomes. For example, the likelihood of having a beta-2 microglobulin level > 2.5mg/L is independent of the myeloma duration observed in that patient.

2. *Patients were distributed into myeloma duration groups of 0–1 year, 1–2 years, 2–3 years and so on, up to 15+ years. The actual duration within each group was assumed to be the midpoint of that group (for example, a patient in the '3–4 years' was assumed to have been diagnosed 3.5 years ago).*

Celgene state that the MOC method predicts higher short-term survival but lower long-term survival compared to the CGP method. They also mention that the KM curves fall between the fitted curves estimated with both methods.

Celgene claim that MOC method is the more conservative approach to estimate long-term survival and therefore is used in the base case analysis, while the CGP method is incorporated in scenario analysis.

HR calculation

To estimate HRs, Celgene compared the median survival estimates (OS and PFS) between studies of interest and MM-010 and MM-009 pooled data and derived a crude approximation of a HR for Len/Dex and each comparator.

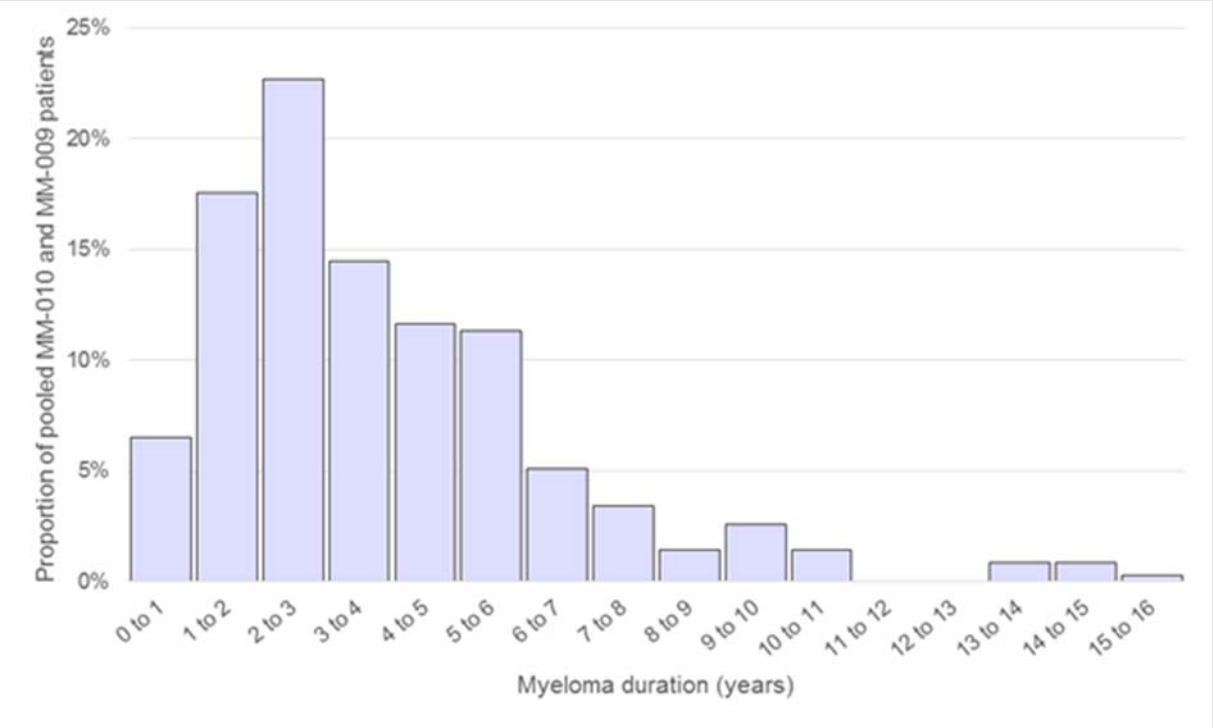
The manufacturer also present the analysis for unadjusted HRs. It is explained that the unadjusted HRs were calculated for OS and PFS using the pooled MM-0019 and MM-009 data with no adjustment for bortezomib patient characteristics.

Celgene mention that estimating the HRs for Bort vs Len/Dex using estimates from the CGP method was constrained by the fact that there were no patient level data in the Taverna (2012) study. As previously mentioned, Celgene assumed a distribution of patients by multiple duration for the comparator studies. They assumed that the shape of the distribution was the same as for the pooled MM-010 and MM-009 pooled data and that it would only be able to shift in relation to the known median or mean input value.

Similarly, where only a median value was available within a bortezomib study, it was assumed that the relationship between the mean and median was constant, meaning the distribution could be shifted according to either the difference in mean or median outcome (between the pooled study data and the comparator evidence).

The distribution of patients in the pooled MM-010 and MM-009 studies by myeloma duration is presented in Figure 5.

Figure 5. Distribution of patients in the pooled MM-010 and MM-009 studies by myeloma duration



Source: Celgene revised submission, Figure 3

Celgene present the adjusted and the unadjusted HRs for the MOC and the CGP methods respectively. We report the respective estimates in Table 9 and Table 10.

Table 9: OS HR estimates

Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data unadjusted	LEN/DEX data adjusted	LEN/DEX data unadjusted
Taverna et al[1]	Exponential	1.86	1.92	1.57	1.61
	Gompertz	1.89	1.94	1.63	1.67
	Weibull	1.86	1.91	1.58	1.61
	Gamma	1.92	1.96	1.62	1.67
	Piecewise exponential	1.83	1.85	1.63	1.65
	Piecewise Gompertz	1.85	1.87	1.66	1.68
Hrusovsky et al[2]	Exponential	2.11	2.04	1.77	1.71
	Gompertz	2.12	2.06	1.83	1.77
	Weibull	2.10	2.02	1.77	1.71
	Gamma	2.14	2.08	1.82	1.77
	Piecewise exponential	2.00	1.96	1.77	1.75
	Piecewise Gompertz	2.02	1.99	1.81	1.78
Key: HR, hazard ratio; LEN/DEX, lenalidomide plus dexamethasone; OS, overall survival.					

Source: Celgene submission, Table 11

Table 10: PFS HR estimates

Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data unadjusted	LEN/DEX data adjusted	LEN/DEX data unadjusted
Taverna et al[1]	Gompertz[5]	1.06	0.90	1.44	1.29
	Log-logistic	1.08	0.96	1.26	1.15
	Log-normal	1.06	0.93	1.26	1.15
	Gamma	1.11	0.99	1.26	1.20
Hrusovsky et al[2]	Gompertz[3]	1.01	1.02	1.36	1.46
	Log-logistic	1.04	1.08	1.20	1.30
	Log-normal	1.02	1.05	1.20	1.30
	Gamma	1.06	1.12	1.20	1.36
Sood et al[19]	Gompertz[3]	1.11	1.43	1.36	2.06
	Log-logistic	1.04	1.52	1.20	1.83
	Log-normal	1.02	1.48	1.20	1.83
	Gamma	1.06	1.57	1.20	1.91
Petrucci et al[12]	Gompertz[3]	1.11	1.13	1.40	1.62
	Log-logistic	1.13	1.20	1.23	1.44
	Log-normal	1.08	1.16	1.26	1.44

[5] This is reported as Exponential in Celgene's revised submission, however this is clearly a mistake and the manufacturer are referring to the Gompertz distribution as per the Excel model and section 3.1 in the revised submission.

Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data unadjusted	LEN/DEX data adjusted	LEN/DEX data unadjusted
	Gamma	*1.11*	*1.24*	*1.26*	*1.50*
Dispenzieri et al[14]	*Gompertz*[3]	*1.42*	*1.67*	*1.20*	*1.26*
	Log-logistic	*1.32*	*1.64*	*1.32*	*1.42*
	Log-normal	*1.35*	*1.67*	*1.32*	*1.42*
	Gamma	*1.32*	*1.61*	*1.32*	*1.42*
***Key:** HR, hazard ratio; LEN/DEX, lenalidomide plus dexamethasone; PFS, progression-free survival.*					

Source: Celgene submission, Table 12

Critique of the method applied in the revised analysis

Celgene present additional evidence regarding an MTC initially undertaken for TA 171, which the manufacturer have updated with a new study (DOXIL). The ERG question the inclusion of the DOXIL study as it seems that this study population does not receive Dex at any point. Given that Dex is the common drug between the other three studies in the MTC, which allows the network of treatment comparisons, it is unclear how Celgene were able to undertake an MTC using the DOXIL study. Nevertheless the ERG haven't focused on the MTC analysis as this is not used in Celgene's base case analysis.

In their revised analysis, Celgene have used MM-010 and MM-009 pooled data, along with the Taverna (2012) study to estimate the effectiveness of Len/Dex compared with Bort/Dex. Pooling the data is appropriate in this case and can potentially provide more precise estimates.

Celgene also claim to have employed a clearer and more transparent process to describe their estimation of OS, PFS and TTF in the economic analysis. Even though Celgene provide a description of the steps undertaken to obtain the different survival models, the ERG are still concerned with the lack of transparency around some of the analysis.

Therefore we provide a brief summary of the process undertaken by Celgene to derive the OS, PFS and TTF models used in the economic analysis:

1. Celgene started by determining which parametric survival models provided the best fit to the MM-010 and MM-009 pooled data. This was done using the unadjusted data for the full ITT population.

2. Celgene then identified potentially relevant covariates to test for prognostic significance in the survival models. The manufacturer claim to have identified these prognostic factors through publications, the CSRs and clinical feedback.

3. After identifying the potential prognostic variables, Celgene report using stepwise regression analysis to determine which of the variables identified in step 2 were statistically significant in predicting the survival outcomes observed in MM-010 and MM-009 trials.

4. After selecting the prognostic variables to be included in the survival models Celgene produced the adjusted survival models. There are essentially two steps involved in this adjustment process:

a. Firstly, the manufacturer adjusted the survival curves to reflect the identified prognostic factors. Pooled data from MM-010 and MM-009 are used to reflect the patients' characteristics with regards to each one of the prognostic factors (using the MOC method as a base case and the CGP method as an alternative). This adjustment process produces the survival curves used in the intervention arm of the economic model.

b. The manufacturer then adjusted the curves to try to improve exchangeability between the MM trials and the comparator studies, so they can establish the size of comparative effectiveness and derive HRs. This was done by using data from the comparator studies to reflect the patients' characteristics with regards to each one of the prognostic factors (using the MOC and the CGP method as before). After adjusting the survival curves to reflect the Taverna population characteristics (or any other comparator study used in scenario analysis), a median survival estimate is derived, reflecting what the median OS or PFS would be if the Taverna patients were the same as the MM-010 and MM-009 patients. Celgene then assume constant hazards so that the median survival estimate derived can be compared with the observed median survival estimate in the Taverna (or any other comparator) study, allowing a HR to be calculated. Finally, the HRs are applied to the survival curves used in the intervention arm of the model therefore producing the survival curves used in the comparator arm of the model.

We now review in turn the aforementioned steps undertaken by Celgene to estimate the OS, PFS and TTF models.

1. Unadjusted models and selection of parametric models

Unadjusted pooled data from MM-010 and MM-009 were used to assess which parametric models provided the best fit to the data. It seems that Celgene used the full unadjusted dataset from both studies to inform the analysis, which is appropriate.

- Overall survival

In the original submission, a piecewise exponential model was used to fit the MM-010 OS data. The piecewise exponential model with survival time split into 6 months intervals was considered to be the best approach to deal with the presence of the found violation of proportional hazards.

After reviewing the original submission, the ERG noted how the OS and the PFS curves crossed each other both in the Len/Dex arm of the model as well as in the Bort arm. Rationally, such crossing is not possible as the OS curve determines the proportion of people alive at each cycle of the model. It is therefore impossible to have a greater number of people free from disease progression than the number of people alive at the same point in time. Similarly, it was noted that the OS and the TTF curves also crossed each other, both in the Len/Dex arm and in the Bort arm of the model.

Subsequently Celgene decided to use a log-logistic distribution to fit the OS curve in their updated submission.

The ERG and the Committee had several concerns with the use of the log-logistic distribution to model mortality in the model as the rationale behind the distribution selection was flawed and the fit

was not very good when compared with MM-010 KM data. Furthermore the Committee concluded that the exponential piecewise model was more appropriate than the log-logistic one to model OS.

In the revised submission, Celgene decide to use a Gompertz distribution to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon. Celgene report undertaking visual inspections of the fitted curves and using AIC and BIC to assess the best model fit.

Based on visual inspection of the curves, the gamma distribution seems to be as good as the Gompertz in terms of fit to the KM unadjusted data and the reason why the AIC and BIC statistics are worse for the gamma distribution can be attributed to the fact that this distribution has more parameters and not necessarily because it is a worse fit to the data.

Furthermore Celgene claim that the use of a piecewise *approach does not seem to be necessary.* This is surprising given the Committee preference for the piecewise exponential distribution to model OS in the economic analysis. It is not clear at all why Celgene consider the piecewise approach to be "unnecessary", given that on their original submission Celgene believed this to be the appropriate method to deal with the proportional hazards violation.

Nonetheless, the manufacturer claim to have incorporated the piecewise exponential model in the undertaken scenario analysis. However, the approach taken by Celgene seems to be closer to a "broken curve approach[6]" than to a piecewise model. The former consists in using KM data up to a point x, then fitting a model beyond that point x to extrapolate data. This extrapolation process is highly sensitive to point x and it often involves fitting the extrapolated curve to a very small amount of data at the tail of the dataset.

Celgene decided to model the initial 100 weeks of the OS curve with KM data based on the argument that there is a large number of observations and that there is a change in gradient of the KM after 100 weeks. After this point, an exponential curve was fitted up to 25 years.

The ERG find this decision surprising given that in their original submission, the manufacturer modelled the piecewise exponential model in a different manner by which they have fitted separate parametric models to different time periods, allowing the hazard to change over time (which is generally what the piecewise approach consists on).

We consider the use of different distributions to fit OS data in the "Adjusted survival models" section below.

- Progression-free survival

A Gamma distribution was used to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon. Celgene report undertaking visual inspections of the fitted curves and using AIC and BIC to assess the best model fit.

In their previous submissions, Celgene have decided to use the log-logistic distribution to fit PFS. The Committee did not raise any concerns related to this approach, however the difference in the revised submission is that instead of using MM-010 data alone, the manufacturer have polled the MM-010 and MM-009 data.

[6] http://www.ispor.org/news/articles/Aug12/Extrapolation-in-Oncology-Modelling.asp

Visual inspection of the curves show that other distributions could potentially be a good fit to the data (especially the Gompertz distribution) and AIC and BIC statistics were in general similar across distributions.

Again, we explore the use of different distributions to fit OS data in the “Adjusted survival models” section below.

- Time to treatment failure

A Gamma distribution was used to fit the MM-010 and MM-009 pooled data in order to extrapolate the study results to a 25 year horizon. Celgene report undertaking visual inspections of the fitted curves and using AIC and BIC to assess the best model fit. The ERG conclusions for PFS apply to TTF.

Overall, the ERG are not very clear why some distributions were considered the best fit while others were deemed inappropriate. Visual inspection of the curves and assessment of the AIC and BIC statistics do not reveal clear preferences within a certain range of distributions. Particularly, we are not sure why the distributions used to model PFS and TTF changed.

Furthermore, it is not clear why the piecewise exponential was deemed inappropriate, why the broken curve approach was undertaken (instead of a traditional piecewise exponential) and especially why did Celgene decide to choose 100 weeks to start extrapolating the curve. As explained above, the point where the extrapolation process begins will highly influence the behaviour of the survival curve.

Additionally, as pointed out by the ERG in their previous report it is not helpful to present extrapolated curves only up to the point where the KM curves end. Despite this, Celgene have presented the extrapolated curves for visual inspection only up to the end of the KM unadjusted data. The extrapolated portion of the curves would have been helpful to inform Celgene’s decision on the selection of distributions. As the raw data used to build the unadjusted survival curves were not reported in the excel model, the ERG could not replicate the curves for visual comparison with the KM data.

2. Identification of relevant covariates

Celgene present a list of the variables considered as potentially relevant determinants of survival outcomes (Table 1). The manufacturer claim that a transparent selection process was undertaken based upon factors previously considered to affect OS, PFS and TTF.

The ERG still have some issues with the transparency of the selection process. Even though the manufacturer claim that the potentially prognostic factors came from analysis presented in the CSRs for MM-010 and MM-009 (as well as other sources), the ERG could not find any analysis in the CSRs related with such variables presented in Table 1. Moreover, some of the variables considered from other sources, like for example lytic bone lesions referenced in Dimopoulos (2009) were not statically significant predictors of clinical outcomes in the original study but were still included in Celgene’s list.

3. Covariates used in the economic model

Similarly to the previous submission, the ERG remain concerned with the lack of transparency of the process undertaken to select the covariates used in the economic analysis.

Celgene report undertaking stepwise regression analysis on the full ITT population data to assess which of the covariates were statistically significant predictors of OS, PFS and TTF and present these together with the respective p-values (Table 2). However in the Excel models submitted by the manufacturer, the p-values for such covariates differ from the ones presented in the submission. Furthermore the manufacturer decided to omit the p-values for the non-significant variables (>0.05).

The ERG present the p-values reported in the Excel models for the ITT populations for OS, PFS and TTF in Table 11, where values in bold identify p-values>0.05, therefore representing not statistically significant variables. The following discrepancies should be noted between Table 11 and Table 2:

- The p-values for the myeloma duration coefficients for PFS and TTF are presented and are **not statistically significant predictors** of PFS and TTF.

- In general all the p-values presented **are different** even though the majority of them still indicate statistically significant predictors.

- Worsening extramedullary plasmacytoma disease is **not a statically significant predictor** of TTF in Table 11.

The ERG also feel that for the purposes of consistency and transparency the p-values for all variables should have been presented. Additionally, upon the submission of the original analysis the ERG have requested from Celgene that the stepwise regression equations were provided to us in full. Nevertheless Celgene have not provided these.

Celgene explain that covariates would be included in the economic model if they were found statistically significant predictors of *OS and one or both of the other outcomes* (i.e. PFS and TTF). The ERG are not clear on this statement as it seems to imply that any covariate found to be statistically significant for OS and at least one other outcome (i.e. PFS or TTF) would be included. However, looking at Table 2, the variable myeloma stage 3 was a statically significant predictors of OS and PFS while hyperplasia was a statically significant predictors of OS and TTF but both variables were not included in the economic model. More importantly, myeloma duration was found a statistically significant predictor only for OS (**and not for PFS and TTF**) and still was included as a covariate in the economic model.

The manufacturer mention that *OS was the preferred outcome as it is the most important clinical outcome and the driver of the cost-effectiveness results*. Celgene explain that this approach prevented the *inclusion of variables that statistically appear to affect only one variable, avoiding clinical inconstancies whereby only one of two related variable is adjusted.* The ERG believe this approach introduces bias to the analysis.

On one hand Celgene are eliminating statistically significant predictors. This is not justifiable given that the OS, PFS and TTF models are three different models which could therefore be adjusted with different covariates. On the other hand from a clinical point of view it seems reasonable that different prognostic factors can influence different survival outcomes (for example, survival and disease progression).

Celgene identified myeloma duration, beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease as being the only statistically significant prognostic factors. Celgene also claim that *the proportion of patients who had received one or two or more therapies was forcibly kept in the stepwise process, as this covariate is required to present outcomes for a second-line population.*

The ERG find Celgene's statement confusing. Looking at Table 2 and Table 11 the number of prior therapies appears to be a statistically significant prognostic factor. Therefore we do not understand why this covariate would have to be forcibly kept in the stepwise process, when it would be logical to include it.

Table 11.Stepwise regression results reported by Celgene in the Excel model

Parameter	P-values		
	OS - Gompertz	PFS - Gamma	TTF - Gamma
Patient age			
BL ECOG = 0			
BL ECOG = 1			
BL ECOG = 2			
Myeloma duration (yrs)	0.006	**0.061**	**0.114**
Sex = Male			
Beta-2 mic >2.5mg/l	0.000	0.000	0.000
2+ prior therapies	0.000	0.025	0.008
Prior HDT/SCT			
BL plasma = high			
BL plasma = low			
BL bone lesions			
Earlier DEX			
Prior doxorubicin		?	
Prior melphalan			?
Prior thalidomide			
Prior bortezomib			
Worsening Ex PD	0.004	0.028	**0.081**
Myeloma stage 1			
Myeloma stage 2			
Myeloma stage 3	?	?	
BMCA biop, normal			
Hyperplasia	?		?
Hypoplasia			

In conclusion, the ERG are concerned with the lack of transparency of the process undertaken to select the covariates used in the economic analysis since:

- P-values reported in the revised submission differ from the ones presented in the excel file, and were not reported anywhere for the variables considered to be not statically significant by Celgene.

- Celgene appear to have established a rule for the selection of variables to be included in the models which is likely to introduce bias. Regardless of this, it seems this rule is not in reality being followed in their selection process.

- It is not clear at all why the different survival models were not adjusted with different covariates, which could have been selected on the basis of being statistically significant predictors of the specific survival outcomes.

- According to Table 2, myeloma stage 3 was a statically significant predictors of OS and PFS while hyperplasia was a statically significant predictors of OS and TTF but these variables were not included in the economic model.

- Myeloma duration is being included as a predictive factor in the OS, PFS and TTF models however this is not a statically significant predictor of PFS and TTF.

Overall it seems that Celgene should have used different prognostic factors to adjust the different survival models (i.e. OS, PFS and TTF models). The approach undertaken is likely to introduce bias in the analysis and influence the estimation of the extrapolated survival data.

The ERG have further issues related to the selection of the covariates used in the model, which will be explored in the next subsection 4 b).

4. Adjusted survival models

Celgene use the MOC method in their base case analysis. Therefore both adjustments a) and b) were undertaken with this method. The manufacturer also presented the results for the adjustments using the CGP method. In this subsection the ERG describe these and provide some insight on each alternative method.

a) Adjustment to baseline population characteristics in MM-010 and MM-009

Here we describe the methods undertaken to derive the survival curves used in the **intervention arm of the model** and explore the appropriateness of alternative distributions to fit survival data.

- Mean of covariates method

The prognostic variables selected to adjust the survival models (myeloma duration, beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease) were used in regression analysis to estimate the adjusted OS, PFS and TTF curves utilized in the economic model.

The curves were reported by Celgene and the ERG have reproduced these in Figure 6. KM curves for the second-line MM-010 and MM-009 pooled data are presented, together with the OS curve fitted with the Gompertz distribution and the PFS and TTF curves, both fitted with the Gamma distribution.

Celgene claim that *although the curves for different clinical outcomes cross over, the Committee noted that this is not itself an issues*. The ERG strongly disagree with this statement. The Committee was concerned with the implausibility of crossing curves and acknowledged this to be a fundamental flaw in the analysis (section 4.6, 4.13 and 4.15 of the ACD).

Furthermore, it can be noted that the curves still cross in the revised submission. In fact, compared with the original one, this can be considered as a retrogression. When OS was fitted with a piecewise exponential distribution, OS curves crossed PFS and TTF curves around week 1000 (approximately 20 years) and week 600 in the intervention and the comparator arms of the model, respectively. However Figure 6 shows how in the Len/Dex arm of the revised model the OS curve crosses the PFS curve by week 500 (around 10 years) and the TTF curve around week 660 (13 years) while Figure 10 (presented in the next subsection) shows how in the comparator arm of the model the OS curve crosses the PFS curve by week 300 (around 6 years) and the TTF curve around week 470 (9 years).

Similarly to the previous submission, the ERG are concerned with the following issues:

- OS curves crossing PFS and TTF curves after 10 years in the intervention arm of the model.
- The appropriateness of the distributions used to fit the data.
- The method used to adjust the survival curves.

Visual inspection of the curves, BIC and AIC statistics do not seem enough to decide which distribution would be more appropriate to fit OS data. Potential candidates would appear to be the Gompertz, the gamma and the piecewise exponential models.

To note is that the survival predictions with the three curves seem clinically reasonable. With the Gompertz distribution, less than 1% of patients are alive in the Len/Dex arm of the model after 12 years. The same survival estimates (i.e. <1%) are achieved by year 15 with the gamma distribution. The “piecewise exponential” predicts that 1.2% of patients are still alive after 25 years, which might be less plausible.

Also to consider is that the Gompertz distribution can only represent monotonically increasing or decreasing hazards, while for example the gamma distribution can represent hazards that initially increase, and then begin to decrease. In this regard, a piecewise exponential model can also be fitted to represent any kind of hazard pattern over time (to note is that this approach was not taken by the manufacturer as they have followed a broken curve approach).

It is also important to consider the regression coefficients for the different distributions used to fit the OS, PFS and TTF curves respectively. These have been previously presented in
Table 4 to Table 6, however the ERG presents these again in Table 12 to Table 15. We decided to present the tables again as:

- The tables initially reported by Celgene (reproduced in Table 4 to Table 6) did not present all relevant information. Therefore we have looked for the data in the excel models, and present the values (highlighted in red) in the tables below.
- These help the interpretation and comparison of the different coefficients across distributions.

Looking at the coefficients presented in Table 12 and Table 13, it is interesting to note that:

- In Table 12, the gamma parameter of the Gompertz distribution is >0. This means that the hazard will increase monotonically over time (hence the extrapolated curve quickly reduces to 0).

- In Table 13 the sigma and kappa parameters for the gamma distribution are very close to 1. In the case where these parameters are both equal to 1, then the gamma model is reduced to an exponential model. This might suggest that beyond 100 weeks the hazards are close to being constant.

The fact that the hazards become close to being constant after 100 weeks, might suggest that OS needs to be fitted with a model that is more flexible than the Gompertz (for example the gamma or a piecewise exponential).

However, as previously mentioned, the estimated survival values seem to indicate that OS is being reasonably estimated in the economic model. Nonetheless, given the clear problem with the crossing curves, the extrapolation process is still fundamentally flawed. This suggests that the problem might be related with the extrapolation of the PFS and the TTF curves, which appear to be flattening off too soon (Figure 6).

Visual inspection of the curves, BIC and AIC statistics do not seem enough to decide which distribution would be more appropriate to fit PFS and TTF data. Potential candidates would appear to be the Gompertz, the gamma and the log-logistic distribution (especially for TTF).

As aforementioned, one of the possible explanations for the OS and the PFS and TTF curves to cross might be related to the PFS and TTF curves flattening too soon, indicating that the hazards will decrease monotonically over time. In fact analysis of Table 14 and Table 15, shows that the gamma parameters for the Gompertz distribution are <0. This is surprising given that the same parameter assumes a value >0 for OS. **This suggests that for PFS and TTF the hazards decrease monotonically over time, while for OS the hazards will increase monotonically**. The clinical plausibility of this is questionable and should therefore be considered by the Committee.

Furthermore, looking at the sigma and kappa parameters of the gamma distribution for PFS and TTF we can note that the kappa parameter is close to 0. If the kappa parameter of a gamma model is 0, then the model is reduced to a log-normal. Also to note is that the p-values for the gamma, sigma and kappa parameters were not provided by Celgene. So it might be the case that the kappa parameter is indeed not statistically significantly different from 0.

These considerations show that there is a substantial amount of uncertainty related to the appropriateness of the model fitted to the survival data, which is concerning especially because of the crossing survival curves so early in the model.

Finally, and possibly more importantly, it should be noted that the extrapolated PFS and TTF curves flatten early in the model due to the KM curves also flattening at the end (whereas the OS KM curve does not flatten) (Figure 6). The number of patients at risk at the end of the PFS and TTF curves would therefore be crucial to understand the KM curve behaviour. If these numbers are very low then the tail of KM curves are exposed to a high level of uncertainty and the flattened shape of the extrapolated curves might be a misrepresentation. Unfortunately the manufacturer did not provide the numbers at risk for the PFS nor the TTF curves, and so the ERG cannot assess these.

If the number of patients at risk are low at the end of the PFS and TTF KM curves, then it can be argued that dropping observations at the end of the dataset could potentially provide a more realistic survival curve, which would continue with a steady slope downwards rather than flattening off early in time as it happens in the presented analysis. However this argument should be interpreted with extreme caution, as dropping observations means losing valuable data, therefore the rationale behind it must be robust.

Figure 6. OS, PFS and TTF adjusted (with MOC) Len/Dex curves and KM second-line data

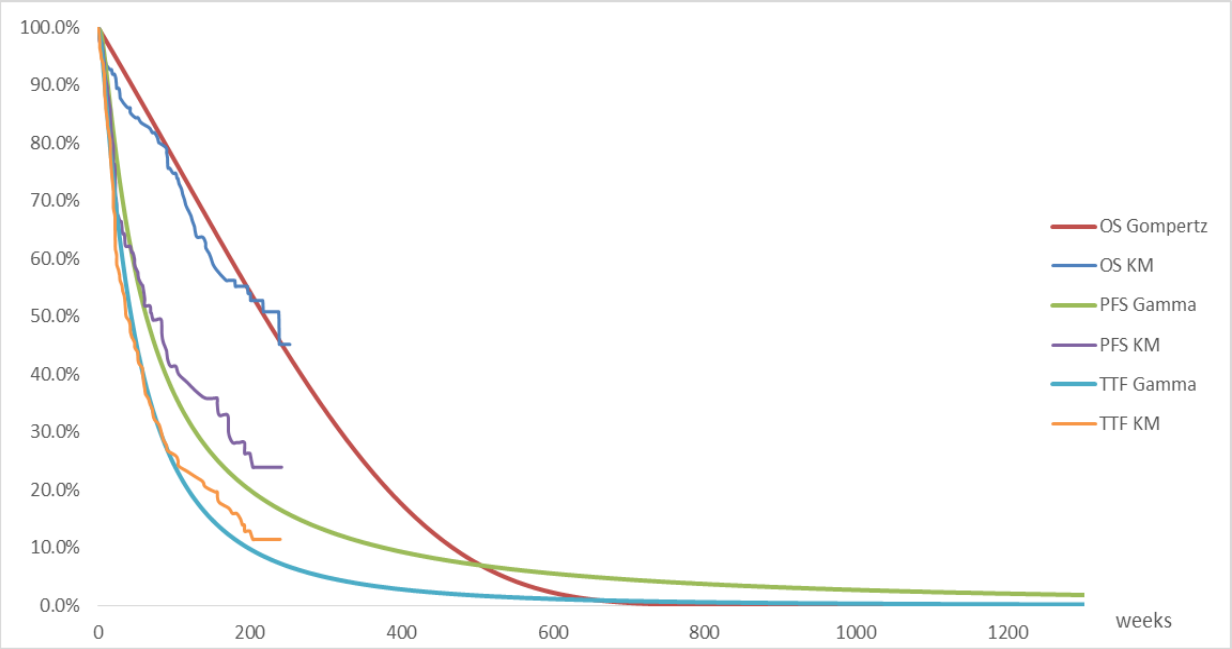


Table 12: Final models – full OS

Parameter	***Coefficients***			
	Exponential	***Gamma***	***Gompertz***	***Weibull***
Myeloma duration (years)	*-0.081*	*0.063*	*-0.086*	*-0.084*
Beta-2 microglobulin >2.5mg/L	*0.745*	*-0.609*	*0.791*	*0.767*
2+ prior therapies	*0.583*	*-0.517*	*0.610*	*0.598*
Worsening Ex PD	*0.902*	*-0.722*	*0.966*	*0.943*
Intercept	*-6.119*	*6.148*	*-6.445*	*-6.711*
Gamma			*0.003*	
Rho				*1.113*
Sigma		*0.651*		
Kappa		*1.658*		
Key: *Ex PD, extramedullary plasmacytoma disease; OS overall survival.*				

Source: Celgene revised submission, Table 5

Table 13: Final models – OS beyond week 100 (for piecewise, Kaplan–Meier and curve scenario analysis)

Parameter	*Coefficients*			
	Exponential	*Gamma*	*Gompertz*	*Weibull*
Myeloma duration (years)	*-0.099*	*0.099*	*-0.099*	*-0.100*
Beta-2 microglobulin >2.5mg/L	*0.726*	*-0.700*	*0.740*	*0.738*
2+ prior therapies	*0.789*	*-0.755*	*0.794*	*0.794*
Worsening Ex PD	*1.011*	*-0.997*	*1.014*	*1.015*
Intercept	*-5.949*	*5.834*	*-6.055*	*-6.260*
Gamma			*0.002*	
Rho				*1.066*
Sigma		*1.010*		
Kappa		*0.862*		
***Key:** Ex PD, extramedullary plasmacytoma disease; OS overall survival.*				

Source: Celgene revised submission, Table 6

Table 14: Final models – PFS with values added by the ERG

Parameter	Coefficients			
	Gamma	Gompertz	Log-logistic	Log-normal
Myeloma duration (years)	0.060	-0.051	0.056	0.058
Beta-2 microglobulin >2.5mg/L	-0.647	0.524	-0.679	-0.653
2+ prior therapies	-0.405	0.476	-0.510	-0.452
Worsening Ex PD	-0.880	1.041	-0.911	-0.951
Intercept	4.349	-4.716	4.569	4.523
Gamma		-0.006	0.743	
Rho				
Sigma	1.308			1.277
Kappa	-0.236			
Key: Ex PD, extramedullary plasmacytoma disease; PFS, progression-free survival.				

Source: Adapted from Celgene revised submission, Table 7

Table 15: Final models – TTF with values added by the ERG

Parameter	Coefficients			
	Gamma	Gompertz	Log-logistic	Log-normal
Myeloma duration (years)	0.043	-0.029	0.041	0.047
Beta-2 microglobulin >2.5mg/L	-0.618	0.533	-0.616	-0.603
2+ prior therapies	-0.409	0.353	-0.390	-0.403
Worsening Ex PD	-0.665	0.727	-0.564	-0.614
Intercept	4.161	-4.319	4.055	4.032
Gamma		-0.006	0.704	
Rho				
Sigma	1.216			1.237
Kappa	0.172			

Key: Ex PD, extramedullary plasmacytoma disease; TTF, time to treatment failure.

Source: Adapted from Celgene revised submission, Table 8

Overall, the ERG are concerned with the outstanding issue of OS curves crossing PFS and TTF curves after 10 years in the intervention arm of the model. The estimated curves predict that, for example, at year 15 (note that the model time horizon is 25 years), 4% of patients alive haven't progressed, while less than 0.1% of patients are supposed to be alive. This is clearly implausible. Given that the manufacturer seem be modelling a clinically plausible OS curve (even though analysis of the data suggests that a more flexible model on the hazards could be more appropriate), it is suggested that the problem is related with the estimation of PFS and TTF curves. This is potentially related with the shape of the KM at the end of the curve.

Additionally, the ERG have expressed several concerns with the covariates used to adjust the curves in the previous subsection. The fact that some parameters which are not statically significant predictors of PFS and TTF are included in the PFS and TTF models, together with the use of the MOC method to adjust the curves, can also be contributing to the problems identified in the extrapolation of the PFS and TTF curves.

In conclusion, there are several sources of uncertainty, which implies that the robustness of the final ICER is questionable.

- Corrected group prognosis method

Due to the ERG concerns regarding the use of the mean of covariates method, Celgene have undertaken scenario analysis using the CGP method.

However the ERG have several concerns with the way the CGP method was implemented. Overall, we found the calculations in the Excel model confusing, nonetheless we try to shed some light on the process used to implement this method. Furthermore, Celgene incorporated the CGP estimates as hardcoded values in the main excel model. Even though the manufacturer provided separate files with these calculations, this added a considerable burden to the ERG work.

Celgene have considered a matrix of possible patients, presented in Figure 7, taken from the manufacturer excel model. For example, patient A in the matrix below has Beta-2 mic >2.5mg/l, is a second-line patient and has worsening Ex PD, patient B has Beta-2 mic >2.5mg/l, is a second-line patient and has no worsening Ex PD…etc). Celgene also set the number of previous therapies to <2, i.e. this is always 0 in the matrix of possible patients to reflect the population of interest.

We also provide an example to demonstrate how the weights were derived for each patient:

To get to the 0.32 value (cell E4), Celgene calculate: 63.7% / sum (E1:H1). So effectively 63.7% / 2. Similarly, to obtain the 0.03 value they calculate 5.6% / sum (E1:H1). Finally, to get to the 0.17 weight they sum E4 + E5+ E6 and then divide by 2.

Figure 7. Celgene calculations for the matrix of possible patients

				E	F	G	H	
Parameter	**Value in MM-010 and MM-009**	**Alternative**		**Matrix of possible patients**				
Myeloma dur (yrs)	2.7			***A***	***B***	***C***	***D***	
Beta-2 mic >2.5mg/l	63.7%	36.3%		1	1	0	0	1
2+ prior therapies	0.0%	100.0%		0	0	0	0	2
Worsening Ex PD	5.6%	94.4%		1	0	1	0	3
Constant	1		Value	0.32	0.32	0.18	0.18	4
			Value	0.00	0.00	0.00	0.00	5
			Value	0.03	0.47	0.03	0.47	6
			Weights	**0.17**	**0.40**	**0.10**	**0.33**	

The main problem with this approach is that, for example, Celgene derive a weight for group C of 0.10. This means that 10% of patients were in group C, therefore they had Beta-2 mic<2.5 and they had Worsening Ex PD. However, only 5.6% had Worsening Ex PD in the trials. Accordingly, it would be impossible for 10% to have Worsening Ex PD AND Beta-2 mic>2.5.

The best way of estimating the weights would have been to use patient level data to categorise what percentage of patients was in group A, what percentage was in group B, etc. Celgene would have been able to do this for MM-010 and MM-009 pooled data as they had the appropriate data.

In the case where patient level data were not available, the second best approach would seem to be multiplying each row to get the weight for each column (Figure 8), so for example, to get to the 0.04 weight for patient A we would calculate E4*E5*E6. This seems to give more plausible weights. To note is that these quite different from the ones obtained through Celgene's approach. We explore the impact of changing this on the final ICER on Section 6.

Figure 8. ERG alternative calculation of weights

E	F	G	H	
Matrix of possible patients				
A	***B***	***C***	***D***	
1	1	0	0	1
0	0	0	0	2
1	0	1	0	3
0.637	0.637	0.363	0.363	4
1.000	1.000	1.000	1.000	5
0.056	0.944	0.056	0.944	6
0.04	**0.60**	**0.02**	**0.34**	**Weights**

The ERG are not clear on the calculations around the myeloma duration categories. The estimated categories are presented in Figure 9 and were taken from Celgene's excel file. These seem reasonable, and would have to be incorporated into the weight calculations. This should bring the total number of patient groups to 64. Even though the ERG could confirm that this was the number of patient groups in Celgene model some of the hardcoded values remain confusing.

Figure 9. Myeloma categories

Myeloma duration			MM010/009	
Duration (years)	Midpoint	%	Pts	%
0 to 1	0.5	7%	9	7%
1 to 2	1.5	34%	42	34%
2 to 3	2.5	28%	35	28%
3 to 4	3.5	14%	17	14%
4 to 5	4.5	6%	7	6%
5 to 6	5.5	6%	7	6%
6 to 7	6.5	4%	5	4%
7 to 8	7.5	2%	2	2%
8 to 9	8.5	0%	0	0%
9 to 10	9.5	0%	0	0%
10 to 11	10.5	0%	0	0%
11 to 12	11.5	0%	0	0%
12 to 13	12.5	0%	0	0%
13 to 14	13.5	0%	0	0%
14 to 15	14.5	0%	0	0%
15 to 16	15.5	0%	0	0%

Given the hardcoded values in the CGP calculations (and the lack of explanations for these), the ERG are not entirely confident that this method has been implemented correctly (despite the weights' calculation issue).The ultimate aim of CGP is to calculate S(t|x) for each possible combination of x (x being a patient group), and then to calculate the adjusted S(t) as the weighted average of these.

b) Adjustment to comparator population characteristics

The models designed to predict PFS, TTF and OS in the Len/Dex arm of the model were also used to improve exchangeability between studies. This was done by adjusting the median pooled survival estimates (PFS and OS) from MM-010 and MM-009 to reflect the characteristics of the population in the comparator study (e.g. Taverna, 2012).

Celgene re-run their base case analysis with the mean of covariates (MOC) method using the MM-010 and the MM-009 pooled data. They also report undertaking the analysis with the corrected group prognosis (CGP) method to adjust OS, PFS and TTF curves to population characteristics.

Here we describe the methods undertaken to derive the survival curves used in the **comparator arm of the model.**

- Mean of covariates method

PFS, TTF and OS curves were adjusted using the MOC method, in which average values of the covariates (myeloma duration, beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease) in the comparator studies (Table 8) are entered into a proportional hazards regression equation.

However, as we can observe from Table 8 not all the relevant prognostic factors had their median or mean values reported in the comparator studies. For example, Taverna (2012) study only reported the median myeloma duration in the baseline population (and did not present data on the beta-2 microglobulin, the number of prior therapies and worsening extramedullary plasmacytoma disease of the population). It is not clear to the ERG how this was dealt with in the model. Celgene do not

provide any explanation on how they have approached this matter (for example, they might have assumed the MM-010 and MM-009 population characteristics to be the same as the ones expected to be observed in the Taverna (2012) population).

Regardless of the method used, which should have been clearly reported by Celgene, the lack of data in the comparator studies implies that some sort of assumptions had to be made, which impacted the final ICER.

The comparator survival curves were estimated by taking the intervention survival curves and exponentiating these to the estimated HR. Therefore the ERG provide more detail on the survival curves in the comparator arm of the model after tackling the HR calculation process.

HR calculation

Celgene compared the median survival estimates between studies of interest and derived a crude approximation of a HR for Len/Dex and each comparator. This approach assumes that progression/mortality occurs at a constant rate across studies and that studies' populations and conditions are exchangeable.

It should be noted that there is a serious theoretical limitation (also raised by the Committee) with the approach taken to derive the HRs.

The ERG provide a detailed explanation of the process undertaken to estimate the HRs to facilitate the understanding of our concern:

- After using parametric models to adjust the OS and PFS survival curves to reflect the Taverna population characteristics (or any other comparator study used in scenario analysis), either with the MOC or the CGP methods, a median survival estimate is derived, reflecting what the median survival would be if the Taverna patients were the same as the MM-010 and MM-009 patients.
- Because the Taverna (2012) study only provides OS and time to progression (TTP) median estimates, Celgene have to do further adjustments between the observed TTP and PFS median estimates in MM-010 and MM-009 pooled data so a relationship can be established between TTP outcomes in the Taverna (2012) study and the MM studies. To note is that the OS HR did not this adjustment as there were OS data available for all studies.
- Celgene then **assume constant hazards** so that the median survival estimate initially derived (and then adjusted to TTP/PFS) can be compared with the observed median survival estimate in the Taverna study, allowing a HR to be calculated.

However, the constant hazards assumption is fundamentally flawed as the parametric models fitted to the MM-010 and MM-009 pooled data (Gompertz for OS and Gamma for PFS) do not assume constant hazards. This assumption can only be applied to exponential models.

Furthermore, the ERG found a mistake in the PFS HR calculation, related with the PFS/TTP adjustment step. This is now explained in detail.

Equation 1 shows how Celgene calculated the PFS HR in the previous submission. The data necessary for the calculations are presented in Table 16.

In Equation 2 the ERG present an alternative equation, nonetheless an equivalent one, which we believe is easier to interpret as it used the median estimates (fourth column) instead of the log of the hazards (eighth column).

Equation 1. Celgene PFS HR calculations in previous submission

$$\exp\left[-2.72 - (-2.86) * \left(\frac{-2.45}{-2.68}\right)\right] = 0.9$$

Equation 2. ERG PFS HR alternative calculations

$$\left(\frac{8.02}{10.5} * \frac{12.05}{10.12}\right) = 0.9$$

Table 16. PFS data in previous submission

Treatment	Description	N*	Median PFS in months	95% LCI	95% UCI	Hazard	ln(Ht)
Len/Dex MM-010 data	Unadjusted observed PFS	176	10.12	6.83	14.92	0.068	-2.68
Len/Dex MM-010 data	Unadjusted observed TTP	176	12.05	9.41	19.69	0.058	-2.86
Len/Dex adjusted MM-010	Estimates adjusted to Taverna MM duration 2.8 years*	176	8.02	8.02	17.47	0.086	-2.45
Bortezomib TTP	Unadjusted TTP Taverna 2012	42	10.50	8.40	12.60	0.066	-2.72

Equation 3 presents the PFS HR calculation undertaken in the revised submission. Comparing Equation 3 with Equation 1 it can be observed how the ratio between the unadjusted observed ln(Ht) PFS and the unadjusted observed ln(Ht) TTP is inverted. This is a mistake and the correct formula is presented in Equation 4 with the Celgene method and in Equation 5 with the ERG equivalent method. The data necessary for the calculations are presented in Table 17.

Therefore, the PFS HR is 1.19 (instead of 1.11). The ERG have corrected all the HRs for the different comparator studies and distributions.

Equation 3. Celgene PFS HR calculations in revised submission

$$\exp\left[-4.188 - (-4.326) * \left(\frac{-4.241}{-4.279}\right)\right] = 1.11$$

Equation 4. Celgene PFS HR calculations in revised submission with ERG correction

$$\exp\left[-4.188-(-4.326)*\left(\frac{-4.279}{-4.241}\right)\right]=1.19$$

Equation 5. ERG PFS HR alternative calculations

$$\left(\frac{52}{45.7}*\frac{52.4}{48.1}\right)=1.19$$

Table 17. PFS data in revised submission (Gamma distribution)

Treatment	Description	Median PFS in weeks	95% LCI	95% UCI	Hazard	ln(Ht)
Len/Dex pooled data	Unadjusted observed PFS	48.1	38.5	57.8	0.014	-4.241
Len/Dex pooled data	Unadjusted observed TTP	52.4	41.9	62.9	0.013	-4.326
Len/Dex adjusted pooled data	Estimates adjusted to Taverna MM duration 2.8 years with Gamma distribution	50	40	60	0.013	-4.279
Bortezomib TTP	Unadjusted TTP Taverna 2012	45.7	36.5	54.8	0.015	-4.188

To note is that the PFS HR in the previous submission was <1, suggesting that Bort was more effective in keeping patients from progressing than Len/Dex. This was one of main concerns raised by the Committee. As it can be observed from Equation 4 and Equation 5, this is no longer the case in the manufacturer revised submission, where the PFS HR is >1.

It should also be noted that in the previous submission the median PFS for Len/Dex was adjusted downwards (8.02 vs 10.12 – fourth column in in Table 16) when accounting for the Taverna (2012) characteristics. In the revised submission the median PFS for Len/Dex is being adjusted upwards (48.1 vs 52 – third column in Table 17) when accounting for the Taverna characteristics.

Celgene presented the adjusted HRs for the MOC and the CGP methods respectively (Table 9 and Table 10). However, as mentioned before, all the OS and PFS adjusted HRs presented are fundamentally flawed since they were all derived assuming constant hazards, which means that theoretically these are the result of using an exponential model (the only exception is for the OS HRs derived with the exponential distribution, which is theoretically correct).

The manufacturer also present the analysis for “unadjusted” HRs. It is mentioned that the unadjusted HRs were calculated for OS and PFS using the pooled MM-0010 and MM-009 with no adjustment to Bort patient characteristics. However, the HRs provided for the “unadjusted” data are in fact adjusted values. As previously explained, Celgene undertook two adjustment processes. An initial one where survival curves were adjusted to MM-010 and MM-009 patient characteristics for the selected prognostic factors and a second one, where survival curves were adjusted for the comparator study populations (for example, Taverna 2012). In this case, Celgene presented the HRs before the second adjustment, therefore the values are only adjusted to the MM-010 and MM-009 population

characteristics with the different distributions and methods. This is also why the “unadjusted” values differ for the MOC and the CGP method.

Furthermore, the ERG find that it would have been more helpful to present the HRs to reflect Len/Dex vs Bort instead of Bort vs Len/De as these are more intuitive and easier to interpret. We have therefore reproduced the HRs to reflect Len/Dex vs Bort, and so **HRs<1 favour Len/Dex**, while **HRs>1 favour Bort**. These are presented in Table 18 for OS and in Table 19 for PFS with the ERG corrections previously mentioned. To note is that the HRs for TTF were assumed to be the same as for PFS since there were no data available to calculate these.

The HRs for OS and PFS used in the economic model are shown in bold in Table 18 and in Table 19 respectively.

Even though Celgene claim that the HRs predicted with both methods consistently favour Len/Dex, after the ERG corrections some of the PFS HRs for the adjusted analysis using the CGP method (highlighted in red in Table 19) produced HRs>1, suggesting that Bort is more effective than Len/Dex in keeping patients in remission. Also to note is that for the Taverna (2012) and other comparator studies, the HRs adjusted with the CGP method are generally close to 1.

Finally, it is the ERG opinion that the HRs for the CGP method reflect a more conservative scenario, as these consistently predict higher HRs than the ones derived with the CGP method. Therefore the CGP method would be considered the conservative approach. To note is that the ERG have issues with the way both methods were implemented in the economic analysis.

Table 18: ERG OS HR estimates

Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data “unadjusted”	LEN/DEX data adjusted	LEN/DEX data “unadjusted”
Taverna 2012	Exponential	0.54	0.52	0.64	0.62
	Gompertz	**0.53**	0.52	0.61	0.60
	Weibull	0.54	0.52	0.63	0.62
	Gamma	0.52	0.51	0.62	0.60
	Piecewise exponential	0.55	0.54	0.61	0.61
	Piecewise Gompertz	0.54	0.53	0.60	0.60
Hrusovsky 2010	Exponential	0.47	0.49	0.56	0.58
	Gompertz	0.47	0.49	0.55	0.56
	Weibull	0.48	0.50	0.56	0.58
	Gamma	0.47	0.48	0.55	0.56
	Piecewise exponential	0.50	0.51	0.56	0.57
	Piecewise Gompertz	0.50	0.50	0.55	0.56

Source: Adapted from Celgene submission, Table 11

Table 19: ERG PFS HR estimates

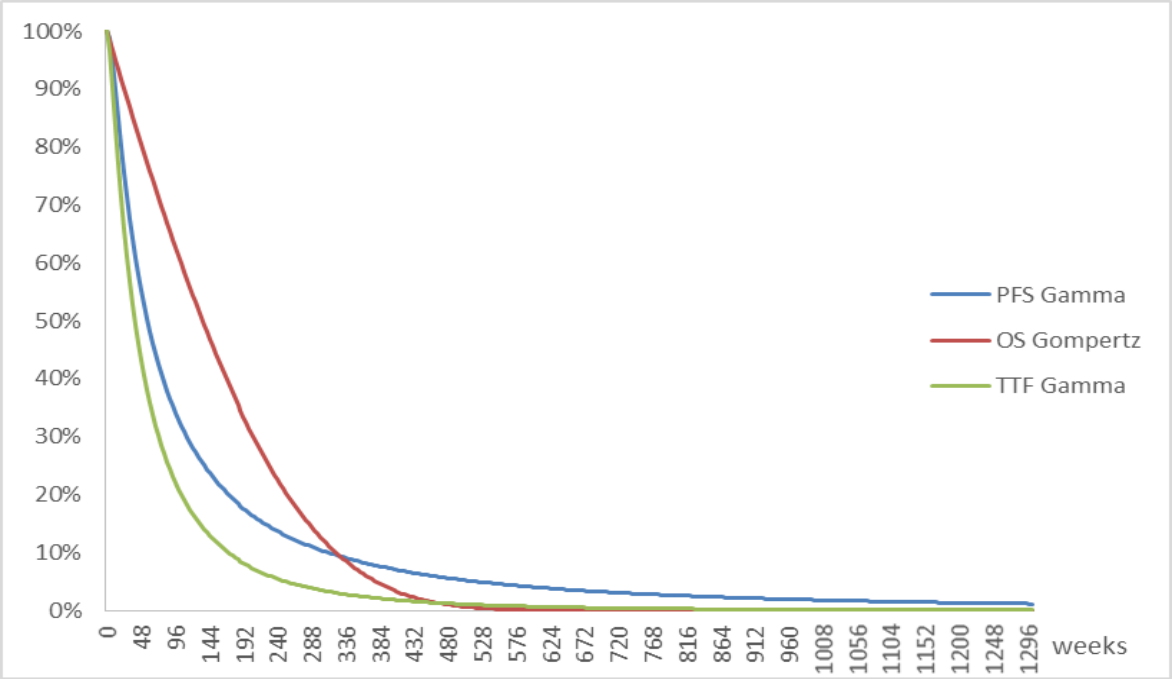
Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data "unadjusted"	LEN/DEX data adjusted	LEN/DEX data "unadjusted"
Taverna 2012	Gompertz	0.80	0.67	1.08	0.98
	Log-logistic	0.82	0.72	0.95	0.87
	Log-normal	0.80	0.70	0.95	0.87
	Gamma	**0.84**	0.75	0.95	0.91
Hrusovsky 2010	Gompertz	0.76	0.60	1.03	0.87
	Log-logistic	0.79	0.64	0.91	0.77
	Log-normal	0.77	0.62	0.91	0.77
	Gamma	0.80	0.66	0.91	0.81
Sood 2009	Gompertz	0.84	0.42	1.03	0.61
	Log-logistic	0.79	0.45	0.91	0.55
	Log-normal	0.77	0.44	0.91	0.55
	Gamma	0.80	0.47	0.91	0.57
Petrucci 2013	Gompertz	0.84	0.54	1.05	0.78
	Log-logistic	0.86	0.58	0.93	0.70
	Log-normal	0.82	0.56	0.95	0.70
	Gamma	0.84	0.60	0.95	0.73
Dispenzieri 2010	Gompertz	0.70	0.60	0.83	0.79
	Log-logistic	0.76	0.61	0.76	0.70
	Log-normal	0.74	0.60	0.76	0.70
	Gamma	0.76	0.62	0.76	0.70

Source: Adapted from Celgene submission, Table 12

After applying the estimated HRs to the intervention survival curves, Celgene obtained the survival curves for the Bort arm of the model.

Figure 10 shows OS, PFS and TTF adjusted curves used in the comparator arm of the model. As discussed previously, the ERG are concerned with the OS curve crossing the PFS (around 6 years) and the TTF curve around week 470 (9 years).

Figure 10. OS, PFS and TTF adjusted Bort curves



The ERG have previously reported their concerns with regards to the first adjustment process used to derive the survival curves used in the intervention arm of the model. Additionally to these, the ERG have several concerns with the adjustment of the curves to the comparator study populations and with the HRs calculations. Thus it can be inferred that the curves used in the comparator arm of the model are even more exposed to uncertainty and are methodological weaker as these were derived by carrying out the two adjustment processes.

- Corrected group prognosis method

The same approach that was used for the MOC method was employed in the CGP method. In order to estimate median survival estimates and compare these to the ones observed in Taverna (2012), Celgene had to make some assumptions given that the Taverna (2012) retrospective study did not provide patient level data.

Celgene assumed that the likelihood of a patient presenting in one particular clinical diagnosis had no impact on the likelihood of the patient falling into any other possible clinical combination of outcomes. For example, the likelihood of having a beta-2 microglobulin level > 2.5mg/L is independent of the myeloma duration observed in that patient. This is a strong assumption, therefore the Committee should assess its clinical validity.

2.2 The effectiveness of Len/Dex compared with chemotherapy

Summary of previous submission and Committee's comment

In their previous submission, Celgene have run scenario analysis where Len/Dex was compared against bendamustine and other chemotherapy agents.

Celgene used Damaj (2012) study to assess the effectiveness of Len/Dex compared with chemotherapy agents. Damaj (2012) looked into the effectiveness of bendamustine given after treatment with bortezomib in a previous treatment line. As the manufacturer claimed there were no studies available for treatment with chemotherapy agents after initial treatment with bortezomib, the effectiveness of bendamustine was used as proxy for the clinical effectiveness of other chemotherapy agents.

However, the Committee decided that bendamustine was not a relevant comparator as clinical experts mentioned it was rarely used and it is not licensed for second-line treatment (ACD, section 4.3).

The Committee also pointed that there were two relevant **subgroups of patients**. Those for whom Bort retreatment would be appropriate, in which case the treatment options considered should include Len/Dex, Bort and standard chemotherapy and those for whom retreating with Bort would not be appropriate, in which case the treatment options include Len/Dex and standard chemotherapy (ACD, section 4.3).

The Committee concluded that, when offered second line, Len/Dex was likely to be more effective than placebo and standard chemotherapy, but that it was unclear whether Len/Dex was clinically more effective than retreatment with Bort (ACD, section 4.4)

Overview of method applied in the revised analysis

The manufacturer states that the same methods used to derive the effectiveness of Len/Dex compared to Bort were used to estimate the effectiveness of Len/Dex compared with chemotherapy agents.

In their revised submission, Celgene didn't use the Damaj (2012) study. They considered this study to be no longer appropriate given that bendamustine was excluded from the analysis. Therefore the manufacturer have broadened their search to identify any evidence for chemotherapy in patients who had at least one prior therapy (i.e. not limiting the prior therapy to Bort).

Two relevant papers have been identified which contained data on the effectiveness of chemotherapy regimens. Celgene used the MOC and the CGP methods to adjust survival curves to the comparator studies. The studies considered relevant for analysis were:

- Celesti (1997): a non-randomised study comparing high-dose cyclophosphamide 1.2 g/m2 + dexamethasone and low-dose cyclophosphamide 0.5 mg/m2 + dexamethasone which reported the overall response rate (ORR), duration of response (DoR), OS and toxicity.

- Petrucci (1989): a single arm trial of melphalan 25 mg/m2 IV + prednisolone 60 mg/m2 which reported ORR, DoR, OS and safety.

Celgene reported the trial outcomes and baseline characteristics from the two trials in Appendix E of their revised submission. We have reproduced these in Table 22 and Table 23. To note is that the outcomes for Petrucci (1989) were not reported by Celgene. The manufacturer state that the outcomes presented in the two chemotherapy studies were significantly worse than the outcomes predicted in Damaj (2012) for Bort.

The manufacturer also claim that the Celesti study reported that “the two schedules (high-dose vs low-dose cyclophosphamide) proved to be equally effective”, therefore they have used pooled outcomes from the study in their analysis.

Celgene state that alternative chemotherapies (such as vincristine and doxorubicin) were not included in the analysis due to lack of evidence and similar costs across all agents.

Adjustment to comparator population characteristics

The models designed to predict PFS, TTF and OS in the Len/Dex arm of the model were used to improve exchangeability between studies. This was done by adjusting the median pooled survival estimates from MM-010 and MM-009 to reflect the characteristics of the population in the comparator study (e.g. Celesti, 1997 and Petrucci, 1989).

Details of patient characteristics for the covariates included in the final models were extracted from the literature and are reported in Table 20.

Table 20: Patient characteristics from the literature identified – chemotherapy

Source	***Characteristic***	***Value***
Celesti et al (1997)	*1 prior therapy (vs 2+ prior therapies)*	*40.0%*
Petrucci et al (1989)	*Myeloma duration*	*Median: 2.25 years*
MM-009 and MM-010 ITT population	*1 prior therapy (vs 2+ prior therapies)* *Myeloma duration*	*35.1%* *Median: 3.9 years*

Source: Celgene revised submission, Table 13

Celgene reports using the available chemotherapy evidence through the same methods as the ones previously described for Bort, to establish the effectiveness of Len/Dex compared with chemotherapy (i.e. to calculate HRs).

Table 21 presents the HRs reported by Celgene in their revised submission. Similarly to the Bort analysis, the manufacturer present adjusted and “unadjusted” HRs for chemotherapy agents vs Len/Dex, using the MOC and the CGP method respectively. HRs>1 favour Len/Dex. Celgene note that the HRs for chemotherapy agents are notably higher than those estimated for Bort, reflecting the clinical superiority of Bort compared with chemotherapy.

Table 21: OS HR estimates – chemotherapy agents compared to Len/Dex

Source of adjustment	***Model type***	***HRs using mean of covariates method***		***HRs using corrected group prognosis method***	
		LEN/DEX data adjusted	***LEN/DEX data unadjusted***	***LEN/DEX data adjusted***	***LEN/DEX data unadjusted***
Celesti et al (1997)[20]	*Exponential*	*5.13*	*5.01*	*4.25*	*4.22*
	Gompertz	*5.19*	*5.07*	*4.39*	*4.36*
	Weibull	*5.10*	*4.98*	*4.28*	*4.22*
	Gamma	*5.25*	*5.13*	*4.39*	*4.36*
	Piecewise exponential	*4.89*	*4.84*	*4.28*	*4.30*
	Piecewise Gompertz	*4.98*	*4.89*	*4.33*	*4.39*
Petrucci et al (1989)[21]	*Exponential*	*4.51*	*4.89*	*3.79*	*4.11*
	Gompertz	*4.66*	*4.94*	*4.00*	*4.25*
	Weibull	*4.54*	*4.86*	*3.85*	*4.11*
	Gamma	*4.71*	*5.00*	*4.00*	*4.25*
	Piecewise exponential	*4.54*	*4.71*	*4.25*	*4.20*
	Piecewise Gompertz	*4.63*	*4.77*	*4.34*	*4.28*
Key: *HR, hazard ratio; LEN/DEX, lenalidomide plus dexamethasone; OS, overall survival.*					

Source: Celgene revised submission, Table 14

To note is that the manufacturer report that chemotherapy evidence was available for OS alone, so Celgene used the OS HR to estimate PFS and TTF in the model.

Celgene conclude that the results reveal a considerable efficacy advantage of Len/Dex over standard chemotherapy agents. Furthermore they point out that the survey of clinicians carried out in England, presented within the original submission, found that such regimens were used in around 5% of patients (and no more than 10%).

The manufacturer state that this was comparable to the reported use of bendamustine, which the Committee have agreed is not a relevant comparator.

Table 22: Key study & patient characteristics of non-RCTs identified for chemotherapy

Study	*Design*	*Drug & dose*	*N*	*Age, median years (range)*	*Sex, % male*	*ISS, % per stage*	*Prior therapies, % per number*	*Primary outcome(s)*	*Secondary outcome(s) of interest*
Celesti, 1997[20]	*Non-randomised study*	*Cyclophosphamide 1.2 g/m^2 + LD dexamethasone*	*13*	*69 (44-74)*	*54*	*II = 31; III = 69*	*1 = 60; 2 = 8; 3 = 24; 4 = 8*	*ORR*	*DoR; OS; toxicity*
		Cyclophosphamide 0.5 mg/m^2 + MD dexamethasone	*15*	*61 (54-73)*	*53*	*I = 7; II = 14; III = 79*	*1 = 20; 2 = 27; 3 = 33; 4 = 13; 5 = 7*		
Petrucci, 1989[21]	*Single-arm trial*	*Melphalan 25 mg/m^2 IV + prednisolone 60 mg/m^2*	*34*	*59*	*59*	*NR*	*NR*	*ORR*	*DoR; OS; safety*

Key: *CR, complete response; DoR, duration of response; HD, high dose; HT, haematological toxicity; LD, low dose; MD, medium dose; NR, not reported; ORR, overall response rate; OS, overall survival; WHO, World Health Organisation*

Source: Celgene revised submission, Table 38

Table 23: Key clinical and safety outcomes from non-RCTs identified for chemotherapy

Study	***Treatment arm***	***Survival outcomes, median months***	***Response outcomes, n (%)***	***Discontinuations, n (%)***	***Adverse events, n (%)***
Celesti, 1997[20]	*HD cyclophosphamide + LD dexamethasone*	*OS = 9.4*	*ORR = 5 (38)* ***Median months*** *DoR = 7.6*	*NR*	***Toxicity (WHO), %*** *0 = 31; 1 = 31; 2 = 31; 3 = 7*
	LD cyclophosphamide + MD dexamethasone	*OS = 6.3*	*ORR = 5 (34)* ***Median months*** *DoR = 5.2*		*0 = 53; 1 = 27; 2 = 20*
	Overall	*OS = 7.8*	*ORR = 10 (36)* ***Median months*** *DoR = 6.3*		

Source: Celgene revised submission, Table 3

Critique of the method applied in the revised analysis

The manufacturer state that the same methods used to derive the effectiveness of Len/Dex compared to Bort were used to estimate the effectiveness of Len/Dex compared with chemotherapy agents. Therefore, all the issues raised by the ERG in the previous section apply for the analysis of chemotherapy agents.

Celgene decided not use the Damaj (2012) study as it looked at the effectiveness of bendamustine and so they have broadened their search to identify any evidence for chemotherapy in patients who had at least one prior therapy (i.e. not limiting the prior therapy to Bort).

Celgene mention that the efficacy of chemotherapy is assumed to be unaffected by the type of prior therapy given. While this might be true in the case where Bort was the drug received at a previous treatment lines, this statement is unlikely to be valid in the case that the previous therapy was based on chemotherapy agents. Similarly to Bort, retreatment with chemotherapy is likely to be less effective than initial treatment with chemotherapy agents. Clinical opinion sought by the ERG confirmed this to be the case.

In the Celesti (1997) and the Petrucci (1989) papers, all patients received chemotherapy treatment regimens as a previous line of treatment. This means that the effectiveness of cyclophosphamide + dexamethasone (CY-DEX) evaluated in Celesti (1997) is likely to be underestimated. The same is valid for the effectiveness of melphalan + prednisolone (MP) analysed in the Petrucci (1988) study.

Celgene fail to report the outcomes for the Petrucci (1989) study. The ERG have therefore reported the main study outcomes in Table 24. To note is that for the patients responding to the drug (reduction >50% of the abnormal protein with disappearance of all clinical symptoms), median OS was not achieved by the end of the study at 28 months.

The manufacturer also claim that the Celesti study reported that “the two schedules (HD vs LD cyclophosphamide) proved to be equally effective”, therefore they have used pooled outcomes from the study in their analysis. While it is true that Celesti (1997) mention both regimens to be effective, there is no evidence presented to support this argument. Statistical significance was not investigated in terms of median OS nor in terms of ORRs.

Adjustment to comparator population characteristics

Again, the concerns raised by the ERG in Section 2.1 with regards to the adjustment of the survival curves to the comparator studies’ population characteristics hold their validity for the chemotherapy analysis. The same is true for the HR calculations.

Table 24. Patient characteristics in Petrucci (1989)

Study	Treatment arm	Survival outcomes, median months	Response outcomes, n (%)
Petrucci (1989)	All treated patients (n=34)	OS = 8	ORR = 12 (35) **Median months** DoR = 16
	Responding patients (n=12)	OS = not achieved	

Table 25 presents the HRs reported by Celgene in their revised submission. Similarly to the Bort analysis the "unadjusted" HRs for chemotherapy agents vs Len/Dex cannot be considered as truly unadjusted estimates since they are adjusted to the predictive variables in MM-010 and MM-009. The ERG have also presented the inverted HRs (corrected for the mistakes previously mentioned), to reflect the effectiveness of Len/Dex vs chemotherapy. HRs<1 show that Len/Dex is more effective than chemotherapy regimens.

To note is that for the reasons explained above, it is likely that the effectiveness of chemotherapy is being underestimated, therefore it is also likely that the HRs produced by Celgene are overestimating the effectiveness of Len/Dex compared with chemotherapy.

The OS HR used for MP (in bold in Table 25) were estimated with a Gompertz model, in the same fashion as the OS HR used for Bort.

Table 25: OS HR estimates – chemotherapy compared to Len/Dex

Source of adjustment	Model type	HRs using mean of covariates method		HRs using corrected group prognosis method	
		LEN/DEX data adjusted	LEN/DEX data unadjusted	LEN/DEX data adjusted	LEN/DEX data unadjusted
Celesti et al (1997)	Exponential	0.19	0.20	0.24	0.24
	Gompertz	0.19	0.20	0.23	0.23
	Weibull	0.20	0.20	0.23	0.24
	Gamma	0.19	0.19	0.23	0.23
	Piecewise exponential	0.20	0.21	0.23	0.23
	Piecewise Gompertz	0.20	0.20	0.23	0.23
Petrucci et al (1989)	Exponential	0.22	0.20	0.26	0.24
	Gompertz	**0.21**	0.20	0.25	0.24
	Weibull	0.22	0.21	0.26	0.24
	Gamma	0.21	0.20	0.25	0.24
	Piecewise exponential	0.22	0.21	0.24	0.24
	Piecewise Gompertz	0.22	0.21	0.23	0.23

Celgene used the OS HR to estimate PFS and TTF in the model as the chemotherapy studies did not provide PFS and TTF data. However OS is very different from PFS and TTF in the MM studies.

The ERG question the utility of modelling PFS and TTF with the OS HR. Instead, Celgene could have used the Damaj (2012) PFS data for bendamustine as a proxy for chemotherapy as previously done.

Clinical opinion sought by the ERG agreed with the manufacturer that only about 5% of patients are prescribed chemotherapy agents at second line and that if there is a basis for excluding bendamustine from the analysis then the same logic could be applied to chemotherapy. Clinical opinion also informed the ERG that there is evidence suggesting that after an initial treatment with Bort failed, patients who are prescribed Bort plus bendamustine present good results.

3.0 Model structure

Celgene seem to have addressed only one of the issues raised by the Committee with regards to the model structure (even though this is not mentioned in the revised submission).

Therefore the ERG assume that all other issues initially raised remain unchanged and therefore are still a source of concern.

We present a brief summary of the issues previously raised:

- There was not a clear separation between second-line treatment outcomes and the beginning of the third-line treatment option and respective outcomes in the Bort arm of the model. Not only this reflects a slight structural inconsistency between intervention and comparator arms of the model, but also from a clinical point of view, this seems to reflect a very unlikely scenario. This issue seems to have been addressed. There is now a clear separation between second and third-line treatments in the Bort arm of the model.

- After second-line of treatment, the manufacturer only considered the utility associated with the PD state. Arguably, these patients would be expected to stay in a "post-relapsed" PFS state for a certain period of time (while experiencing a higher utility) and then progress again (experiencing a lower utility). This has not been addressed in the revised submission.

- In the Len/Dex arm of the model, subsequent treatment options are not evaluated on their effectiveness. Only costs of third and fourth-line treatment baskets are considered. However, in the Bort arm of the model, patients receive Len/Dex as third-line treatment, following the discontinuation of Bort. Hence these patients are exposed to the OS and PFS hazards associated with Len/Dex at second-line. To model fourth-line treatment in the comparator arm, a "real-world" treatment basket is used which only evaluates costs and excludes the effectiveness of the treatment basket from the analysis.

- Celgene assumed that the second-line effectiveness of Len/Dex is replicated at the third line of treatment. Celgene also argued that Len/Dex is shown to be more effective at earlier lines. The ERG had some issues with the data used to back up these assumptions.

Therefore the ERG believe that the following issues should be considered:

1. The likelihood of MM patients receiving third and fourth-line treatment regimens:

   Clinical opinion sought by the ERG explained that on average, once patients fail treatment with Len, they frequently live for a short period of time. This is due to the AEs related with the use of Len, especially a very low bone marrow function. This is, however, closely related to the duration of treatment.

   Furthermore clinical opinion revealed that some of the drugs considered in the treatment basket like cisplatin, doxorubicin and etoposide are generally out of use in current clinical practice.

2. The value of including subsequent treatment options in the economic analysis given that mainly only cost data is used:

The ERG understand the value of including subsequent treatment options to reflect the MM complex clinical pathway as accurately as possible. However, in this case and given that the available data does not allow for the evaluation of effectiveness and quality of life resulting from further treatment options in the Len/Dex arm we question the value of only costing these options. Additionally, the treatment mix might not accurately reflect current practice as mentioned in the point above.

Furthermore including Len/Dex as a third-line treatment option for Bort patients and assuming that the effectiveness of Len at third-line is the same as the effectiveness of Len/Dex at second-line raises some concerns.

4.0 Costs

Summary of previous submission and Committee's comment

The Committee noted four key issues with regards to the costs used in the economic model.

The fact that Celgene didn't include the costs of Dex in the comparator arm of the model in the base case analysis was considered inappropriate as clinical opinion informed that corticosteroids are taken concomitantly with Bort in clinical practice. The Committee therefore agreed that it would have been more appropriate to include the costs of dexamethasone in the model (ACD, section 4.11).

The inclusion of patient access schemes (PAS) was also discussed. The Committee noted that the manufacturer had not offered a PAS for second line use of Len and agreed that the PAS for Len used at third line should be included in the economic analysis. The Committee also agreed that the patient access scheme for Bort should be included in the analysis, although it acknowledged that precise information about the number of patients for whom the rebate would be applied is not available (ACD, section 4.11).

The duration of retreatment with Bort was also discussed. The Committee noted that in the ERG's scenario analyses, the ICER considerably increased (from being a dominant ICER to nearly £30,000) when the number of Bort cycles was reduced from 'treatment until progression' to 8 cycles. The Committee heard from the clinical specialists that clinical practice in the UK varies in the number of Bort treatment cycles offered to patients, from 6–8 cycles to until treatment progression. The Committee agreed that, because bortezomib was given until treatment progression in the manufacturer's model, this overestimated the cost of treatment in the comparator arm. The Committee concluded that the true number of cycles was likely to be lower, and would lie between 6 cycles to until treatment progression (ACD, section 4.11).

Overview of method applied in the revised analysis

Inclusion of the cost of corticosteroids in the comparator arm of the model

Celgene state that the cost of Dex was included in the base case model for the proportion of patients receiving Dex in the Taverna (2012) study. In the retrospective study, 64% of patients received the corticosteroid together with Bort.

Len and Bort PASs

- Third-line Len PAS

Celgene pointed to the fact that the PAS for Len at third-line had been included in the base case analysis in their previous submission.

The manufacturer also explained that in the revised analysis the PAS within the model was changed in order to simplify the analysis and improve consistency between the methodology used in the intervention and comparator arms of the model. As a result the level of discount received was derived from the mean time on treatment with Len before and after the start of the PAS (7 cycles and 20 cycles, respectively), using the curves fitted to patients receiving third-line therapy in MM-010 and MM-009.

It was estimated that approximately 16% of patients remain on treatment at 26 cycles. Celgene mention that this estimate is consistent with available literature.

- Bort PAS

Celgene state that the Bort PAS was included in their base case revised model. They have applied a 15% discount, however it is explained that the uptake of Bort PAS is low in clinical practice. The manufacturer explain that the reason for this is that the scheme requires close clinician monitoring and record-keeping, and that the refund must be claimed within 60 days of the qualifying assessment.

Furthermore, Celgene indicate that available literature shows that less than half of patients who would be eligible actually receive this discount, therefore only 55% of patients in the revised model receive the discount.

Duration of Bort retreatment

Celgene claim that published evidence indicates that Bort retreatment is given for longer durations than 8 cycles. They claim that both in the Taverna (2012) and Sood (2009) studies, patients receive up to 19 cycles of Bort.

Celgene also claim that patients receiving Bort for the first time in the APEX study were allowed to continue on Bort as maintenance after receiving an initial 8 cycles. They explain that patients could receive up to 11 cycles and that responding patients received a median of 10 cycles.

Celgene also mention NICE TA 129, where they claim that the SPC *does not specify a maximum number of cycles in patients who receive monotherapy*. They report the wording in the SPC (Table 26) to back up this statement.

Table 26. Bort SPC

Indication / dosing regimen (not specific to retreatment)	***Wording of duration of treatment from the SPC***
Progressive multiple myeloma (patients who have received at least one prior therapy) monotherapy	*"It is also recommended that responding patients who do not achieve a complete remission receive a total of 8 cycles of VELCADE therapy."*
Progressive multiple myeloma (patients who have received at least one prior therapy) combination with dexamethasone	*"Patients achieving a response or a stable disease after 4 cycles of this combination therapy can continue to receive the same combination for a maximum of 4 additional cycles."*
Note: *There is no guidance in the SPC relating to the duration of treatment for bortezomib when used as a retreatment.*	

Source: Celgene revised submission, Table 17

Celgene explain that the Bort SPC does not contain specific information on posology for retreatment and that some centres provide the drug until progression and some provide only a maximum of 8 cycles. Furthermore they claim that only the Petrucci (2012) study reported a maximum of 8 cycles but that this study did not present OS, PFS or TTP for all patients.

Celgene mention that clinical opinion supports that the effectiveness of Bort is directly related to the number of cycles administered. The manufacturer say that this idea is reinforced by the Petrucci (2012) study, *which reports a TTP for only the 40% of patients who responded. This is still lower than that of the overall population (including non-responding patients) in the Taverna et al data (8.4 months vs 10.5 months).*

The manufacturer also state that data from MM-010 and MM-009 show that patients who respond to Len/Dex experience considerably longer TTP than those who don't. They go on to the conclude that as both OS and TTP/PFS are only available from the Taverna and the Hrusovsky studies, *the most reasonable comparison with bortezomib retreatment is one which assumes that treatment continues for longer than 8 cycles, provided the patient is continuing to receive benefit.*

Celgene therefore decide to model the use of Bort until disease progression. The manufacturer mention that scenario analysis explored the use of Bort for fewer cycles, but that the impact on the treatment effectiveness could not be assessed due to lack of evidence. Therefore only costs were reduced as a result of less treatment cycles, which Celgene deem to bias the results in favour of Bort retreatment.

Critique of the method applied in the revised analysis

Inclusion of the cost of corticosteroids in the comparator arm of the model

Celgene included the cost of Dex for 64% of the Bort patients, which is in agreement with the Taverna (2012) study.

From analysing the cost calculations in the excel model it seems that Celgene assumed the following dose regimen: 38mg per day, for 12 days every cycle, in the initial 4 cycles of treatment and then 4 days ever cycle after the fifth cycle. However, this was left to the ERG interpretation, by observing the calculations in the excel model, when it should have been clearly stated in the revised submission.

Len and Bort PASs

- Third-line Len PAS

The manufacturer explained that in the revised analysis the PAS within the model was changed in order to simplify the analysis and improve consistency between the methodology used in the intervention and comparator arms of the model. Overall the ERG are confused with the cost calculations associated with Len/Dex at third line in the comparator arm of the model.

- Bort PAS

Celgene have applied a 15% discount for 55% of Bort patients in the base case analysis. Clinical opinion sought by the ERG confirmed that the Bort PAS is hardly ever used in clinical practice.

Nonetheless, the ERG have identified an error in the calculation of the Bort costs with the PAS. Taking the PAS into consideration, the cost of Bort per cycle should be calculated as:

(cost of Bort per cycle * 15% * number of people getting Bort per cycle * 55%) + (cost of Bort per cycle * number of people getting Bort per cycle * 45%) = cost related to patients received PAS + cost related to the remaining patients

However Celgene are calculating this as:

cost of Bort per cycle * number of people getting Bort per cycle * [1 – (55%* 15%)]

The two formulas are not equivalent, with Celgene's formula overestimating the cost of Bort per cycle. The ERG have corrected this mistake in the economic model and present the results in Section 6.

Duration of Bort retreatment

Celgene claim that published evidence indicates that Bort retreatment is given for longer durations than 8 cycles. This statement should be interpreted with caution.

The manufacturer claim that both in the Taverna (2012) and Sood (2009) studies, patients receive up to 19 cycles of Bort. This should be clarified, as only a very small number of patients actually receive more than 8 cycles.

In Taverna (2012), the median duration of Bort retreatment is 3 cycles, with **90.4% of paints receiving 1 – 6 cycles** and only 2% of patients receiving more than 10 cycles (max 19 cycles). In the analysis undertaken by Sood (2009), 50% of patients received 5 cycles of Bort retreatment, with the range being 1 to 12 cycles.

Celgene also claim that patients receiving bortezomib for the first time in the APEX study were allowed to continue on bortezomib as maintenance after receiving an initial 8 cycles. The ERG have looked into the APEX study and 56% of patients completed 5 cycles of Bort while 29% completed 8 cycles of Bort.

Celgene also mention that the Bort SPC *does not specify a maximum number of cycles in patients who receive monotherapy*. Even though the wording in the SPC (Table 26) is reported by the manufacturer to back up this statement, it seems that the SPC clearly defines a maximum of 8 cycles of treatment with Bort.

Furthermore the manufacturer claim that only the Petrucci (2013a) study reported a maximum of 8 cycles but that this study did not present OS, PFS or TTP for all patients. The ERG believe this argument backs up our view on the matter and that the duration of Bort treatment is approximately 8 cycles. Furthermore Petrucci (2013b) which reports a maximum of 9 cycles of Bort presents OS and PFS (even though Bort is administrated along with chemotherapy).

The manufacturer state that data from MM-010 and MM-009 show that patients who respond experience considerably longer TTP than those who don't. They go on to the conclude that as both OS and TTP/PFS are only available from the Taverna and the Hrusovsky studies, *the most reasonable comparison with bortezomib retreatment is one which assumes that treatment continues for longer than 8 cycles, provided the patient is continuing to receive benefit.* Celgene therefore decide to model the use of Bort until disease progression.

The ERG find this argument confusing. Taverna (2012) reports a median duration of Bort retreatment of 3 cycles, with 90.4% of paints receiving 1 – 6 cycles and only 2% of patients receiving more than 10 cycles (max 19 cycles). Hrusovsky (2010) reports that 43% of patients received 1-3 cycles, 41% of patients received 4-6 cycles and only 6% received more than 10 cycles (max 14 cycles).

In conclusion the majority of studies point to a treatment duration between 1 - 6 cycles. In that sense the ERG suggestion of 8 cycles to model Bort retreatment could be seen as conservative. It is also evident from literature that Bort is not given to patients until disease progression but instead for a fixed time. Further to this, only a very small percentage of patients across studies (usually around 2%) undertake Bort retreatment for more than 8 cycles. The decision to model the duration of Bort retreatment until disease progression in therefore questionable.

The manufacturer have run scenario analysis, where they use 19 cycles as an alternative to the unlimited treatment duration of Bort. The ERG feel that based on the available literature this is an extreme scenario and it should have been considered together with a scenario where the duration of treatment was 8 cycles.

The duration of treatment is one of the key drivers of the final ICER therefore we present exploratory analysis around this parameter is Section 6.

Finally, some of the issues mentioned in the ERG's previous report don't seem to have been addressed. For example, for the purpose of calculating transportation costs Celgene assumed that 50% of patients require transportation for their treatment administration and also that if more than one treatment occurs during one week the patient will be kept in the hospital for up to one week. Clinical opinion sought by the ERG did not believe this to be a reasonable assumption. Firstly the percentage of MM patients requiring transportation to the hospital was considered to be significantly lower than 50%. Secondly, the assumption that patients in need of more than one treatment per week would stay in the hospital was believed to be unrealistic. These assumptions are likely overestimating the cost of Bort in the economic analysis as only one administration visit is considered for Len.

5.0 Use of utilities in the cost-effectiveness model

Summary of previous submission and Committee's comment

Health-related quality of life data (HRQoL) were not collected in MM-010 and MM-009 clinical trials. As such, in their previous submission the manufacturer used the utility estimates from TA171 which were taken from a cost-utility carried out by van Agthoven (2004) in patents with previously untreated multiple myeloma. The utility estimate for the pre-progression state was 0.81, which decreased after 2 years to 0.77 and the value for the post-regression state was 0.64. The utility values were adjusted for age, based upon published UK EQ-5D values by Kind (1999).

The Committee was concerned that the utility values were derived from a younger population and were higher than the average population of the same age. The Committee was also aware that utility decrements for adverse events were taken from several sources. Therefore the Committee concluded that although the values had been used in previous appraisals there was significant uncertainty surrounding the values of utility and disutility used in the modelling (ACD, section 4.11).

The Committee also noted that it was not clinically plausible that as an individual's disease progresses, the utility value remained constant despite receiving third- and fourth-line treatments (ACD, section 4.10).

Overview of method applied in the revised analysis

Celgene have undertaken three scenario analyses using different sources for the utility values applied in the economic model. Two scenarios use utility values presented in the TA228 ERG report while the third one uses a utility decrement for patients entering fourth-line therapy.

These scenarios are described in turn.

Utility values from alternative sources presented in TA228

The TA228 ERG looked into a systematic review of HRQoL studies in the population of interest i.e. previously untreated MM who are not candidates for high-dose chemotherapy with autologous stem cell transplantation (SCT). The systematic review did not find any generic preference-based QoL studies but identified two studies which used the disease specific questionnaire EORTC QLQ-C30: Gulbrandsen (2004) and Strasser-Weipl and Ludwig (2008). In order to present comparative data, the ERG then mapped the QoL scores obtained from the EORTC QLQ-C30 questionnaire onto the EQ-5D using two different algorithms: one described by McKenzie and van der Pol (2009) and one by Kontodimopoulos and colleagues (2009).

The ERG suggested that the most appropriate source of HRQoL data for the treatment and post-treatment periods was Gulbrandsen (2004) which used the mapping by McKenzie and van der Pol. They suggested using the utility value estimated from the 1-month time point for the treatment period (0.58 utility value), and the average of the 6 to 36-month time points for the post treatment (and post progression) period (0.68 utility value).

Celgene have used the same utility estimates, but the manufacturer decided to use the different estimates for different points in time instead of associating these with specific health states. Utility values used by Celgene are presented in Table 27.

Table 27. EQ-5D utilities derived by mapping EORTC QLQ-C30 scores as presented by the ERG in TA228

Algorithm used to map to EQ-5D	Time in Months				
	0	1	5	12	24
McKenzie and van der Pol	0.55	0.58	0.68	0.68	0.68
Kontodimopolous	0.52	0.58	0.69	0.69	0.69

Source: Adapted from Celgene revised submission, Table 15.

Reduced utility for subsequent treatment lines

Celgene also ran a scenario where a lower utility value was applied to patients starting fourth-line treatment, which remained constant until death. The utility values used in this scenario are presented in Table 28.

The utility value used for the fourth-line treatment was taken from a recent poster by Palumbo (2013) reporting a baseline utility of 0.59 for MM patients (at a median of 5 prior lines of treatment). QoL was collected within the MM-003 study using the EQ-5D questionnaire.

Table 28. EQ-5D utilities by health state – lower utility from fourth-line treatment scenario

State	Utility	Source
Pre-progression	0.81	van Agthoven et al. (2004)
Pre-progression after 2 years	0.77	van Agthoven et al. (2004)
Post-progression	0.64	van Agthoven et al. (2004)
4th line onwards	0.59	Palumbo et al. (2013)

Source: Adapted from Celgene revised submission, Table 16

Critique of the method applied in the revised analysis

Utility values from alternative sources presented in TA228

Celgene do not explain why they have decided to take a different approach from the one suggested in TA228. Even though the manufacturer are using alternative utility values these are not associated with any disease state (as suggested in TA228) but instead with time.

Celgene present two alternative scenarios using utility values derived from the same population presented in Gulbrandsen (2004). Gulbrandsen and colleagues analysed QoL scores from two prospective Nordic Myeloma Study Group trials: one of them with a population of 221 patients aged <60 years and the other with a population of 203 patients aged >60 years. No mean age was presented in the study.

Even though the utility values applied in the two scenarios are lower than the average population values for the same age in an healthy population (as reported in Table 41 of the previous ERG

report, 63-year old healthy patients are associated with a mean utility of 0.80), the manufacturer do not address the Committee's concern regarding the age of the population used to derive the utility values. The population used to derive the utility values is still a younger population.

Reduced utility for subsequent treatment lines

The ERG appreciate the fact that Celgene tried to improve the clinical plausibility of the modelled utility in subsequent treatment lines by applying a utility decrement to fourth-line treatments. However the fundamental problem still remains. After second-line of treatment patients would be expected to stay in a "post-relapsed" PFS state for a certain period of time (while experiencing a higher utility) and then progress again (experiencing a lower utility). This has not been entirely addressed in the revised analysis.

Furthermore, it is unclear why the manufacturer changed the fourth-line treatment utility but did not change the third-line estimates.

Finally, since the 0.59 utility estimate was derived from a different population, the Committee might want to consider the clinical plausibility of this value. To note is that there is no information available regarding the age of patients used for this QoL study.

In conclusion, there is some value to the additional analysis presented by Celgene and the impact of this analysis on the final ICER is presented in the next section of the report. However, for the reasons explained above, the ERG are not convinced that the undertaken scenario analysis addressed the Committee concerns in full.

6.0 Economic modelling results

In this section the ERG present the base case ICERs as well as the outputs of different scenario analyses.

To note is that a wide range of results could have been presented, resulting from all the possible combinations of alternative methodologies and assumptions used in the economic model. Each scenario considered could be presented for all the following alternatives:

- Duration of Bort retreatment (until disease progression, 19 cycles and 8 cycles)
- Inclusion of subsequent treatment lines vs considering second-line treatment only
- MOC method vs the CGP method

Since the number of potential scenarios is vast, we have presented the ones deemed more relevant. As a result, we do not present some of the scenarios ran by Celgene and we report some additional ones ran by the ERG.

For consideration of all the scenarios presented by the manufacturer, the reader should consult Section 5 in Celgene's revised submission.

Furthermore, the ERG emphasise that all ICERs presented in this section should be interpreted with extreme caution. As we have explained throughout the report, there are several flaws to the analysis undertaken by Celgene and the impact of these on the final ICERs is not easily predictable.

All values presented are discounted.

Base case

Celgene's revised cost-effectiveness analysis produced a dominant ICER over 25 years (Table 29).

The ERG corrected the errors related to the HR calculations and the Bort PAS aforementioned and present the results in Table 30. It can be noted how the originally dominant ICER increased to **£21,787** per QALY gained after the ERG corrections and that this increase was due to changes in the comparator arm of the model (the costs and QALYs for Len/Dex remained the same after the ERG corrections).

Both Table 29 and Table 30 present the base case ICERs including third and fourth-treatment lines and assuming that Bort is given until disease progression. In Table 31 we present the base case ICER with the ERG corrections but using the CGP method. We can see that in comparison with the MOC method, the CGP analysis produced lower ICERs.

However the ERG have several concerns with the inclusion of subsequent treatment lines in the economic model and with the use of Bort until disease progression. Therefore in the next sections we present the ICERs for second-line patients only and consider different treatment durations for Bort.

Table 29. Len/Dex vs bortezomib retreatment – Celgene base case

Treatment	Costs	QALYs	LYs	ICER
Len/Dex	£104,864	3.05	4.69	Lenalidomide dominates
Bort retreatment	£118,498	2.24	3.28	
Incremental	-£13,634	0.81	1.41	

Table 30. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS)

Treatment	Costs	QALYs	LYs	ICER
Len/Dex	£104,864	3.05	4.69	**£21,782**
Bort retreatment	£84,314	2.11	3.08	
Incremental	£20,550	0.94	1.62	

Table 31. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS)

Treatment	Costs	QALYs	LYs	ICER
Len/Dex	£102,361	2.90	4.47	**£17,565**
Bort retreatment	£88,385	2.10	3.05	
Incremental	£13,976	0.80	1.42	

In Table 32 we present the base case ICER for Len/Dex compared with chemotherapy (MP) reported by Celgene. The reported ICER is **£54,898** including all treatment lines. To note is that after 22 cycles, the cost of MP is <£1. Comparing Table 32 with Table 29, the ERG find the ICERs reported for chemotherapy quite surprising given the size of the HRs reported for Len/Dex vs chemotherapy (section 2.2). The following issues should be considered:

- The HRs reported for Len/Dex vs chemotherapy are four to five times fold the HRs reported for Len/Dex vs Bort. This suggests that Bort is significantly superior to chemotherapy with regards to OS and PFS. It is therefore surprising that the difference in QALYs between Bort and chemotherapy is only 0.18 QALYs.

- The total costs of chemotherapy are lower than the costs associated with Bort by around £68,000.

- The fact that the total treatment cost with chemotherapy is half of the total treatment cost with Len/Dex (when Bort is more expensive than Len/Dex) but that the QALY gain associated with chemotherapy compared to Len/Dex is similar to the QALY gain associated with Bort compared to Len/Dex produced very dissimilar ICERs.

It is therefore counterintuitive that chemotherapy, which supposedly performs much worse than Bort (according to the difference in HRs) in relation to Len/Dex, would produce ICERs above **£50,000** (potentially indicating that Len/Dex was not cost-effective) while the ICERs comparing Len/Dex with Bort suggest that Len/Dex dominates Bort.

In Table 33 and Table 34 we present the ICERs with the ERG corrections (related to the HRs calculations) and with the CGP method. As above, these ICERs include all treatment lines.

Table 32: Summary results vs MP – Celgene base case

Treatment	Costs	QALYs	LYs	ICER
LEN/DEX	£104,864	3.05	4.69	**£54,898**
Melphalan + prednisolone	£50,450	2.06	3.35	
Incremental	£54,414	0.99	1.34	

Table 33. Len/Dex vs MP - ERG corrections (HRs)

Treatment	Costs	QALYs	LYs	ICER
LEN/DEX	£104,864	3.05	4.69	**£34,076**
Melphalan + prednisolone	£36,796	1.05	1.57	
Incremental	£54,414	2.00	3.12	

Table 34. Len/Dex vs MP - CGP method and ERG corrections (HRs)

Treatment	Costs	QALYs	LYs	ICER
LEN/DEX	£102,361	2.90	4.47	**£35,222**
Melphalan + prednisolone	£34,578	0.97	1.44	
Incremental	£67,782	1.92	3.03	

ICERs for second-line treatment

Here we present the ICERs for second-line treatment. Third and fourth-treatment lines have therefore been excluded from the analysis.

Celgene's revised cost-effectiveness analysis produced a **£4,126** ICER over 25 years (Table 35) for second-line patients.

The ERG corrected the errors related to the HR calculations and the Bort PAS aforementioned and present the results in Table 36. This produced a final ICER of **£38,841** per QALY gained. Again, this increase was due to changes in the comparator arm of the model (the costs and QALYs for Len/Dex remained the same after the ERG corrections).

Both Table 36 and Table 37 present the second-line ICERs assuming that Bort is given until disease progression. In Table 37 we present the second-line ICER with the ERG corrections but using the CGP method. We can see that in comparison with the MOC method, the CGP analysis produces lower ICERs.

To note is that when only second-line treatment is included in the analysis, the resulting ICERs are significantly higher than when all treatment lines are included, with the ERG corrected ICER being **£38,841**.

Table 35. Len/Dex vs bortezomib retreatment – second-line

Treatment	Costs	QALYs	ICER
Len/Dex	£98,231	3.05	**£4,126**
Bort retreatment	£94,893	2.24	
Incremental	£3,338	0.81	

Table 36. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) - second-line

Treatment	Costs	QALYs	ICER
Len/Dex	£98,231	3.05	**£38,841**
Bort retreatment	£61,586	2.11	
Incremental	£36,645	0.94	

Table 37. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS) - second-line

Treatment	Costs	QALYs	ICER
Len/Dex	£96,117	2.90	**£34,847**
Bort retreatment	£68,391	2.10	
Incremental	£27,726	0.80	

In Table 38 we present the ICER derived with the MOC method for Len/Dex compared with chemotherapy (MP) for second-line patients with the ERG corrections. The reported ICER is **£43,708**. To note is that after 22 cycles, the cost of MP is <£1. In relation to the ICER presented in Table 36, the chemotherapy ICER is still considerably higher than the Bort ICER, even though the difference is less drastic than the one observed when all treatment lines are included.

Table 38. Len/Dex vs MP with ERG corrections (HRs) – second-line

Treatment	Costs	QALYs	ICER
LEN/DEX	£98,231	3.05	**£43,708**
Melphalan + prednisolone	£10,922	1.05	
Incremental	£87,309	2.00	

Duration of Bort treatment

In Table 39 we present Celgene's reported ICERs resulting from reducing the duration of Bort to 19 cycles. The ERG are confused with the calculations in the model as it seems that costs are still incurring after 19 cycles. We present the results for all treatment lines and for the second-line treatment alone. No note is that the ICERs reported in Table 39 do not include the ERG corrections.

In Table 40 we present the ICERs with the ERG corrections for 19 and 8 cycles for the second-line treatment only. We have not presented the corresponding ICERs for all treatment lines as we have focused our additional analysis on the second-line population and producing the ICERs for all treatment lines would have required further manipulation of Celgene's model. However, we have used the ERG corrected model to test the variation in the final ICER if all treatment lines were included and for 19 cycles of Bort given the ICER was above **£60,000**.

Table 41 presents the ICERs with the ERG corrections for 19 and 8 cycles of Bort respectively, for the second-line treatment only and using the CGP method. To note is that when the duration of Bort is fixed (instead of administering the drug until disease progression) the CGP method produces

higher ICERs than the MOC method, whilst when Bort is given until disease progression the CGP ICERs were lower than the ones produced with the MOC method.

In the analysis presented in Table 40 and Table 41 we have considered that the total costs for 8 and 19 cycles in the model were simply the sum of costs from cycle 0 to cycle and 8 and 19 respectively.

As we can observe, diminishing the treatment duration for Bort has a massive impact on the final ICERs. Even if Bort is given for 19 cycles (note that only 2% of patients received Bort for 19 cycles in the Taverna, 2012 study) the ICERs for the second-line treatment option are above **£50,000** per QALY gained.

When we consider the 8-cycle treatment regimen (to note is that more than 50% of patients in most comparator studies received Bort for less than 6 cycles) the final ICERs are generally above **£70,000** per QALY gained.

Table 39. Len/Dex vs bortezomib retreatment – 19 cycles

All treatment lines	Costs	QALYs	ICER all treatment lines
Len/Dex	£104,864	3.05	**£15,409**
Bort retreatment	£92,335	2.24	
Incremental	£12,528	0.81	
Second-line treatment only	**Costs**	**QALYs**	**2nd line ICER**
Len/Dex	£98,231	3.05	**£36,465**
Bort retreatment	£68,731	2.24	
Incremental	£29,500	0.81	

Table 40. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) -19 and 8 cycles

Second-line treatment only	Costs	QALYs	2nd line ICER
Len/Dex	£98,231	3.05	**£57,481**
Bort retreatment 19 cycles	£43,999	2.11	
Incremental	£54,232	0.94	
Bort retreatment 8 cycles	£30,601	2.11	**£71,682**
Incremental	£67,629	0.94	

Table 41. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS) - 19 and 8 cycles

Second-line treatment only	Costs	QALYs	2nd line ICER
Len/Dex	£96,117	2.90	**£65,133**
Bort retreatment 19 cycles	£44,293	2.10	
Incremental	£51,824	0.80	
Bort retreatment 8 cycles	£30,250	2.10	**£82,783**
Incremental	£65,867	0.80	

To note is that the effectiveness of Bort has not been changed in this exploratory analyses. As only costs were changed, we are potentially skewing the final ICERs upwards (as the incremental effectiveness of Len/Dex compared with Bort would theoretically increase).

Scenario analysis

Corrections applied to the CGP method

As mentioned in Section 2.1, the ERG have some concerns regarding the calculation of the weights used in the CGP method. We have corrected this accordingly. However, we could only apply this correction to the intervention arm of model, in that we could only change the survival curves used for OS, PFS and TTF for Len/Dex. We could not update the HR calculation, as this was not flexible in the excel model hence we could not update the survival curves in the comparator arm of the model.

Additionally, we cannot predict the impact on the final ICER resulting from not adjusting the curves in the comparator model. We also identified other problems with the CGP method which we could not address. For these reasons, this analysis is presented as scenario analysis, and was not incorporated in the base case values under the ERG corrected ICERs.

Table 42 and Table 43 present the ICERs for Len/Dex vs Bort with all ERG corrections (HRs, Bort PAS and the CGP weights corrections). Table 42 presents the ICER for all treatment lines while Table 43 shows the ICER for second-line treatment only. Both ICERs assume that Bort is given until progression.

To note is that when the treatment duration with Bort was reduced to 19 cycles, the ICER for second-line patients was above **£70,000** per QALY gained.

Table 42. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS and CGP)

Treatment	Costs	QALYs	ICER
LEN/DEX	£108,722	3.11	**£16,020**
Bort retreatment	£96,005	2.31	
Incremental	£12,716	0.79	

Table 43. Len/Dex vs bortezomib retreatment with CGP method and ERG corrections (HRs and Bort PAS and CGP) – second-line

Treatment	Costs	QALYs	ICER
LEN/DEX	£102,296	3.11	**£36,721**
Bort retreatment	£73,148	2.31	
Incremental	£29,148	0.79	

Alternative chemotherapy regimens

Celgene present the ICERs for Len/Dex compared with alternative chemotherapy regimens (CY-DEX). Even though this alternative analysis was not functioning in the excel model, once the ERG corrected this, the resulting ICERs were the same ones reported by Celgene in their revised

submission (Table 44). The ICERs presented are for all treatment lines. The ERG have not corrected these ICERs as the resulting values would only be higher than the ones reported in Table 44.

Table 44. Len/Dex vs CY-DEX

Parameter	Base case assumption	Scenario analysis	Incremental		ICER
			Costs	QALYs	
Comparator	Defined by user	LD cyclophosphamide + MD-Dex	£56,548	1.00	**£56,738**
		HD cyclophosphamide + LD-Dex	£58,197	1.00	**£58,393**

Alternative distributions and comparator sources

As previously mentioned, the ERG are not very clear why some distributions were considered the best fit while others were deemed inappropriate. Visual inspection of the curves and assessment of the AIC and BIC statistics do not reveal clear preferences within a certain range of distributions. Therefore we have presented the final ICERs when different distributions are used. Table 45 reports the ICERs with the ERG corrections (HRs and Bort PAS), include all treatment lines and assume that Bort is given until disease progression. The equivalent base case ICER is £21,782 (Table 30).

Again, it should be noted that the ICER estimates presented are based on what the ERG consider to be a flawed analysis. Therefore the alternative values here presented should also be interpreted with caution. The ICERs reported in Table 45 serve the main purpose of illustrating the variation in the base case ICER when different distributions are used.

It can be observed that when different distributions are used the ICER estimates range from **£12,031** to **£23,673**. Additionally, all these changes have been considered in isolation (i.e. changing one distribution at a time). If these were changed simultaneously, the impact on the final ICER would also be different. For example, if the PFS and TTF distributions are changed to a log-logistic, the final ICER increases to **£24,563**.

As before, we also present the ICERs relating to second-line treatment and assuming Bort is given for 19 cycles (Table 46). The base case equivalent ICER is £57,481. When different distributions are considered separately, these estimate can go up to **£69,000**.

The ERG also present the final ICER using the alternative sources suggested by Celgene and implemented in the model as scenario analysis. We present the estimates for all treatment lines and assuming Bort is given until disease progression. After the ERG corrections were applied to the model, the final ICER increase up to **£32,766** (even with unlimited Bort and all treatment lines considered) when the MTC was used to inform Bort effectiveness. However as we have previously explained, it is still not clear how Celgene derived the MTC estimates.

Table 45. Use of alternative distributions in the economic model – all treatment lines included and Bort given until progression

Parameter	Base case assumption	Scenario analysis	Incremental		ICER
			Costs	QALYs	
Curve fit for OS	Gompertz with MOC method	Piecewise exponential MOC	£21,443	1.31	**£16,294**
		Gamma MOC	£21,070	1.11	**£18,908**
		Piecewise exponential CGP	£14,445	1.20	**£12,031**
		Gamma CGP	£13,533	0.95	**£14,279**
Curve fit for PFS	Gamma with MOC method	Lognormal MOC	£22,206	0.95	**£23,423**
		Log-logistic MOC	£21,398	0.94	**£22,645**
		Lognormal CGP	£13,985	0.79	**£17,668**
		Log-logistic CGP	£13,980	0.79	**£17,608**
Curve fit for TTF	Gamma with MOC method	Lognormal MOC	£21,959	0.94	**£23,278**
		Log-logistic MOC	£22,328	0.94	**£23,673**
		Lognormal CGP	£15,196	0.80	**£19,097**
		Log-logistic CGP	£16,083	0.80	**£20,215**

Table 46. Use of alternative distributions in the economic model – second-line treatment and Bort given for 19 cycles

Parameter	Base case assumption	Scenario analysis	Incremental		ICER
			Costs	QALYs	
Curve fit for OS	Gompertz with MOC method	Piecewise exponential MOC	£56,680	1.32	**£43,071**
		Gamma MOC	£55,576	1.11	**£49,873**
		Piecewise exponential CGP	£54,249	1.20	**£45,183**
		Gamma CGP	£53,224	0.95	**£56,158**
Curve fit for PFS	Gamma with MOC method	Lognormal MOC	£54,928	0.95	**£57,939**
		Log-logistic MOC	£54,585	0.94	**£57,767**
		Lognormal CGP	£51,833	0.79	**£65,482**
		Log-logistic CGP	£51,828	0.79	**£65,278**
Curve fit for TTF	Gamma with MOC method	Lognormal MOC	£56,487	0.94	**£59,878**
		Log-logistic MOC	£56,316	0.94	**£59,709**
		Lognormal CGP	£54,909	0.80	**£69,018**
		Log-logistic CGP	£54,171	0.80	**£68,079**

Table 47. Use of alternative sources for Bort effectiveness– all treatment lines included and Bort given until progression

Parameter	Base case assumption	Scenario analysis	Incremental		ICER
			Costs	QALYs	
Source of Bort retreatment evidence	OS & PFS: Taverna 2012	OS & PFS: Celgene MTC	£19,753	0.60	**£32,766**
		OS & PFS: Hrusovsky 2010	£24,948	1.09	**£22,959**

Bort PAS

As mentioned before, one of the key drivers of the final ICER is the total cost of Bort in the analysis (whether through the length of treatment administration or the cost of the drug).

When the Bort PAS is removed from the economic analysis, we can observe that for all treatment lines considered, Len/Dex dominates Bort. When only second-line treatment is considered, the ICER using the MOC method is £5,428 and dominant with the CGP method.

When 19 cycles of Bort are considered, the MOC method produces an ICER of £34,883 and when 8 cycles of Bort are considered the equivalent ICER is £57,322.

It is noticeable how the cost of Bort (changed through the PAS) and the length of the treatment have a major impact on the final ICER. Therefore these two issues should be carefully considered by the Committee.

Table 48. Len/Dex vs bortezomib retreatment with ERG corrections (HRs) with no Bort PAS

Treatment	Costs	QALYs	ICER
Len/Dex MOC	£104,864	3.05	**Lenalidomide dominates**
Bort MOC	£115,837	2.11	
Incremental MOC	-£10,973	0.94	
Len/Dex CGP	£108,722	3.11	**Lenalidomide dominates**
Bort CGP	£133,935	2.31	
Incremental CGP	-£25,213	0.79	

Table 49. Len/Dex vs bortezomib retreatment with ERG corrections (HRs) with no Bort PAS – second-line treatment

Treatment	Costs	QALYs	ICER
Len/Dex MOC	£98,231	3.05	**£5,428**
Bort MOC	£93,109	2.11	
Incremental MOC	£5,121	0.94	
Len/Dex CGP	£102,296	3.11	**Lenalidomide dominates**
Bort CGP	£111,078	2.31	
Incremental CGP	-£8,781	0.79	

Alternative utilities

As previously mentioned, the ERG are not very confident that the scenario analysis undertaken by Celgene addresses the Committee issues in full. However we present the final ICERs when different sources are used. Table 50 reports the ICERs with the ERG corrections (HRs and Bort PAS), include all treatment lines and assume that Bort is given until disease progression. The equivalent base case ICER is £21,782 (Table 30).

Table 50. Len/Dex vs bortezomib retreatment with ERG corrections (HRs and Bort PAS) using different utilities sources

Treatment	Costs	QALYs	ICER
Len/Dex McKenzie & van der Pol	£104,864	2.89	**£22,005** McKenzie & van der Pol **£21,417** Kontodimopoulos **£24,153** Palumbo
Len/Dex Kontodimopoulos		2.95	
Len/Dex Palumbo		2.96	
Bort retreatment McKenzie & van der Pol	£84,314	1.95	
Bort retreatment Kontodimopoulos		1.99	
Bort retreatment Palumbo		2.11	
Incremental McKenzie & van der Pol	£20,550	0.93	
Incremental Kontodimopoulos		0.96	
Incremental Palumbo		0.85	

HRs used in the analysis

Since the effectiveness of Len/Dex compared with Bort raised several concerns during the previous Committee meeting the Committee suggested that Celgene presented a model that assumed Len/Dex and Bort to have equal effectiveness (that is, assuming a HR of 1 throughout the analysis). However, Celgene did not present such analysis.

The ERG have replaced the HRs used in the economic model by HR = 1, to understand the impact of assuming the same effectiveness for Len/Dex and Bort. As expected, the incremental QALY gain was close to 0, and so the final ICER depends entirely on the difference between the total costs of the two treatment regimens. The total costs, in their turn, will depend on the duration of Bort treatment. So the final ICERs obtained with HR=1 vary:

- When Bort is given until disease progression: Len/Dex dominates as the total cost of Len/Dex is £104,864, while the total cost of Bort is £118,146.

- When Bort is given for 19 or less cycles: the final ICER is >£1,000,000.

Unadjusted analysis

One of the ERG main concerns with Celgene's analysis is the series of adjustment processes applied to the survival models. Therefore it would have been extremely helpful if Celgene had run an unadjusted analysis, removing all potential bias from the different models. More specifically, Celgene could have run a model where intervention and comparator curves were unadjusted as well as with unadjusted HRs.

To note is that Celgene claim to use unadjusted HRs as scenario analysis. However, as the ERG have pointed in their report, these HRs are not truly unadjusted as they are still adjusted for the MM studies population characteristics.

Len vs placebo

The Committee also recommended that Celgene run a model that compared lenalidomide with placebo because it would be based on high-quality data from MM-009 and MM-010. Unfortunately, Celgene did not run such analysis.

Sequence of treatments

The Committee recommended including sequences of Len/Dex followed by Bort compared with Bort followed by Len/Dex (without asymmetry between the Len/Dex and comparator arms).

After the ERG corrections (HRs and Bort PAS) were applied to the model, comparing Len/Dex followed by Bort compared with Bort followed by Len/Dex produced an ICER of £35,415.

To note is that running this scenario was one of the options provided in the excel model. However, due to time constraints, the ERG could not verify how this scenario was run in the model.

Sensitivity analysis

Celgene ran deterministic and probabilistic sensitivity analysis. The respective outputs can be found in Section 5 of their revised submission.

Due to time constraints, the ERG did not have the time to analyse how sensitivity analysis was undertaken in the excel model. However, and similarly to the previous submission, the revised sensitivity analysis consistently reports dominant ICERs, which is somewhat questionable.

7.0 Conclusion

Having considering the previous submission made by Celgene, the Committee concluded that the ICER resulting from the analysis (likely to be >£30,000) was not reliable based on the uncertainties around model parameters and methods used in the economic analysis.

In this section of the report we summarise the uncertainties associated with the previous economic analysis and present the implications of the revised analysis.

The main uncertainties in Celgene' analysis were related to the following matters:

- The estimated **clinical effectiveness of lenalidomide compared with bortezomib.** Even though the evidence seems clear in establishing that lenalidomide is more effective than placebo and standard chemotherapy, it was unclear whether lenalidomide was clinically more effective than retreatment with bortezomib.

- The Committee pointed that there were two relevant **subgroups of patients**. Those for whom bortezomib retreatment would be appropriate, in which case the treatment options considered should include lenalidomide, bortezomib and standard chemotherapy and those for whom retreating with bortezomib would not be appropriate, in which case the treatment options include lenalidomide and standard chemotherapy.

- The **model structure**. Third and fourth-treatment lines were modelled inconsistently. Also, in the bortezomib arm of the model, the approach taken to modelling second and third-line treatment outcomes reflected an unlikely scenario from a clinical point of view.

- Several issues related with the **data extrapolation process** were identified. More specifically, the estimated progression-free survival hazard ratio of 0.9 favouring bortezomib compared to lenalidomide was found to be clinical implausible. This also raised concerns related to the method used to adjust the survival curves to reflect baseline population characteristics and with the process undertaken to fit overall survival data.

- Finally, the Committee noted several issues with **data inputs** related with the costs assumptions made in the economic model, such as patient access schemes being excluded from the base-case analysis. It also noted that utility values used in the analysis were based on some uncertainty.

Implications of the revised analysis

Having considered Celgene's revised submission the ERG view is that:

- The estimated **clinical effectiveness of lenalidomide compared with bortezomib** is still a main source of concern**.** The undertaken process to estimate the HRs is methodologically weak and is potentially biased.

- Analysis undertaken for the **subgroups of patients** for whom treatment options include lenalidomide and standard chemotherapy present inexplicably high ICERs.

- There are outstanding issues related with the **model structure**. After second-line of treatment, the manufacturer only consider the utility associated with the disease progression

state also the ERG question the value of including third and fourth-treatment lines, especially in the intervention arm of the model, as only cost data is available and the basket of drugs considered might not accurately reflect current clinical practice.

- There are still several issues related with the **data extrapolation process**. The adjustment processes undertaken to estimate survival models are fundamentally flawed and are potentially introducing a high level of noise to the analysis. OS seems to be modelled in a more plausible fashion however PFS and TTF survival curves seem to flatten off too soon in the model. OS curves cross and PFS and TTF curves after 10 years in the intervention arm of the model.

- Finally, the ERG are still concerned with the assumption of keeping patients on Bort until disease progression in the economic model. Literature and clinical opinion consistently report that Bort is given for a fixed period of time to MM patients. Furthermore, the duration of treatment is one of the key drivers of the final ICER. With regards to the utility values used in the analysis, even though Celgene tried to address some of the issues raised by the Committee, the values are still based on some uncertainty.

Overall there is a high level of uncertainty embedded in several steps of the analysis and it is extremely difficult to assess to which extent the issues raised impact the final ICER.

For all the reasons provided, the ERG lack confidence in the final ICERs presented, therefore all estimates presented in the previous section should be interpreted with extreme caution.

When the ERG changed the duration of treatment with Bort, the final second-line ICER increased to £57,481. Whilst we do not suggest that this is a reliable alternative ICER it shows the sensitivity of the model outcomes to these parameters. The key determinants of the final ICERs seem to be the **duration of Bort treatment** and the **inclusion (or exclusion) of subsequent treatment lines.**

8.0 References

Bacelar M, Durand A, Cooper C, Hyde C, Latimer N, Murray, D. 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.

Celesti L, Clavio M, Poggi A, et al. The association of cyclophosphamide and dexamethasone in advanced refractory multiple myeloma patients. Haematologica. 1997; 82(3):351-3.

Celgene Ltd. (2007), Protocol CC-5013-MM-009. A multicenter, randomised, parallel group, double-blind, placebo-controlled study of CC-5013 plus dexamethasone versus dexamethasone alone in previously-treated subjects with multiple myeloma. (Clinical Study Report).

Celgene Ltd. (2007). Protocol CC-5013-MM-010. A multicenter, randomised, parallel group, double-blind, placebo-controlled study of CC-5013 plus dexamethasone versus dexamethasone alone in previously-treated subjects with multiple myeloma, (Clinical Study Report).

Damaj, G. Malard, F. Hulin, et al. (2012), Efficacy of bendamustine in relapsed/refractory myeloma patients: results from the French compassionate use program, Leuk Lymphoma, 53(4):632-4.

Dispenzieri A, Jacobus S, Vesole DH, et al. Primary therapy with single agent bortezomib as induction, maintenance and re-induction in patients with high-risk myeloma: results of the ECOG E2A02 trial. Leukemia. 2010; 24(8):1406-11.

Hrusovsky, I. Emmerich, B. von Rohr, A. et al. (2010), Bortezomib retreatment in relapsed multiple myeloma - results from a retrospective multicentre survey in Germany and Switzerland. Oncology. 79(3-4):247-54.

Janssen-Cilag Ltd. Velcade®. Summary of product characteristics. Available at: http://www.medicines.org.uk/emc/medicine/17109/SPC/ Accessed: 30 October 2013.

National Institute for Health and Care Excellence, (2013), Single Technology Appraisal (STA) of lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy with bortezomib (partial review of TA171): Celgene submission of evidence.

National Institute for Health and Care Excellence. Appraisal consultation document (ACD) – Lenalidomide for treating multiple myeloma after 1 prior treatment with bortezomib (part-review of TA171). March 2014.

National Institute for Health and Clinical Excellence (NICE). TA171: Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior treatment with bortezomib (part-review of TA171). Celgene's response. April 2014

Petrucci MT, Avvisati G, Tribalto M, et al. Intermediate-dose (25 mg/m2) intravenous melphalan for patients with multiple myeloma in relapse or refractory to standard treatment. European journal of haematology. 1989; 42(3):233-7.

Petrucci MT, Giraldo P, Corradini P, et al. A prospective, international phase 2 study of bortezomib retreatment in patients with relapsed multiple myeloma. Br J Haematol. 2013a; 160(5):649-59.

Petrucci MT, Levi A, Bringhen S, Scotti S, Gentilini F, Russo S, Siniscalchi A, Larocca A, Grammatico S, Boccadoro M, Foa` R, Palumbo A. Bortezomib, Melphalan, and Prednisone in Elderly

Patients With Relapsed/Refractory Multiple Myeloma. A Multicenter, Open Label Phase 1/2 Study. Cancer 2013a; 10.1002/cncr.27820.

Sood R, Carloss H, Kerr R, et al. Retreatment with bortezomib alone or in combination for patients with multiple myeloma following an initial response to bortezomib. American journal of hematology. 2009; 84(10):657-60.

Stadtmauer, EA. Weber, DM. Niesvizky, R. et al. (2009), Lenalidomide in combination with dexamethasone at first relapse in comparison with its use as later salvage therapy in relapsed or refractory multiple myeloma. European Journal of Haematology, 82(6):426-32.

Taverna C, Voegeli J, Trojan A, et al. Effective response with bortezomib retreatment in relapsed multiple myeloma--a multicentre retrospective survey in Switzerland. Swiss Med Wkly. 2012; 142:w13562.

Van Agthoven, M. Segeren, CM. Buijt, I. et al. (2004), A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study, European Journal of Cancer, 40(8):1159-69.