The clinical and cost-effectiveness of lenalidomide for people who have received at least one prior therapy with bortezomib (partial review of TA171)

A critique of the submission from Celgene

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 Knowledge
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 Cooper, Senior Information Specialist, PenTAG
Chris Hyde, Professor of Public Health and Clinical Epidemiology, PenTAG
Nick Latimer, Research Fellow in Health Economics, University of Sheffield
David Murray, Senior Associate, Public Health & Economics, Matrix Knowledge

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, 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.

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 Cooper: Commented on the searches provided by the manufacturer and contributed to report writing.

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.

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

Correspondence to:

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

Date completed:

January 2014

Contents

1.0 Summary 9
1.1 Scope of the submission 9
1.2 Summary of submitted clinical effectiveness evidence 9
1.3 Summary of submitted cost effectiveness evidence 10
1.4 Commentary on the robustness of submitted evidence 10
1.5 Key issues 10
2.0 Background 12
2.1 Critique of manufacturer's description of underlying health problem 12
2.2 Critique of manufacturer's overview of current pathway of care and service provision 12
3.0 Critique of manufacturer's definition of decision problem 13
3.1 Population 13
3.2 Intervention 13
3.3 Comparators 14
3.4 Outcomes 15
3.5 Time frame 15
4.0 Clinical effectiveness 16
4.1 Critique of manufacturer's approach 16
4.2 Summary of submitted evidence 32
5.0 Economic evaluation 49
5.1 Overview of manufacturer's economic evaluation 50
5.2 Critique of approach used 75
5.3 Results included in manufacturer's submission 111
5.4 Comment on validity of results presented with reference to methodology used 118
6.0 Additional work undertaken by the ERG 118
6.1 Correction for errors in Celgene's model 119
7.0 Summary of clinical and cost-effectiveness issues 124
8.0 References 126

List of tables

Table 1. Updated incidence of MM in England and Wales 12
Table 2. Eligibility criteria used of study selection 17
Table 3. List of relevant primary publications 19
Table 4. Clinical appraisal of relevant RCTs 22
Table 5. Myeloma response determination criteria 29
Table 6. Time to Progression for ITT population 33
Table 7. Response rates, TTP, PFS and OS 34
Table 8. Summary of PFS (ITT population) – MM-009 and MM-010 36
Table 9. PFS of second-line patients in studies MM-009 and MM-010 38
Table 10. Overall survival 40
Table 11. OS of second-line patients in studies MM-009 and MM-010 42
Table 12. Response rates at unblinding 42
Table 13. Outcomes in patients by number of prior therapies 44
Table 14. Time to first worsening of ECOG performance status 45
Table 15. Pooled duration of treatment in studies MM-009 and MM-010 45
Table 16. Grade ≥3 adverse events occurring in more than 5% of patients 47
Table 17. Most important errors in versions of the economic model sent to ERG by Celgene. . 49
Table 18. Regression results for PFS with log-logistic distribution 57
Table 19. Regression results for TTF with log-logistic distribution 57
Table 20. Regression results for OS with log-logistic distribution 59
Table 21. Sources used to model HRs 60
Table 22. Health states utility values 62
Table 23. AE rates applied in the economic analysis 63
Table 24. AEs utility decrements 64
Table 25. Costs of AEs included in the economic model 65
Table 26. Proportions of AEs included in the economic model 67
Table 27. Health state costs 68
Table 28. Proportion of patients receiving Len and G-CSF per model cycle (based on MM-010). 69
Table 29. Len and unit cost 70
Table 30. Unit cost of Dex 70
Table 31. Other comparators acquisition costs 71
Table 32. Treatment administration costs 72
Table 33. Third-line therapy mix in the economic model 73
Table 34. Third-line treatment cost per cycle 74
Table 35. Fourth-line treatment cost per cycle 74
Table 36. Critical appraisal checklist based on NICE Reference Case (NICE, 2008) 76
Table 37. Critical appraisal checklist from Drummond and colleagues (Drummond et al. 1997) 78
Table 38. Critical appraisal checklist of Philips et al (2004) for model-based analysis 79
Table 39. Third-line therapy mix in the economic model 85
Table 40. Baseline characteristics in Stadtmauer analysis 87
Table 41. Weighted health state index by age and sex 110
Table 42. Base case outputs per patient at 25 years in the original analysis 112
Table 43. Base case outputs per patient at 25 years in the updated analysis 112
Table 44. Base case model outputs (Len/Dex) compared with trial data 112
Table 45. Scenario analysis outcomes – updated model. 115
Table 46. Scenario analysis using clinical inputs for the second-line population 116
Table 47. Scenario analysis using clinical inputs from MM-009 117
Table 48. Second-line base case results with corrections from Celgene's model 123

List of figures

Figure 1. KM curve for TTP at study unblinding for MM-010. 33
Figure 2. Kaplan–Meier estimate of TTP in the ITT population 34
Figure 3. Kaplan–Meier estimate of TTP of patients treated with Len/Dex by number of prior therapies 35
Figure 4. Kaplan–Meier estimate of PFS in the ITT population 37
Figure 5. Kaplan–Meier estimate of PFS of patients treated with Len/Dex by number of prior therapies 37
Figure 6. KM curves for OS for all patients in MM-009 - May 2006, ITT population 39
Figure 7. KM curves for OS for all patients in MM-010-May 2006, ITT population 39
Figure 8. Kaplan–Meier curves of OS for all patients 40
Figure 9. KM curve of OS of patients treated with Len/Dex by number of prior therapies 41
Figure 10. Celgene's model structure 51
Figure 11.Treatment pathway conisdered in the economic model 52
Figure 12. Modelling process for OS, PFS and TTF 53
Figure 13. KM plot and fitted log-logistic model for PFS 56
Figure 14. KM plot and fitted log-logistic model for TTF 57
Figure 15. KM plot and fitted exponential piecewise model for OS 59
Figure 16. Simplified structure of Celgene's model structure – intervention and comparator arms 82
Figure 17. Alternative representation of Celgene's mode 83
Figure 18. Disease progression in MM 84
Figure 19. KM plot and fitted log-logistic curve for PFS over 25 years produced by the ERG ... 93
Figure 20. KM plot and fitted log-logistic curve for TTF over 25 years produced by the ERG ... 95
Figure 21. Markov traces from original submission 99
Figure 22. PFS and OS curves in the Len/Dex arm of the original model 101
Figure 23. PFS and OS curves in the Bort arm of the original model 101
Figure 24. TTF and OS curves in the Len/Dex arm of the original model 102
Figure 25. TTF and OS curves in the Bort arm of the original model 102
Figure 26. KM plot and fitted log-logistic curve for OS over 25 years –Len/Dex 104
Figure 27. KM plot and fitted piecewise exponential curve for OS over 25 years – Len/Dex ... 106
Figure 28. PFS and OS curves in the Bort arm of the model 107
Figure 29. TTF and OS curves in the Bort arm of the model 107
Figure 30. Tornado diagram with top 10 parameters in terms of ICER sensitivity – original analysis 113
Figure 31. Tornado diagram with top 10 parameters in terms of NMB sensitivity – updated analysis 114
Figure 32. Cost-effectiveness scatter plot – original submission 117
Figure 33. Cost-effectiveness scatter plot – updated submission 118
Figure 34. Markov traces in the updated model with ERG corrections 122

LIST OF ABBREVIATIONS

AE Adverse event
AFT Accelerated failure time
ASH American Society of Hematology
CDF Cancer Drugs Fund
CI Confidence interval
CR Complete response
CSR Case study report
Dex Dexamethasone
EBMT European Society for Blood and Marrow Transplantation
ECOG Eastern Cooperative Oncology Group
EMEA European Medicines Agency
EORTC European Organisation for Research and Treatment of Cancer
FDA Food and Drug Administration
G-CSF Granulocyte-colony stimulating factor
HR Hazard ratio
HRQoL Health related quality of life
ICER Incremental cost effectiveness ratio
IDMC Independent Data Monitoring Committee
ITT Intention to treat
Len Lenalidomide
Len/Dex Lenalidomide plus dexamethasone
MM Multiple myeloma
nCR Near complete response
NE Not estimable
ORR Overall response rate
OS Overall survival
PAS Patient access scheme
PD Progressive disease
PFS Progression-free survival
PR Partial response
QALY Quality adjusted life year
QLQ Quality of Life Questionnaire
SCT Stem cell transplant
SD Stable disease
SE Standard error
SPC Summary of product characteristics
TTF Time to treatment failure
TTP Time to progression

1.0 Summary

Text cited directly from the submission by Celgene (hereafter referred to as 'the submission') is presented in italic and cross referenced.

The ERG found several important logical errors in the economic model first sent to us by Celgene. 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 (Table 17). In result of this, Celgene submitted an updated analysis, with the goal to address the problems identified in the original submission. This report mainly discusses the final model version.

Given the nature of the STA process, the ERG was bound to time constraints. Most of the initial review process was dedicated to finding the methodological and logical errors in the submission and providing the manufacturer with some time and suggestions to address these. Therefore, and as discussed with NICE, in order to stay within the agreed timeframe we have focused in depth on certain aspects of the submission and only provide some insight on others.

The primary focus of this critique is on the second-line treatment for multiple myeloma (MM). Second-line treatment in the economic analysis compares lenalidomide taken concomitantly with dexamethasone with bortezomib. Considerations are also made for subsequent treatment options.

1.1 Scope of the submission

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

1.2 Summary of submitted clinical effectiveness evidence

The evidence is based on two identically designed RCTs (MM-009 and MM-010) in people with MM who had received at least one prior therapy. Both trials evaluate the efficacy of lenalidomide taken concomitantly with dexamethasone with dexamethasone alone. Individual trial outcomes and results from a pooled analysis showed an increase in time to disease progression (TTP), progression-free survival (PFS) and overall survival (OS) for patients receiving lenalidomide/dexamethasone compared to patients receiving dexamethasone alone.

Evidence for the comparators is based on retrospective studies. While the Taverna (2012) study was used to assess the effectiveness of bortezomib, Damaj (2012) was used to drive the efficacy of all other comparators.

1.3 Summary of submitted cost effectiveness evidence

Celgene developed a cost-utility model as a partitioned survival structure. The model describes four health states: Pre-progression on treatment (PFS-T), pre-progression off treatment (PFS-OT), progressive disease (PD) and death.

Celgene's updated model produced a dominant ICER, favouring lenalidomide/dexamethasone.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

- It is likely that the cost-effectiveness systematic review of the literature undertaken by Celgene contains all relevant studies.
- The MM-010 RCT on which the cost-effectiveness analysis is based is of high quality.

1.4.2 Weaknesses

The ERG's main concerns are regarding the **model structure** and with the **data extrapolation process** employed by Celgene.

The ERG lack confidence in the final ICER presented. Celgene's revised economic model reports base case dominant ICERs, which significantly depart from the ICERs presented in TA171. Furthermore the undertaken sensitivity analysis consistently report dominant ICERs, which is somewhat questionable.

It is the ERG conclusion that the approach taken to modelling the cost-effectiveness of lenalidomide/dexamethasone compared with bortezomib for MM patients presented in this submission needs to be fundamentally reconsidered.

1.5 Key issues

The ERG is overall concerned with the **model structure** used by Celgene. The approach undertaken raises the following concerns:

- There is not a clear separation between second-line treatment outcomes and the beginning of a third-line treatment option and respective outcomes in the bortezomib arm of the model.
- After second-line of treatment, the manufacturer only consider the utility associated with the disease progression state.
- 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.

More importantly, the ERG is generally concerned with the **data extrapolation process** employed by Celgene. The approach taken raises the following issues:

- The use of the progression-free survival hazard ratio of 0.9, which is likely underestimating the effectiveness of lenalidomide/dexamethasone compared with bortezomib.
- Use of a log-logistic distribution to fit overall survival data, which appears to be a very poor fit to MM-010 trial data.
- Likely overestimation of lenalidomide overall survival.
- The use of the mean of covariates method to adjust the progression-free survival and overall survival curves to reflect MM-010 population characteristics, which might potentially be skewing these survival estimates.

2.0 Background

2.1 Critique of manufacturer's description of underlying health problem

In Sections 2.1 to 2.3 of their submission, Celgene describe the underlying health problem. They provide a summary of the characteristics and progression of MM. Incidence of MM in England and Wales is presented and is based on reliable sources, however the ERG noticed that incidence figures used in the submission have been updated by Cancer Research UK in October.

Table 1. Updated incidence of MM in England and Wales

Description	England estimate	Wales estimate	Source
Myeloma incidence (per 100,000) – Males	7.1	7.1	Cancer Research UK (2013)
Myeloma incidence (per 100,000) – Females	4.4	4.1	

Information on life expectancy of patients with MM by cancer stage category was also provided.

2.2 Critique of manufacturer's overview of current pathway of care and service provision

In Sections 2.5 to 2.7 of their submission, Celgene outline how the management of MM is generally determined on an individual basis, depending on several factors like age, prior therapies and bone marrow function, amongst others.

Celgene make reference to TA129 where NICE recommend bortezomib as a second-line treatment option for patients who have undergone, or are unsuitable for bone marrow transplantation. TA228 is also mentioned as it subsequently recommended bortezomib (in combination with an alkylating agent and a corticosteroid) as an option for the first-line treatment of MM in the cases where high dose chemotherapy with stem cell transplantation is considered inappropriate and the patient is contraindicated for thalidomide.

Additionally, TA171 recommends the use on lenalidomide for patients who had at least two prior therapies.

Celgene suggest that the anticipated place for lenalidomide in the treatment pathway is likely to be in those patients who have received one prior treatment. The rationale provided is that patients with MM will only be treated with bortezomib as first line therapy if they are contraindicated for thalidomide and so, in the case where retreatment with bortezomib is not appropriate, lenalidomide will be administered as a second line treatment.

Though Celgene justify their selection of comparator treatments it is unclear if this reflects

clinical practice, particularly given its rapidly evolving nature in relation to new combination regimens. This is further discussed below in Section 3.3.

3.0 Critique of manufacturer's definition of decision problem

3.1 Population

The population considered by the submission is adults with MM for whom thalidomide is contradicted and whose disease has progresses after at least one prior treatment with bortezomib.

Celgene claim that the population defined in the submission is in line with the one defined in the NICE Scope[1].

Even though the population considered by the submission is defined as patients who have received at least one prior treatment with bortezomib, in the trial used to inform the economic analysis only 4% of patients have been previously treated with bortezomib. However, clinical advice sought by the ERG revealed that this is unlikely to affect the effectiveness of lenalidomide. This means that, had the majority of patients in the trial received bortezomib as a first-line therapy, the effectiveness of lenalidomide as a second line drug is expected to be similar.

As per expert opinion, the drug used in previous MM therapies is not as crucial to determine the effectiveness of posterior treatments as the duration of response to the previous treatment.

Whilst it is not specified in the submission why patients are contraindicated for thalidomide, our clinical expert also pointed to the fact that a patient contraindicated for thalidomide early in MM treatment (for example, due to poor renal function) might still be treated with thalidomide later on and after relapse.

3.2 Intervention

The intervention under assessment is lenalidomide (Revlimid®) in combination with dexamethasone[2]. Lenalidomide[3] is an immunomodulating agent, belonging to a class of immunomodulatory derivatives. Len is a structural derivative of thalidomide.

Len was launched in the UK in June 2007. In June 2009 NICE recommended the use of Len/Dex for patients who have received two or more prior therapies however, since then, Len was approved for funding through the Cancer Drugs Fund (CDF) as a second line treatment.

[1] Referred to as "the scope" in the remainder of this report.
[2] Referred to as "Len/Dex" in the remainder of this report.
[3] Referred to as "Len" in the remainder of this report.

The EMEA recommended starting dose of Len is 25 mg orally once daily on days 1-21 of repeated 28-day cycles, while the recommended dose of dexamethasone[4] is 40 mg orally once daily on days 1-4, 9-12, and 17-20 of each 28-day cycle for the first 4 cycles of therapy and then 40 mg once daily on days 1-4 every 28 days. This is in line with the treatment regimen followed in MM-010.

Treatment was continued until disease progression occurred (defined in Section 4.2.6) or unacceptable toxicity emerged.

3.3 Comparators

The comparator used was bortezomib[5]. It is not clear in the submission if the manufacturer considered the base case comparator to be Bort or Bort with concomitant dexamethasone[6]. However, the fact that the cost-effectiveness of Bort/Dex has been included in the scenario analysis suggests that the base case comparator was originally considered to be Bort taken alone.

It should be noted that in the retrospective study used to inform the economic analysis, Taverna (2012), 64.3% of patients received concomitant Dex. Furthermore, expert opinion suggests that the use of Bort with concomitant Dex represents current practice in the UK. This is explored in detail in Section 5.1.2.

The Taverna (2012) study did not include a detailed description of the drug regimen administered. It is known that the median number of prior therapies is 2 (range 1-11) and that 31% of the study population had undergone SCT.

Other comparators were specified in the initial scope. However, these were not included in the base case analysis but were instead included in the scenario analysis. This was the case for bendamustine and chemotherapy agents (including regimens based on melphalan, vincristine, cyclophosphamide and doxorubicin).

It is not clear to the ERG why bendamustine was not included in the base case analysis. Since this was one of the comparators included in the scope and data were available to model the cost-effectiveness of Len/Dex compared with bendamustine, it seems to be appropriate to include this comparator in the base case analysis.

Similarly, the cost-effectiveness of alternative chemotherapy agents was evaluated in the scenario analysis. It is stated that all evidence identified for the treatment of MM with chemotherapy agents included combination regimens with either Bort or Len, thus it was considered to be unsuitable for modelling purposes. As a result, the hazard ratios used to model the cost-effectiveness of bendamustine were applied in the chemotherapy agents' scenario analysis. Clinical opinion revealed that this is a reasonable assumption.

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

Finally, as mentioned in Section 3.1, clinical advice sought by the ERG pointed to the fact that thalidomide could potentially be considered for second or third-line treatment even if deemed contraindicated as a first-line treatment option (due to reasons like poor renal failure). Therefore, thalidomide could potentially be a relevant comparator for second or third-line treatment options.

3.4 Outcomes

The outcomes considered in the submission include:

- Progression-free survival (PFS)
- Overall survival (OS)
- Response rates
- Adverse effects of treatment
- Health related quality of life (HRQoL)

This departs from the outcome measures considered in the scope, which included time to next treatment as an outcome. It is stated that time to next treatment was not included in the analysis since it was not reported in trial MM-010.

3.5 Time frame

The time horizon for the economic analysis was 25 years. The proportion of patients alive at this point was about 10%.

4.0 Clinical effectiveness

4.1 Critique of manufacturer's approach

In this chapter we assess the clinical evidence provided by Celgene in their submission.

We start with a description and critique of Celgene's literature search strategy, followed by a description of the main studies selected for clinical effectiveness and their quality assessment. We then look at the manufacturer's selection of outcomes and the statistical approach they used. This is followed by a summary of their submitted evidence for clinical effectiveness and our commentary on their validity.

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate

- Clinical Effectiveness Searches

Celgene ran two literature searches to locate clinical effectiveness studies. In the initial submission, their search attempted to limit the population to second-line treatment (using terms such as relapse and recurrence to indicate failure at first-line). This was deemed inadequate by the ERG.

The ERG raised their concerns at the clarification stage and Celgene ran a second search which effectively removed this limitation.

The ERG accept these second searches as the primary searches in this submission though Celgene have not indicated if their second searches located any additional studies for consideration, which the ERG consider to be crucial information.

The second effectiveness search syntax took the following form:
(Terms for Population) AND (Intervention terms + terms for comparators) AND (Methods terms (RCTs/ SRs/ Case-Control Studies)) AND (Limit to Human only populations)

The searches were run from database inception in the following bibliographic databases:

- MEDLINE & MEDLINE In-Process (OVID);
- EMBASE (OVID);
- The Cochrane Library (CDSR, CENTRAL, HTA & DARE); and,
- CINAHL (EBSCO HOST).

The following conference proceedings were searched:

- American Society of Hematology (ASH);
- American Society of Clinical Oncology (ASCO); and,

- European Hematology Association (EHA).

Celgene's searches for non-RCT evidence are limited by study design or by specific keywords (such as follow-up.) The ERG do not consider these to be adequately sensitive searches for this type of evidence.

Finally, in clarification, Celgene confirmed that, in their opinion, no additional evidence is likely to become available for this indication and position in the treatment pathway relating to this appraisal in the next 12 months.

- Adverse events

Celgene did not run separate adverse event (AE) literature searches in their submission. Given the noted AE profile related with this intervention, the ERG were surprised by this decision and raised this in clarification.

Celgene replied that MM-009 and MM-010 studies had been designed to capture AE data and as such, the AE profile presented in the submission is adequate since it is comparable to that listed in the current SPC updated in 2013 and confirmed by clinical expert opinion sought by Celgene.

Given the noted AE profile, the ERG would still have preferred that separate searches were conducted to look beyond the two studies which have driven this submission.

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate

In their review of clinical effectiveness of Len, Celgene applied the inclusion/exclusion criteria listed in Table 2 below:

Table 2. Eligibility criteria used of study selection

	Inclusion criteria	***Exclusion criteria***
Population	• *Adult patients with relapsed or refractory multiple myeloma (rrMM) with ≥ 1 prior treatment with bortezomib*	• *Newly diagnosed multiple myeloma or treatment naïve patients* • *Studies that investigated both newly diagnosed and rrMM, but did not segregated the results* • *Studies on children and other blood cancer* • *Studies in which patients had no prior bortezomib*
Intervention	• *Lenalidomide/ dexamethasone*	• *Lenalidomide monotherapy* • *Lenalidomide/any other interventions*
Comparator	• *Bortezomib (Bor)*	• *Any other type of mono-chemotherapy and/*

	Inclusion criteria	Exclusion criteria
	monotherapy • *Bortezomib/high-dose dexamethasone (Bor/Dex)* • *Regimens based on mephalan, vincristine, cyclophosphamide and doxorubicin* • *Bendamustine*	*or combination chemotherapy* • *Stem cell transplantation*
Outcomes	• *Progression-free survival (PFS)* • *Overall survival (OS)* • *Overall response rate (ORR)* • *Best response, including complete response (CR), partial response (PR), and very good partial response (VGPR) Minimal response (MR) if part of an ORR summation.* • *Time to next treatment (TNT)* • *Time to progression (TTP)* • *Adverse events (only grade 3 and 4, or serious AEs)* • *Health related quality of life (HRQOL)*	• *Studies that did not report data on at least one of the outcomes of interested*
Study type	• *Randomised and non-randomised controlled trial of ≥ 5 patients* • *SR/MA of RCTs and non-RCTs*	• *Letter, secondary analysis with no new/relevant data, expert opinions, commentaries, non-systematic reviews*
Language of publication	• *English language*	• *Non -English language*
Date	• *None for full text publication* • *2011-2013: Conference proceedings and SR/MA*	

Source: Submission Table 81 Appendix 2

The inclusion criteria reflect the final scope issued by NICE and the licensed indication; that is to include studies of patients with MM for whom thalidomide is contraindicated and whose disease has progressed after at least 1 prior treatment with Bort.

Overall these criteria seem appropriate to identify all relevant evidence on the clinical effectiveness of Len. Despite this, the ERG requested clarification on some aspects of the search.

The submission includes a flow diagram that shows the number of studies identified through the database searches and the number of studies included and excluded at each stage of the review and the reasons for exclusion.

In addition, as none of the identified studies showed evidence of direct comparison between Len and the comparators, the search strategy was widened to the individual comparators. Eligibility criteria were the same as for Len/Dex with the exception of single case studies.

The submission therefore relies on indirect comparison between Len and Bort.

4.1.3 Studies included and excluded

- Len/Dex

The search strategy identified 44 studies in 53 publications. Four of these studies, presented in Table 3, were randomised clinical trials (RCTs). Six subgroup and two update analyses for two of these RCTs were identified. Three of the RCTs examined the efficacy of Len/Dex while the fourth RCT examined the efficacy of Bort treatment (AMBER).

Table 3. List of relevant primary publications

Author, year (reference)	***Intervention***	***Comparator***	***Population***	***Publication type***
Richardson 2006[39]	*Len 30mg/ once daily/Dex*	*Len 15mg/ twice daily/Dex*	*rrMM patients with ≥1 prior anti-myeloma therapy (at least one prior chemotherapy)*	*Full paper*
MM-010[40]	*Len/Dex*	*Placebo/Dex*	*rrMM patients with ≥1 prior anti-myeloma therapy*	*Full paper*
MM-009[41]	*Len/Dex*	*Placebo/Dex*	*rrMM patients with ≥1 prior anti-myeloma therapy*	*Full paper*
AMBER[42]	*Bor/Placebo*	*Bor/Bev*	*rrMM patients with ≥1 prior anti-myeloma therapy*	*Full paper*
Bev, bevacizumab; Bor, bortezomib; Dex, dexamethasone; Len, lenalidomide; rrMM, relapsed or refractory multiple myeloma.				

Source: Submission Table 7

Both MM-009 and MM-010 trials were sponsored by Celgene.

The Richardson (2006) study was excluded as it is a phase II study and was superseded by the two phase III studies MM-009 and MM-010.

The AMBER study was also excluded by Celgene on the basis of the proportion of patients receiving Bort as second-line treatment being low, and because it compared Bort monotherapy to bevacizumab.

Two additional studies were included by Celgene:

- A post-hoc analysis by Dimopoulos (2009) of MM-009 and MM-010 trial data to assess the impact of prior treatment history.
- A paper from Mateos (2010) which consisted of an updated follow-up and impact assessment of subsequent therapy in the Phase III VISTA trial, which compared Bort plus melphalan and prednisone with melphalan and prednisone.

• Comparators

As previously mentioned, due to lack of evidence to conduct a direct comparison between Len and the comparators, the submission included separate studies to inform the clinical effectiveness for the comparator treatments.

Celgene could not identify relevant RCT evidence for Bort and bendamustine. Six non-RCTs studies were identified for Bort however only the Taverna (2012) study was deemed relevant by Celgene. The only evidence identified for bendamustine was also a non-RCT study by Damaj (2012).

The ERG is not completely convinced by the reasons given for selecting Taverna (2012) as the main evidence source and excluding Hrusovsky (2010). The latter also reports OS in a European population (Germany and Switzerland) and of 100% patients had previously received Bort.

Therefore we now provide a brief description of the Hrusovsky (2010) study for comparison purposes:

Overview

Hrusovsky (2010) is a retrospective survey which was conducted in Switzerland and Germany, involving relapsed MM patients who had responded to initial Bort treatment.

Results

Initial Bort treatment and response: Patients in the per-protocol population (n = 60) received a median of 4.7 cycles of Bort (range 1–12 cycles) as the initial treatment, with the majority (85%) having received 1–6 cycles.

Bort retreatment: Patients received a median of 4.1 cycles of Bort retreatment (range 1–14), with 85.0% patients receiving 1–6 cycles.

Bort retreatment had an overall response rate (complete response (CR) + near complete response (nCR) + partial response (PR)) of 63.3% and a clinical benefit rate (CR + nCR + PR + stable disease) of 80%. Median TTP following Bort retreatment was 9.3 months.

Median OS from first diagnosis, initial Bort and after Bort retreatment was 1.1, 3.3 and 1.7 years, respectively. At the time of data cut-off following Bort retreatment, 30 patients had died.

Overall, the type of outcomes reported are fairly similar to Taverna (2012), however the rationale for excluding Hrusovsky (2010) is not transparent.

4.1.4 Details of any relevant studies that were not included in the submission

The ERG considers that all studies relevant to the direct comparison of Len/Dex with Bort were included in the submission.

4.1.5 Description and critique of manufacturer approach to validity assessment

In this section the two main studies presented in Celgene submission are assessed for their validity. MM-009 and MM-010 were initially assessed by Celgene (Table 82 in the submission). We present our independent comments alongside their assessment.

Table 4 provides the quality assessment of study MM-009 and MM-010.

Table 4. Clinical appraisal of relevant RCTs

Assessment question	Celgene response	ERG comments
Was randomisation carried out appropriately?	*Yes. In trials MM-009 and MM-010, A stratified randomisation list was independently generated before the study was initiated, which randomised the subjects in a 1:1 ratio to either the Len/Dex group or the placebo/Dex group. Randomisation was done centrally using an integrated voice-response system (IVRS). Randomisation was centralized and stratified by three factors: baseline serum beta-2 microglobulin, prior treatment with high-dose chemotherapy or SCT or no prior treatment, and number of prior anti-myeloma regimens.*	This is appropriate.
Was the concealment of treatment allocation adequate?	*Yes. MM-009 and MM-010 were double-blind studies. The lenalidomide and placebo capsules were identical in appearance, and the subjects, investigators, other study site personnel, and Celgene personnel who were responsible for the study were blinded to each subject's treatment assignment until the study was unblinded. An IVRS was used and all medication allotments were assigned by the IVRS. The clinical sites enrolled the patients and did so by accessing the central IVRS*	This method is adequate.
Was a justification of the sample size provided?	*The sample size was based on 85% power to detect a hazard ratio of 1.5 for TTP between the two arms (an increase of 6 to 9 months) and 80% power to detect a hazard ratio of 1.5 for OS (an increase of 12-18 months).*	These assumptions on which the sample size is established are justified by the reported outcomes presented in the submission (TTP HR of 2.8 and 2.9 and PFS HR of 3.0 and 2.6 for MM-009 and MM-010 respectively), however these HRs could not be found in either MM-009 or MM-010 clinical study reports (CSRs).

Assessment question	Celgene response	ERG comments
Was follow-up adequate?	*All patients were followed in the active phase of the study until disease progression or treatment was discontinued for any other reason. Subjects were contacted every 6 months during the follow-up phase.*	Follow-up was adequate. 353 and 351 patients were followed up to the extended follow-up cut-off period of 23 Jul 2008 and 02 Mar 2008 for MM-009 and MM-010 respectively, as described in the CSRs. Additionally, follow-up data collected for OS have been updated and reflect a follow-up period until 11 December 2008.
Were the individuals undertaking the outcomes assessment aware of allocation?	*No, all review of outcomes by the adjudication committee were conducted in blinded fashion.*	The ERG could not find any reference to the "adjudication committee" throughout the submission or the CSRs. The primary efficacy endpoint, TTP, is calculated as the time from randomization to the first occurrence of disease progression, which was assessed by a battery of tests based on various objective (haematological) and subjective (bone and soft tissue lesion) criteria. However, it is not clear who assessed this outcome and it would be important to ensure that allocation was concealed from the assessor. The knowledge of treatment allocation would introduce potential for bias, most likely in favour of Len/Dex. OS is unlikely to be affected in this way.

Assessment question	Celgene response	ERG comments
Was the design parallel-group or crossover? Indicate for each crossover trial whether a carryover effect is likely.	*It was a parallel-group design. Patients in the placebo/Dex group were only allowed to roll over to receive lenalidomide after disease progression, or cross over to receive Len/Dex after the Independent Data Monitoring Committee (IDMC) had declared the studies could be unblinded. Carry-over effect is not applicable in these two trials.*	The number of patients who crossed over in each trial, MM-009 and MM-010, was not clearly reported in the submission. Results from the pooled analysis show that 167/351 (47.6%) patients who received placebo/Dex chose to receive Len after unblinding of the study (submission p 77). However, CONSORT flow charts on the number of patients who crossed over in each trial, show 101/176 and 63/175 patients for MM-009 and MM-010 respectively. This is inconsistent with the results from the pooled analysis, as the CONSORT flow data adds up to a total of 164/351 patients. (submission p 62 and p 63)
Was the RCT conducted in the UK (or were one or more centres of the multinational RCT located in the UK)? If not, where was the RCT conducted, and is clinical practice likely to differ from UK practice?	*MM-009 took place in the USA and Canada, while MM-010 took place in Europe, Israel and Australia. Specifically, MM-010 included sites in The study was conducted in Australia (6 sites), Austria (1 site), Belgium (2 sites), France (5 sites), Germany (6 sites), Greece (1 site), Ireland (1 site), Israel (3 sites), Italy (6 sites), Poland (3 sites), Spain (6 sites), Switzerland (2 sites), Ukraine (5 sites), and the United Kingdom (3 sites; 2 in London and 1 in Bristol). A total of 15 patients across three UK sites were enrolled into MM-010. These countries are all representative of the clinical practice of lenalidomide use in Western countries and are relevant to that used in the England and Wales.* *See **Error! Reference source not found.** for comparison of second-line patient characteristics in the MM-010 trial and UK practice.*	Both trials were multinationals. The manufacturer justify using data from trial MM-010 alone to model the cost-effectiveness of Len as this included a European patient population.

Assessment question	Celgene response	ERG comments
How do the included in the RCT participants compare with patients who are likely to receive the intervention in the UK? Consider factors known to affect outcomes in the main indication, such as demographics, epidemiology, disease severity, and setting.	*There is no reason to suspect that the trial patient characteristics and outcomes would differ significantly from those seen in UK practice. However, since MM-009 and MM-010 were initiated, thalidomide and bortezomib have been licensed in Europe for first and second-line treatment, respectively. Therefore the proportion of patients in the UK receiving either of these drugs as prior therapies may be greater in clinical practice than was seen in the trials. In MM-009, 41.8% and 10.7% of patients in the Len/Dex arm had, respectively, received prior treatment with thalidomide and bortezomib. In MM-010, the respective proportions of patients previously treated with these agents were 30.1% and 4.5% in the Len/Dex arm.40, 41 The patients enrolled in the trials are slightly younger and have a better performance status at baseline than those that might be seen in UK clinical practice. However, the trial data show Len/Dex significantly improves outcomes over Dex regardless of age and performance status.*	Both trials were initiated in 2004. Since then the management of MM has undergone profound changes particularly with the introduction of novel agents such as Bort, therefore trials MM-009 and MM-010 do not accurately reflect the current clinical practice in the UK. The mean age of patients in both trials (63 years) reflect a slightly younger population than the typically Presenting UK population, which is approximately 70 years-old as reported in the guidelines for the diagnosis and management of multiple myeloma Bird (2013). Expert opinion sought by the ERG confirmed that MM has a higher incidence amongst Afro-Caribbean ethnic groups. MM-009 study contains more patients from this ethnic group than MM-010 (submission Table 9). Furthermore, there is a significant proportion of patients (submission Table 10) enrolled in both trials who received more than 3 prior stem cell transplant (SCT). This does not seem to reflect current clinical practice in the UK, where a smaller number of prior SCT would be expected as per our expert clinical advice. Finally, the proportion of patients who received 2 or 3 prior anti-myeloma therapies in both MM-009 and MM-010 is higher than the percentage of patients who received just 1 prior therapy. Clinical advice sought by the ERG revealed that in current clinical practice, most patients would have received one prior therapy.

Assessment question	Celgene response	ERG comments
For pharmaceuticals, what dosage regimens were used in the RCT?	*Dosage regimens were the same as those detailed in the SPC.*	The dosage was as per EMEA recommendation: • Len: 25 mg orally once daily on days 1-21 of repeated 28-day cycles • Dex: 40 mg orally once daily on days 1-4, 9-12, and 17-20 of each 28-day cycle for the first 4 cycles of therapy and then 40 mg once daily on days 1-4 every 28 days
Were the study groups comparable?	*Yes. The demographic and baseline characteristics of the study groups are comparable.*	Patient characteristics presented on Table 9 and Table 10 of the submission show comparable study groups.
Were the statistical analyses used appropriate?	*Yes. The statistical analyses used are considered appropriate. The protocol for both studies, including the statistical methods section, went through a Special Protocol Assessment by FDA and was agreed upon by the agency.*	The approach to the statistical analysis of MM-009 and MM-010 was generally sound (see section.4.2.7)
Was an intention-to-treat analysis undertaken?	*Yes.*	Yes, intention-to-treat analysis was undertaken, however the ERG noted some discrepancies in the data presented for MM-009 trial. The number of patients in the ITT group for MM-009 is reported to be 353 in the CONSORT flow chart (submission p 62), however, patient characteristics are only reported for 340 patients (submission p 51-53). We understand that this might be related with the 12 case report forms claimed to have been missing from the analysis, however, it was the manufacturer's decision to later include these 12 patients once data were recovered. Therefore, the submission should have included a presentation of their characteristics.

Assessment question	Celgene response	ERG comments
Were there any confounding factors that may attenuate the interpretation of the results of the RCT(s)?	*In the MM-009 and MM-010 trials, patients in the placebo/Dex arm were allowed to cross-over to the Len/Dex arm when there was a documented progression or at unblinding by the IDMC. This cross-over confounded the measurement of OS in favour of the placebo/Dex group in general, and is likely to explain the decreasing difference in OS between the study groups over time.* *TTP in the placebo/Dex arms is relatively unaffected by the treatment crossover, because most patients had developed PD when the studies were unblinded – 75.0% in MM-009 and 81.1% in MM-010.*	Detail on the number of cross-over patients for the pooled analysis is given on page 77 of the submission, however the equivalent information is not provided for trials MM-009 and MM-010 separately. Despite of this, any impact is likely to be in favour of the control arm.

Response source: Submission Table 9 and Table 10.

To the best of the ERG's knowledge, even though several tools exist for the assessment of non-RCT evidence, there is no standard validated tool.

The manufacturer reported the quality assessment of non-RCTs in Appendix 7 of the submission using the same tool as for RCTs evidence. This is not very informative.

The ERG note that the Centre for Reviews and Dissemination (CRD) quality assessment tool (Chambers 2009) for case series could have been used for the quality assessment of non-RCTs evidence.

4.1.6 Description and critique of manufacturers outcome selection

The outcome selection in Celgene's submission is a direct reflection of those included in the two main RCTs.

Primary efficacy endpoints

Primary and secondary efficacy analyses were conducted on data from all patients on MM-009 and MM-010 trials separately, as well as on data from the pooled analyses and on long-term safety data.

The primary outcome in both studies was time to disease progression (TTP). This was measured from the date of the first assessment in the series of tests required to determine progression and calculated as the time from randomisation to either:

- The first occurrence of disease progression according to the myeloma response assessment data developed by EBMT in Blade (1998) and Durie (2006) (Table 5) or,
- discontinuation from treatment due to disease progression (whether or not confirmed by response criteria) or,
- death due to disease progression during the treatment period.

Observations were censored at the date of last response assessment for subjects who had either:

- Not progressed at the time of analysis or,
- withdrew from the treatment phase before documented progression or,
- died of causes not related to multiple myeloma or,
- received another anti-myeloma therapy without documented progression or intolerable AEs.

Additional sensitivity analyses were undertaken to assess:

- Time to progression (FDA definition: counting subjects who withdrew from the study for any reason or who received antimyeloma therapy during the treatment period as having events on the last assessment day prior to withdrawal from the study or to receiving antimyeloma medication); and
- time to treatment failure.

Secondary efficacy endpoints

Secondary outcomes analysed are:

- Progression-free survival (PFS), conducted as part of a supportive analysis for the primary endpoint. This was calculated as the time from randomisation to documented progression or death due to any cause during the treatment period, whichever occurred first.
- Overall survival (OS) defined as the time from randomisation until death from any cause.
- Response rate assessed using the myeloma response determination criteria developed by EBMT (Table 5).

- Functioning and quality of life, consisting of:
    - Time to first skeletal-related event.
    - Time to first decrease in ECOG performance status, calculated as the time from randomisation to the date of the first worsening compared with the last ECOG evaluation obtained prior to randomisation.
- Adverse events.

Table 5. Myeloma response determination criteria

Outcome	Criteria for Classification [a]
Complete response (CR)	*A CR required:* • *Disappearance of M-paraprotein in serum and/or urine by electrophoresis maintained for ≥6 weeks.* • *Documentation of the following findings within ±2 weeks of the confirmatory electrophoresis studies:* – *Absence of M-paraprotein confirmed by immunofixation studies of serum and urine.* – *Less than 5% plasma cells in the bone marrow aspirate or biopsy.* – *Disappearance of soft tissue plasmacytomas.* – *No increase in size of number of lytic bone lesions (the development of bone fractures did not exclude a response).* *If some, but not all, of the criteria for a CR were fulfilled, the response was classified as a PR or RR, provided that all other requirements were satisfied.*
Remission response (RR)	*An RR required:* • *A 75% to 99% reduction from baseline in serum M-paraprotein and, if present, a 90% to 99% reduction from baseline in 24-hour urinary light chain excretion or a reduction in the 24-hour urinary light chain excretion to <200mg by electrophoresis, which was maintained for ≥6 weeks.* • *Documentation of the following findings within ±2 weeks of the confirmatory electrophoretic studies:* – *If present, at least a 50% reduction from baseline in the sum of the products of perpendicular diameters of measurable soft tissue plasmacytomas by radiography or clinical examination [b]. If present, there must be no clear progression of evaluable soft tissue plasmacytomas or non-evaluable disease [c, d].* – *No increase in the size or number of lytic bone lesions (the development of bone fractures did not exclude a response).* – *No evidence of disease progression by bone marrow aspirate/biopsy findings (see PD, below).*
Partial response (PR)	*A PR required:* • *A 50% to 74% reduction from baseline in serum M-paraprotein and, if present, a 50% to 89% reduction from baseline in 24-hour urinary light chain excretion by electrophoresis, which was maintained for ≥6 weeks.* • *Documentation of the following findings within ±2 weeks of the confirmatory electrophoretic studies:* – *At least a 50% reduction from baseline in the sum of the products of perpendicular diameters of measurable soft tissue plasmacytomas by radiography or clinical examination [b]. If present, there must be no clear progression of evaluable soft tissue plasmacytomas or non-evaluable disease [c, d].*

Outcome	Criteria for Classification [a]
	– *No increase in the size or number of lytic bone lesions (the development of bone fractures did not exclude a response).* – *No evidence of progressive disease (PD) by bone marrow aspirate/biopsy findings (see PD, below).*
Stable disease (SD)	*Criteria for PR or PD were not met.*
Plateau phase of response	*For subjects who achieved at least a confirmed PR, plateau phase of response was defined by stable M-paraprotein values (within 25% above or below nadir value) and, if present, stable measurements for measurable soft tissue plasmacytomas (sum of the products of perpendicular diameters within 25% above or below the nadir value) maintained for at least 3 months without evidence of PD or further response.*
Progressive disease (PD)	*PD for subjects in CR required at least one of the following:* • *Reappearance of serum or urinary M-paraprotein on immunofixation or electrophoresis on 2 consecutive occasions at least 1 week apart.* • *Increase in the percentage of plasma cells in bone marrow aspirate or biopsy to ≥5%.* • *Development of at least one new lytic bone lesion or soft tissue plasmacytoma.* • *Clear increase in size of residual bone lesions (the development of a bone fracture, including a vertebral compression fracture, did not, in of itself, constitute PD).* • *Development of hypercalcaemia (serum calcium level, corrected for albumin concentration, >11.5mg/dL [2.8 mmol/L]) not attributable to any other cause.*
Progressive disease (PD)	*PD for subjects not in CR required at least one of the following:* • *Compared with the nadir value, a >25% increase in the level of serum M-paraprotein, which represented an absolute increase of ≥500mg/dL (5g/L), on 2 consecutive occasions at least 1 week apart.* • *Compared with the nadir value, a >25% increase in the level of the 24-hour light chain excretion, which represented an absolute increase of ≥200mg/dL/24 hours, on 2 consecutive occasions at least 1 week apart.* • *Compared with the lowest marrow plasma cell percentage achieved during study treatment, a >25% increase in plasma cells in bone marrow aspirate or biopsy, which represented an absolute increase of ≥10%.* • *Development of at least one new lytic bone lesion or soft tissue plasmacytoma.* • *Clear increase in size of existing bone lesions (the development of a bone fracture, including a vertebral compression fracture, did not, in itself, constitute PD).* • *Compared with the nadir value achieved, a >25% increase in the sum of the products of existing measurable soft tissue plasmacytomas.* • *Clear PD of evaluable soft tissue plasmacytomas or non-evaluable disease.* • *Development of hypercalcaemia (serum calcium level, corrected for albumin concentration, >11.5mg/dL [2.8 mmol/L]) not attributable to any other cause.*
[a] Response criteria for both serum and urine myeloma paraprotein (M-paraprotein) must be met in subjects in whom both are present. *[b] Measurable soft tissue plasmacytomas have defined borders and have perpendicular diameters that measure ≥1 cm x ≥1 cm.* *[c] Evaluable soft tissue plasmacytomas have poorly defined borders or are measurable in only one dimension.* *[d] Non-evaluable disease comprises malignant pleural or pericardial effusions, ascites, and previously irradiated lesions.*	

Source: Submission Table 88 Appendix E

The selection of outcomes appears reasonable to provide a sensible range of dimensions to assess the clinical effectiveness of Len.

Unfortunately there was no questionnaire to directly capture patients' HRQoL in any of the two trials.

4.1.7 Description and critique of the statistical approach used

Statistical analysis was performed on the main study outcomes i.e. time-to-event and time-to-first worsening of ECOG-PS for the ITT population of studies MM-009 and MM-010. Statistical analysis was also performed for second-line patient's subgroup but only for PFS and OS.

It is stated in Celgene's submission (p 59) that *formal statistical hypothesis tests of the superiority of Len/Dex relative to placebo/Dex were conducted at the 2-sided, 0.05 level of significance*. However, the manufacturer also states (p 44), that they powered their trial to detect a difference measured by a one-sided log-rank test at the 0.025 level. Additionally, all log-rank tests were specified to be one-tailed.

It should be noted that in most instances highly significant p-values are generated by these tests and that the overall approach to the statistical analysis of MM-009 and MM-010 was generally sound.

4.1.8 Summary statement

Following responses to the ERG's questions for clarification in relation to the effectiveness and health related quality of life searches, we are content that the searches presented in this submission are broadly suitable for the task.

The ERG opinion is that the manufacturer's search strategy on clinical effectiveness was generally appropriate. However to note is that:

- The fact that no separate searches were undertaken for AEs is perceived as a weakness.
- The ERG is not clear as to why some papers providing evidence for comparator treatments were excluded from the economic analysis.

The methodology used to assess the quality of the included RCT was adequate.

The ERG consider that the evidence submitted generally reflects the decision problem outlined in the final scope of the submission.

4.2 Summary of submitted evidence

4.2.1 Summary of results

The ERG found that in the original submission there was lack of clarity as to which set of the presented outcomes were used to inform the economic analysis.

After the request for clarification by the ERG, Celgene stated that the outcomes presented in Section 6 of the submission for MM-010 were based on an early analysis of the trial data up to unblinding (3 August 2005), whereas model inputs use the extended, open-label follow-up data (2 March 2008). Furthermore Celgene claim that outcomes from the extended follow-up period, stratified by number of prior anti-myeloma therapies received are provided on page 79-81 of the submission and in Appendix D. However, these are provided only for the subset analysis conducted by Stadtmauer (2009).

Additionally, as per Celgene's response to clarification (question B8), the economic analysis was conducted using the full MM-010 dataset. Therefore, the results presented in Appendix D are not very informative, as they only report a subset analysis.

- #### Primary outcome: Time to Progression

ITT population at unblinding

Time to progression (TTP) is the primary outcome for both RCTS MM-009 and MM-010. In Table 6 we report the results at study unblinding (28 June 2005 [MM-009] / 3 August 2005 [MM-010]). The manufacturer also present results for TTP at protocol-defined interim analysis (Table 14 p 67 in the submission).

Table 6. Time to Progression for ITT population

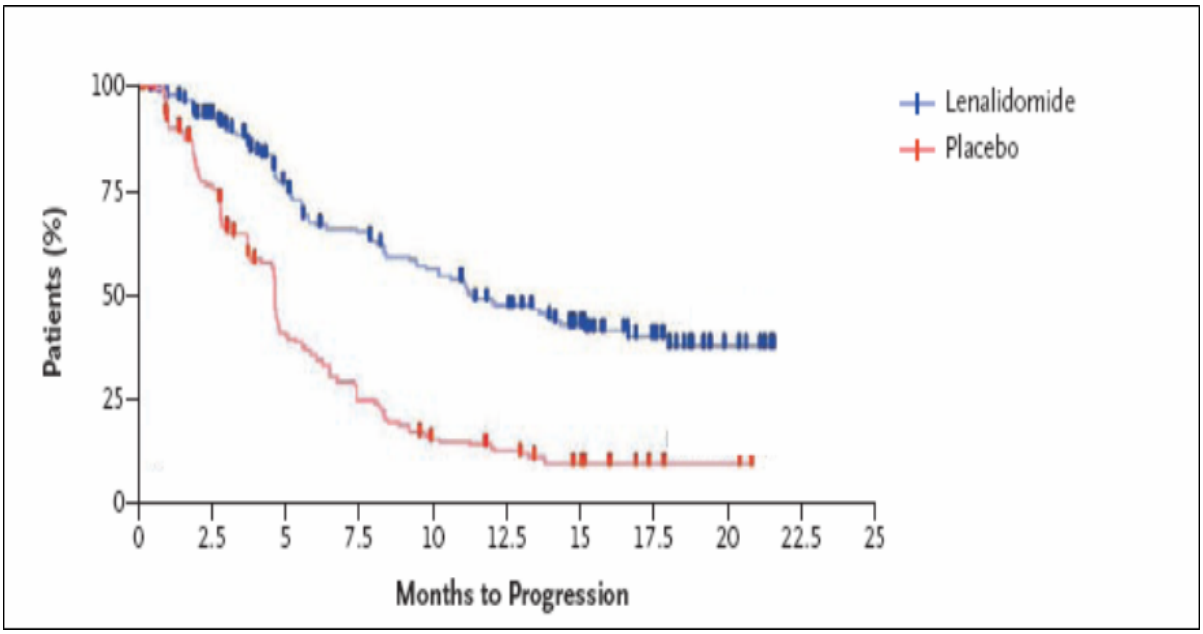
	Statistic	Study MM-009		Study MM-010	
		Len/Dex	Dex	Len/Dex	Dex
TTP *Progressed* *Censored*	*N* *n (%)* *n (%)*	*177* *92 (52.0)* *85 (48.0)*	*176* *132 (75.0)* *44 (25.0)*	*176* *82 (46.6)* *94 (53.4)*	*175* *142 (81.1)* *33 (18.9)*
Overall TTP (weeks)	*Median* *[95% CI]* *[a]*	*48.1* *[36.9, 1.4]*	*20.1* *[16.7, 23.1]*	*48.7* *[40.9, 72.1]*	*20.1* *[18.1, 20.7]*
Hazard Ratio [95% CI] [b]		*2.822 [2.146, 3.701]*		*2.850 [2.159, 3.762]*	
Log-rank Test p-Value [c]		*< 0.001*		*< 0.001*	

CI, confidence interval.
Percentages are based on the number of treated subjects. The median in this table is based on Kaplan–Meier estimate, and the mean is the univariate mean without adjusting for censoring.
[a] 95% confidence intervals about the median overall time to progression.
[b] Based on a proportional hazards model comparing the hazard functions associated with the treatment groups (Len/Dex:Dex)
[c] The p value is based on the a one-tailed unstratified log rank test of survival curve differences between the treatment groups.
Median follow up: 17.1 months for MM-009 (n=76), 16.7 months for MM-010 (n=74), 16.9 months for combined (n=150).

Source: Submission Table 15

Kaplan-Meier (KM) curves for TTP at the time of unblinding are presented for each trial in the submission (Figure 6 and Figure 7 for MM-009 and MM-010, respectively). Figure 1 presents the KM curve for TTP in the MM-010 population.

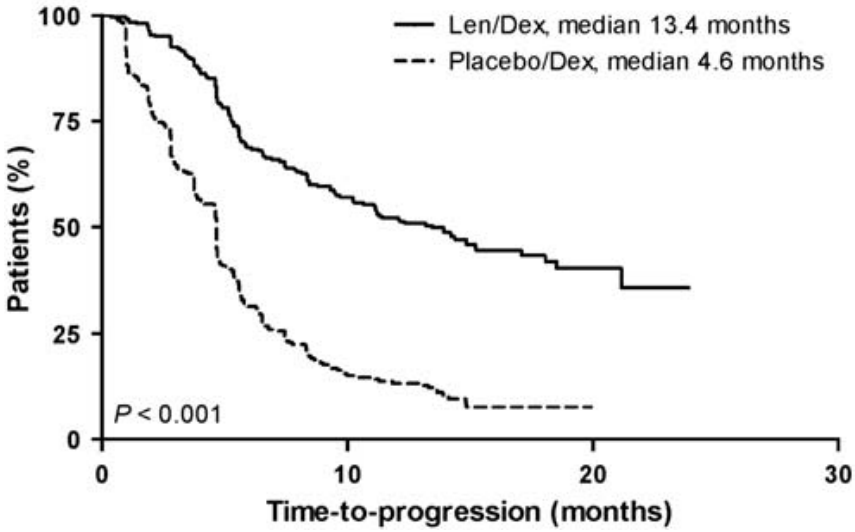
Figure 1. KM curve for TTP at study unblinding for MM-010.



Pooled analysis

Dimopoulos (2009) conducted a pooled analysis of efficacy data from both trials with long-term findings, which is presented in Celgene submission. They report efficacy data up to study unblinding, with follow-up OS data of 48 months (up to July 2008). Figure 2 shows the KM estimate of TTP for the ITT population using data up to unblinding (June 2005 for MM-009 and August 2005 for MM-010).

Figure 2. Kaplan–Meier estimate of TTP in the ITT population



Source: Submission Figure 8

The pooled efficacy results from Dimopoulos (2009) are summarised in Table 7.

Table 7. Response rates, TTP, PFS and OS

	Len/Dex (n=353)	***Dex (n=351)***	***p value***
Up to unblinding (median 17.5 months) *Response rate, %*			
ORR	*60.6*	*21.9*	*<0.001*
CR	*15.0*	*2.0*	*<0.001*
VGPR	*17.3*	*2.8*	
PR	*28.3*	*17.1*	
Median TTP, months	*13.4*	*4.6*	*<0.001*
Median DoR, months	*15.8*	*7.0*	*<0.001*
Median PFS, months	*11.1*	*4.6*	*<0.001*
Extended FU (median 48 months) *Median OS, months*	*38.0*	*31.6*	*0.045*

CR, complete response; DoR, duration of response; FU, follow-up; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; TTP, time-to-progression; VGPR, very good partial response.

Source Celgene Submission Table 16

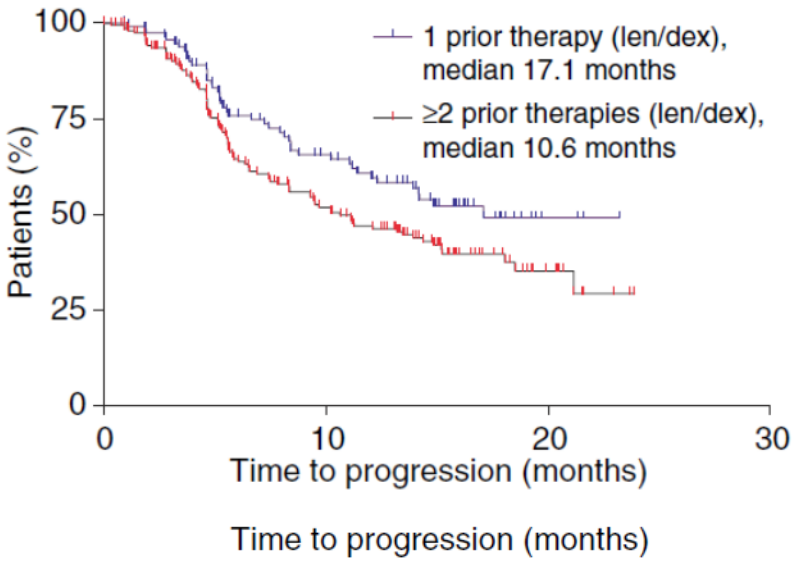
As shown in Table 6, it is noticeable that the results from the two different studies agree extremely well, therefore one would expect a similar estimate when using pooled data. However the results for the median TTP for the treatment arm from the pooled efficacy are 13.4 months (58 weeks) vs. 4.6 months (19.9 weeks). This result differ somewhat from the results observed on Table 6 for Len/Dex: 48.1 vs. 20.1 weeks and 48.3 vs. 20.1 weeks for MM-009 and MM-010 respectively.

Stratified according to relapse phase

The manufacturer present separate TTP results for first relapse and second/subsequent relapse for the updated analysis from Stadtmauer (2009). This analysis encompassed 353 Len/Dex patients, of whom 133 had received one prior therapy and 220 had received two or more prior therapies.

Figure 3 represents the KM curves of TTP patients for trial participants who were at first relapse at baseline (133 patients) compared to those who had received two or more prior therapies (220 patients).

Figure 3. Kaplan–Meier estimate of TTP of patients treated with Len/Dex by number of prior therapies



Time to progression (months)

Despite the observed discrepancy regarding the median TTP for the treatment arm from the pooled analysis, a significant TTP benefit was observed for patients taking Len/Dex compared to those taking Dex only.

- ## Progression-free survival

ITT population at unblinding

Celgene report PFS data up to unblinding of the two phase III studies. This data is reproduced in Table 8.

Table 8. Summary of PFS (ITT population) – MM-009 and MM-010

	Statistic	Study MM-009		Study MM-010	
		Len/Dex	Dex	Len/Dex	Dex
PFS [a] *Progressed* *Censored*	*N* *n (%)* *n (%)*	*170* *46 (27.1)* *124 (72.9)*	*171* *99 (57.9)* *72 (42.1)*	*176* *51 (29.0)* *125 (71.0)*	*175* *104 (59.4)* *71 (40.6)*
Overall PFS (wk)	*Median [b]* *[95% CI] [c]*	*41.1* *[29.4, NE]*	*20.1* *[16.7, 24.1]*	*NE* *[34.1, NE]*	*20.1* *[19.7, 21.7]*
	Mean [d] *SD* *Min, Max*	*21.2* *13.39* *0.0, 60.1*	*15.7* *11.17* *0.0, 57.0*	*19.8* *10.93* *0.0, 44.7*	*16.4* *10.03* *0.3, 48.1*
Hazard ratio [95% CI] [d]		*2.970 [2.089, 4.222]*		*2.567 [1.834, 3.592]*	
Log-rank Test p-value [e]		*<0.001*		*<0.001*	

PFS, progression-free survival; NE = not estimable.
[a] Calculated as the time from randomisation to documented progression or death due to any cause, whichever occurred first. If withdrawal due to adverse events or change of therapy occurred before documented progression or death, then these observations were censored at the last progression assessment date.
[b] The median is based on the Kaplan–Meier estimate, and the mean is the univariate mean without adjusting for censoring (i.e., the mean values represent mean PFS documented to date as of the data cut-off date, without consideration of the fact that a substantial number of subjects who had not yet progressed were continuing in the study).
[c] Ninety-five percent confidence intervals about the median overall PFS.
[d] Based on a proportional hazards model comparing the hazard functions associated with the treatment groups (Len/Dex: placebo/Dex).
[e] The p-value is based on a one-tailed unstratified log rank test of survival curve differences between the treatment groups.

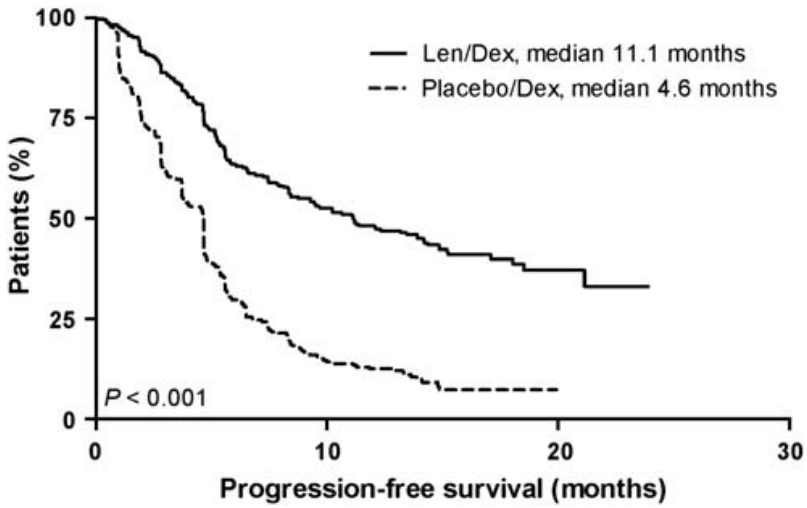
Source: Celgene Submission Table 17

It is not clear to the ERG why the number of subjects considered here is different from that considered in the TTP analysis, i.e. 170 for Len/Dex and 171 for Dex for PFS and 177 for Len/Dex and 176 for Dex for TTP. Celgene explained that 12 subjects from one investigation site were not reviewed at the interim analysis, however they were included in the results later on at study unblinding and therefore results for these patients should have been included.

Pooled analysis

Celgene also present the KM curves of the pooled analysis from Dimopoulos (2009), which is reproduced in Figure 4.

Figure 4. Kaplan–Meier estimate of PFS in the ITT population

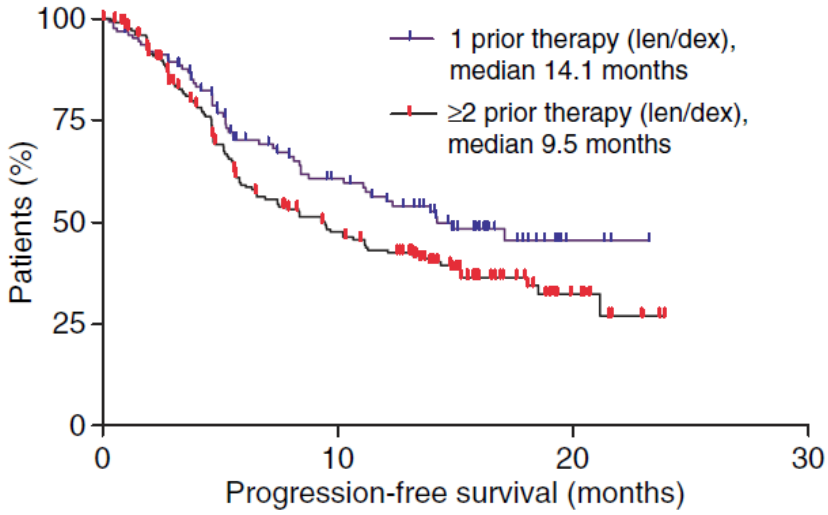


Source: Submission Figure 9

Stratified according to relapse phase

The KM curve depicting observed PFS for patients with one prior therapy after first relapse compared with those who received two or more therapies is reproduced in Figure 5.

Figure 5. Kaplan–Meier estimate of PFS of patients treated with Len/Dex by number of prior therapies



Source: Submission Figure 14

Subgroup analysis in second line patients

Celgene present a patient level analysis of data from studies MM-009 and MM-010 to evaluate OS and PFS in second-line patients only.

Table 9. PFS of second-line patients in studies MM-009 and MM-010

		Study MM-009		*Study MM-010*	
		Len/Dex	*Dex*	*Len/Dex*	*Dex*
PFS	*n*	*68*	*67*	*56*	*57*
Died	*n (%)*	*37 (54.4)*	*48 (71.6)*	*31 (55.4)*	*48 (84.2)*
Censored	*n (%)*	*31 (45.6)*	*19 (28.4)*	*25 (44.6)*	*9 (15.8)*
	Median (months)	*16.6*	*4.6*	*13.3*	*4.5*
	[95% CI]	*[11.0, 36.8]*	*[4.0, 5.7]*	*[5.1, 26.9]*	*[2.8, 5.6]*
Hazard ratio [95% CI]		*0.30 [0.19, 0.47]*		*0.39 [0.24, 0.62]*	
Log-rank test p-value		*<0.0001*		*<0.0001*	

Source: Celgene Submission Table 22

Kaplan-Meier (KM) curves for PFS among second-line patients only are presented for each trial in the submission (Figure 16 and Figure 17 for MM-009 and MM-010, respectively).

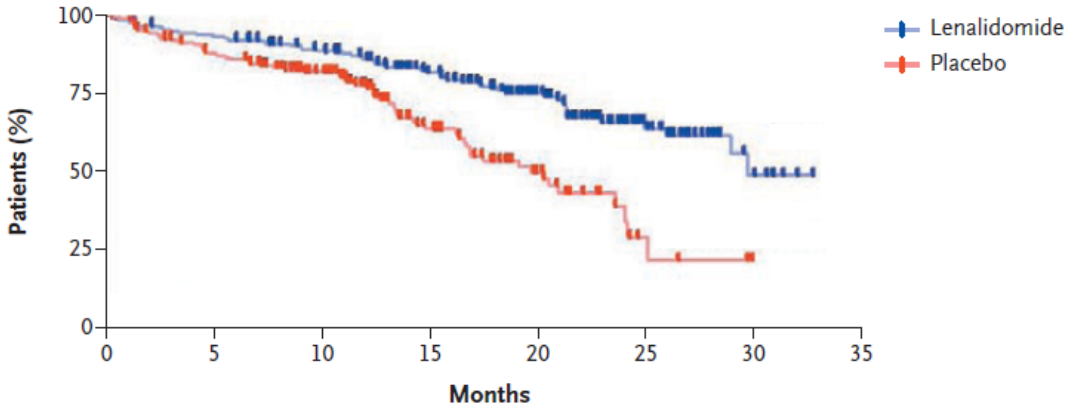
- ## Overall survival

ITT population

In their submission, Celgene present KM data for OS data measured in May 2006. Kaplan-Meier curves depicting observed OS for the ITT population in each of the two RCTs are reproduced in Figure 4 and Figure 5.

OS data were provided for patients treated with one prior therapy only, and results are presented in the next subsection.

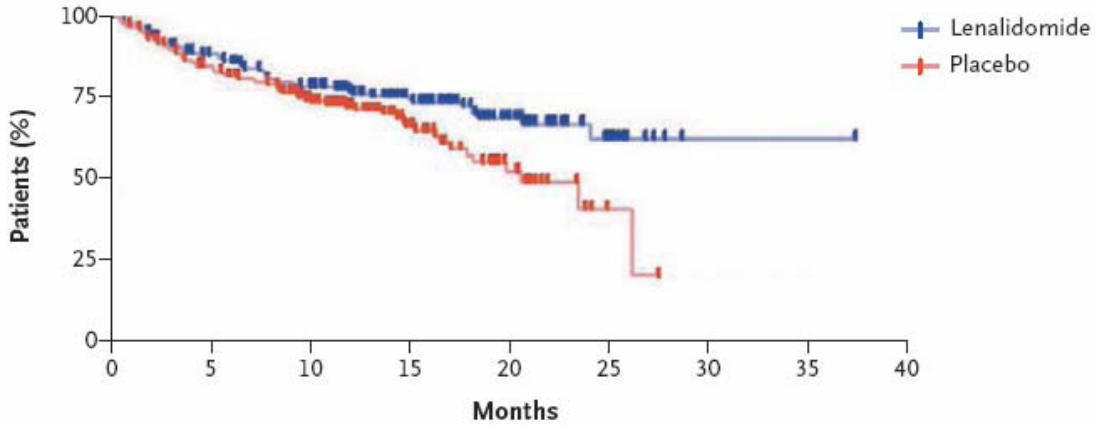
Figure 6. KM curves for OS for all patients in MM-009 - May 2006, ITT population



No. at Risk							
Lenalidomide	177	164	144	109	74	34	7
Placebo	175	144	115	51	26	5	1

Source: Celgene Submission Figure 10

Figure 7. KM curves for OS for all patients in MM-010-May 2006, ITT population



Source: Celgene Submission Figure 11

It can be noted that Figure 7 does not provide the number of patients at risk throughout time. We requested clarification from the manufacturer regarding the number of patients at risks in MM-010, however the manufacturer failed to provide these data.

Comparing Figure 6 with Figure 7 we can observe that curves differ substantially particularly towards the later stages of follow-up for the intervention arm in MM-010. The fair degree of inconsistency in these results is probably to be expected, given the very high proportion of

censorship in the underlying dataset (fewer than 30% of participants in each trial had died at the time the data was analysed – information obtained from TA171).

One prior therapy

In the submission, data were provided for patients treated with one prior therapy only. This is presented in Table 10.

Table 10. Overall survival

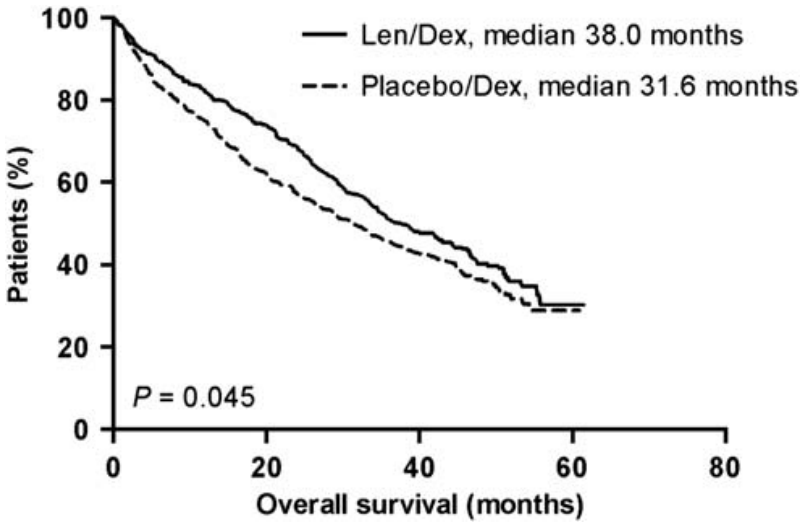
Characteristic	*MM-009*		*MM-010*	
	Len/Dex	*Dex*	*Len/Dex*	*Dex*
Died, n (%)	*49 (27.7)*	*63 (35.8)*	*47 (26.7)*	*60 (34.3)*
Median OS (months)	*29.6*	*20.2*	*NE*	*20.6*
Hazard ratio	*0.44*		*0.66*	
95% CI	*0.30–0.65*		*0.45–0.96*	
P	*<0.001*		*0.03*	
CI, confidence interval: NE, not estimable; ITT. Intention to treat; OS, overall survival. Data analysed as of May 2006 for both studies – a time from study initiation of 3 years and 3 months for MM-009 and 2 years and 8 months for MM-010. Median follow-up at this time-point is 17.1 months for MM-009 and 16.5 months for MM-010.				

Source: Celgene Submission Table 18

Pooled data

Follow-up analysis was presented in the submission on page 76 as a pooled analysis for a median follow-up of 48 months: *199 (56.4%) Len/Dex patients had died, compared with 219 (62.4%) placebo/Dex patients.*

Figure 8. Kaplan–Meier curves of OS for all patients



Source: Submission Figure 12

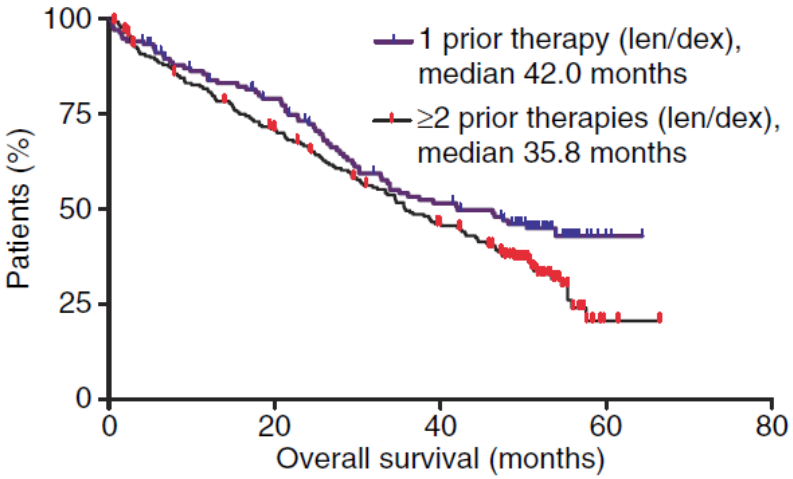
It should be mentioned that follow-up data are more mature but also more susceptible to confounders due to post-unblinding crossover in the placebo arm.

Celgene provide two different figures reporting patient crossover, however we have not been able to identify which is the most up to date. Indeed, on page 75 of the submission Celgene explain that *170 out of 351 patients in the placebo/Dex arms subsequently went on to receive additional Len*. However, the manufacturer state on page 77 that *of the 351 placebo/Dex patients, 167 (47.6%) received Len-based therapy after unblinding of the study or following disease progression*

Stratified according to relapse phase

KM curves depicting observed OS stratified according to first relapse v. second/subsequent relapses are reproduced in Figure 9.

Figure 9. KM curve of OS of patients treated with Len/Dex by number of prior therapies



Source: Celgene Submission Figure 15

Subgroup analysis in second line patients

Celgene also present OS results from the patient level analysis in second-line patients. Results from both RCTs are reproduced in Table 11.

Table 11. OS of second-line patients in studies MM-009 and MM-010

		Study MM-009		Study MM-010	
		Len/Dex	Dex	Len/Dex	Dex
OS	n	68	67	56	57
Died	n (%)	34 (50.0)	41 (61.2)	21 (37.5)	31 (54.4)
Censored	n (%)	34 (50.0)	26 (38.8)	35 (62.5)	26 (45.6)
	Median (months)	50.1	37.6	NE	37.2
	[95% CI]	[8.3, 32.5]	[6.5, 21.5]	[34.3, NE]	[2.1, 23.5]
Hazard ratio [95% CI]		0.70 [0.44, 1.10]		0.71 [0.41, 1.23]	
Log-rank test p-value		0.1179		0.2175	
CI, confidence interval; OS, overall survival; NE, not estimable.					

Source: Celgene Submission Table 23

KM curves for OS among second-line patients are presented in the submission (Figure 18 and Figure 19 for MM-009 and MM-010, respectively).

It is worth noting that this subgroup analysis did not generate significant results. Crossover may have contaminated results, however it is difficult to confirm and quantify the degree of confounding without employing a method to adjust for crossover such as the inverse probability of censoring weighting or the rank preserved structural failure time.

- ## Response to therapy

ITT population at unblinding

The response rates of patients treated with Len/Dex vs. placebo/Dex using the determination criteria developed by EBMT (see Section 4.1.6) are reproduced in Table 12.

Table 12. Response rates at unblinding

	Study MM-009		Study MM-010	
	Len/Dex n=177	Dex n=176	Len/Dex n=176	Dex n=175
Response				
CR [c]	25 (14.1%)	1 (0.6%)	28 (15.9%)	6 (3.4%)
RR	52 (29.4%)	16 (9.1%)	46 (26.1%)	16 (9.1%)
PR	31 (17.5%)	18 (10.2%)	32 (18.2%)	20 (11.4%)
SD	54 (30.5%)	102 (58.0%)	53 (30.1%)	97 (55.4%)
PD	5 (2.8%)	25 (14.2%)	3 (1.7%)	25 (14.3%)
NE [a]	10 (5.6%)	14 (8.0%)	14 (8.0%)	11 (6.3%)

	Study MM-009		Study MM-010	
	Len/Dex n=177	***Dex n=176***	***Len/Dex n=176***	***Dex n=175***
p-value [e]	*<0.001*		*<0.001*	
Dichotomised response				
CR, RR or PR	*108(61.0%)*	*35 (19.9 %)*	*106 (60.2%)*	*42 (24.0 %)*
SD, PD or NE	*69 (39.0%)*	*141 (80.1 %)*	*70 (39.8%)*	*133 (76.0 %)*
p-value [f]	*<0.001*		*<0.001*	
Odds ratio [g] [95% CI]	*6.31 [3.91, 10.17]*		*4.80 [3.03, 7.59]*	

CI, confidence interval; CR, complete response; NE, not evaluable; PD, progressive disease, ; PR, partial response; RR, remission response, SD, stable disease.
[a] Response in this table is based on the review of all myeloma assessment data using Blade criteria.
[b] Response is the highest assessment of response during the treatment phase of the study.
[c] Comparison of the CR rate shows that the CR rate is significantly higher in the Len/Dex group than in the placebo/Dex group (p<0.003 continuity corrected Pearson chi square).
[d] Including subjects who did not have any response assessment data at the data cut-off point, or whose only assessment was 'response not evaluable'.
[e] Probability from Wilcoxon rank sum test.
[f] Probability from continuity-corrected Pearson chi square test.
[g] Calculated based upon the reported response rates
The median follow-up was 17.6 months for MM-009 and 16.4 months for MM-010.

Source: Celgene Submission Table 20

At study unblinding, the overall response rate (defined as complete, near-complete or partial response) are significantly higher in the Len/Dex group than in the placebo/Dex group for both RCTs,

Pooled data - stratified according to relapse phase

Celgene present response results stratified according to first relapse vs. second/subsequent Relapse. The data is reproduced in Table 13.

Table 13. Outcomes in patients by number of prior therapies

	Len/Dex		
	1 prior therapy (n=133)	***≥2 prior therapies (n=220)***	***p-value***
Response rates, n (%)			
Overall response	*89 (66.9)*	*125 (56.8)*	*0.060*
CR	*27 (20.3)*	*26 (11.8)*	*0.028*
VGPR	*26 (19.5)*	*35 (15.9)*	
CR + VGPR	*53 (39.8)*	*61 (27.7)*	*0.025*
Partial response	*36 (27.1)*	*64 (29.1)*	
Stable disease	*30 (22.6)*	*77 (35.0)*	
Progressive disease	*6 (4.5)*	*2 (0.9)*	
Response not evaluable	*8 (6.0)*	*16 (7.3)*	
Median duration of treatment, months (range)	*12.5 (0.3–24.1)*	*9.2 (0.03–24.8)*	*<0.001*
Median duration of response, months (range)	*NR (11.4–NR)*	*13.0 (8.4–NR)*	*0.21*
Patients who relapsed, %	*34.5*	*44.4*	*0.16*
Patients who had a dose reduction [a], %	*33.1*	*38.0*	*0.36*
Patients who discontinued due to toxicity, %	*14.3*	*14.5*	*0.54*
[a] With or without interruption in lenalidomide treatment. *CR, complete response; NR, not reached; VGPR, very good partial response.*			

Source: Celgene Submission Table 21

- ## Health related quality of life outcomes

Time to first worsening of ECOG-PS

Celgene provide time to first worsening of ECOG performance status results from the two trials for the overall population. The data is reproduced in Table 14.

Table 14. Time to first worsening of ECOG performance status

		Study MM-009		Study MM-010	
	Statistic	Len/Dex N=177	Dex N=176	Len/Dex N=176	Dex N=175
Time to first worsening Worsened Censored	N n (%) n (%)	171 88 (51.5) 83 (48.5)	174 101 (58.0) 73 (42.0)	173 111 (64.2) 62 (35.8)	172 97 (56.4) 75 (43.6)
Overall time to first worsening (wk)	Median [95% CI]	36.3 [16.1, NE]	12.1 [8.3, 16.4]	10.1 [8.1, 16.1]	12.3 [10.1, 24.1]
	Mean SD Min, Max	30.6 31.11 0.0, 104.3	15.2 17.25 0.0, 80.9	20.6 23.36 0.0, 93.0	17.9 18.13 0.0, 88.4
Hazard ratio [95% CI]		1.448 [1.083, 1.937]		0.858 [0.653, 1.128]	
Log-rank Test p-value		0.012		0.271	
CI, confidence interval; ECOG-PS, Eastern Cooperative Oncology Group performance status; NE, not estimable.					

Source: Celgene Submission Table 24

Time to first skeletal-related event

Results for time to first skeletal-related event are not available as there were too few events for both studies and no analysis could be done (submission p 86).

- ## Adverse events

The manufacturer report patient exposure from a pooled study of MM-009 and MM-010 as of 31 December 2005 with a median duration of treatment of 44.0 weeks in the Len/Dex arm. Data is reproduced in Table 15.

Table 15. Pooled duration of treatment in studies MM-009 and MM-010

	Len/Dex N=353		Dex N=350	
Treatment phase duration				
	N	%	n	%
<1 week	1	0.3	2	0.6
1 to <4 weeks	14	4.0	14	4.0
4 to <8 weeks	14	4.0	38	10.9
8 to 12 weeks	27	7.6	42	12.0
12 to <16 weeks	15	4.2	28	8.0
16 to <20 weeks	18	5.1	31	8.9
20 to <24 weeks	16	4.5	23	6.6
24 to <28 weeks	19	5.4	38	10.9
28 to <32 weeks	19	5.4	27	7.7
32 to <36 weeks	10	2.8	12	3.4

	Len/Dex N=353		***Dex N=350***	
36 to <40 weeks	*11*	*3.1*	*15*	*4.3*
40 to <44 weeks	*12*	*3.4*	*13*	*3.7*
44 to <48 weeks	*8*	*2.3*	*8*	*2.3*
48 to <52 weeks	*6*	*1.7*	*4*	*1.1*
≥52 weeks	*163*	*46.2*	*55*	*15.7*
Duration of exposure (weeks)				
n	*353*		*350*	
Mean	*53.9*		*29.7*	
SD	*38.76*		*26.41*	
Median	*44.0*		*23.1*	
Min, Max	*0.1, 161.7*		*0.3, 124.0*	
Min, minimum value; Max, maximum value; SD, standard deviation.				

Source: Celgene Submission Table 28

According to the EPAR (EMEA scientific discussion page 25), which was reported in Celgene submission page 105, *anaemia, neutropenia, thrombocytopenia, constipation, pneumonia, decreased weight, hypokalaemia, hypocalcaemia, tremor, rash, and deep vein thrombosis (DVT) were reported significantly more frequently in the Len/Dex group than in the placebo/Dex group.*

This statement does not tally perfectly the numerical data of incidence. For instance the incidence of Anaemia NOS appears to be relatively similar across the two groups: 119/353 = 33.7% vs. 83/350 = 23.7% for grade 1-4 and 38/353 = 10.8% vs. 21/350 = 6.0% for grade 3-4. In addition, a number of grade 3-4 AEs that are not mentioned are significantly more common in the Len/Dex arm compared to Dex:

- Nausea 7/353 = 2.0% vs. 2/350 = 0.6%
- Abdominal pain NOS 5/353 = 1.4% vs. 1/350 = 0.3%
- Insomnia 7/353 = 1.1% vs. 1/350 = 0.3%

Serious Adverse Events

Serious AEs were reported in Celgene submission (Table 30) according to the EPAR (EMEA scientific discussion p 27). The incidence of serious AEs was fairly high in the trial, with individuals receiving Len/Dex being more likely than those on Dex alone to experience at least one serious (57.2% vs. 46.6%).

Celgene report that *neutropenia and thrombocytopenia were the primary reasons for dose reductions in the Len/Dex groups, and the frequency of discontinuation was low – for MM-009, neutropenia (2.4%; 4/170) and thrombocytopenia (0.6%; 1/170) and for MM-010, neutropenia or thrombocytopenia (0.6% and 0.6%, respectively).*

Pooled analyses

The manufacturer report results from the pooled analysis of safety data from MM-009 and MM-010. Results are reproduced in Table 16.

Table 16. Grade ≥3 adverse events occurring in more than 5% of patients

Adverse event, n (%)	*Len/Dex (n=353)*	*Dex (n=351)*
Neutropenia	*125 (35.4)***	*12 (3.4)*
Thrombocytopenia	*46 (13.0)***	*22 (6.3)*
Anaemia	*38 (10.8)**	*21 (6.0)*
Pneumonia	*32 (9.1)*	*19 (5.4)*
All thromboembolic events	*56 (15.9)***	*19 (5.4)*
Hyperglycaemia	*27 (7.6)*	*27 (7.7)*
Fatigue	*23 (6.5)*	*17 (4.9)*
Muscle weakness	*20 (5.7)*	*11 (3.1)*
Hypokalaemia	*20 (5.7)*	*5 (1.4)*
Asthenia	*17 (4.8)*	*18 (5.1)*
** p<0.001; ** p<0.05.*		

Source: Celgene Submission Table 32

The results confirmed that thromboembolic events were significantly higher in patients treated with Len/Dex in the absence of a prophylactic use of an anticoagulant.

4.2.2 Critique of submitted evidence synthesis

One of the weaknesses of the clinical effectiveness evidence is that there is no direct trial-based comparison between Len and the primary comparators defined in the scope, therefore the submission relies on indirect comparison. Additionally the only evidence found for comparator treatments are non RCTs.

Surprisingly, despite time to treatment failure (TTF) being an endpoint of MM-010 and MM-009 trials and more importantly, being one of the inputs used in the cost- effectiveness analysis, the effectiveness summary in the manufacturer submission does not present this outcome.

According to MM-009 and MM-010 CSRs, TTF is defined as the time from randomization to treatment failure. In addition to counting progressions and deaths as events and calculating the time to event as for PFS, subjects who withdrew from the study for any reason or who received antimyeloma therapy during the treatment period were counted as having events on the day of last adequate assessment prior to withdrawal from the study or prior to receiving antimyeloma medication.

TTF was reported in both CSRs for the ITT population up to study unblinding and given its relevance to the economic analysis, should have been included in the clinical evidence synthesis reported by Celgene.

4.2.3 Summary of clinical effectiveness

It is likely that the systematic review on the literature for clinical effectiveness undertaken by Celgene contains all the relevant studies. The evidence submitted generally reflect the decision problem outlined in the final scope of the submission.

Len/Dex effectiveness relied on evidence drawn from the two identically designed, good quality RCTs (MM-009 and MM-010). The efficacy evaluation of the two individual trials and results from a pooled analysis showed increased TTP, PFS and OS with Len/Dex compared to Dex alone.

The manufacturer could not identified RCT that examine the retreatment with Bort or other comparator treatments used as second line and evidence submitted comes from the Taverna (2012) study for Bort and Damaj (2012) for bendamustine and other comparators.

5.0 Economic evaluation

In this chapter, we assess the cost-effectiveness analysis submitted by Celgene. The ERG found several important logical errors in the economic model first sent to us by Celgene. On the 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 (Table 17). As a result of this, Celgene submitted an updated analysis[7], with the goal to address the problems identified in the original submission. This section discusses the final model version.

We start with a summary of the systemic review of cost-effectiveness studies presented by Celgene and the methods used in the economic evaluation (Section 5.1). Then we present a critique of the methods they used (Section 5.2). This is followed by a description of Celgene's results (Section 5.3) and our comment on their validity (Section 5.4).

Table 17. Most important errors in versions of the economic model sent to ERG by Celgene.

Model version	Original model	Updated model
Date received by ERG	13th November 2013	19th December 2013
ICER (Len/Dex vs. Bort/Dex)	£14,535	**Dominant**
Errors found in the model	Overall survival (OS) curves crossed progression-free survival (PFS) and time to treatment failure (TTF) curves in both the intervention and the comparator arms of the model.	OS curves still cross PFS and TTF curves in the comparator arm of the model.
	The hazard ratio (HR) for PFS was not adjusted for patient's characteristics in the Taverna (2012) study.	The HR was adjusted for patients characteristics and the updated estimate was 0.9, favouring Bort.
	There were inconstancies in the model structure between the intervention and the comparator arms of the model. Also some structural problems were found in the evaluation of third and fourth line treatment options.	These problems were still found in the updated model.

Source: Table produced by the ERG

[7] Hereafter referred to as "updated model"

5.1 Overview of manufacturer's economic evaluation

5.1.1 Summary of Celgene's systematic review of cost-effectiveness studies

- #### Description of manufacturer's search strategy and comment on whether the search strategy was appropriate

The ERG are happy to accept Celgene's cost-effectiveness searches. The search terms and databases used were appropriate to the task.

The cost effectiveness search syntax took the following form:

(Terms for Population) AND (Intervention terms + terms for comparators) AND (Terms for cost effectiveness) AND (Limit to Human only populations)

The following bibliographic databases were used:

- MEDLINE & MEDLINE In-Process (OVID);
- EMBASE (OVID);
- NHS EEDS (The Cochrane Library);
- Econlit (OVID); and,
- CINAHL (EBSCO HOST).

The following conference proceedings were searched:

- American Society of Hematology (ASH) 2011-2012;
- American Society of Clinical Oncology (ASCO) 2011-2012; and,
- European Hematology Association (EHA) 2011-2012.

- #### Search results

A range of studies were identified and their relevance assessed according to the inclusion/exclusion criteria described in Table 37 in Celgene's submission.

While at the primary review the majority of the studies failed to meet the inclusion criteria, at the secondary review, no studies met the inclusion criteria. Therefore, no relevant cost-effectiveness studies were found. For this reason, a de-novo analysis was undertaken.

5.1.2 Celgene's economic model submitted to NICE

We now turn to the economic evaluation that Celgene presented to NICE. Celgene report costs per QALY estimates for Len plus Dex versus Bort in MM patients. Different treatment options were considered for third and fourth line treatments.

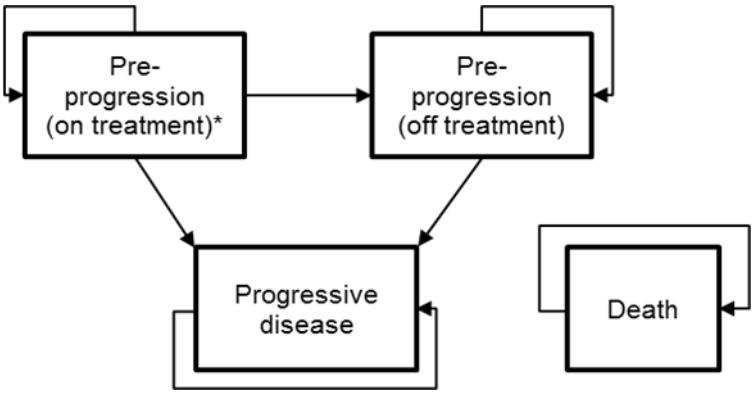
The model was built in Microsoft Excel©. Here, we summarise the main features of the model. In general, we found some significant problems with the model structure. These issues are discussed in the following sections of this report.

Celgene cost-effectiveness analysis relied greatly on the extrapolation of MM-010 survival data. It is the ERG opinion that the data extrapolation process has several serious flaws, which makes us question the validity of the final ICERs presented. We are particularly concerned with the process used to model overall survival in the economic analysis. The appropriateness of the methods employed by Celgene is explored in Section 5.2.3.

- ### Model structure

Celgene's cost-utility model was developed in a partitioned survival structure. The structure of the model, illustrated in Figure 10, is argued to be appropriate and reflective of the clinical pathway of MM.

Figure 10. Celgene's model structure



Source: Figure taken from submission Figure 21.

The model describes four health states:

- Pre-progression on treatment (PFS-T): all patients enter the model in the PFS-T state. All patients are assumed to have been treated with Bort at first-line and are on their second-line treatment.
- Pre-progression off treatment (PFS-OT): pre-progression patients can also be off second-line therapy.
- Progressive disease (DP): this health state captures disease progression.
- Death: this is the absorbing state of the model.

All patients enter the model in the PFS-T state. The patient population consists of adults with MM for whom thalidomide is contraindicated and whose disease has progressed after at least 1

prior treatment with Bort. All patients are assumed to be on second-line treatment in this health state.

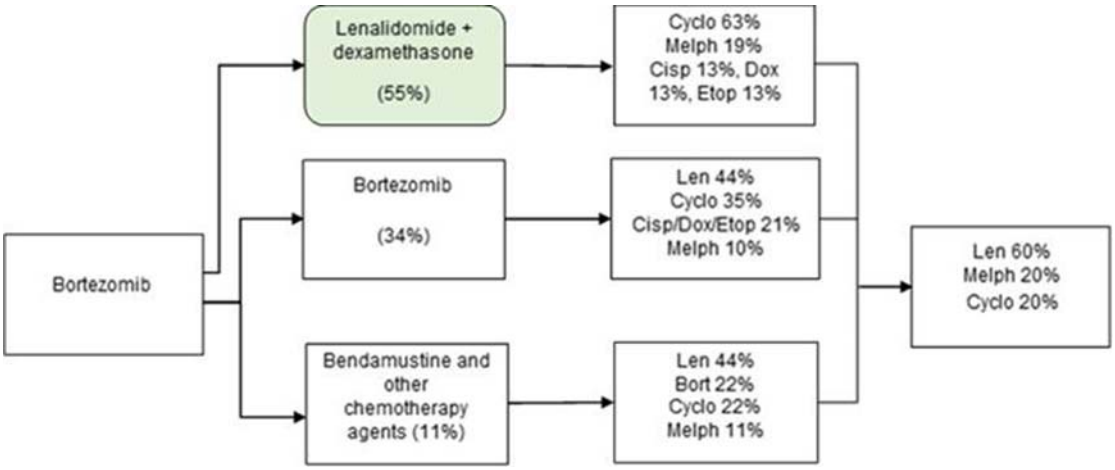
Patients can then stop second-line treatment but still be in the pre-progression state. For these patients the disease has not progressed yet but the treatment failed due to study withdrawal, for example.

Patients can also move from the pre-progression state (either from PFS-T or PFS-OT) to the PD state. This is the case for all patients whose disease has progressed. These patients are also assumed to have stopped receiving second-line treatment. Patients can die while in the PFS-T, PFS-OT or in the PD states.

Once patients are off second-line treatment, a range of treatment options are considered. These are mentioned in the submission as “real-world” treatment options and are claimed to have been taken from Haematological Malignancy Research Network (HMRN) data.

Celgene show the clinical pathway followed by patients in the model, represented in Figure 11. However the “real-world” treatment basket in the model does not exactly match the pathway represented below. This is explored in Section 5.2.2.

Figure 11.Treatment pathway conisdered in the economic model



Source: Celgene Submission Figure 22

The cycle length in the economic model is 28 days and a half-cycle correction was applied.

The time horizon considered in the economic model was 25 years (1300 weeks).

- ### Treatment effectiveness within submission

It is stated that the main aims of MM therapy are to prolong survival and maintain a good quality of life by controlling the disease and relieving symptoms. MM is characterised by a sequence of

relapses where treatments eventually cease to be effective and the patient ends up suffering disease progression and respective decrement in quality of life.

Treatment effectiveness within the model works essentially through transition probabilities between the health states presented in the previous section.

In the Len/Dex arm of the model, transition probabilities between health states were derived from survival functions based on MM-010 patient-level data. Data from MM-009 trial were not used as it is stated that MM-010 data are more suited to use in the economic model. The reasons provided are:

- This trial has a European patient population and is therefore the most relevant to the decision problem.
- Pooling results from separate studies is not appropriate as this breaks randomisation and as data are only available from single arms of trials no meta-analysis, indirect or mixed treatment comparison is possible.
- The results of MM-009 and MM-010 are comparable.

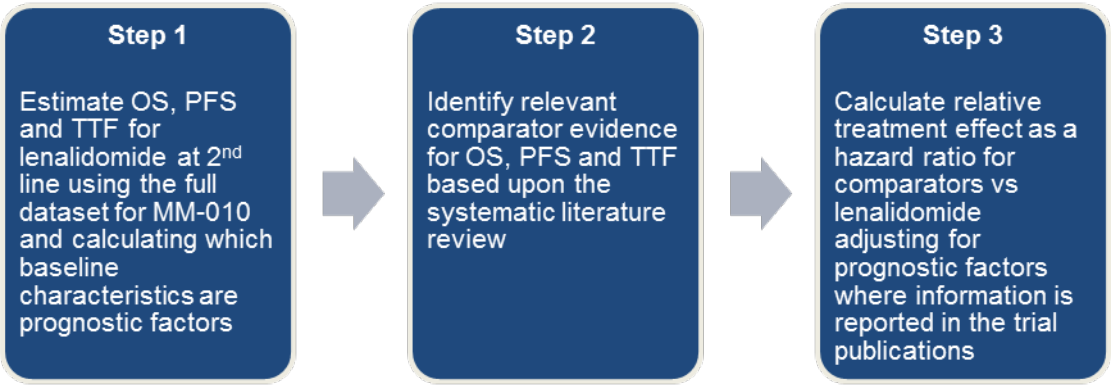
The key survival data used from MM-010 were overall survival (OS), progression-free survival (PFS) and time to treatment failure (TTF).

In the comparator arm of the model, transition probabilities between health states were computed by using the transition probabilities from the intervention arm of the model and applying a hazard ratio (HR) for each comparator.

For the base case comparator, Bort, Taverna (2012) data was used to compute all HR estimates for each outcome. For all other model comparators, HRs were obtained from Damaj (2012). Both are retrospective studies.

Figure 12 was taken form Celgene submission and it is claimed to represent the process used to model OS, PFS and TTF for Len/Dex vs. comparator treatments.

Figure 12. Modelling process for OS, PFS and TTF



Source: Submission Figure 23.

We now describe how transition probabilities between health states were estimated within the different arms of the economic model:

Len/Dex arm

It is stated in the submission that multivariate equations were developed to model OS, PFS and TTF for Len. MM-010 data were used to fit the multivariate parametric curves.

The justification provided for using multivariate equations is that these allow parametric curves to be adjusted to reflect differences in baseline population characteristics (i.e. risk factors), which was deemed essential in the submission since:

- It allowed adjustment of model inputs for different baseline characteristics across MM-010, and population characteristics available in the studies used to model the comparators effectiveness (for example Taverna, 2012).
- It allowed the use of all relevant information from MM-010 due to the small sample sizes available for a second-line population (available comparator information did not provide subgroup analysis by line of treatment and sample sizes were limited within the MM-009 and MM-010 trials)

A range of baseline risk factors were considered to be potentially prognostic of clinical outcomes in MM-010. After this hypothesis was tested for each risk factor, the latter was included (or excluded) from the multivariate parametric models accordingly. The risk factors considered are summarized as follows:

- Age (years)
- Sex
- Disease stage (I, II or III)
- Number of prior anti-myeloma regimens
- Time since diagnosis of multiple myeloma
- ECOG performance score (0,1,2+)
- Beta-2 microglobulin count (mg/L)
- Number of prior stem cell transplants
- Presence or absence of bone lesions

It is stated that variables were initially reviewed for missing data.

A separate category (unknown) was created for categorical variables with missing estimates. Binary and categorical variables were reviewed to confirm whether the existing categorisations were satisfactory and to ensure there were sufficient numbers of patients in each group to permit analysis. Continuous variables were reviewed to confirm whether they showed evidence of a linear relationship with the clinical outcome of interest. An initial set of variables were identified for each endpoint using backwards stepwise elimination (using $p<0.05$), cross

validated using forwards stepwise selection (using p<0.1). Selected variables were analysed for evidence of collinearity using a review of variance inflation factor (VIF) statistics.

If collinearity was evident (VIF>10), only the variable that showed the strongest relationship with the outcome and greatest face validity was retained. Cox-Snell residuals were examined to assess the final model goodness of fit.

Celgene explain that the effect of each independent categorical variable was assessed using proportional hazard (PH) and accelerated failure time (AFT) assumptions. Alternative approaches would be considered if a standard parametric approach provided a poor fit to observed data, or given evidence of PH or AFT violation for the variables included in the multivariable regression model.

It is claimed that Beta-2 microglobulin count, time since diagnosis of multiple myeloma, number of prior therapies, baseline presence or absence of bone lesions, and ECOG performance score were all found to be significant predictors of PFS, TTF and OS in MM-010.

However, Celgene decided to exclude the duration of multiple myeloma and number of prior therapies from all models. The reasons provided for this decision are that multiple myeloma duration was centred on the mean duration (4.5 years in MM-010), therefore the baseline hazard predicts survival for this duration and that the population of interest is treated in the second-line setting. As such these two terms were not required to model the population of relevance.

Once significant predictors were identified, PFS, TTF and OS curves seem to have been adjusted using the mean of covariates method, in which average values of covariates (like for example the beta-2 microglobulin count and presence of bone lesions for the baseline MM-10 population) are entered into a proportional hazards regression equation.

After employing the multivariate equations used to model OS, PFS and TTF, the resulting survival estimates were used in different manners to compute transition probabilities between health states. These are briefly described below and are further explored in Section 2.2.3.

We initially focus on the transition probabilities used to model second-line treatment and then briefly describe the process employed for deriving transition probabilities for subsequent treatment lines.

<u>Pre-progression on treatment to pre-progression off treatment (PFS-T to PFS-OT) – second-line</u>

Patients in the PFS-T health state are those for whom the disease has not progressed and who are still on Len/Dex treatment. This condition is captured on one hand by progression-free survival (PFS[8]) individual-level data in MM-010, which defines disease progression, and on the

[8] Defined in Section 4

other hand, by time to treatment failure (TTF[8]) individual-level data in MM-010, which defines treatment continuation/failure.

Patients in the PFS-OT heath state are those for whom the disease has not progressed but are not on Len/Dex treatment anymore (for example due to study withdrawal). As before, this condition is captured by both PFS and TTF individual-level data in MM-010.

It is therefore crucial how PFS and TTF were extrapolated in the economic model:

- Progression-free survival

A log-logistic distribution was used to fit the MM-010 PFS data in order to extrapolate the study results to a 25 year horizon. Figure 13 was taken from the original submission and it shows the KM PFS curve for Len/Dex as well as the fitted PFS curve.

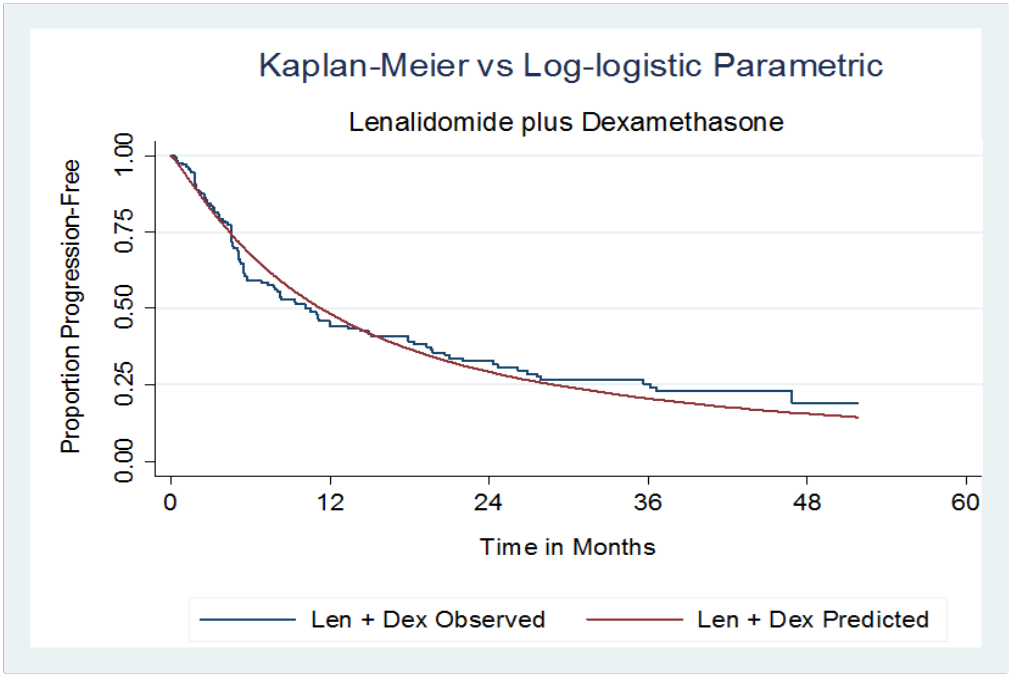
Table 18 was provided to the ERG by Celgene and it presents the regression results for the prediction of PFS with a log-logistic model, where beta-2 microglobulin count was found to be the only significant predictor of PFS.

- Time to treatment failure

Similarly to PFS, a log-logistic distribution was used to fit the MM-010 TTF data in order to extrapolate the study results to a 25 year horizon. Figure 14 was taken from the original submission and it shows the KM TTF curve for Len/Dex as well as the fitted TTF curve.

Table 19 was taken from the original submission and it presents the regression results for the prediction of TTF with a log-logistic model where, similarly to PFS, beta-2 microglobulin count was found to be the only significant predictor of TTF.

Figure 13. KM plot and fitted log-logistic model for PFS



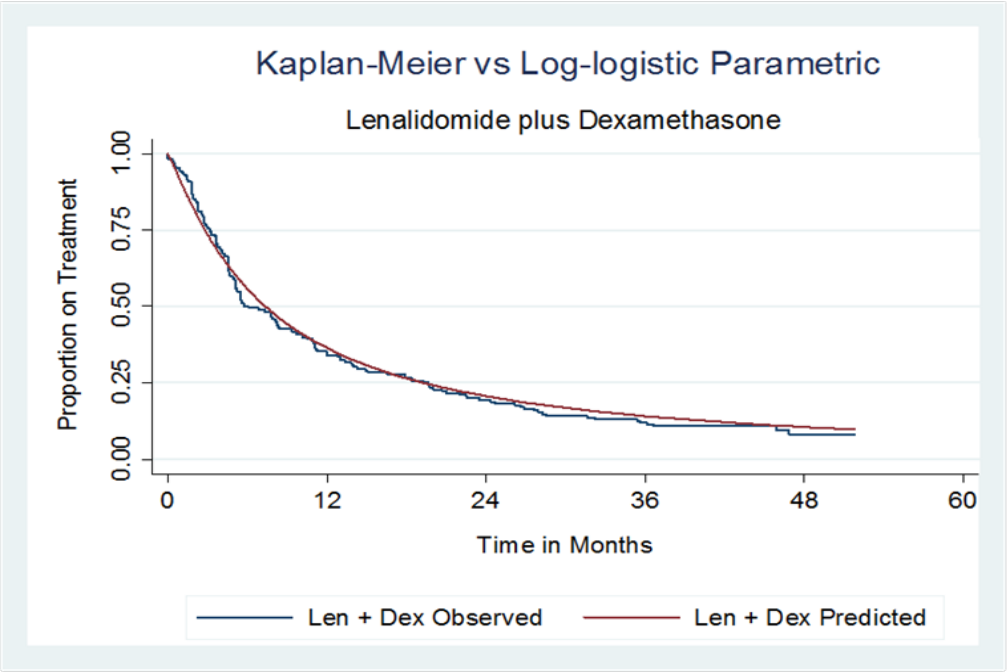
Source: Submission Figure 25

Table 18. Regression results for PFS with log-logistic distribution

Variable	Coefficient	Standard Error	P value
Constant term	3.116	0.184	0.000
Beta 2M count ≥2mg/L	-0.779	0.143	0.000
Ln (gamma)	-0.218	0.080	0.007

Source: Adapted from Clarification request response 12-12-13 letter –response to B7

Figure 14. KM plot and fitted log-logistic model for TTF



Source: Celgene Submission Figure 26

Table 19. Regression results for TTF with log-logistic distribution

Variable	Coefficient	Standard Error	P value
Constant term	2.678	0.169	0.000
Beta 2M count ≥2mg/L	-0.731	0.130	0.000
Ln (gamma)	-0.153	0.067	0.022

Source: Adapted from Clarification request response 12-12-13 letter –response to B7

<u>Pre-progression on treatment to progressive disease (PFS-T to PD) - second-line</u>

As mentioned above, patients in the PFS-T health state are those for whom the disease has not progressed and who are still on Len/Dex treatment. This condition is captured on one hand by PFS data from MM-010, which defines disease progression, and on the other hand, by TTF data from MM-010, which defines treatment continuation/failure.

Patients in the PD state are those for whom the disease has already progressed. These patients are assumed to be off second-line treatment. Therefore, this condition can be captured by PFS data alone.

Pre-progression off treatment to progressive disease (PFS-OT to PD) - second-line

As previously mentioned, patients in the PFS-OT heath state are those form whom the disease has not progressed but are not on Len/Dex treatment anymore. As before, this condition is captured by both PFS and TTF data from MM-010.

Patients in the PD state are those for whom the disease has already progressed.

Death - second-line

Mortality in the model is captured by overall survival (OS[8]) individual-level data in MM-010.

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. Figure 15 was taken from the original submission and it shows the KM OS curve for Len/Dex as well as the fitted OS curve.

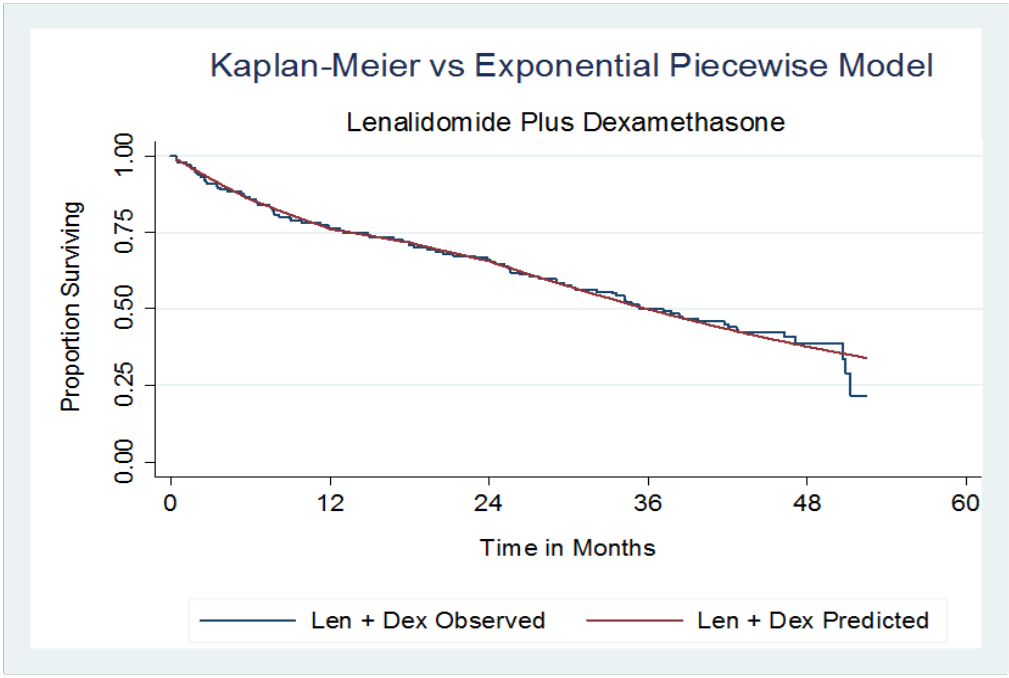
The piecewise exponential model with survival time split into 6 months intervals was considered to be the best approach to deal with the presence of the found PH violation.

However, in the updated submission Celgene changed the distribution used to fit the MM-010 OS data to a log-logistic. This decision was in response to the issues raised by the ERG in the initial submission regarding OS curves (fitted with the exponential piecewise model) being found to cross PFS and TTF curves in the original model. This is explored in detail in Section 5.2.3.

Table 20 was provided to the ERG by Celgene and it presents the regression results for the prediction of OS with the log-logistic distribution. We can observe that only beta-2 microglobulin count, ECOG score = 2, 3 and the presence of bone lesions were found to be statistically significant.

In the updated submission, Celgene did not provide the KM OS curve with the log-logistic fitted OS curve for visual inspection. However the ERG have produced these curves and present them, together with a discussion on the appropriateness of the fit, in Section 5.2.3.

Figure 15. KM plot and fitted exponential piecewise model for OS



Source: Celgene Submission Figure 28

Table 20. Regression results for OS with log-logistic distribution

Variable	Coefficient	Standard Error	P-value
Beta-2m >2mg/L	-0.824	0.209	0.000
ECOG score = 1	-0.183	0.197	0.352
ECOG score = 2	-0.898	0.275	0.001
ECOG score unknown	-0.014	0.765	0.985
Bone lesions present	-0.549	0.230	0.017
Constant	5.124	0.337	0.000
Ln (sigma)	-0.138	0.088	0.116

Source: Celgene clarification document to the ERG

Subsequent treatment lines

Celgene claim that following second and third relapse, a “real-world” treatment basket (Figure 11) is used to take into account all relevant costs over the lifetime of MM patients. Hence third and fourth-line treatment options are defined by a combination of different drugs.

It should be made clear however, that subsequent treatment options are not evaluated on their effectiveness. Only costs of third and fourth-line treatment baskets are considered.

This means that once patients fail Len/Dex at second-line, they move on to a third-line treatment received for a fixed period of time of 4 cycles. After the 4 cycles all patients are assumed to move to a fourth-line treatment, which will end after 4 cycles as well. Mortality is said to be accounted for in the fixed length of the subsequent treatment duration.

While on third and fourth-line treatment, patients are assumed to always be on the PD state, therefore experiencing the associated utility.

Bort arm and other comparators

It is mentioned in the submission that there is a lack of good quality published evidence for second-line patients who have been previously treated with Bort and subsequently received any of the comparators considered (Bort, bendamustine and other chemo agents).

Taverna (2012 and Damaj (2012) were the relevant studies identified hence median OS and PFS outcomes were taken from these sources and used to calculate HRs for each comparator relative to Len/Dex in MM-010. Celgene compared the median survival estimates (OS and PFS) 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.

Table 21 shows the sources used to model effectiveness in the comparator arms of the model. Sources in bold were used for the base case analysis, while the others were used in scenario analysis. The ERG found some problems with the PFS HR originally presented by Celgene (presented in Section 5.2.3) thus the value provided in Table 21 is the updated one.

Table 21. Sources used to model HRs

Treatment	Variable	Evidence source	Hazard ratio
Bortezomib retreatment	OS	**Taverna 2012**	**1.70**
		White 2013	1.42
	PFS	**Taverna 2012**	**0.90***
		White 2013	1.76
		Hrusovsky 2010	1.09
		Dispenzieri 2010	1.28
		Petrucci 2013	1.26
		Min 2007	0.84
Bendamustine (and chemotherapy agents)	OS	Damaj 2012	3.00
	PFS	Damaj 2012	1.09

*estimate from the updated model

Source: Submission Table 47

The estimated HRs were then applied to the transition probabilities used in the Len/Dex arm of the model. This was done by exponentiating the Len/Dex transition probability to the HR in each cycle of the economic model.

As explained before, the multivariate parametric models designed to predict PFS and OS in the Len/Dex arm of the model (explained above) 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).

Finally, it was assumed that the HR for TTF would be the same as for PFS. The reasons provided for this were that the two survival endpoints are similar and that no information on TTF was presented in the evidence found for the comparators.

Transition probabilities between health states in the comparator arm of the model were thus taken from the intervention arm. The difference is that the transition probabilities in the comparator model were estimated by applying an HR to the Len/Dex probabilities.

Subsequent treatment lines

It is stated in the submission that in the base case analysis, patients on the comparator model arm receive Len/Dex as third-line treatment, following the discontinuation of Bort. Hence these patients are exposed to the OS, PFS and TTF hazards associated with Len/Dex at second-line (described above and explored in detail in Section 5.2.2).

Celgene claim this to be a conservative assumption as it implies that the second-line effectiveness of Len/Dex is replicated at the third line of treatment, despite Celgene's statement that Len/Dex is shown to be more effective at earlier lines.

To model fourth-line treatment, the "real-world" treatment basket (presented in Figure 11) is claimed to be used. The mean fourth-line treatment duration is 16.8 weeks. As explained for Len/Dex, the fourth-line treatment option for Bort only evaluates costs and excludes the effectiveness of the treatment basket from the analysis.

- ### Health related quality of life

As health-related quality of life data (HRQoL) data were not collected as part of the MM-009 and MM-010 trials, the manufacturer carried out a systematic review to identify health state utility value.

Celgene ran two literature searches for HRQoL. The first search omitted key search terms and, on the basis of the scoping under-taken by the ERG, the ERG asked Celgene to clarify why these terms had been omitted. In response to this question, Celgene undertook new searches.

Celgene's new searches used a HRQoL search filter (which includes the HRQoL terms the ERG felt were missing previously) and the ERG is content to accept these second searches as the

primary searches, noting that a new paper was identified which report the assessment of the validity of mapping QLQ-C30 onto EQ-5D: Crott (2013)

The search identified seven studies, which are reported in Appendix K. On p159 of the submission, the manufacturer state that: *one additional paper was added as a result of reference searching. This gave eight studies that met all the inclusion criteria after both primary and secondary filtering. These studies are reviewed below.* However, only seven studies are reported in Celgene Submission Table 50.

Celgene also claim that *three of these studies directly measure utilities*, however amongst the seven studies reported in Table 50 of the submission, only two are identified as primary study were utility values were directly calculated from QoL analysis: Khanna (2006) and van Agthoven (2004). The reason for not reporting the third primary study i.e. Goss (2006) in this table is not clear.

Although it is not clearly stated by Celgene, the ERG assumed that the additional study added as a result of reference searching is the van Agthoven (2004) study, which was used in the five secondary utility studies.

In the absence of relevant utility data identified in patients who have received at least 1 prior treatment, the manufacturer used the same utilities as within the model submitted for TA171, which were taken from a cost-utility carried out by van Agthoven (2004) in patents with previously untreated multiple myeloma. Utility values used are presented in Table 22.

Table 22. Health states utility values

Variable	***Value***	***CI (distribution)***	***Justification***
Utility value: pre-progression	*0.810*	*0.63 to 0.94 (beta*)*	*van Agthoven et al., 2004*
Utility value: pre-progression after 2 years	*0.770*	*0.60 to 0.90 (beta*)*	*van Agthoven et al., 2004*
Utility value: post-progression	*0.640*	*0.51 to 0.76 (beta*)*	*van Agthoven et al., 2004*

Source: Adapted from Celgene Submission Table 51

The utility values change over time for the pre-progression health state and patients in the pre-progression state for longer than 2 years were assumed to have a reduced utility weight, from 0.81 to 0.77.

The additional paper found with the new search reports utility values obtained by mapping EORTC QLQ-C30 questionnaire responses into EQ-5D. Those utilities are obtained from the same study by The Dutch-Belgian Haemato-Oncology Cooperative Study Group (HOVON), as

in the van Agthoven (2004) paper, except that patients received intensive chemotherapy followed by myeloablative therapy with autologous stem-cell rescue, typically given to young and fit patients.

The utility value associated with each health state is then multiplied by the length of time spent in that state. The total QALYs over the lifetime of a patient were calculated by aggregating the estimated QALYs from each health state.

- ## Adverse events

The manufacturer claim to have incorporated the same grade 3 and 4 treatment related AEs as for TA171. Therefore the following AEs were modelled: anaemia, constipation, diarrhoea, deep-vein thrombosis, hypercalcaemia, neutropenia, peripheral neuropathy, pneumonia and thrombocytopenia.

Celgene claim that the event rates, as shown in Table 23, for the Len/Dex are derived from MM-010, with the total number of events being divided by the total number of patient years on treatment. However the ERG was not able to trace back the number of events associated each of the AEs in the CSRs.

The AE rates fort Bort and all other comparator arms are claim to be taken from NICE TA228 for the bortezomib plus melphalan plus prednisolone/prednisone (VMP) and melphalan plus prednisolone/prednisone (MP) arms respectively, in the absence of evidence elsewhere for other comparators.

Table 23. AE rates applied in the economic analysis

Adverse event	*Grade*	*Annual rate (cycle rate)*		
		Len/Dex	*Bortezomib*	*Other comparators*
Anaemia	*3*	*8.3% (0.6%)*	*17.7% (1.3%)*	*26.2% (2.0%)*
	4	*0.5% (<0.1%)*	*3.0% (0.2%)*	*10.3% (0.8%)*
Constipation	*3*	*1.4% (0.1%)*	*0.7% (0.1%)*	*0.0% (0.0%)*
	4	*0.0% (0.0%)*	*0.0% (0.0%)*	*0.0% (0.0%)*
Diarrhoea	*3*	*2.3% (0.2%)*	*7.7% (0.6%)*	*0.8% (0.1%)*
	4	*0.0% (0.0%)*	*0.7% (0.1%)*	*0.0% (0.0%)*
Deep vein thrombosis	*3*	*3.2% (0.5%)*	*1.0% (0.1%)*	*0.8% (0.1%)*
	4	*0.5% (<0.1%)*	*0.0% (0.0%)*	*0.0% (0.0%)*
Hypercalcaemia	*3*	*0.5% (<0.1%)*	*0.0% (0.0%)*	*0.0% (0.0%)*
	4	*0.0% (0.0%)*	*0.0% (0.0%)*	*0.0% (0.0%)*
Neutropenia	*3*	*59.3% (4.4%)*	*34.0% (2.6%)*	*31.4% (2.4%)*
	4	*5.1% (0.4%)*	*11.3% (0.9%)*	*19.5% (1.5%)*
Peripheral neuropathy	*3*	*1.4% (0.1%)*	*14.3% (1.1%)*	*0.0% (0.0%)*
	4	*0.0% (0.0%)*	*0.3% (<0.1%)*	*0.0% (0.0%)*
Pneumonia	*3*	*4.6% (0.4%)*	*5.3% (0.4%)*	*5.2% (0.4%)*

	4	*0.5% (<0.1%)*	*2.0% (0.2%)*	*1.6% (0.1%)*
Thrombocytopenia	*3*	*11.0% (0.8%)*	*22.7% (1.7%)*	*21.8% (1.7%)*
	4	*1.4% (0.1%)*	*19.4% (1.5%)*	*18.7% (1.4%)*

Source: Celgene Submission Table 48

Using AE rates, the manufacturer then calculated the weighted average decrement per cycle for each treatment arm, associated with AEs to apply to patients on treatment at every model cycle.

Patients on the comparator arm go on to receive Len/Dex following treatment discontinuation and therefore are subjected to the AE rates associated with Len/Dex, as in Table 23. Celgene claim that due to the paucity of evidence and lack of impact of AEs on the ICER, the AEs are not modelled after treatment discontinuation for other subsequent therapies.

Utility decrements for AEs are included in the model. These are based on a paper by Brown (2013), a cost-effectiveness analysis of Len/Dex vs. Dex, and then applied on a per-cycle basis as one utility decrement for patients on treatment, or for patients on Len/Dex as a third-line treatment. Those decrements are presented in Table 24.

Table 24. AEs utility decrements

Variable	***Value***	***CI (distribution)***	***Justification***
Anaemia	*0.310*	*0.196 to 0.437 (beta**)*	*Only decrement obtained through systematic search. Brown et al, 2013*
Hypercalcaemia	*0.000*	*Not included in SA*	*Only decrement obtained through systematic search. Brown et al, 2013*
Pneumonia	*0.190*	*0.121 to 0.270 (beta**)*	*Only decrement obtained through systematic search. Brown et al, 2013*
Thrombocytopenia	*0.310*	*0.196 to 0.437 (beta**)*	*Only decrement obtained through systematic search. Brown et al, 2013*
Neutropenia	*0.145*	*0.093 to 0.206 (beta**)*	*Only decrement obtained through systematic search. Brown et al, 2013*
Diarrhoea	*0.000*	*Not included in SA*	*Only decrement obtained through systematic search. Brown et al, 2013*
Constipation	*0.000*	*Not included in SA*	*Only decrement obtained through systematic search. Brown et al, 2013*
Peripheral neuropathy	*0.065*	*0.042 to 0.093 (beta**)*	*Only decrement obtained through systematic search. Coffey et al, 2002*[132]
Deep vein thrombosis	*0.150*	*0.096 to 0.213 (beta**)*	*Only decrement obtained through systematic search. Brown et al, 2013*

Source: Adapted from Celgene Submission Table 51

The manufacturer modelled costs associated with the AEs listed above. *The unit cost of treating adverse events depends on the setting in which it is treated. Four possible settings are included in the model: inpatient, hospital day case, outpatient, and general practice. The unit cost of*

treatment in primary care is that of a GP visit as presented in Curtis (2012). Other costs are obtained from NHS reference costs for 2011/12. These costs are presented in Table 25.

Table 25. Costs of AEs included in the economic model

Adverse Event	***Unit Cost by Setting***			
	Inpatient	***Day case***	***Outpatient***	***Primary care***
Anaemia	*£489.05*	*£372.16*	*£159.56*	*£36.00*
NHS reference cost code (2011/12)	*NEI_S SA04D: Iron deficiency anaemia with CC*	*DC SA04D: Iron deficiency anaemia with CC*	*Total OPATT 303: Clinical haematology*	*Curtis 2012. GP 11.7 minute contact, including direct care staff costs, excluding qualification costs.*
Constipation	*£496.00*	*£375.00*	*£128.00*	*£36.00*
NHS reference cost code (2011/12)	*NEI_S PA26B: Other Gastrointestinal Disorders without CC*	*DCRA: Other Gastrointestinal Disorders without CC*	*Total OPATT 301: Gastroenterology*	*Curtis 2012, as above.*
Diarrhoea	*£496.00*	*£375.00*	*£128.00*	*£36.00*
NHS reference cost code (2011/12)	*Same as constipation*	*Same as constipation*	*Same as constipation*	*Curtis 2012, as above*
Deep vein thrombosis	*£463.00*	*£132.00*	*£150.32*	*£36.00*
NHS reference cost code (2011/12)	*NEI_SQZ20Z: Deep vein thrombosis*	*DC QZ20Z: Deep vein thrombosis*	*Total OPATT 300: General medicine*	*Curtis 2012, as above*
Hypercalcaemia	*£598.19*	*£400.36*	*£159.56*	*£36.00*
NHS reference cost code (2011/12)	*NEI_S SA08D: Other Haematological or Splenic Disorders, with CC*	*DC SA08D: Other Haematological or Splenic Disorders, with CC*	*Total OPATT 303: Clinical haematology*	*Curtis 2012, as above*
Neutropenia	*£598.19*	*£400.36*	*£159.56*	*£36.00*
NHS reference cost code (2011/12)	*Same as hypercalcaemia*	*Same as hypercalcaemia*	*Same as hypercalcaemia*	*Curtis 2012, as above*
Peripheral neuropathy	*£555.92*	*£272.79*	*£150.32*	*£36.00*
NHS reference cost code (2011/12)	*NEI_S WA21W: Other Procedures or Health Care Problems, with CC*	*DC WA21W: Other Procedures or Health Care Problems, with CC*	*Total OPATT 300: General medicine*	*Curtis 2012, as above*
Pneumonia	*£1,274.51*	*£480.69*	*£142.80*	*£36.00*
NHS reference cost code	*NEI_L DZ19A: Other Respiratory*	*DC DZ19A: Other Respiratory*	*Total OPATT 340:*	*Curtis 2012, as above*

Adverse Event	***Unit Cost by Setting***			
	Inpatient	***Day case***	***Outpatient***	***Primary care***
(2011/12)	*Diagnoses with Major CC*	*Diagnoses with Major CC*	*Respiratory medicine*	
Thrombocytopenia	*£538.63*	*£400.18*	*£159.56*	*£36.00*
NHS reference cost code (2011/12)	*NEI_S: Thrombocytopenia with CC*	*DC SA12D: Thrombocytopenia with CC*	*Total OPATT 303: Clinical haematology*	*Curtis 2012, as above*

Source: Celgene Submission Appendix M Figure 110

The unit cost of treating an event in each setting is multiplied by the proportion of events treated in each setting, obtained from NICE technology appraisal 171. The resulting weighted average cost is multiplied by the overall proportion of events actively treated.

Table 26. Proportions of AEs included in the economic model

Adverse Event	Grade	Proportion actively treated	Proportion treated by setting			
			In-patient	Day case	Out-patient	Primary care
Anaemia	3	91.9%	5.7%	73.2%	15.4%	5.7%
	4	100.0%	19.6%	69.6%	5.4%	5.4%
Constipation	3	100.0%	37.5%	21.4%	35.4%	5.7%
	4	100.0%	100.0%	0.0%	0.0%	0.0%
Diarrhoea	3	95.7%	57.5%	12.5%	28.6%	1.4%
	4	100.0%	100.0%	0.0%	0.0%	0.0%
Deep vein thrombosis	3	100.0%	12.9%	16.1%	68.9%	2.1%
	4	100.0%	81.2%	3.5%	15.4%	0.0%
Hypercalcaemia	3	100.0%	50.4%	27.5%	22.1%	0.0%
	4	100.0%	77.5%	11.8%	10.7%	0.0%
Neutropenia	3	44.1%	5.0%	55.6%	39.4%	0.0%
	4	70.7%	12.3%	40.4%	43.5%	3.9%
Peripheral neuropathy	3	79.3%	0.0%	4.6%	94.6%	0.8%
	4	83.9%	9.1%	15.5%	71.8%	3.6%
Pneumonia	3	100.0%	98.6%	1.4%	0.0%	0.0%
	4	100.0%	100.0%	0.0%	0.0%	0.0%
Thrombocytopenia	3	28.9%	6.2%	81.5%	12.3%	0.0%
	4	96.4%	17.1%	80.0%	2.1%	0.7%
Adverse event treatment inputs were obtained from the Evidence Review Group report as part of the TA171 appraisal process.[78]						

Source: Celgene Submission Appendix M Figure 111

The total cost per cycles associated with AEs was calculated by multiplying the weighted average costs by the AE rates and estimated at £17.11, £29.26 and £29.74 per cycle on treatment for Len/Dex, Bort and other comparators, respectively.

- ## Resources and costs

The model submitted by Celgene used costs based on the NHS & PSS perspective. Costs included in the model are drug costs and disease management costs (such as monitoring costs and outpatient visits).

Estimates of resource use were obtained from literature searches and previous guidelines. Resource use in the model was dependent on whether the patient had experienced disease progression or not. The model heath states are claimed to have been costed in a similar manner to NICE TA171 and TA228 and the respective costs are presented in Table 27.

AEs were costed using NHS reference costs and are addressed in the AEs subsection of this report.

Other miscellaneous costs were considered in the economic analysis. This included the cost of terminal care.

Table 27. Health state costs

Health states	***Items***	***Cost per cycle***
Pre-progression (typically on treatment)	*Technology*	*Lenalidomide: £3,773* *Dexamethasone (cycles 1-4): £7.76* *Dexamethasone (cycles 5+): £2.59* *Bortezomib: £4,067.30*
	Concomitant G-CSF and administration	*With lenalidomide: £473.62*
	Monitoring and tests	*£153.34*
	Administration	*Lenalidomide: £161.85 in first cycle only* *Bortezomib: £1,065.76*
	Transport	*Lenalidomide: £6.39 in first cycle only* *Bortezomib: £17.04*
	Adverse events	*Lenalidomide: £17.11* *Bortezomib: £29.26*
Post-progression	*3rd line treatment*	*Following lenalidomide: Therapy: £70.20* *IV administration: £69.63* *Transport: £3.06* *Following bortezomib: Therapy: £1,716.99* *IV administration: £49.45* *Transport: £2.20*
	4th line treatment	*Therapy: £2,277.28* *IV administration: £0.00* *Transport : £0.00*
	Monitoring and tests	*£175.86*

	Adverse events	*In receipt of 3rd line lenalidomide: £17.11* *Otherwise: £0.00*
	Terminal care	*£1,235 on death*

Source: Adapted from submission Table 61.

It should be noted that following the request from the ERG, Celgene updated the cost of Bort in the model, so it would reflect the fact that 64% of patients in the comparator arm are also taking Dex. Therefore the cost of Dex was included as part of a scenario analysis.

Len acquisition costs

The cost of Len was calculated as a weighted average of daily doses across all patient days in the MM-010 study. As per study protocol, treatment with Len could be interrupted and the dose regimen could also be reduced.

Table 28 shows the proportion of patients days spent on each Len dose per model cycle. The cost of Len per cycle was estimated based on these.

It is stated in the submission that the dose used for G-CSF given together with the Len treatment regimen is not considered in the Len SPC at any specific dosing level, but that G-CSG was used concomitantly with Len in MM-010. Therefore an assumption was made on the drug dose administered as well as on the duration of G-CSF treatment.

It was assumed that all patients went to receive 25mg of Len with concomitant G-CSF after their first dose interruption. A 300µg vial of G-CSF was considered as the daily dose, applied for 1 week in patients who required concomitant G-CSF. This dosing regimen was claimed to be obtained from a Celgene UK Physician Survey conducted 2011-2012.

Table 28. Proportion of patients receiving Len and G-CSF per model cycle (based on MM-010).

Daily dose of lenalidomide	***Proportion of patient days spent in receipt of this dose***
25mg	*43.2%*
25mg + G-CSF	*26.8%*
15mg	*10.5%*
10mg	*5.0%*
5mg	*3.1%*
0mg (interruption)	*11.4%*

Source: Adapted from submission Table 54

The unit cost for Len was obtained from the British National Formulary (BNF) and is presented in Table 29.

The resulting weighed cost of Len per cycle is £3,773.

Table 29. Len and unit cost

Therapy	*Standard unit*	*Source*	*Unit price*
Lenalidomide	*21 tab pack: 25mg*	*BNF*	*£4368.00*
	21 tab pack: 15mg		*£3969.00*
	21 tab pack: 10mg		*£3780.00*
	21 tab pack: 5mg		*£3570.00*
	Dose interruption		*£0.00*

Source: Adapted from submission Table 56

The cost of a 300 µg vial of G-CSF is £52.70 (BNF).

Dex acquisition costs

Len arm

The cost of Dex is calculated as the cost per milligram. It is stated that this approach was applied to simplify cost calculations and that given the low price of Dex this was unlikely to have an impact on study results.

The cost of Dex per pack was taken from the Department of Health Electronic Market Information Tool (eMit) and is presented in Table 30.

Table 30. Unit cost of Dex

Therapy	*Standard unit*	*Source*	*Unit price*
Dexamethasone	*500 tab pack: 2mg*	*eMit*	*£11.97*
	100 tab pack: 2mg		*£3.23*
	50 tab pack: 2mg		*£1.80*

Source: Adapted from submission Table 56

A weighted average cost per milligram was estimated, providing a cost per milligram of £0.02.

No information was provided about the dose of Dex used in MM-010. In the excel model, a dose of 38.6mg per day was assumed, therefore resulting in a £0.65 cost per day.

Bort arm

The ERG requested that Celgene included the cost of Dex in the Bort arm of the model, since 64% of the study population in Taverna (2012) used concomitant Dex. In their updated model, the cost of Dex was thus included in the Bort arm as a scenario analysis.

Again, no information was provided about the dose of Dex used in Taverna (2012). In the excel model, a dose of 38.6mg per day was assumed, therefore resulting in a £0.65 cost per day. It seems like the same dose regimen assumed for Dex in MM-010 was used to model the cost of Dex in the Bort arm.

Bort acquisition costs

For Bort, only the cost per 3.5mg vial is presented. This was taken from the BNF and is reported to be £762.38.

No information is provided about the dose administered in the Taverna (2012) study and how the final cost of Bort per cycle was estimated. This is explored in Section 5.2.2.

It is mentioned how a patient access scheme (PAS) is modelled by undertaking scenario analysis.

Table 31. Other comparators acquisition costs

Therapy	*Standard unit*	*Source*	*Unit price*
Bendamustine	*100mg vial*	*BNF 65*	*£275.81*
	25mg vial		*£69.45*
Melphalan	*50mg vial*	*BNF 65*	*£129.81*
	25 tab pack: 2mg		*£42.88*
Prednisone	*100 tab pack: 50mg*	*BNF 65*	*£89.00*
Prednisolone	*28 tab pack: 5mg*	*eMit*	*£0.30*
	28 tab pack: 1mg		*£0.16*
Cyclophosphamide	*1g vial*	*BNF 65*	*£17.60*
	500mg vial		*£9.20*
	100 tab pack: 50mg		*£20.20*
Cisplatin	*100mg vial*	*eMit*	*£14.10*
	50mg vial		*£7.16*
Doxorubicin	*200mg vial*	*eMit*	*£32.38*
	50mg vial		*£4.87*
	10mg vial		*£1.73*
Vincristine	*2mg vial*	*eMit*	*£8.49*
	1mg vial		*£3.42*
Etoposide	*500mg vial*	*eMit*	*£73.29*
	100mg vial		*£22.38*

Source: Adapted from submission Table 56

Treatment administration costs

Treatment administration costs were obtained from NHS reference costs. These included possible transportation costs to the hospital and are presented in Table 32.

Table 32. Treatment administration costs

Item	Len/Dex	Bort/other comparators	Assumptions
Treatment administration	£161.85 once	£199.83 every administration appointment	While the cost of administration for Len/Dex was assumed to occur only for the first appointment (after which it is assumed that the patient self-administers oral treatment), the administration cost of Bort was applied for every administration appointment in the hospital.
Transportation to the hospital	£12.78	£12.78	This assumed that 50% of patients will 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 to receive full treatment.

Source: Adapted from submission Table 53

Monitoring costs

It is stated in the submission that monitoring frequency depends primarily on whether the patient has experienced disease progression or not. Additionally Len treatment is associated with an increased monitoring requirement during the initial treatment phase.

It is not clear in the submission which costs have been assumed to be related with the disease state (i.e. progression or progression-free) or with the initial monitoring phase associated with Len treatment.

It is stated how monitoring costs by health state have been taken from previous NICE TAs 171 and 228 however, the list of monitoring testes shown in TA171 seems to be much more extensive than the one reported in Celgene's submission.

The initial Len monitoring regimen is claimed to be taken form the SPC.

This is further explored in Section 5.2.2.

Subsequent treatment costs

Len/Dex arm

Celgene claim that following second and third relapse, a “real-world” treatment basket (Figure 11) is used to take into account all relevant costs over the lifetime of MM patients. Hence third and fourth-line treatment options are defined by a combination of different drugs.

However, the treatment basket shown in Figure 11 does not seem to be reflective of the treatment combinations used as third-line treatment in the economic model.

Instead, the treatment combination used in the economic model to cost third-line treatment options is shown in Table 33 which was adapted by the ERG from Table 57 in the submission.

The fourth-line treatment basket in the comparator arm is the one presented in Figure 11.

Table 33. Third-line therapy mix in the economic model

Drug	Base case 3rd line treatment (Len/Dex arm)	Base case 3rd line treatment (Bort arm)	Scenario analysis 3rd line treatment (Bort arm)
Bortezomib	0.0%	0.0%	0.0%
Dexamethasone	56.3%	0.0%	31.3%
Melphalan	18.8%	0.0%	10.4%
Cyclophosphamide	62.5%	0.0%	34.7%
Cisplatin	12.5%	0.0%	6.9%
Doxorubicin	12.5%	0.0%	6.9%
Etoposide	31.3%	0.0%	17.4%
Prednisolone	6.3%	0.0%	3.5%
Prednisone	6.3%	0.0%	3.5%
Lenalidomide	0.0%	100.0%	44.4%

Source: Adapted from submission, Table 57.

The third-line therapy mix provided in Table 33 is then used to estimate a weighted average cost for the third-line treatment. Similarly, weighted average transportation and administration costs were also computed. The total cost per cycle is presented in Table 34 and it is based on the unit costs presented previously in this subsection.

The same approach was followed for estimating fourth-line treatment costs. The final cost per cycle are presented in Table 35.

Table 34. Third-line treatment cost per cycle

Cost item	***Cost per cycle***			
	Len/Dex arm	***Base case comparator arm***	***Following bortezomib***	***Other 2nd line comparator***
Drug cost	*£70.20*	*£3,772.88*	*£1,716.99*	*£2,592.00*
IV administration	*£68.63*	*£0.00*	*£49.45*	*£203.00*
Transport	*£3.06*	*£0.00*	*£2.20*	*£3.25*

Source: Submission, Table 58.

Table 35. Fourth-line treatment cost per cycle

Cost Item	***Cost per cycle***
Drug cost	*£2,277.28*
IV administration	*£0.00*
Transport	*£0.00*

Source: Submission, Table 60.

Bort arm

It is stated in the submission that for the base case analysis, patients on the comparator model arm receive Len/Dex as third-line treatment, following the discontinuation of Bort. Hence these patients are exposed to the costs associated with Len/Dex at second-line (described in this section).

To model fourth-line treatment, the “real-world” treatment basket (presented in Figure 11) is claimed to be used. The mean fourth-line treatment duration is 16.8 weeks and the costs are the same as the ones presented in Table 35.

Miscellaneous costs

The cost of terminal care was also estimated. In order to calculate this, it was assumed that 20% of MM patients will likely need end-of-life care. This was then applied to a unit cost of hospice care (£6,177) for 8 weeks. The unit cost was reported to have been taken from a King’s Fund report.

- ### Discounting

All costs and health benefits were discounted at a 3.5% rate as recommended by NICE.

- ### Sensitivity analysis

Deterministic and probabilistic sensitivity analysis (PSA) were undertaken by Celgene. The outputs are reported in Section 5.3 of this report. A list was provided with the model parameters included in sensitivity analysis. Distributions used to run PSA were also reported in the submission (Table 62).

- ### Model validation

It is stated by Celgene that the model was checked for internal quality at the company who built the economic model, however the ERG discovered several important logical errors in the economic model first sent to us by Celgene.

For external validation, the OS estimated by the economic model was validated against the results in TA228 and registry data. Cost data was also compared to costs estimates from a different source (Bruce et al, 1999). The trial population was also compared with registry data.

It is also mentioned in the submission that expert opinion was sought by Celgene to validate the treatment pathway, resource use and terminal care costs.

5.2 Critique of approach used

In this section, we comment on Celgene approach and methodology. First, we consider the model against checklists of good practice. Then we critically appraise the model structure and data as well as the methods used in the cost effectiveness analysis.

The primary focus of the critique is on the second-line treatment for MM. Second-line treatment in the analysis compares Len/Dex with Bort. Considerations are also made for subsequent treatment options.

5.2.1 Critical appraisal frameworks

Celgene's economic analysis was assessed against three widely used study quality checklists for economic models:

- NICE Reference Case (NICE, 2008).
- Drummond assessment criteria (Drummond et al., 1997).
- Criteria for decision model-based economic evaluations (Philips et al., 2006).

Table 36. Critical appraisal checklist based on NICE Reference Case (NICE, 2008)

NICE reference case requirement		Critical appraisal	Reviewer comment
Defining the decision problem	The scope developed by NICE	✓	
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	?	In the original submission the manufacturer considered the base case comparator to be Bort. However in the trial used to inform the economic analysis, 64.3% of patients received concomitant Dex. Other comparators were specified in the initial scope, however, these were not included in the base case analysis but instead took part in the scenario analysis. This was the case for bendamustine and chemotherapy agents (including regimens based on melphalan, vincristine, cyclophosphamide and doxorubicin). It is not clear to the ERG why bendamustine was not included in the base case analysis since data were available to model the cost-effectiveness of Len/Dex compared with bendamustine. It seems to have been appropriate to include this comparator in the base case analysis.
Perspective on costs	NHS and PSS	✓	NHS & PSS
Perspective on outcomes	All health effects on individuals	✓	
Type of economic evaluation	Cost-effectiveness analysis	✓	
Synthesis of evidence on outcomes	Based on a systematic review	?	Based primarily on single trial (RFHE3001) evidence
Measure of health benefits	QALYs	✓	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	✓	EQ- 5D survey

NICE reference case requirement		Critical appraisal	Reviewer comment
Source of preference data for valuation of changes in HRQL	Representative sample of the public	✓	EQ- 5D survey
Discount rate	3.5% pa for costs and health effects	✓	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓	

Note: ✓ indicates 'clear'; X indicates 'concerns'; ? indicates 'some concerns'.

Table 37. Critical appraisal checklist from Drummond and colleagues (Drummond et al. 1997)

Item	Critical appraisal	Reviewer comment
Is there a well-defined question?	•	
Is there a clear description of alternatives (i.e. who did what to whom, where, and how often?)	•	
Has the correct patient group/population of interest been clearly stated?	?	There are some differences between the trial population and the typically presenting UK population in terms of: • Number of prior stem cell transplants • Number of prior antimyeloma therapy
Is the correct comparator used?	?	In the updated model, 64.3% of Bort patients received concomitant Dex. It would have been appropriate to include bendamustine in the base case analysis since it is a relevant comparator and data were available to conduct the economic analysis.
Is the study type reasonable?	✓	A Markov structure for the cost-utility analysis was used
Is the perspective of the analysis clearly stated?	✓	UK NHS PSS
Is the perspective employed appropriate?	✓	NHS Reference Costs
Is the effectiveness of the intervention established?	✓	Quality of MM-010 is good.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	✓	A 25-year time horizon was used. After 25 years, virtually 90% of patients modelled are dead.
Are the costs and consequences consistent with the perspective employed?	✓	All costs are presented from the UK NHS & PSS perspective
Is differential timing considered?	✓	All future costs and benefits are discounted with a 3.5% rate.
Is incremental analysis performed?	✓	
Is sensitivity analysis undertaken and presented clearly?	✓	Deterministic and probabilistic sensitivity analyses are reported.

Note: ✓ indicates ‘clear’; X indicates ‘concerns’; ? indicates ‘some concerns’.

Source:

Table 38. Critical appraisal checklist of Philips et al (2004) for model-based analysis

Dimension of quality		Critical appraisal	Comments
Structure			
S1	Statement of decision problem/objective	✓	Len/Dex versus Bort (64.3% of Bort patients received concomitant Dex) for MM patients who have received at least one prior therapy with bortezomib
S2	Statement of scope/perspective	✓	NHS & PSS perspective was implemented. Cost and benefit inputs were consistent with this. Scope of the model stated.
S3	Rationale for structure	✓	A cost-utility model using a state transition Markov approach is appropriate.
S4	Structural assumptions	X	Generally, the ERG are not convinced by some of the structural assumptions. These are explored in Section 5 of this report.
S5	Strategies / comparators	?	It is not clear to the ERG why bendamustine was not included in the base case analysis since data were available to model the cost-effectiveness of Len/Dex compared with bendamustine.
S6	Model type	✓	A Markov model is appropriate in this case.
S7	Time horizon	✓	A 25-year time horizon was used. After 25 years, virtually 90% of patients modelled are dead.
S8	Disease states / pathways	✓	The health states used are: Pre-progression (on treatment), pre-progression (off treatment), post-progression and death. These are appropriate to capture disease progression over time.
S9	Cycle length	✓	Cycle length is 28 days .This is appropriate to capture disease progression and treatment regimens.
Data			
D1	Data identification	?	Data identification methods were generally well described. For Len, overall survival (OS), progression-free survival (PFS) and time to treatment failure (TTF) data were taken from MM-010 however patient level data was not provided. The effectiveness of Bort and bendamustine (as well as other chemo agents) was modelled with data from Taverna (2012) and Damaj (2012) respectively. These are retrospective studies which do not provide complete information of patients' characteristics and treatment regimens administered.
D2	Pre-model data analysis	?	More details on the extrapolation method and transition probabilities calculation could have been provided.

Dimension of quality		Critical appraisal	Comments
D2a	Baseline data	?	Baseline data from MM-010I, which is appropriate. Taverna (2012) baseline data was used to adjust PS, OS and TTF curves however not much detail was provided on patient characteristics.
D2b	Treatment effects	X	Base case relative treatment effect was estimated with survival analysis. The ERG do not feel confident in the estimation of OS and PFS effectiveness in the model.
D2c	Quality of life weights (utilities)		HRQoL was not recorded in MM-010. Utilities were the same utilities as within the model submitted for TA171, which were taken from a cost-utility carried out by van Agthoven (2004).
D3	Data incorporation	?	Data inputted in the model is generally poorly referenced in the submission.
D4	Assessment of uncertainty	?	A PSA is presented but the results always report dominant ICERs.
D4a	Methodological	?	Results always report dominant ICERs.
D4b	Structural	?	Results always report dominant ICERs.
D4c	Heterogeneity	?	Subgroup analysis was not clearly reported and the ERG is not clear if/how this was conducted.
D4d	Parameter	?	Results always report dominant ICERs.
Consistency			
C1	Internal consistency	X	Even though Celgene claim to have sought validation for the excel model, the model contained several serious logical errors.
C2	External consistency	✓	Only expert opinion was sought for external validation.

Note: ✓ indicates 'clear'; X indicates 'concerns'; ? indicates 'some concerns'.

5.2.2 Critique of the modelling approach and structure

The ERG found several problems in the original and updated model structures.

Original submission

Upon receipt of the original submission, the ERG pointed to an overall lack of consistency between the calculations in the intervention and the comparator arms of the economic model. It also noted some initial concerns with regards to third and fourth-line treatment calculations.

Manufacturer's approach

The manufacturer revised the calculations related with third and-fourth line treatment options. It was stated that minor amendments were performed.

Furthermore Celgene claimed to have made the excel flow sheets consistent across intervention and comparator arms of the model. However, to the best of the ERG knowledge, no changes in the model structure were undertaken.

ERG critique of the updated model

Having revised the updated economic model, the ERG still found some structural problems. More specifically, the ERG noted again the previously found inconstancies in the model structure across treatment and comparator arms and also a structural problem with the evaluation of third and fourth line-treatment options. These are discussed below.

Figure 16 is a simplification of the model structure presented in the previous section and it focus only on the second-line treatment option, therefore comparing Len/Dex with Bort as second line drugs. Death is also a possible heath state (the absorbing one) but hasn't been included in the diagrams below for simplification purposes. The model structure for the intervention and the comparator arms is presented separately.

The use of the PFS-T state as starting point in both arms of the model is appropriate for the disease pathway. Patients can then progress (PD), in which case they stop the second line drug or they can stop treatment but still be in the PFS state. This seems sensible considering disease progression.

In the intervention arm of the model patients can go to the PFS-OT and the PD health states and accrue the corresponding costs and QALYs, and then move to the third-line treatment. The economic analysis of subsequent treatments only evaluated costs and not drug effectiveness.

However, in the comparator arm of the model, as soon as patients stop treatment (whether in the PD or the PFS-OT state) they are assumed to immediately start a subsequent Len/Dex third-line treatment. Therefore the costs and mortality benefits related to the third-line treatment option (in this case Len/Dex) start accruing in the same cycle. This means that there is no clear separation between second-line treatment outcomes and the beginning of the third-line treatment option and respective outcomes.

To illustrate this with an example, in the same model cycle (28 days) Bort patients can fail second-line treatment, move to a third-line treatment option (in this case Len) and also experience the mortality benefits associated with Len/Dex treatment. This does not seem clinically plausible as it represents a situation where within 28 days, patients who have just stopped Bort treatment can experience the same mortality rate as a Len/Dex patient.

Clinical opinion sought be the ERG informed that after a patient stops Bort treatment he will, on average, be off any kind of MM treatments during 1 or 2 months. More importantly, it takes

Superseded

around 4 to 6 months (on average) until the mortality benefits associated with Len treatment to be visible. Clinical opinion pointed to the fact that the reduction in mortality associated with Len treatment is mainly related with the patient's immune system being able to recover and this process usually takes a few months.

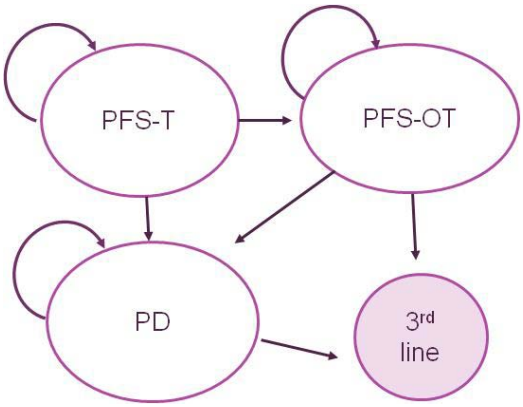
Therefore, it seems that a more reasonable scenario would be to assume that once patients stop Bort treatment they will, for a certain period of time, accrue the outcomes related with Bort therapy (like mortality rates) and then initiate a third-line treatment.

In Figure 17, the ERG present and alternative structure, which we believe to be more accurate and could be adopted in both arms of the economic model. Again, death has been excluded from the diagram for simplification reasons.

He structure presented in Figure 17 would allow for the calculation of a second, third and fourth line treatment ICER, respectively. This is further explored in Section 6.

Figure 16. Simplified structure of Celgene's model structure – intervention and comparator arms

Intervention arm – Len/Dex



Comparator arm - Bort

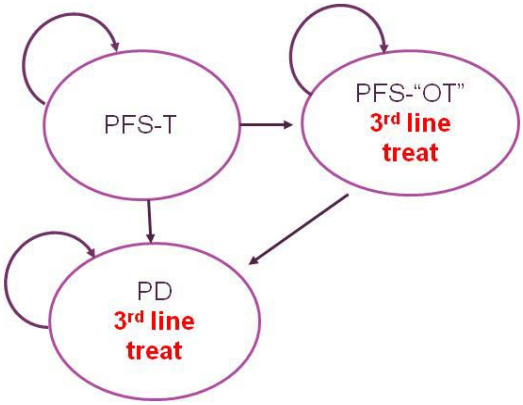
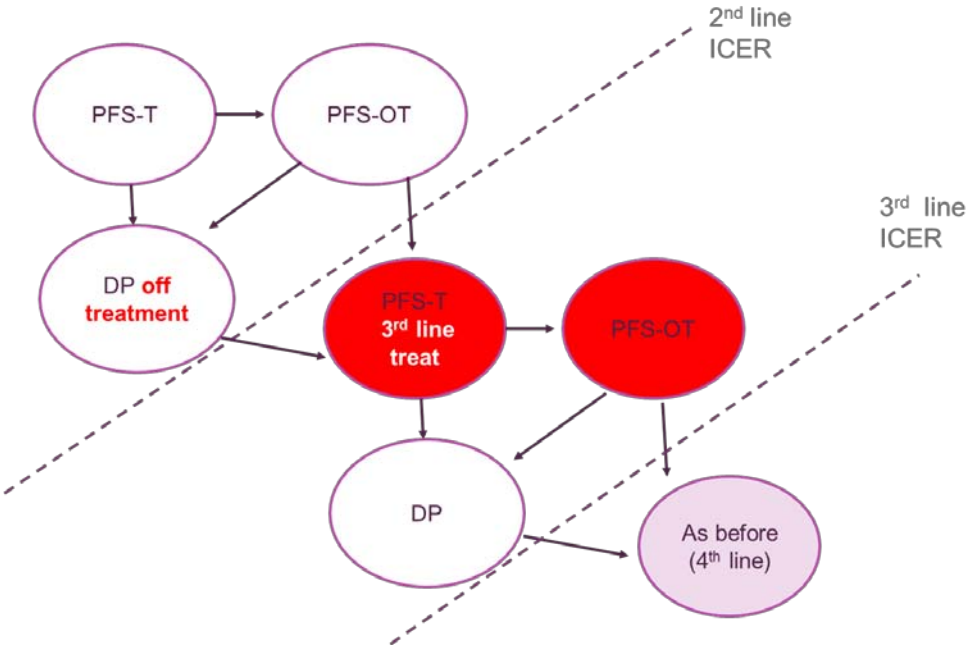


Figure 17. Alternative representation of Celgene's mode



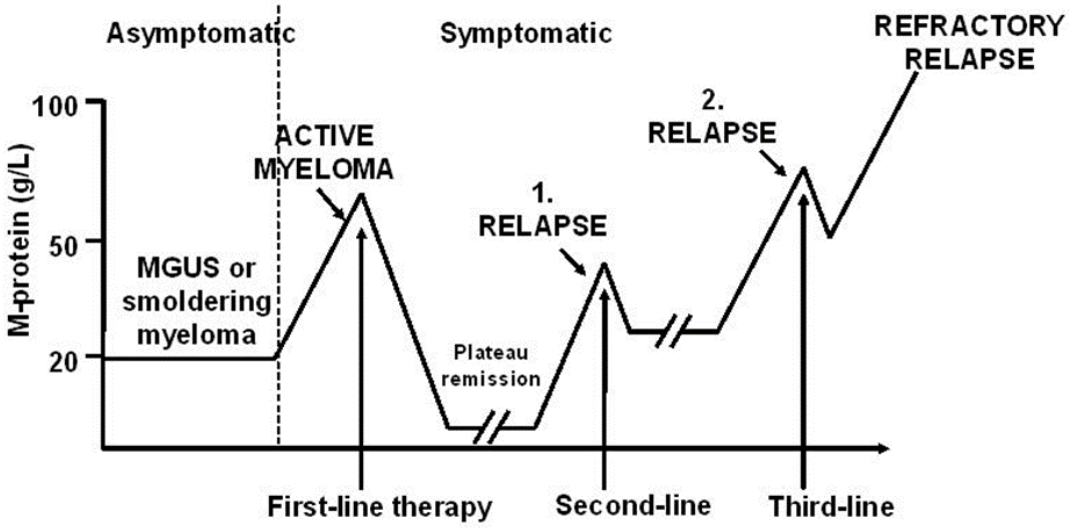
Source: Figure produced by the ERG

It should also be noted that in the manufacturer model, both PFS on and off third-line treatment health states (in red in Figure 17) do not exist. This means that once patients are off the second-line treatment drug they are assumed to always be in 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 perception is reinforced by Figure 18 taken from the original submission, which shows that after a relapse and the beginning of a subsequent line of treatment, patients will achieve the remission state for a while before relapsing again.

The ERG understand that this modelling alternative would require additional effectiveness data to understand TTF and PFS when patients are on the "real-world" treatment baskets.

Figure 18. Disease progression in MM



Source: Submission, Figure 1

In conclusion, for subsequent treatment lines, the manufacturer only consider the QALYs associated with the PD state and the costs of the mixed treatment options. This is further explored in the next subsection.

A 25-year time horizon was used in the model. The time horizon seems reasonable to capture all relevant outcomes for MM patients entering the model at 63 years of age. However around 11% of patients were still alive in the intervention arm of the model at year 25.

The ERG believe that this is related with a possible overestimation of OS, which is discussed in Section 5.2.3.

Furthermore, given that data for patients receiving Len/Dex is available for a follow-up of only 54 months (4.5 years), by which time over half of patients taking Len/Dex are still alive, the 25-year time horizon represents a very large extrapolation. There is therefore a great deal of uncertainty in the survival times of patients in the model. This introduces considerable uncertainty in the estimates of cost-effectiveness.

Cycle length in the model was 28 days. A half-cycle correction was applied. The ERG are generally satisfied with this.

The manufacturer considered Bort/Dex (or Bort alone) to be the base case comparator. However, it should be noted that in the trial used to inform the economic analysis (MM-010), 64.3% of patients received concomitant Dex. Therefore from a conceptual point of view, the comparator in the economic analysis is Bort with concomitant Dex for 64.3% of patients.

Subsequent treatment lines

Celgene claim that in the Len/Dex arm of the model, following second and third relapse, a "real-world" treatment basket (Figure 11) is used to take into account all relevant costs over the lifetime of MM patients. Hence third and fourth-line treatment options are defined by a combination of different drugs.

However, the treatment basket shown in Figure 11 does not seem to match the third-line treatment combinations used in the intervention arm of the excel model. Instead, the treatment combination used to model third-line treatment options is shown in Table 39, which was adapted by the ERG from Table 57 in the submission.

Patients on the Bort arm of the model received Len/Dex as third-line treatment.

The fourth-line treatment basket both in the comparator and in the intervention arms is presented in Figure 11.

Table 39. Third-line therapy mix in the economic model

Drug	Base case 3rd line treatment (Len/Dex arm)	Base case 3rd line treatment (Bort arm)	Scenario analysis 3rd line treatment (Bort arm)
Bortezomib	0.0%	0.0%	0.0%
Dexamethasone	56.3%	0.0%	31.3%
Melphalan	18.8%	0.0%	10.4%
Cyclophosphamide	62.5%	0.0%	34.7%
Cisplatin	12.5%	0.0%	6.9%
Doxorubicin	12.5%	0.0%	6.9%
Etoposide	31.3%	0.0%	17.4%
Prednisolone	6.3%	0.0%	3.5%
Prednisone	6.3%	0.0%	3.5%
Lenalidomide	0.0%	100.0%	44.4%

Source: Adapted from submission, Table 57.

It should be made clear however, that for the Len/Dex arm, subsequent treatment options are not evaluated on their effectiveness. Only costs of third and fourth-line treatment baskets are considered.

Once patients fail Len/Dex at second-line, they move on to a third-line treatment received for a fixed period of time of 4 cycles. After the 4 cycles all patients are assumed to move to a fourth-line treatment, which will similarly end after 4 cycles. Mortality is said to be accounted for in the fixed length of the subsequent treatment duration.

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.

Celgene claim this to be a conservative assumption as it implies that the second-line effectiveness of Len/Dex is replicated at the third line of treatment. Celgene also argue that Len/Dex is shown to be more effective at earlier lines.

To model fourth-line treatment in the comparator arm, the "real-world" treatment basket (presented in Figure 11) is used. The mean fourth-line treatment duration is 16.8 weeks. The fourth-line treatment option for Bort only evaluates costs and excludes the effectiveness of the treatment basket from the analysis.

Three points in Celgene's approach are worth further discussion:

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. Firstly, this approach departs from the one followed in the intervention arm of the model, where there was no consideration for

the effectiveness of subsequent treatment lines and only costs were evaluated. Secondly, Celgene claim this to be a conservative approach as Len has "*shown to be more effective at earlier lines*". However the analysis presented to justify this assumption suggests otherwise.

Table 40 was taken from the original submission and it presents the baseline characteristics of the population included in the Stadtmauer (2009) study. Celgene mention that this study showed that Len/Dex is more effective at earlier lines. Nonetheless, looking at the patient's baseline characteristics (i.e. before they received the Len/Dex treatment) the median time from diagnosis shows that patients receiving >2 therapies have been sicker for a longer time, which is reasonable. Therefore, this group of patients is sicker at baseline so when they receive the Len/Dex treatment at third line, it might be that the drug is not as effective because the patient baseline heath status is worse rather than the drug being less effective at later stages.

Table 40. Baseline characteristics in Stadtmauer analysis

	Lenalidomide plus dexamethasone		
	1 prior therapy (n=133)	***≥2 prior therapies (n=220)***	***p-value***
Median age, years	*62.1*	*63.1*	*0.34*
Male sex, n (%)	*82 (61.7)*	*128 (58.2)*	*0.58*
Baseline beta-2 microglobulin ≤2.5mg/L, n (%)	*47 (35.3)*	*56 (25.5)*	*0.054*
Baseline beta-2 microglobulin >2.5mg/L, n (%)	*86 (64.7)*	*164 (74.5)*	*0.054*
ECOG score 0–1, n (%)	*119 (89.5)*	*188 (85.5)*	*0.77*
Median time from diagnosis, years (range)	*2.2 (0.4–9.7)*	*4.1 (0.5–15.7)*	*<0.001*
Prior ASCT, n (%)	*89 (66.9)*	*117 (53.2)*	*0.014*
Prior treatment with thalidomide, n (%)	*13 (9.8)*	*114 (51.8)*	*<0.001*
Prior treatment with bortezomib, n (%)	*2 (1.5)*	*25 (11.4)*	*<0.001*

ASCT, autologous stem cell transplantation; ECOG, Eastern Cooperative Oncology Group.

Source: Submission, Table 11.

3. Calculations in the excel model

Overall, the ERG found the calculations in the excel model used to simulate subsequent treatment lines to be confusing. Some mistakes were found in the allocation of the number of patients receiving the drug.

Finally, the sequencing of subsequent treatments is not considered and it is currently unclear to the ERG weather this is important in terms of drug response.

In conclusion, the ERG is overall concerned with the **model structure** used by Celgene. The approach undertaken raises the following concerns:

- There is 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 it also makes the evaluation of a second line ICER impossible. Furthermore, from a clinical point of view, this seems to reflect a very unlikely scenario.
- After second-line of treatment, the manufacturer only consider 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).

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

Finally, upon request from the ERG, Celgene adjusted the Dex costs in the Bort arm of the model to reflect the fact that 64.3% of Bort patients receive concomitant Dex in the Taverna (2012) study. However, this was not considered to be the base case analysis but instead included as a scenario analysis.

5.2.3 Data inputs

• Patient group

The modelled intervention patient group is reflective of the population on the MM-010 study on which the analysis is based.

The modelled population in the comparator arm takes data from Taverna (2012) in the base case analysis.

MM-010 population

As noted before, there are some differences between the trial population and the typically presenting UK population. Clinical opinion sought by the ERG informed that the number of prior stem cell transplants (SCT) verified in MM-010 was higher than what would be expected in typical UK practice. Similarly the distribution of patients across the number of prior anti-myeloma therapies was different from the expected in the UK population.

In MM-010, 46% of the population received 0 prior SCT, while 25% received more than 3 and 13% received 2 prior SCT. The remaining 16% received either 1 prior SCT (12%) or 3 (4%). However the clinical expert opinion is that the highest percentage of the MM population in the UK receives on average 2 SCT (which is the case for only 13% of the MM-010 population). This suggests that the modelled patient group received, on average, less prior SCT than the average MM population in the UK.

In the same fashion, 33% of the MM-010 population received 1 prior anti-myeloma therapy while 67% received 2 or 3. Clinical opinion sought by the ERG informed that the inverse scenario would be expected in current practice (i.e. most patients would be expected to have received 1 prior therapy). This suggests that the modelled patient group received, on average, less prior SCT than the average MM population in the UK.

In addition only 4% of the MM-010 study population had received prior Bort, while 38% of the population had previously received thalidomide. This is not reflective of the population defined in the scope, which is includes adults with MM for whom thalidomide is contradicted and whose disease has progresses after at least one prior treatment with bortezomib.

However, clinical advice sought by the ERG revealed that this is unlikely to affect the effectiveness of Len. This means that, had the majority of patients in the trial received Bort, the effectiveness of Len as a second line drug is expected to be similar.

Taverna (2012) population

Celgene compared the median survival estimates (OS and PFS) between Taverna (2012) and MM-010 and derived a crude approximation of a HR between Len/Dex and Bort. OS and PFS estimates from MM-010 were adjusted to reflect the baseline characteristics of the Taverna (2012) population.

The Taverna (2012) study did not include a detailed description of its population baseline characteristics. It is known that the median number of prior therapies is 2 (range 1-11) and that 31% of the study population had undergone SCT. However no detail is provided on the number of prior SCT.

- ### Clinical effectiveness data

The main source for clinical effectiveness data was 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 ERG have several concerns with the data extrapolation process, which is explored in the next sub-section.

Celgene decided to use only MM-010 to inform the economic model and exclude MM-009 from the analysis. The reasons provided to substantiate this decision are:

- This trial has a European patient population and is therefore the most relevant to the decision problem.
- Pooling results from separate studies is not appropriate as this breaks randomisation and as data are only available from single arms of trials no meta-analysis, indirect or mixed treatment comparison is possible.
- The results of MM-009 and MM-010 are comparable.

The ERG appreciate that the MM-009 population was mainly enrolled from sites in the USA and Canada. Therefore some population characteristics (like ethnicity) are potentially different from the average UK population. MM-010 enrolled patients mainly from Europe hence this typically reflects the UK population in a better fashion.

However, as Celgene explain, the study results are comparable and very similar which suggests that the baseline population characteristics might not be too relevant.

Also, Celgene claim that pooling results in this case would break randomisation. The ERG feel that this might not be a relevant argument in this case. MM-010 and MM-009 compared the use of Len/Dex with the use of Dex alone. As only the Len/Dex arms of MM-010 and MM-009 trials would be pooled and there was no cross over from the Dex arms into the intervention arms of the model, it seems like pooling the data could have been a valid approach in this case.

Therefore the ERG have asked Celgene to explore the impact of using MM-009 data alone and also the pooled data from MM-010 and MM-009. The results are presented in Section 5.3.1.

- ### Data extrapolation process

Throughout the model, treatment effectiveness is mainly represented by the transition probabilities between different heath states (Figure 10). To estimate these in the base case analysis, Celgene used data from MM-010 and Taverna (2012) through different methods.

These are now presented in turn. As before, we initially focus on second-line treatment and then provide some details on subsequent treatment lines.

Len/Dex arm

To obtain transition probabilities across heath states over 25 years in the Len/Dex arm of the model, Celgene extrapolated trial data from study MM-010.

PFS, TTF and OS curves seem to have been 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) are entered into a proportional hazards regression equation.

The use of the mean of covariates method to adjust the PFS, TTF and OS curves to reflect MM-010 population characteristics might potentially be skewing these survival estimates. In fact this method has been criticized for the validity of the resulting estimated curves (Ghali, 2001).

One of the underlying reasons is that using baseline mean characteristics to adjust survival curves might skew the curve if the mean values are also skewed. Other reasons include the assignment of mean covariate values between 0 and 1 to dichotomous variables (for example, gender) which are meaningless at the individual level and the fact that the method calculates the hazard for a hypothetical average individual rather than a population-averaged value. Alternative approaches could have been used by the manufacturer to adjust for baseline characteristics (Ghali, 2001; Bradburn, 2003).

Additionally the choice of relevant predictors of PFS, TTF and OS (like, the beta-2 microglobulin count) is not very transparent in the submission. For OS for example, the p-values for each potential predictor suggest that the ECOG score of 1 is not a statistically significant predictor (see Section 5.1.2 and Table 20). However, in the excel model this is included as a predictor in the multivariate analysis. Also, for PFS and TTF it appears that only a few possible variables were evaluated for their predictive relationship with survival data. All potential predictors (listed in Section 5.1.2) should have been included in the analysis, otherwise a pre-selection will likely bias the analysis.

Furthermore, Celgene decided to exclude the number of prior therapies as a potential outcome predictor from all models. The reason used to substantiate this decision was that "*the population of interest is treated in the second-line setting*".

This is a very surprising argument given that in their initial request for clarification, the ERG asked Celgene to clarify if for the original economic analysis:

1. The full MM-0010 dataset had been used, with resulting outcomes being adjusted with covariate estimates for the second-line setting,
2. or if the dataset used in the analysis had been stratified and so only the second-line treatment population was included in the economic analysis.

Celgene clarified that the second approach had been taken. However, Celgene's explanation for excluding the number of prior therapies as a covariate would only make sense if the first approach had been taken.

<u>Pre-progression on treatment to pre-progression off treatment (PFS-T to PFS-OT) - second-line</u>

Patients in the PFS-T health state are those for whom the disease has not progressed and who are still on Len/Dex treatment. This condition is captured on one hand by progression-free survival (PFS) individual level data in MM-010, which defines disease progression, and on the other hand, by time to treatment failure (TTF) individual level data in MM-010, which defines treatment continuation/failure.

Superseded

Patients in the PFS-OT heath state are those form whom the disease has not progressed but are not on Len/Dex treatment anymore (for example due to study withdrawal). As before, this condition is captured by both PFS and TTF individual level data in MM-010.

It is therefore crucial how PFS and TTF were extrapolated:

- Progression-free survival

A log-logistic distribution was used to fit the MM-010 PFS 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. Although these are common steps in the assessment of fit process, they should not be the only ones used (for example, to ensure external validity, the plausibility of the extrapolated portion of the curves should also be assessed).

Even though in the original submission other distributions were used in sensitivity analysis (for example the lognormal distribution), this was no longer the case for the updated model, where only Gompertz and Gamma curves were used in sensitivity analysis due to other reasons.

Furthermore, the ERG have the following concerns with Figure 13, presented in the previous section (reported again below) and taken from the original submission which shows the KM PFS curve for Len/Dex as well as the fitted PFS curve:

1. It is not very informative to show the curves only to the point where the KM curve ends. The time period of the graph should be wide enough so the shape of the fitted curve is observed in the longer term and a judgment can be made of the appropriateness of the fitted curve in estimating PFS. Figure 19 shows the graph produced by the ERG, with a time horizon of 25 years (1300 weeks).
2. The ERG could not replicate Figure 13. In the graph produced by the ERG (Figure 19) the fitted curve does not seem to overlap the KM curve as perfectly as in the graph produced by Celgene. It seems that the fitted curve overestimates PFS from around week 10 to week 80.

Superseded

Reproduction of Figure 13. KM plot and fitted log-logistic model for PFS (Celgene submission)

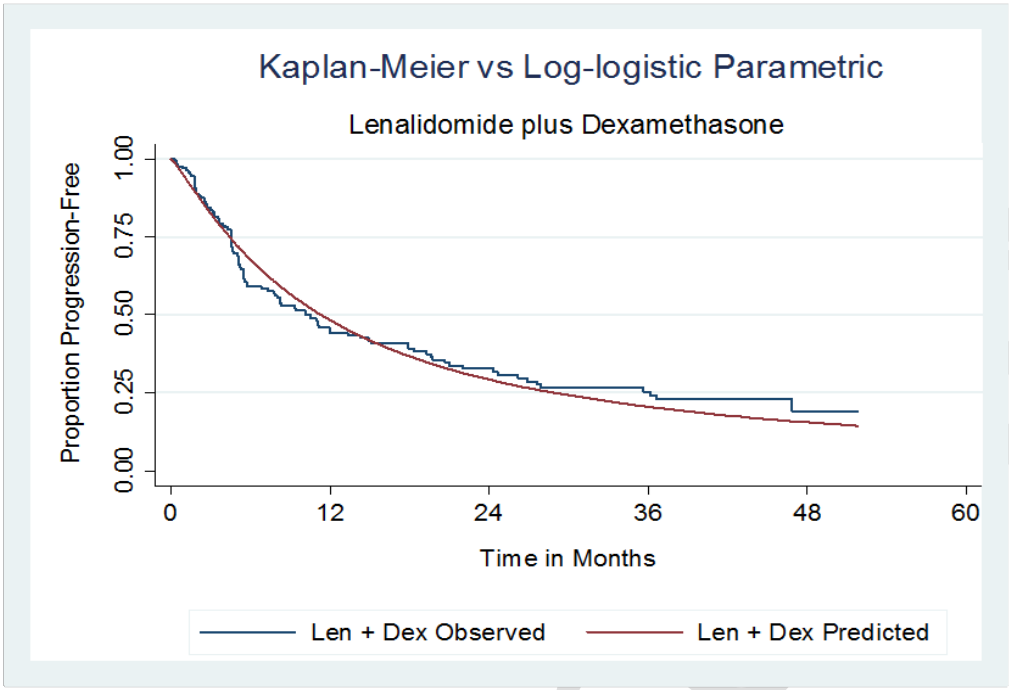
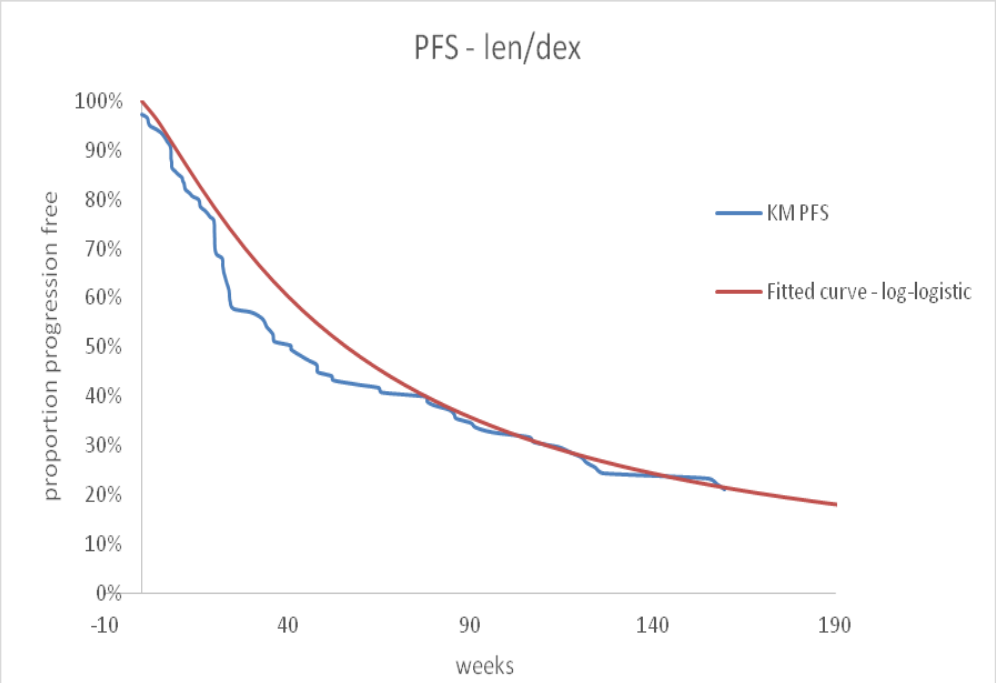


Figure 19. KM plot and fitted log-logistic curve for PFS over 25 years produced by the ERG



Source: produced by the ERG

- Time to treatment failure

Similarly to PFS, a log-logistic distribution was used to fit the MM-010 TTF 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. Again, other steps could have been taken to assess the appropriateness of the distribution used.

As for PFS, other distributions should have been included in the sensitivity analysis. More specifically, the ones that appeared to also be a good fit to MM-010 data (for example the lognormal distribution).

The ERG also identified problems for Figure 14 (reported again below) which was taken directly from the submission and presents the KM TTF curve for Len/Dex as well as the fitted TTF curve:

1. It is not very informative to show the curves only to the point where the KM curve ends. The time period of the graph should be wide enough so the shape of the fitted curve is observed in the longer term and a judgment can be made of the appropriateness of the fitted curve in estimating TTF. Figure 20 shows the graph produced by the ERG, with a time horizon of 25 years (1300 weeks).
2. The ERG could not replicate Figure 14. In the graph produced by the ERG (Figure 20) the fitted curve does not seem to overlap the KM curve as much as in the graph produced by Celgene. It seems that the fitted curve overestimates TTF from around week 30 and onwards.

Superseded

Reproduction of Figure 14. KM plot and fitted log-logistic model for TTF (Celgene submission)

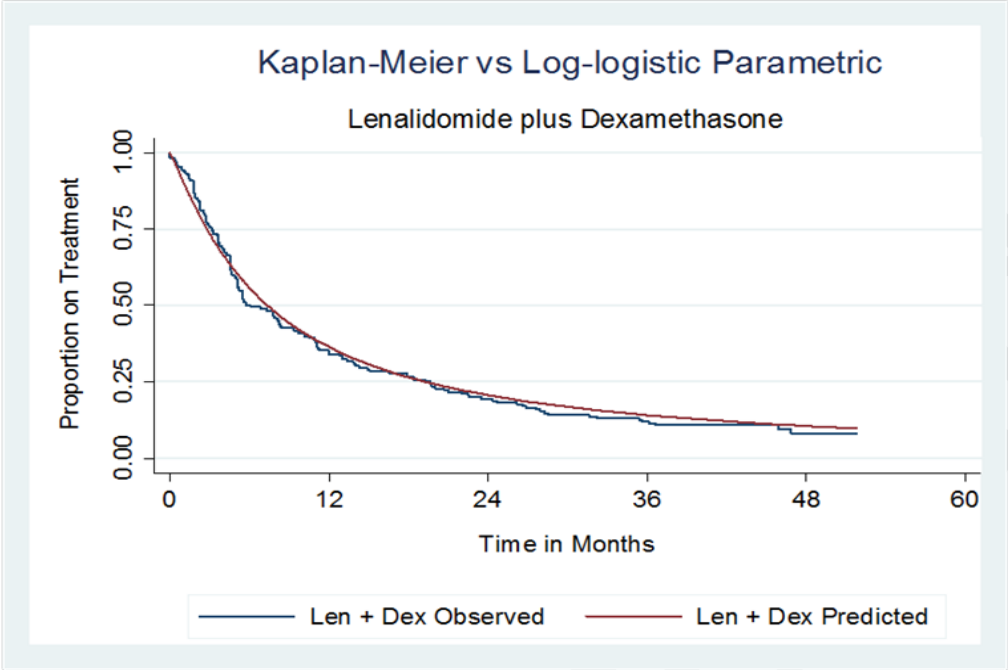
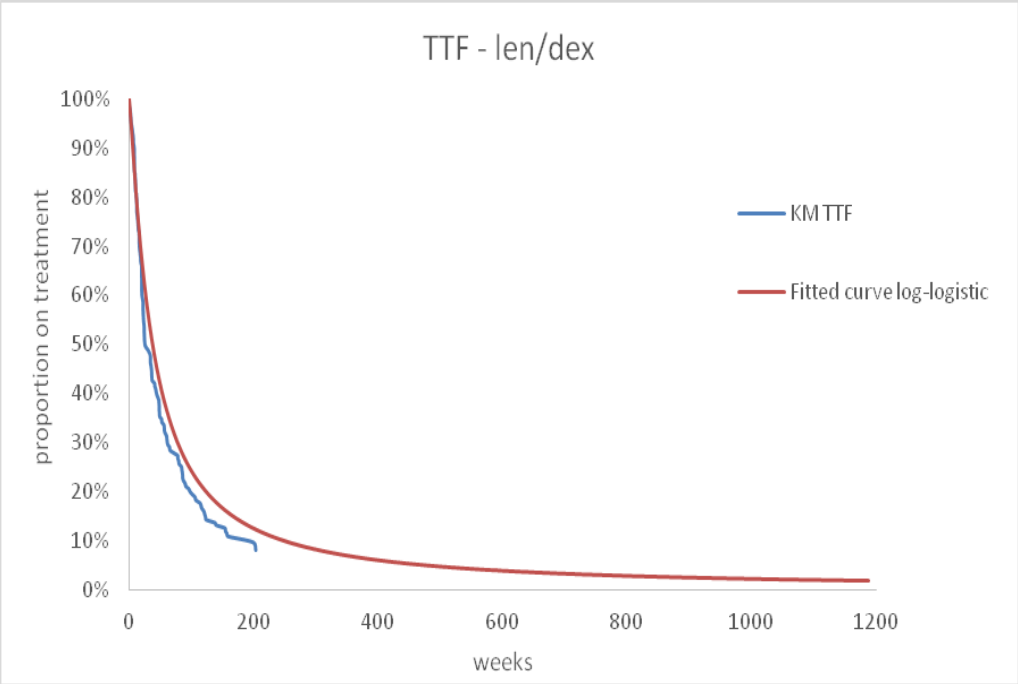


Figure 20. KM plot and fitted log-logistic curve for TTF over 25 years produced by the ERG



Source: produced by the ERG

Pre-progression on treatment to progressive disease (PFS-T to DP) - second-line

As mentioned above, patients in the PFS-T health state are those for whom the disease has not progressed and who are still on Len/Dex treatment. This condition is captured on one hand by PFS individual level data in MM-010, which defines disease progression, and on the other hand, by TTF individual level data in MM-010, which defines treatment continuation/failure.

Patients in the PD state are those for whom the disease has already progressed. These patients are assumed to be off second-line treatment. Therefore, this condition can be captured by the PFS individual level data in MM-010 alone.

Pre-progression off treatment to progressive disease (PFS-OT to DP) – second-line

As previously mentioned, patients in the PFS-OT heath state are those form whom the disease has not progressed but are not on Len/Dex treatment anymore (for example due to study withdrawal). As before, this condition is captured by both PFS and TTF individual level data in MM-010.

Patients in the PD state are those for whom the disease has already progressed.

For the transition probabilities just described (PFS-T to PD and PFS-OT to PD) only PFS and TTF survival estimates are necessary. These have been explained above and we have also provided detail on the distributions used to fit these statistics. We now focus on the Death state, defined by the OS survival data, which is the main source of concern for the ERG.

Death – second-line

Mortality in the model is captured by overall survival (OS) individual level data in MM-010. This is explored in detail in the subsection “mortality data” of the report.

Bort arm

For the bortezomib arm of the model, the same transition probabilities from the intervention arm were used, only exponentiated to computed HRs. To estimate HRs, Taverna (2012) data were used.

As explained in Section 5.1, median OS and PFS outcomes were taken from the sources identified for each comparator and were used to calculate HRs relative to Len/Dex in MM-010.

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. This approach assumes that progression/mortality occurs at a constant rate across studies and that studies’ populations and conditions are exchangeable.

The estimated HRs were then applied to the transition probabilities used in the Len/Dex arm of the model. This was done by exponentiating the Len/Dex transition probability to the HR in each cycle of the economic model. This seems like a reasonable approach.

It is also mentioned in the submission how the multivariate parametric models designed to predict PFS and OS (explained above) 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), which seems like a sound approach.

Finally, it was assumed that the HR for TTF would be the same as for PFS. The reasons provided for this were that the two survival estimates are similar and that no information on TTF was presented in the evidence found for the comparators.

The ERG found a mistake in the calculation of the PFS HR. This will now be discussed in detail.

Original submission

The ERG noted an inconstancy in the 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.

Manufacturer's approach

The manufacturer acknowledged this as an oversight and adjusted the HR of Bort retreatment relative to Len/Dex. The PFS adjusted HR obtained was 0.9.

The manufacturer explained the HR below 1 by providing the following argument:

"*The clinical explanation for this result, with a PFS hazard ratio <1 and OS hazard ratio>1, is that while it appears using the Taverna paper that patients are more likely to experience an initial response to bortezomib having already responded previously, the response duration and therefore resulting OS is shorter (as these patients have been previously treated with and lost response to bortezomib). Bortezomib patients remain on treatment for longer and incur the associated acquisition and administration costs, and fewer patients go on to receive subsequent therapy, and therefore subsequent lenalidomide which provides a survival benefit."*

ERG critique of the updated model

The ERG appreciate the fact that the manufacturer adjusted the HR. The resulting estimate was below 1, suggesting that patients receiving Bort are in the PFS health state for a longer period of time than the patients receiving Len, who supposedly progress faster.

Surprisingly, after the HR adjustment, the final ICER became dominant (while before it was around £14,000) despite the fact that Len actually became less effective in keeping patients from progressing.

The ERG feel that two issues should be taken into consideration in this case:

1. Clinical opinion sought by the ERG revealed that the clinical explanation provided by Celgene is reasonable in the sense that patients receiving subsequent Bort as a second-line therapy are expected to experience a shorter response duration whilst patients on Len/Dex would be expected to remain in the pre-progression state for longer.
2. The clinical rationale does not seem to agree with a PFS HR below 1. The PFS HR of 0.9 in favour of Bort suggests that patients receiving Bort are in the PFS health state for a longer period of time than the patients receiving Len, who supposedly progress faster.

Therefore it appears that the estimated adjusted HR does not accurately reflect the effectiveness of Len/Dex compared with Bort. Clinical expert opinion sought by the ERG reiterated the clinical explanation provided by Celgene thus suggesting that the plausible scenario would be to have PFS HR above 1, favouring Len/Dex.

One possible reason for this might be the use of the mean of covariates method to adjust the PFS curve. Again, the ERG question the validity such method used in the analysis.

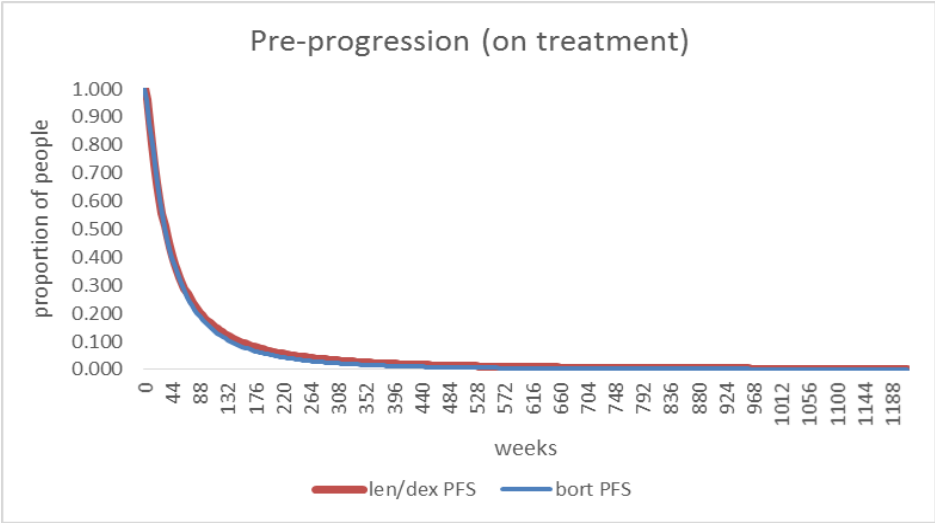
- ### Cohort distribution across health states in the model

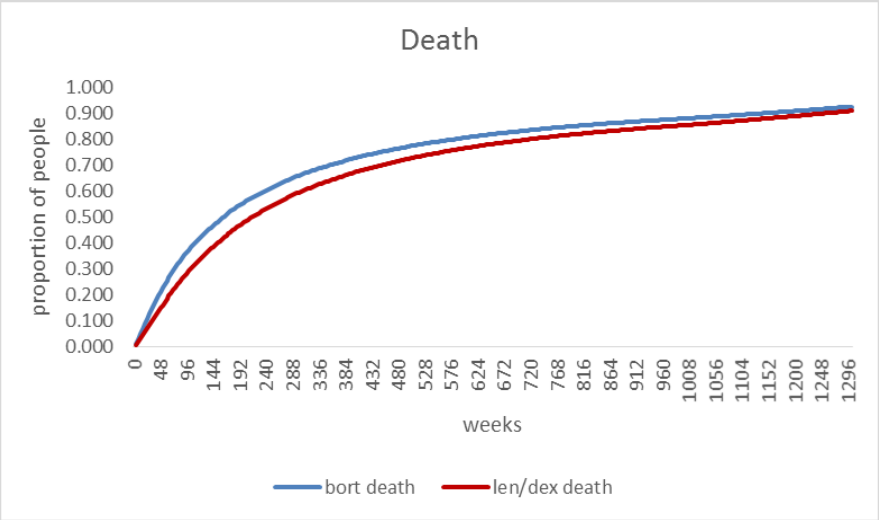
The ERG found some mathematical and conceptual mistakes in the allocation of patients to the different heath states in the Markov model.

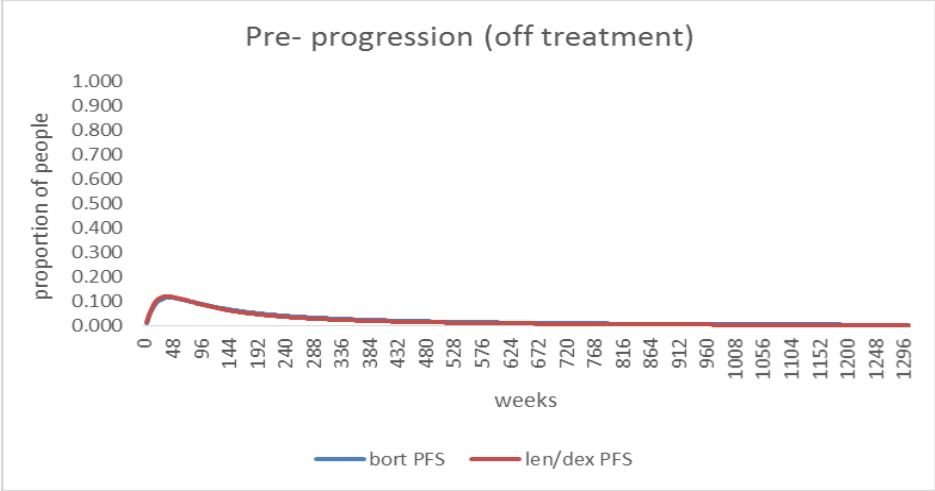
Figure 21 shows the Markov traces for all the heath states in the manufacturer updated model, with regards to the second line treatment option, therefore comparing Len/Dex with Bort. In the y axis we can read the proportion of the cohort allocated to the specific heath state while the x axis presents time in weeks.

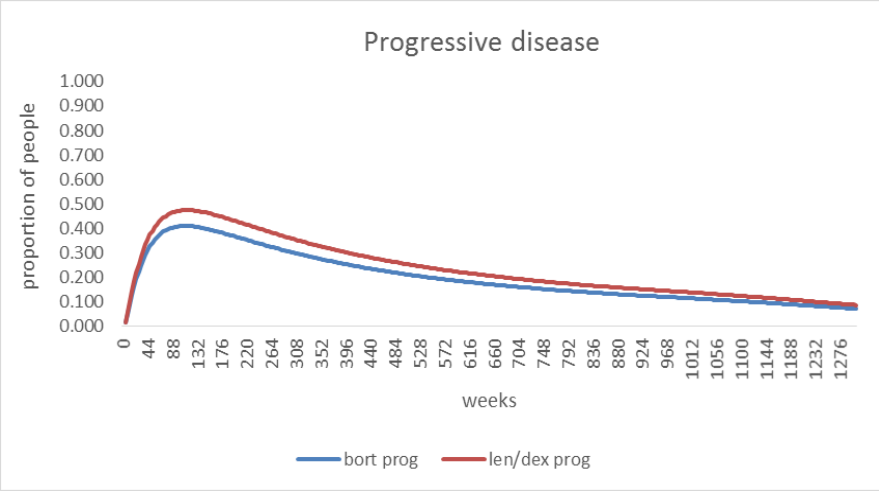
Looking at the pre-progression (on treatment) graph in Figure 21 we can observe that the PFS Len curve is above the PFS Bort curve. This is surprising, giving the HR of 0.9 between the drugs, favouring Bort. The ERG found some mathematical mistakes in the allocation of the cohort to the Markov heath states and corrected these. Results are discussed in Section 6.

Figure 21. Markov traces from original submission









- ## Mortality data

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. 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 PH violation.

However, in the updated submission Celgene changed the distribution used to fit MM-010 OS data to a log-logistic. This decision was in response to the issues raised by the ERG with regards to OS curves (fitted with the exponential piecewise model) crossing PFS and TTF curves in the original model. We now turn to this issue.

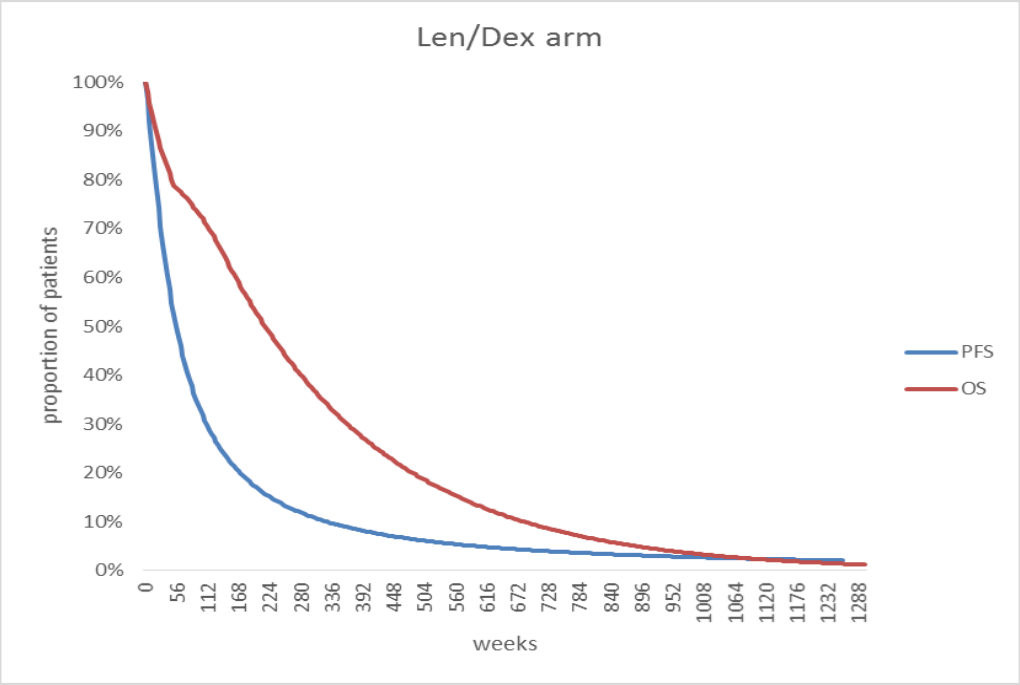
As before, we focus on the second-line treatment and the incremental mortality benefits estimated by the model with regards to Len/Dex vs. Bort.

Original submission

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 (Figure 22 and Figure 23). 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.

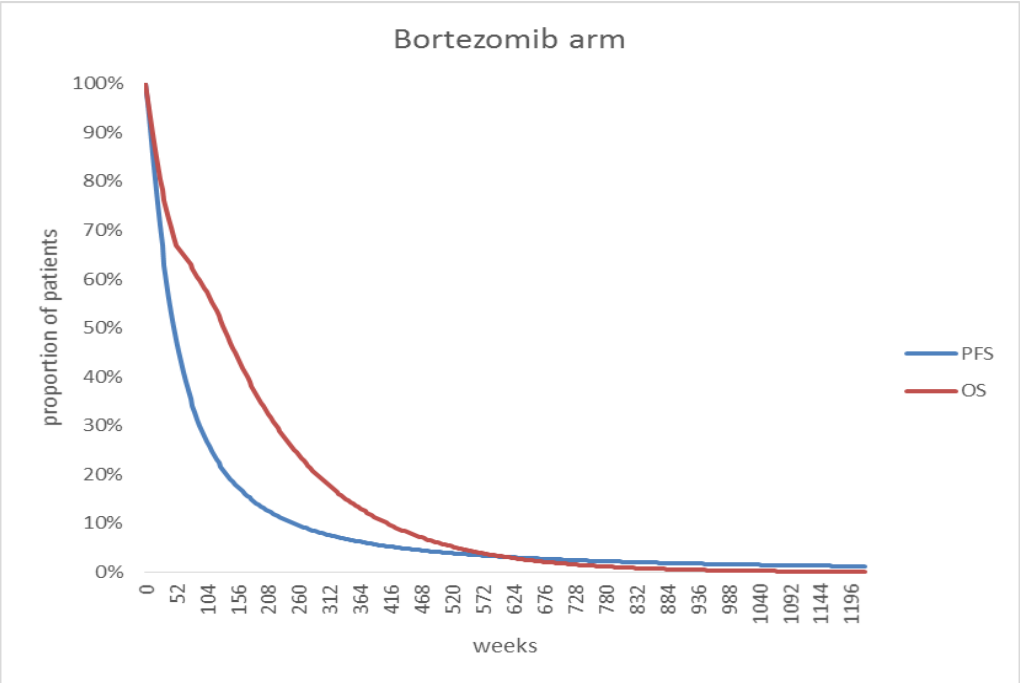
Similarly, it was noted that the OS and the TTF curves also crossed each other, both in the Len/Dex arm and in the Bort arm of the model (Figure 24 and Figure 25). Again, such crossing should not be possible as the TTF curve determines the number of people alive and still on treatment at different points in time. Therefore it is impossible to have a smaller total number of people alive (determined by the OS curve) than the number of people alive and still on treatment at the same point in time.

Figure 22. PFS and OS curves in the Len/Dex arm of the original model



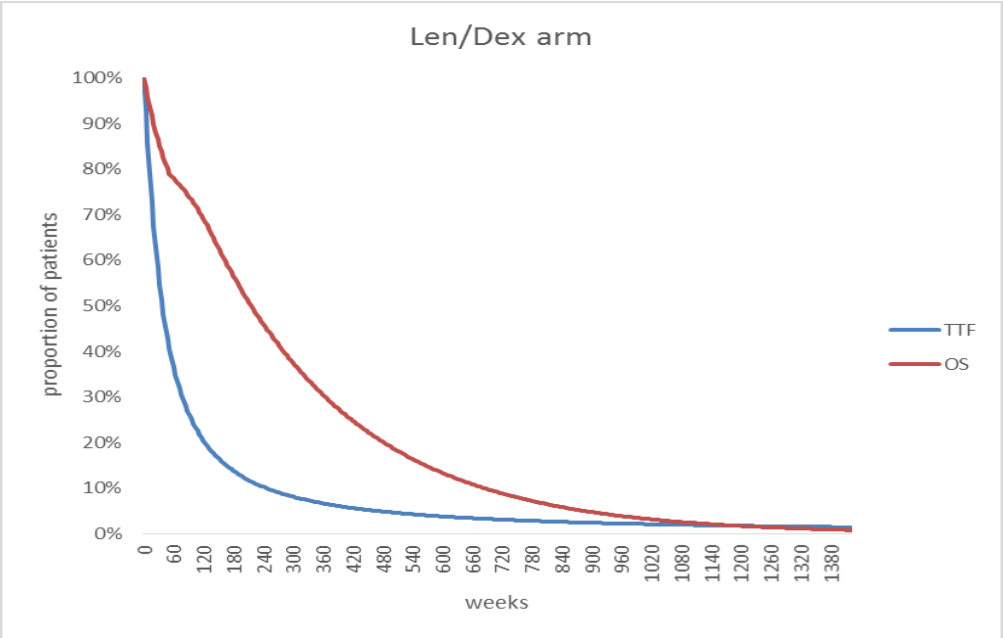
Source: produced by the ERG

Figure 23. PFS and OS curves in the Bort arm of the original model



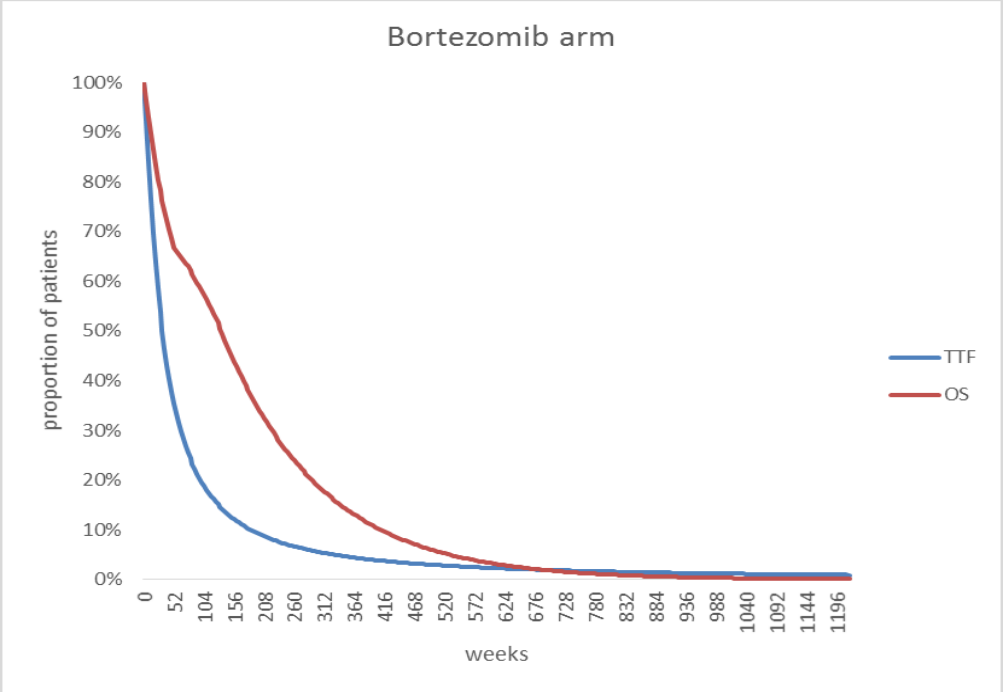
Source: produced by the ERG

Figure 24. TTF and OS curves in the Len/Dex arm of the original model



Source: produced by the ERG

Figure 25. TTF and OS curves in the Bort arm of the original model



Source: produced by the ERG

This suggested that the extrapolation process had some flaws. Therefore the ERG provided some possible explanations for this problem and suggested some approaches to deal with these.

Overall, it seemed that OS was being overestimated. It was suggested by the ERG that the use of the mean of covariates method to adjust the OS curve might potentially be overestimating survival.

Manufacturer's approach

The manufacturer agreed with the ERG that the OS curve crossing the PFS and the TTF curves was clearly implausible. However, it claimed that the use of a Markov structure prevented the curves from crossing.

Furthermore, Celgene claim that crossing of parametric curves is possible when different parametric models are selected for different clinical outcomes. The manufacturer state that in the originally submitted model a piecewise exponential curve was fitted to OS, while PFS and TTF were fitted with a log-logistic distribution, pointing to the fact that log-logistic models typically exhibit a 'long tail', with extended survival in the long term, while this is not commonly observed in exponential models.

The manufacturer claimed that *"importantly the KM plots for PFS and OS do not cross at any point*" and that crossing is the result of different fitted parametric models with different long-term characteristics.

Therefore, Celgene provided different scenarios where the same type of parametric curves are selected to fit the OS, the PFS and the TTF curves as it is claimed that fitting curves to distributions with similar characteristics would prevent the curves from crossing. Subsequently, AIC and BIC values are provided to justify the use a log-logistic distribution to fit the OS curve.

Finally, Celgene claim that it is unlikely that censoring affected the curve crossing seen in the model, noting that most of the OS censoring is found towards the end of the KM chart, while censoring of PFS is spread more evenly over time.

ERG critique of the updated model

The ERG appreciate that the manufacturer see the implausibility of the OS curve crossing with the PFS and the TTF curves.

Nevertheless, the ERG feel that the Markov approach is not a method for preventing survival curves from crossing. Survival curves should inform the transition probabilities in a Markov structure. Even though it is possible to prevent Markov traces from crossing each other (by making sure that the minimum transition probability value is always imputed in each cycle as Celgene do in their submission) this does not solve the fundamental issue that the survival curves informing the Markov transition probabilities cross.

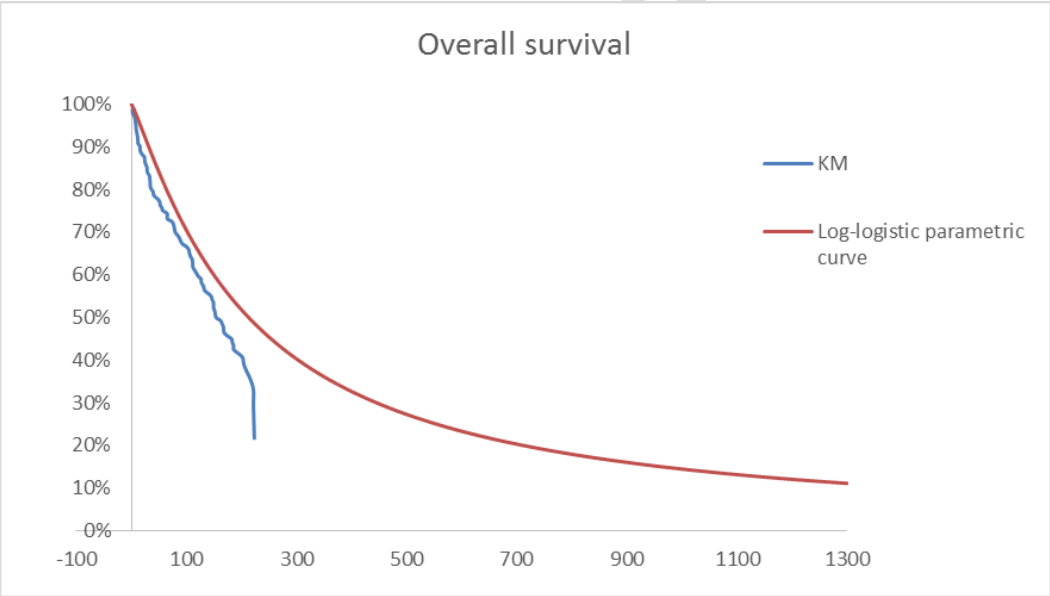
The ERG understand that crossing survival curves are a possible complication arising from fitting data to different distributions. However, when this is observed, a different approach needs to be taken which prevents the curves from crossing for example, using flexible models on the hazard ratios (e.g. fractional polynomials). To note is that the piecewise exponential originally used to fit OS data, would be more flexible in this sense than the log-logistic model.

It is the ERG opinion that Celgene's decision to change the distribution used to model OS from a piecewise exponential to a log-logistic distribution needs to be based on a stronger justification than avoiding survival curves crossing. In fact, the distribution used to model OS should be selected based on the criteria of best fit to the actual survival data and consider all potential complications.

Figure 26 (produced by the ERG) shows the KM curve for MM-010 overall survival data as well as the fitted curve, produced by fitting a log-logistic distribution to OS data in MM-010.

Based on visual inspection of the curves, the log-logistic distribution seems to be a very poor fit to the OS data in MM-010.

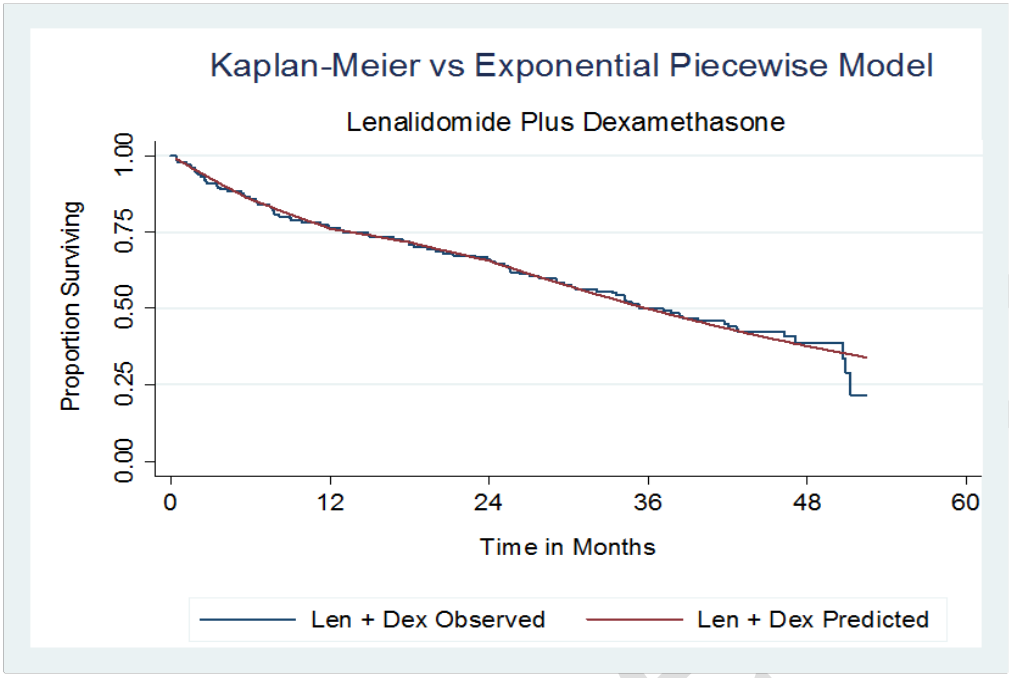
Figure 26. KM plot and fitted log-logistic curve for OS over 25 years –Len/Dex



Source: produced by the ERG

Additionally, the ERG tried to replicate Figure 15, which was taken from the original submission (replicated below) and shows the KM curve for OS as well as the extrapolated curve produced by fitting an exponential piecewise model to OS data. The resulting curves are presented in Figure 27. Unfortunately it was not possible to replicate Figure 15 (the same problem was found for PFS and TTF original graphs) and based on Figure 27 produced by the ERG, even though the exponential piecewise curve seems a better fit until around week 50 it seems to be a poor fit as time progresses.

Reproduction of Figure 15. KM plot and fitted exponential piecewise model for OS (Celgene submission)

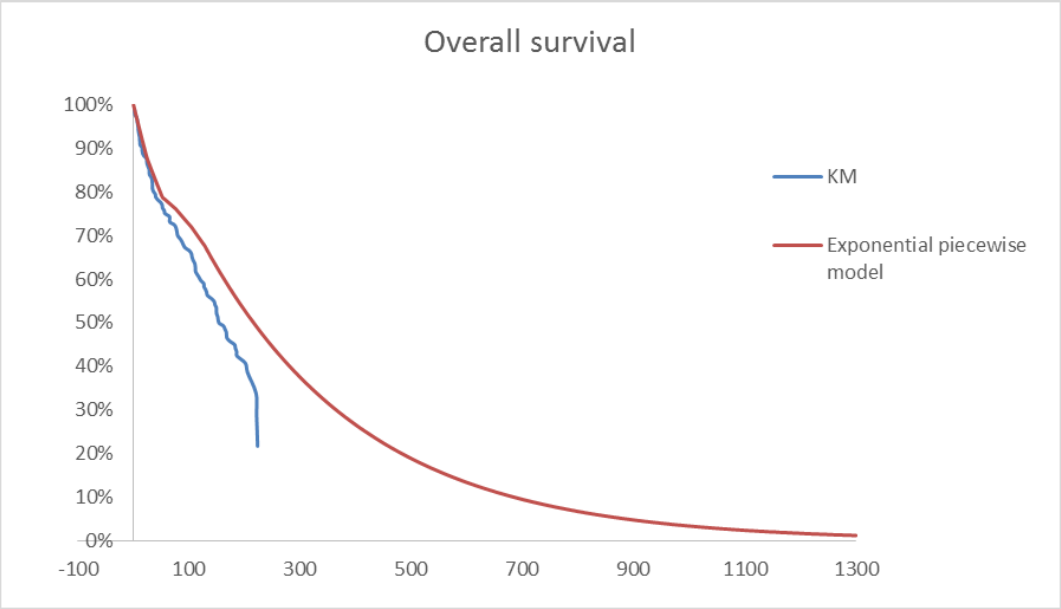


Both Figure 27 and Figure 26 suggest that OS is overestimated in the economic model, especially later in time. The economic model runs for approximately 25 years (1300 weeks) and we can observe that when using the log-logistic distribution to fit OS data, by week 1300 around 11% of patients are still alive. As the population entering the economic model is 63 years old, this would mean that approximately 11% of the MM population lives until the age of 88.

Furthermore, in the submission it is stated that for patients with stage I MM the median expected survival is 62 months, while for patients with stage III disease the median survival is reduced to 29 months. Again, these estimates reinforce the likelihood of the overestimation of predicted survival in the economic model.

Superseded

Figure 27. KM plot and fitted piecewise exponential curve for OS over 25 years – Len/Dex



Source: Produced by the ERG

Celgene also argue that "importantly the KM plots for PFS and OS do not cross at any point" and that crossing is the result of different fitted parametric models with different long-term characteristics.

The ERG question the validity of this argument as it would be truly impossible for KM curves to cross in any case. As KM curves represent real data (instead of extrapolated data) having a PFS KM curve crossing a OS KM curve would mean that in real life, the number of progression-free patients would be higher than the number of patients alive, which is obviously implausible.

Celgene claim that it is unlikely that censoring affected the curve crossing seen in the model. However, the ERG do not have enough evidence to assess this statement.

In summary, the ERG do not feel confident that the explanations and approaches followed by the manufacturer truly addressed the initial problems raised.

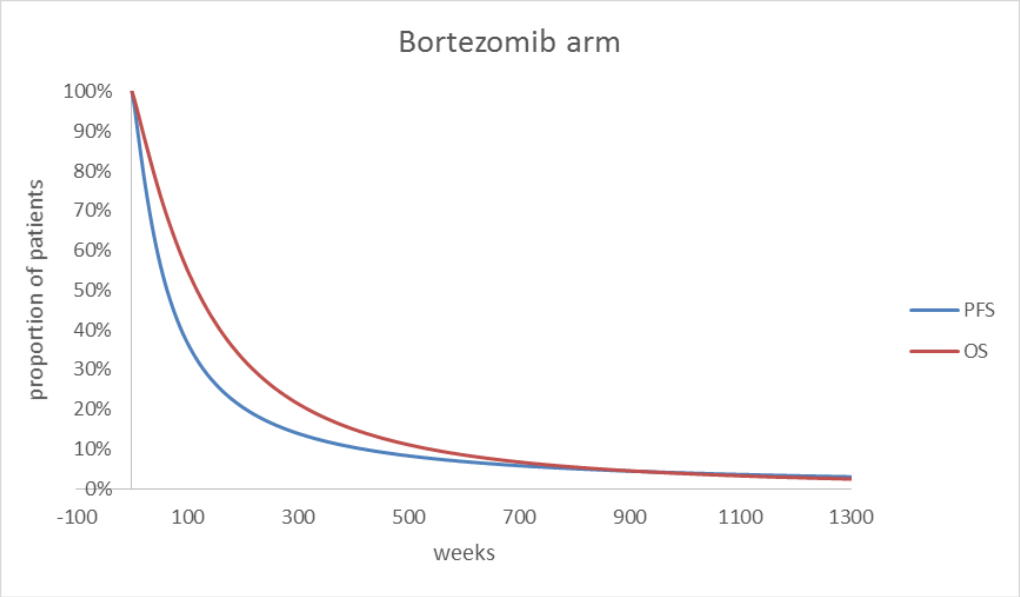
The decision to change the distribution used to model OS from a piecewise exponential to a log-logistic distribution is not based on a sound argument (i.e. preventing the survival curves from crossing) and more importantly, does not solve the problem of the curves crossing.

Even though the OS curves do not cross the PFS and the TTF curves in the intervention arm of the model anymore, Figure 28 and Figure 29 show how this is still a problem in the comparator arm of the model.

The curves now cross later in time (in the original submission the curves crosses around week 600) with PFS and OS curves crossing each other around week 900 (19 years) and TTF and

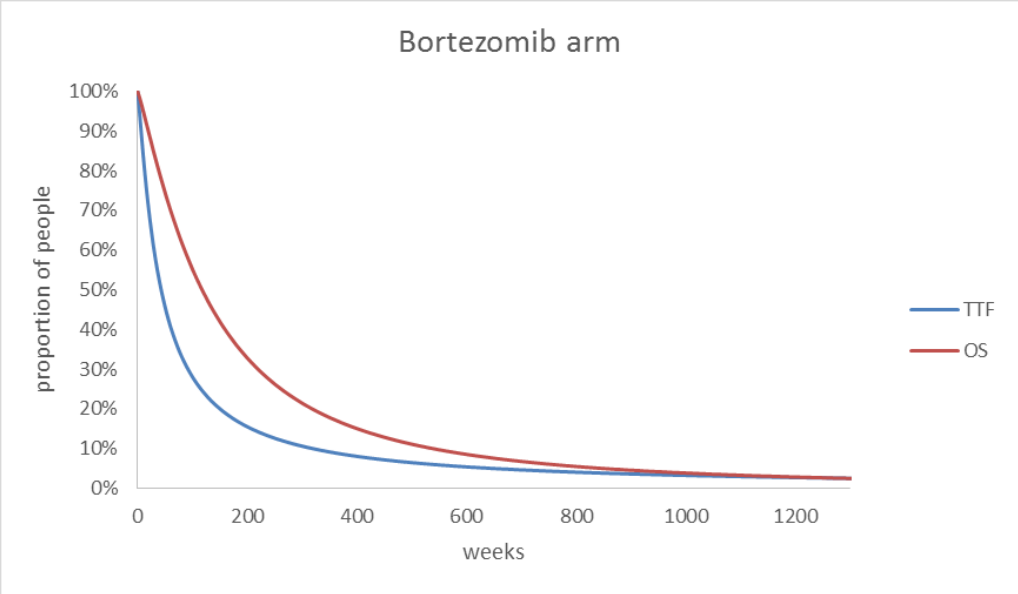
OS curves crossing each other around week 1290, which corresponds to approximately 25 years (note that the economic analysis lasts for 25 years). However this is still an implausible scenario and a not acceptable one, for the reasons explained before.

Figure 28. PFS and OS curves in the Bort arm of the model



Source: produced by the ERG

Figure 29. TTF and OS curves in the Bort arm of the model



Source: produced by the ERG

- ## Drug costs

Len costs

Overall the calculation of the Len costs in the model was satisfactory.
However clinical opinion sought by the ERG revealed that the inclusion of G-CSF in the Len treatment regimen is rather uncommon within current clinical practice. Furthermore, assumptions were made as to the drug dose administered per day. It seems that Celgene could have used real trial data (as G-CSF was administered during MM-010) to model the cost of G-CSF.

Dex costs

Overall the calculation of the Dex costs in the model was satisfactory with the exception of the inclusion of the cost of Dex in the Bort arm of the model. As mentioned in Section 5.1, 64% of patients in the Taverna (2012) study took concomitant Dex and therefore the respective cost should have been included in the base-case ICER.

Celgene have changed this in the updated economic model and a final ICER considering the cost of Dex for 64% of the Bort patients was presented in scenario analysis. However this was not considered to be the base-case ICER.

Bort costs

The ERG had some problems tracing back the calculations undertaken to estimate the cost of Bort per cycle. There are no data in the submission or the Taverna study specifying the exact dose of Bort administered. The ERG understand that the dose regimen of Bort is also related to the patient body mass surface.

From analysing the cost calculations in the excel model it seems that Celgene assumed the following dose regimen: 1 vial per administration and 16 administrations every 3 cycles, which accounts for an average of 5.3 vials per model 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 submission.

The only information available in the Taverna (2012) study with regards to the Bort dose administered was that a median number of 3 cycles of Bort was received by patients (with a range going from 1 to 19 cycles of treatment received).

Other fundamental issue related with the cost of Bort is the duration of retreatment with the drug at second-line. In their submission, the manufacturer assume that patients are kept on treatment until they progress or until treatment fails. However, clinical opinion sought by the ERG informed that Bort treatment in current clinical practice only lasts for a fixed period of time, usually 8 cycles (corresponding to 8 months). This assumption has a great impact on the final ICER. This is further discussed in Section 6 of the report.

Treatment administration costs

While the cost of administration for Len/Dex was assumed to occur only for the first appointment (after which it is assumed that the patient self-administers oral treatment), the administration cost of Bort was applied for every administration appointment in the hospital. This seems a reasonable assumption.

Potential transportation costs to the hospital for administration of the drugs were also considered. This 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 to receive full treatment.

Clinical opinion sought by the ERG did not believe this to be a reasonable assumption. Firstly the percentage of patients requiring transportation to the hospital was considered to be significantly lower than 50% of MM patients. 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.

- ### Disease management costs

The ERG found the calculations related with monitoring costs to be overall confusing.

It is stated in the submission that monitoring frequency depends primarily on whether the patient has experienced disease progression or not. Furthermore Len treatment is associated with an increased monitoring requirement as per the SPC.

It is not clear in the submission which costs have been assumed to relate with the disease state (i.e. progression or progression-free) or with the initial monitoring costs associated with Len.

It is stated how monitoring costs by health state have been taken from previous NICE TAs 171 and 228 however, the list of monitoring testes shown in TA171 seems to be much more extensive than the one reported in Celgene's submission.

Finally it seems like for the period of additional monitoring associated with the Len treatment (first eight weeks of treatment), the ongoing monitoring costs associated with the heath state the patient is in (i.e. disease progression or not) are not being considered.

Again, since no clear explanation is provided as to how these costs were calculated, it seems to the ERG that this might be leading to an underestimation of the monitoring costs associated with Len.

- ### Other costs

Due to the time constraints explained previously in Section 1, the ERG could not cover in detail the calculations of subsequent treatment and terminal care costs.

- ### Adverse events

The incidence of AEs for Len/Dex seem to be appropriately derived from MM-010. Due to a paucity of data the incidence of AEs for other comparators was taken from patients who received melphalan plus prednisolone/prednisone in the NICE TA228 submission. Expert opinion sought by the ERG confirmed that the safety profile of prednisolone/prednisone follow other comparators and therefore this approach seems reasonable since no other data is available.

In comparison with TA171, AE utility decrements are included in the cost-effectiveness analysis with values taken from Brown (2013). Diarrhoea and constipation were not associated with decrements in utility. Our clinical expert suggested that this is an underestimation of the impact on quality of life of those AEs. Conversely, the disutility value attached to experiencing thrombocytopenia was considered to be too high.

- ### Health related quality of life

As in TA171 and TA228, Celgene used utility values reported for patients who underwent intensive chemotherapy in the cost-utility carried out by van Agthoven (2004) for patents with previously untreated multiple myeloma.

Celgene therefore assume a utility value of 0.81 for patients in progression-free survival, 0.77 for patient in pre-progression after 2 years, and 0.64 for progressive disease patients. Utility values were adjusted by an age-dependent factor to reflect decreased utility with age, based upon published UK EQ-5D values by Kind (1999).

Patients enter the model at 63 years, which is associated to a mean utility of 0.80 for a healthy member of the UK population. A decreasing age-dependent weighting factor is applied as patients moved in the model in order to reflect the detrimental effect on HRQoL directly associated with age.

Table 41. Weighted health state index by age and sex

Age	Average	Males	Females
55-64	0.80	0.78	0.81
65-74	0.78	0.78	0.78
75+	0.73	0.75	0.71

Source: Adapted from Table A in Kind (1999)

Based on expert opinion, patients with multiple myeloma in progression-free survival have a lower HRQoL than a member of the general public at the same age. Therefore, we suggest that it may be more appropriate to use a value lower than 0.81 for the utility in progression-free survival.

Furthermore, some mathematical mistakes were found in the estimation of the pre-progressive state QALYs. This will be explored in Section 6.

5.2.4 Sensitivity analysis

A range of deterministic sensitivity analysis was provided in Celgene's submission. A tornado diagram was generated to demonstrate the effect of varying some of the individual parameters in the model.

Deterministic sensitivity analysis aimed to cover parameter uncertainty by varying the values used in the model by their upper and lower confidence interval values and structural uncertainty by using different data to model some of the parameters.

In addition to deterministic sensitivity analysis, Celgene also presented probabilistic sensitivity analysis (PSA).

5.3 Results included in manufacturer's submission

This section presents a summary of the results of Celgene's model.

In this section we focus on the updated model results, however whenever deemed necessary we will make reference to the results in the original submission. For more details on the originally presented results, please see Section 7.7 of Celgene's submission.

5.3.1 Deterministic results

- Base case

The base case outputs originally presented by Celgene are presented in Table 42. Presented in Table 43 are the base case results produced by the updated analysis, subsequently submitted by Celgene.

To note is that in the original analysis, the Len/Dex arm presented higher costs than the comparator (Bort) arm. However in the updated model, the inverse is observed, with the Bort arm yielding higher costs than the intervention (Len/Dex) over the 25 years of analysis.

QALYs were consistently higher in the Len/Dex arm of the model. In fact the incremental gain in QALYs remained the same from the original to the updated analysis (0.53 QALY gain).

The fact that the ICER became dominant in favour of Len/Dex (when previously it was £14,535 per QALY) it is nonetheless surprising given that the initial corrections made to the model by Celgene would actually suggest that Bort was more effective in keeping patients from progress than Len/Dex (corrected PFS HR of 0.9). This is further explored in Section 5.4.

Celgene also presented the median OS and PFS estimates derived from the economic analysis compared with the ones reported in the trial. It should be noted that the estimates used in the excel model (presented in Table 44) are not the ones presented in the economic model. Again the values presented suggest an overestimation of OS.

Table 42. Base case outputs per patient at 25 years in the original analysis

Cost-effectiveness results per patient	Len/Dex (1)	Bort (2)	Incremental value (1-2)
Total costs £	121,422	113,740	**7,682**
QALYs	3.42	2.89	**0.53**
ICER			**£14,535**

Source: Adapted from submission, Table 69.

Table 43. Base case outputs per patient at 25 years in the updated analysis

Cost-effectiveness results per patient	Len/Dex (1)	Bort (2)	Incremental value (1-2)
Total costs £	92,774	131,111	**-38,337**
QALYs	3.98	3.45	**0.53**
ICER			**Len dominates**

Source: Adapted from Celgene reply to ERG further request for clarification, Table 8,

Table 44. Base case model outputs (Len/Dex) compared with trial data

	Clinical trial result	Model result
Median OS lenalidomide (years)	3.10	4.12
Median PFS lenalidomide (years)	0.84	1.09

Source: Celgene excel model

- ## Deterministic sensitivity analysis

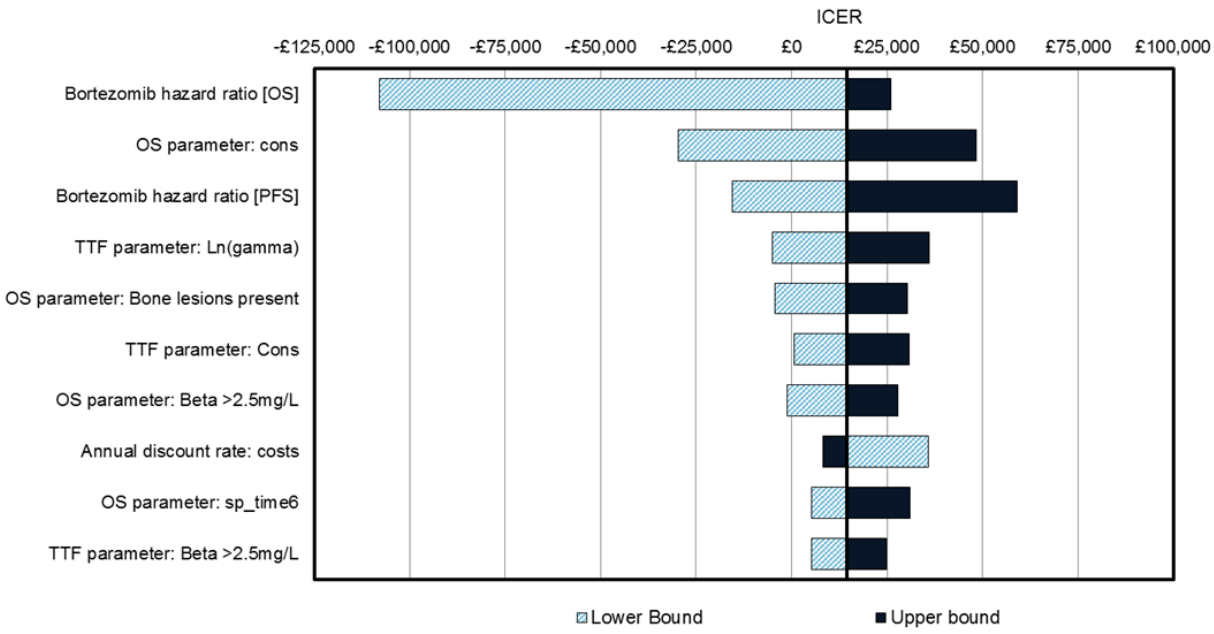
Figure 30 was taken from the original submission and presents the 10 most relevant parameters tested through deterministic sensitivity analysis.

Figure 31 was provided to the ERG by Celgene upon the submission of the updated analysis. However the figure shows the sensitivity of the Net Monetary Benefit (NMB), instead of the ICER, to the model inputs.

Comparing the two figures we can notice how the model is most sensitive to changes in the OS and PFS parameters (i.e. the distribution parameters used to fit the OS and the PFS data). This is not surprising. What is remarkable is how in the updated analysis none of the changes applied to model inputs drives the NMB below 0.

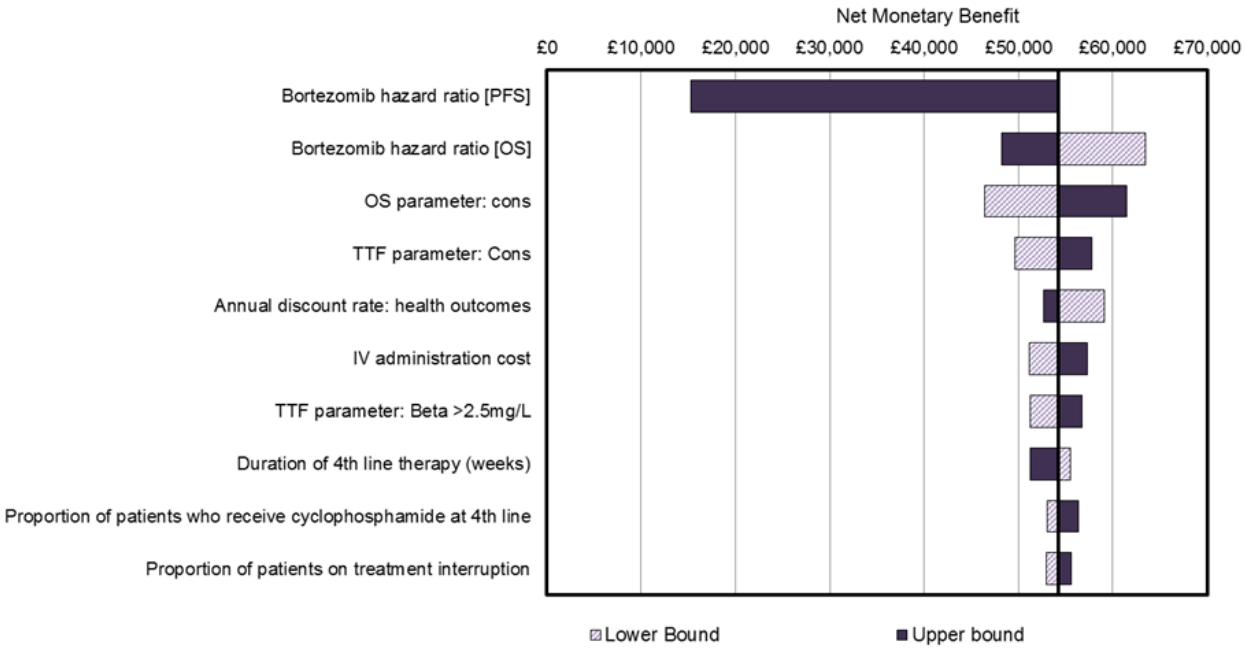
While in the original analysis, ICERs reached values above £50,000, in the updated analysis the MNB is always above 0, which at the used threshold of £30,000 suggests that the intervention is always cost-effective regardless of changes in the model inputs.

Figure 30. Tornado diagram with top 10 parameters in terms of ICER sensitivity – original analysis



Source: Submission, Figure 32.

Figure 31. Tornado diagram with top 10 parameters in terms of NMB sensitivity – updated analysis



Source: Updated model, Figure 2.

Structural uncertainty was addressed using a number of scenarios. These included varying the time horizon of the model, the source used to model the effectiveness of Bort and using the Bort PAS among others.

Scenario analysis was also undertaken to compare Len/Dex with:

- Bendamustine/ prednisolone
- Bendamustine/ Dex
- Melphalan/ prednisone
- High-dose cyclophosphamide/ Dex
- Low-dose cyclophosphamide/ Dex
- Doxorubicin
- Vincristine

Table 45 presents the ICERs resulting from running different scenario analysis. This was provided to the ERG by Celgene upon submission of the updated model.

Similarly to the deterministic sensitivity analysis, most of the outputs of the scenario analysis produced a dominant ICER in favour of Len.

The use of alternative comparators produced a considerable range of ICERs (£23,435 - £36,718). However, the value of this analysis is relative as the source used to model the

effectiveness of bendamustine, Damaj (2012), was also used to model the effectiveness of all other comparators (since no data was found to model these).

As previously mentioned, the ERG question why bendamustine was not considered in the base case analysis as data were available to model the cost-effectiveness of the drug compared with Len/Dex and it made part of the initial scope. Note that the ICER resulting from selecting bendamustine as a comparator is around £23,500.

Table 45. Scenario analysis outcomes – updated model.

Parameter	***Base Case Setting***	***Scenario Setting***	***ICER***
Base case	*N/A*	*N/A*	*Lenalidomide Dominates*
Varying the time horizon			
Time horizon	*25 years*	*5 years*	*Lenalidomide Dominates*
		10 years	*Lenalidomide Dominates*
		15 years	*Lenalidomide Dominates*
		20 years	*Lenalidomide Dominates*
Type of comparison			
Use of 3rd line lenalidomide	*As per NICE recommendations i.e. earlier use of lenalidomide*	*As per historical BSC*	*£26,665*
Choice of comparator	*Bortezomib retreatment*	*Bortezomib retreatment + dexamethasone*	*Lenalidomide Dominates*
		Bendamustine + Dex	*£23,435*
		Bendamustine + presnisolone	*£23,424*
		Melphalan + prednisone	*£28,516*
		HD Cyclophosphamide + LD-Dex	*£36,718*
		LD Cyclophosphamide + MD-Dex	*£33,088*
		Doxorubicin	*£35,836*
		Vincristine	*£33,013*

Include dexamethasone with 3rd and 4th line treatments	*Yes*	*No*	*Lenalidomide Dominates*
Varying modelling of lenalidomide efficacy			
Parameter used to model treatment failure	*Time to treatment failure*	*Progression-free survival*	*Lenalidomide dominates*
Varying treatment efficacy assumptions			
Comparative efficacy of bortezomib	*Overall survival from Taverna 2012*	*White 2013*	*Lenalidomide Dominates*
	Progression free survival from Taverna 2012	*Petrucci 2013*	*Lenalidomide Dominates*
		Hrusovsky 2010	*Lenalidomide Dominates*
		Dispenzieri 2010	*Lenalidomide Dominates*
		White 2013	*Lenalidomide Dominates*
Varying cost assumptions			
Bortezomib PAS	*Not included*	*15% discount received by all patients*	*Lenalidomide Dominates*
		15% discount received by 55% of patients	*Lenalidomide Dominates*

Source: Table provided to the ERG by Celgene upon submission of the updated model

The ERG suggested that Celgene presented a scenario analysis using only patients treated in the second-line setting. Results are presented in Table 46.

Table 46. Scenario analysis using clinical inputs for the second-line population

Model Arm	Total Costs	Total LYG	Total QALYs	Incremental Costs	Incremental LYG	Incremental QALYs	ICER
Bortezomib	£141,359	8.25	4.11	-	-	-	-
Len/Dex	£107,708	9.96	4.92	-£33,651	1.71	0.81	Lenalidomide dominates

Source: Analysis provided to the ERG by Celgene upon submission of the updated model

Finally, as the justification behind the choice of MM-010 raised some concerns, the ERG have also requested that the base case analysis was run with MM-009 trial data. Results are presented in Table 47.

Table 47. Scenario analysis using clinical inputs from MM-009

Model Arm	Total Costs	Total LYG	Total QALYs	Incremental Costs	Incremental LYG	Incremental QALYs	ICER
Bortezomib	£125,576	6.09	3.34	-	-	-	-
Len/Dex	£96,013	6.60	3.67	-£29,563	0.50	0.33	Lenalidomide dominates

Source: Analysis provided to the ERG by Celgene upon submission of the updated model

Due to the time constraints explained in Section 1, the ERG did not have the time to analyse how the deterministic sensitivity analysis was undertaken in the excel model and to explore all the possible implications of the scenario analysis ran.

5.3.2 Probabilistic sensitivity analysis results

Figure 32 was taken from the original submission and presents the output of the PSA using 1,000 model runs.

Figure 33 was provided to the ERG by Celgene upon the submission of the updated analysis and presents the outputs of the PSA ran in the updated analysis.

Similarly to the deterministic sensitivity analysis, we can notice that while in the original analysis, ICERs reached values above £50,000, in the updated analysis the reported ICERs are always negative.

Due to the time constraints explained in Section 1, the ERG did not have the time to analyse how the PSA was undertaken in the excel model.

Figure 32. Cost-effectiveness scatter plot – original submission

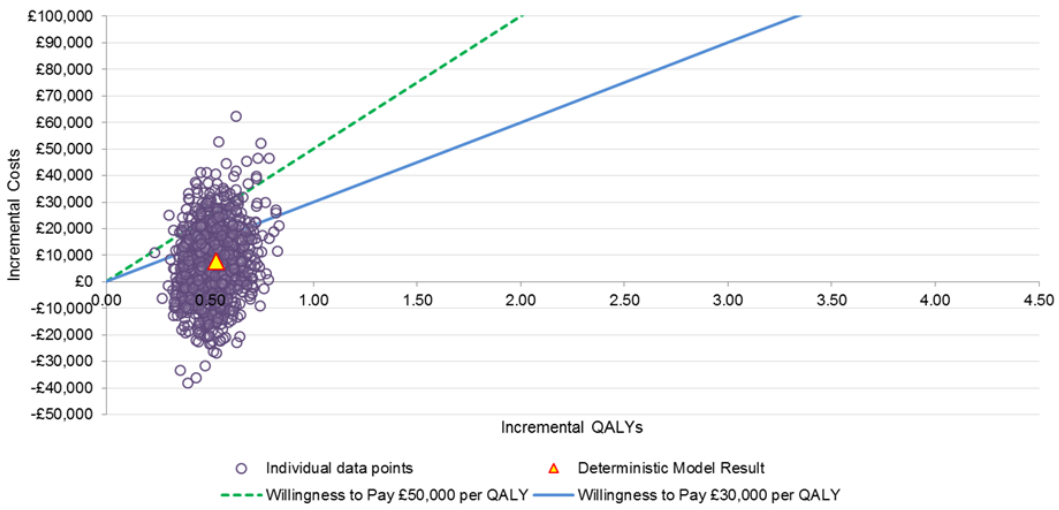
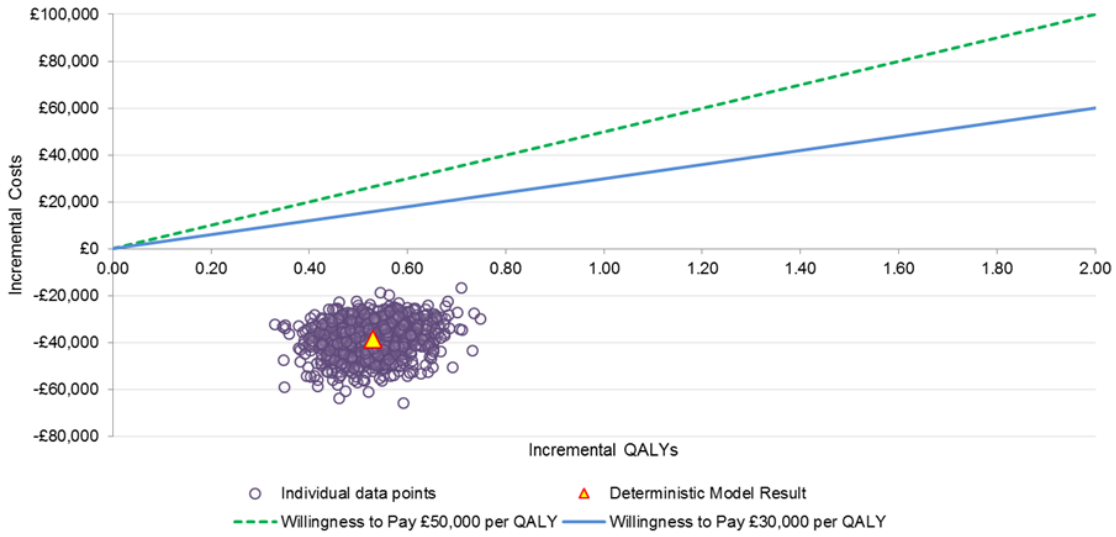


Figure 33. Cost-effectiveness scatter plot – updated submission



5.4 Comment on validity of results presented with reference to methodology used

In Sections 5.2.2 and 5.2.3 we provided a detailed description of the modelling approach and structure adopted by Celgene and of the data and estimations used to populate the model.

The ERG have several concerns with the model structure and essentially with the methods used to estimate OS and the PFS HR in the model.

Seeing that these are the key components of the cost-effectiveness analysis submitted by Celgene, the ERG have little confidence in the final ICERs presented.

As previously mentioned, the ERG question why bendamustine was not considered in the base case analysis as data were available to model the cost-effectiveness of the drug compared with Len/Dex and it made part of the initial scope. Note that the ICER resulting from selecting bendamustine as a comparator is around £23,500.

6.0 Additional work undertaken by the ERG

In this section we explore the implications of some of the errors found in Celgene's model. However, given that two of the major ERG concerns are methodological and relate to the data extrapolation process for OS and the method used to adjust HRs in the model, the alternative ICERs presented should be interpreted with extreme caution.

6.1 Additional searches

Scoping searches were undertaken to assess three queries in the initial submission:

1. Scoping searches were conducted to assess the effect of Celgene's initial decision to limit their effectiveness searches to failure at first-line treatment. The ERG have therefore:

    a. Re-run Celgene's searches with the limiting terms included (taking the submission as presented)
    b. Re-run Celgene's searches with the relapse and recurrence terms removed.

   The unique items between these two searches were then sampled. This led the ERG to clarify the rationale behind the use of the relapse and recurrence search terms with Celgene and ask them to re-run their effectiveness searches.

2. Celgene appear to have made a spelling mistake in their effectiveness searches. The effect of this on retrieval was assessed. The error relates to bendamustine, where Levact has been misspelt.

3. The ERG ran brief scoping searches to assess Celgene's omission of specific health related quality of life (HRQoL) terms in their initial submission. This led the ERG to raise a query at the clarification stage and Celgene responded by re-running their HRQL searches to include a standard HRQoL filter. The ERG are satisfied with this response.

6.2 Correction for errors in Celgene's model

After identifying some technical errors, we have decided to make the following adjustments to Celgene's base case model:

1. The ERG found some mathematical and conceptual mistakes in the allocation of patients to the different heath states in the Markov model. We have corrected these and the results are presented in Section 6.1.2.
2. The manufacturer assumed that patients are kept on Bort treatment until they progress or until treatment fails. However, clinical opinion sought by the ERG informed that bortezomib treatment in current clinical practice only lasts for a fixed period of time. We have changed this assumption in the model so it reflects a treatment duration of 8 cycles (as suggested by clinical expert opinion). The impact of this is presented in Section 6.1.2.
3. The ERG found a mistake in the QALY calculation. This was related with the adverse events' disutility considered in the overall QALY estimation.

To note is that the ICERs presented in Section 6.1.2 are for the second-line treatment pathway. Therefore, the structural problems identified in Section 5.2.2 were not addressed here as these concern the evaluation of subsequent treatment options.

6.2.1 Corrected base case outputs

Allocation of patients to model heath states

Comparing Figure 34 with Figure 29, we can observe that the pre-progression (on treatment) graph in Figure 34 now shows a PFS curve for Bort above the PFS curve for Len, which is in line with the reported HR below 1.

As mentioned in Section 5.2.2, clinical opinion sought be the ERG revealed that patients receiving subsequent Bort as a second-line therapy are expected to experience a shorter response duration whilst patients on Len/Dex would be expected to remain in the pre-progression state for longer.

The clinical rationale does not seem to agree with a PFS HR below 1. The PFS HR of 0.9 in favour of Bort suggests that patients receiving Bort are in the PFS health state for a longer period of time than the patients receiving Len, who supposedly progress faster. However, given that the estimated value is 0.9, the Markov traces should be in conformity with the inputs used and therefore should show that Bort patients are in the PFS state for longer than Len patients.

Again, it is the ERG opinion that the estimated adjusted HR does not accurately reflect the effectiveness of Len/Dex compared with Bort. Clinical expert opinion sought by the ERG, reiterated the clinical explanation provided by Celgene, which suggested a PFS HR above 1, favouring Len/Dex.

By analysis Figure 34 it is also noticeable how the marginal benefits in the analysis are essentially derived from the reduction in mortality, where Len patients seem to have a much higher OS than Bort patients.

Even though a higher number of patients remain in the PFS state while on Bort treatment then on Len treatment, this incremental difference is much smaller than the mortality benefits accrued by Len/Dex.

This, again, raises the question of OS being overestimated in the economic model, favouring Len.

Treatment duration of Bort

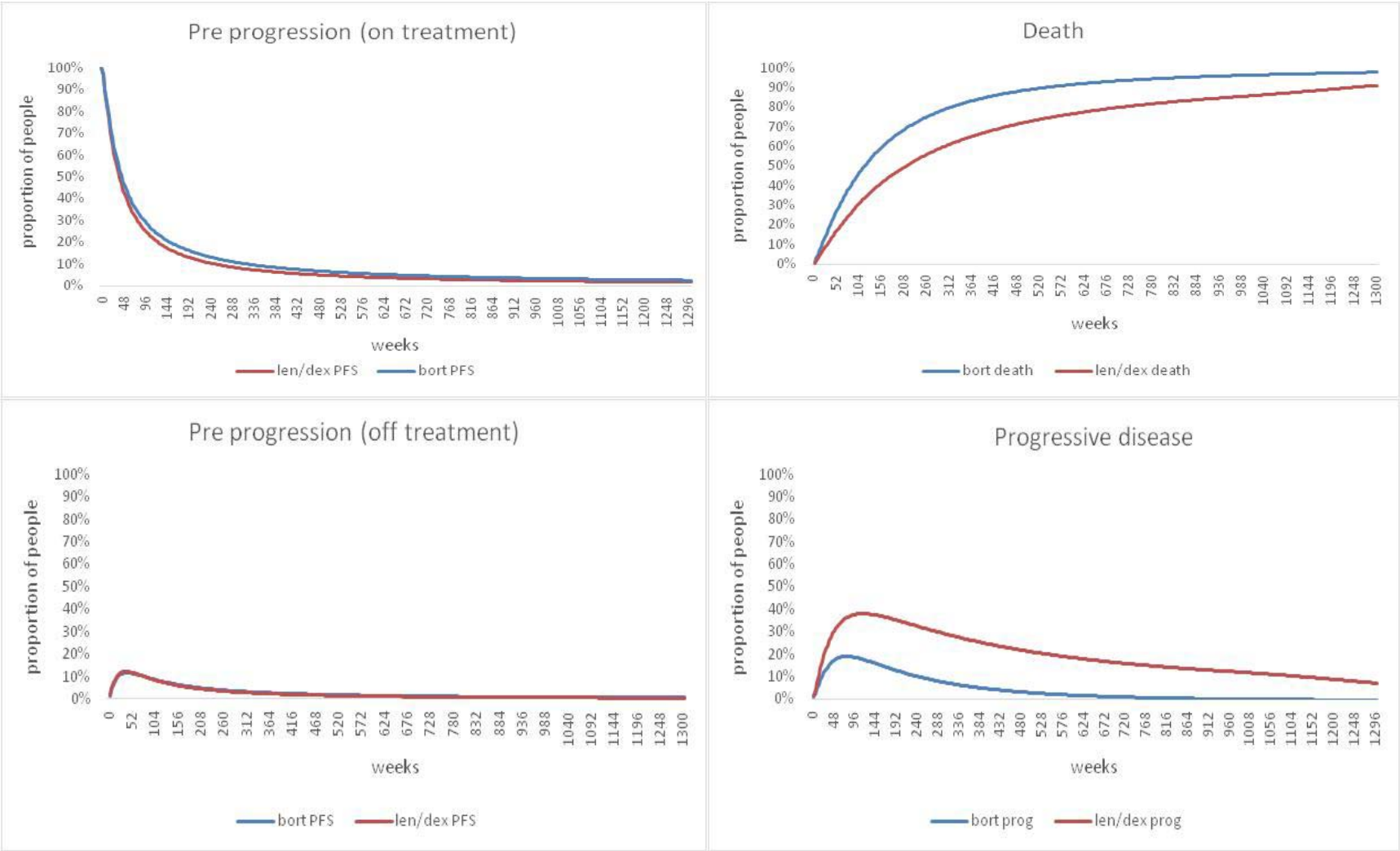
The manufacturer assumed that patients are kept on Bort treatment until they progress or until treatment fails .However, clinical opinion sought by the ERG informed that bortezomib treatment in current clinical practice only lasts for a fixed period of time

The ERG is aware that this might vary according to clinical practice, nonetheless clinical opinion informed us that the average Bort treatment lasts for 8 cycles and not indeterminably.

Adverse events disutility

The ERG corrected the mathematical mistakes found in the calculation of QALYs associated with the pre-progression state. Results are presented in Table 48.

Figure 34. **Markov traces in the updated model with ERG corrections**



With the changes explained in the previous section, Celgene's base case outputs changed as reported in Table 48.

Table 48. Second-line base case results with corrections from Celgene's model[9]

Cost-effectiveness results per patient	Len/Dex (1)	Bort (2)	Incremental value (1-2)
Base Case – second-line only			
Total costs £	£85,546	£86,432	**£-885**
QALYs	3.98	2.49	**1.48**
ICER			**Len dominates**
Allocation of patients to model cycles correction			
Total costs £	£120,268	£158,420	**£-3,8152**
QALYs	4.09	2.66	**1.42**
ICER (compared with base case)			**Len dominates**
ICER with all changes incorporated			**Len dominates**
Duration of Bort treatment correction			
Total costs £	£85,546	£42,839	**£42,707**
QALYs	3.98	2.50	**1.48**
ICER (compared with base case)			**£28,789**
ICER with all changes incorporated			**£54,535**
AE disutility correction			
Total costs £	£85,546	£86,432	**£-885**
QALYs	3.98	2.49	**1.48**
ICER (compared with base case)			**Len dominates**
ICER with all changes incorporated			**£54,369**

[9] All values presented are considering the cost of dex for 64% of patients in the bort arm of the model

7.0 Summary of clinical and cost-effectiveness issues

The ERG is overall concerned with the **model structure** used by Celgene. The approach undertaken raises the following concerns:

- There is 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 it also makes the evaluation of a second line ICER impossible. Furthermore, from a clinical point of view, this seems to reflect a very unlikely scenario.
- After second-line of treatment, the manufacturer only consider 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).
- The ERG question the value of including third and fourth-treatment lines, especially in the intervention arm of the model, since that only cost data is available and that the basket of drugs considered might not accurately reflect current clinical practice.
- Finally, upon request from the ERG, Celgene adjusted the Dex costs in the Bort arm of the model to reflect the fact that 64.3% of Bort patients receive concomitant Dex in the Taverna (20102) study. However, this was not considered to be the base case analysis but instead included as a scenario analysis.

The ERG is overall concerned with the **data extrapolation process** employed by Celgene. The approach taken raises the following concerns:

- Use of PFS HR of 0.9 to estimate the effectiveness of Len/Dex compared to Bort: Clinical opinion sought be the ERG revealed that patients receiving subsequent Bort as a second-line therapy are expected to experience a shorter response duration whilst patients on Len/Dex would be expected to remain in the pre-progression state for longer. This clinical rationale does not seem to agree with a PFS HR below 1. The PFS HR of 0.9 in favour of Bort suggests that patients receiving Bort are in the PFS health state for a longer period of time than the patients receiving Len, who supposedly progress faster. Therefore it appears that the estimated adjusted HR does not accurately reflect the effectiveness of Len/Dex compared with Bort.
- Use of a log-logistic distribution to fit OS data: The decision to change the distribution used to model OS from a piecewise exponential to a log-logistic distribution is not based on a sound argument (i.e. preventing the survival curves from crossing) and more importantly, does not solve the problem of the curves crossing. Furthermore, based on visual inspection of the curves the fitted curve appears to be a very poor fit to MM-010 trial data.
- Overestimation of Len/Dex arm: Both Figure 20 and Figure 19 suggest that OS is overestimated in the economic model, especially later in time. The economic model runs for approximately 25 years, by when around 11% of patients are still alive.

- Overall, it seems like the use of the mean of covariates method to adjust the PFS, and OS curves to reflect MM-010 population characteristics might potentially be skewing these survival estimates. In fact this method has been criticized for the validity of the resulting estimated curves (Ghali 2001).

For all the reasons provided, the ERG lacks confidence in the final ICER presented (for second, third and fourth-line treatments). The ERG lack confidence in the final ICER presented. Celgene's revised economic model reports base case dominant ICERs, which significantly depart from the ICERs presented in TA171. Furthermore the undertaken sensitivity analysis consistently reports dominant ICERs, which is somewhat questionable.

Finally, when the ERG changed the duration of treatment with Bort, hence reducing the costs associated with the drug, the final second-line ICER increased to £54,535. Whilst we do not suggest that this is a reliable alternative ICER it shows the sensitivity of the model outcomes to this parameter. To be noted is that no changes were made to Bort effectiveness. If treatment only lasts for 8 months, as the ERG is suggesting, the effectiveness of Len/Dex would increase (marginally) likely driving the ICER down.

It is the ERG conclusion that the approach taken to modelling the cost-effectiveness of lenalidomide/dexamethasone compared with bortezomib for MM patients presented in this submission needs to be fundamentally reconsidered.

8.0 References

Blade, J. Samson, D. Reece, D. et al. (1998), Criteria for evaluating disease response and progression in patients with multiple myeloma treated by high-dose therapy and haemopoietic stem cell transplantation, Myeloma Subcommittee of the EBMT: European Group for Blood and Marrow Transplant, Br J Haematol. 102(5):1115-23.

Bradburn, MJ. Clark, TG. Love, SB. Altman, DG. (2003), Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. Br J Cancer, 89(4): 605–611.

Brown, RE. Stern, S. Dhanasiri, S. et al. (2013), Lenalidomide for multiple myeloma: cost-effectiveness in patients with one prior therapy in England and Wales. The European Journal of Health Economics : HEPAC : health economics in prevention and care, 14(3):507-14.

Bruce, NJ. McCloskey, EV. Kanis, JA. et al. (1999), Economic impact of using clodronate in the management of patients with multiple myeloma. Br J Haematol, 104(2):358-64.

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).

Crott, R. Versteegh, M. Uyl-de-Groot, C. (2013), An assessment of the external validity of mapping QLQ-C30 to EQ-5D preferences, Qual Life Res, 22:1045-54.

Curtis L. (2012), Unit Costs of Health and Social Care, Personal Social Services Research Unit, [Online], Available: http://www.pssru.ac.uk. [Accessed: 27 January 2014].

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.

Department of Health, (2012), NHS reference costs: financial year 2011 to 2012, [Online], Available: https://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012 [Accessed: 27 January 2014].

Dimopoulos, MA. Chen, C. Spencer, A et al. (2009), Long-term follow-up on overall survival from the MM-009 and MM-010 phase III trials of lenalidomide plus dexamethasone in patients with relapsed or refractory multiple myeloma. Leukemia. 23(11):2147-52.

Drummond, MF. Jefferson, TO. (1996), Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party, British Medical Journal, 313(7052): 275–83.

Durie, BG. Harousseau, JL. Miguel, JS. et al. (2006), International uniform response criteria for multiple myeloma. Leukemia, 20(9):1467-73.

Ghali, WA. Quan, H. Brant, R. van Melle, G. Norris, CM. Faris, PD. Galbraith, PD. Knudtson, ML. (2001), Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models, JAMA, 286(12):1494-7.

Goss, TF. Szende, A. Schaefer, C. et al. (2006), Cost effectiveness of lenalidomide in the treatment of transfusion-dependent myelodysplastic syndromes in the United States, Cancer control : journal of the Moffitt Cancer Center, 13 Suppl:17-25.

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.

Khanna, AJ. Reinhardt, MK. Togawa, D. et al. (2006), Functional outcomes of kyphoplasty for the treatment of osteoporotic and osteolytic vertebral compression fractures, Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA, 17(6):817-26.

Kind, P. Hardman, G. Macran, S. (1999), UK population norms for EQ-5D. Discussion paper 172, [Online], Available: http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20172.pdf [Accessed: 28 January 2014].

Mateos, MV. Richardson, PG. Schlag, R. et al. (2010), Bortezomib plus melphalan and prednisone compared with melphalan and prednisone in previously untreated multiple myeloma: updated follow-up and impact of subsequent therapy in the phase III VISTA trial. J Clin Oncol. 28(13):2259-66.

National Institute for Health and Care Excellence (NICE), (2007), TA129: Bortezomib monotherapy for relapsed multiple myeloma, [Online], Available: http://www.nice.org.uk/nicemedia/pdf/TA129Guidance.pdf [Accessed: 24 January 2014].

National Institute for Health and Care Excellence (NICE), (2009), TA171: Lenalidomide for the treatment of multiple myeloma in people who have received at least one prior therapy, [Online], Available: http://guidance.nice.org.uk/TA171/Guidance/pdf/English [Accessed: 24 January 2014].

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, (NICE), (2011), TA228: Bortezomib and thalidomide for the first-line treatment of multiple myeloma, [Online], Available: http://publications.nice.org.uk/bortezomib-and-thalidomide-for-the-firstline-treatment-of-multiple-myeloma-ta228/guidance [Accessed: 24 January 2014].

Philips, Z. Ginnelly, L. Sculpher, M. et al. (2004), Quality assessment in decision-analytic models: a suggested checklist (appendix 3). In: Review of guidelines for good practice in decision-analytic modelling in health technology assessment, Health Technology Assessment, 8:36.

Richardson, PG. Blood, E. Mitsiades, CS. et al. (2006), A randomized phase 2 study of lenalidomide therapy for patients with relapsed or relapsed and refractory multiple myeloma. Blood, 108(10):3458-64.

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, et al. (2012), Effective response with bortezomib retreatment in relapsed multiple myeloma--a multicentre retrospective survey in Switzerland. Swiss Medical Weekly. 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.