THE CLINICAL- AND COST-EFFECTIVENESS OF LENALIDOMIDE FOR MULTIPLE MYELOMA IN PEOPLE WHO HAVE RECEIVED AT LEAST ONE PRIOR THERAPY: AN EVIDENCE REVIEW OF THE SUBMISSION FROM CELGENE

Report commissioned by: NHS R&D HTA Programme

On behalf of: NICE

Produced by: Peninsula Technology Assessment Group (PenTAG)

Authors:

Martin Hoyle, Research Fellow, Peninsula Technology Assessment Group

Gabriel Rogers, Research Fellow, Peninsula Technology Assessment Group

Ruth Garside, Research Fellow, Peninsula Technology Assessment Group

Tiffany Moxham, Information Scientist, Peninsula Technology Assessment Group

Ken Stein, Professor of Public Health, Peninsula Technology Assessment Group

Correspondence to: Ruth Garside

PenTAG

Noy Scott House

Barrack Road

Exeter EX2 5DW

Ruth.Garside@pms.ac.uk

Date completed : 1 September 2008

About the Peninsula Technology Assessment Group (PenTAG)

- The Peninsula Technology Assessment Group is part of the Institute of Health Service Research at the Peninsula Medical School. PenTAG was established in 2000 and carries out independent Health Technology Assessments for the UK HTA Programme, systematic reviews and economic analyses for the NICE Centre for Public Health Excellence, and systematic reviews as part of the Cochrane Collaboration Heart Group, as well as for other local and national decision-makers. The group is multi-disciplinary and draws on individuals' backgrounds in public health, health services research, computing and decision analysis, systematic reviewing, statistics and health economics. The Peninsula Medical School is a school within the Universities of Plymouth and Exeter. The Institute of Health Research is made up of discrete but methodologically related research groups, among which Health Technology Assessment is a strong and recurring theme. Projects to date include:

- The Effectiveness And Cost-Effectiveness Of Imatinib (STI 571) In Chronic Myeloid Leukaemia - A Systematic Review (2002)
- Screening For Hepatitis C Among Injecting Drug Users And In Genitourinary Medicine (GUM) Clinics - Systematic Reviews Of Effectiveness, Modelling Study And National Survey Of Current Practice (2002)
- Systematic Review Of Endoscopic Sinus Surgery For Nasal Polyps (2003)
- The Effectiveness And Cost-Effectiveness Of Imatinib For First Line Treatment Of Chronic Myeloid Leukaemia In Chronic Phase (2003)
- The Effectiveness And Cost-Effectiveness Of Microwave And Thermal Balloon Endometrial Ablation For Heavy Menstrual Bleeding - A Systematic Review And Economic Modelling (2004)
- Do The Findings Of Case Series Studies Vary Significantly According To Methodological Characteristics?(2005)
- The Effectiveness And Cost-Effectiveness Of Pimecrolimus And Tacrolimus For Atopic Eczema - A Systematic Review And Economic Modelling (2005)

- The Effectiveness And Cost-effectiveness Of Dual Chamber Pacemakers Compared To Single Chamber Pacemakers For Bradycardia Due To Atrioventricular Block Or Sick Sinus Syndrome - Systematic Review And Economic Evaluation (2005)
- The Effectiveness And Cost-Effectiveness Of Surveillance Of Barrett's Oesophagus: Exploring The Uncertainty (2005)
- The Cost-Effectiveness of testing for hepatitis C (HCV) in former injecting drug users. Systematic Review And Economic Evaluation. (2006)
- The Effectiveness And Cost-Effectiveness Of Cinacalcet for Secondary Hyperparathyroidism in end stage renal disease patients on dialysis. Systematic Review And Economic Evaluation (2007)
- The effectiveness and cost-effectiveness of Carmustine Implants and Temozolomide for the treatment of newly-diagnosed High Grade Glioma. Systematic Review And Economic Evaluation (2007)
- The Effectiveness And Cost-Effectiveness of Cardiac Resynchronisation Therapy for Heart Failure. Systematic Review And Economic Evaluation (2007)
- The Effectiveness And Cost-Effectiveness Of Carmustine Wafers And Temozolomide For Newly Diagnosed High Grade Glioma (2007)
- Inhaled Corticosteroids and Long-Acting Beta2-Agonists for The Treatment of Chronic Asthma in Adults and Children Aged 12 Years and Over: a Systematic Review and Economic Analysis (2007)
- Inhaled Corticosteroids and Long-Acting Beta2-Agonists for The Treatment of Chronic Asthma an Children Under the Age of 12 Years: a Systematic Review and Economic Analysis (2007)
- The Effectiveness and Cost-Effectiveness of Cochlear Implants for Severe to Profound Deafness in Children and Adults: A Systematic Review and Economic Model (2007)
- The Effectiveness and Cost-effectiveness of Methods of Storing Donated kidneys from deceased donors: A Systematic Review and Economic Model (2008)

Source of funding

This report was commissioned by the NHS R&D HTA programme.

Competing Interests of Authors

None.

Acknowledgements

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

With many thanks to Joanne Perry and Sue Whiffin (PenTAG) for administrative project support.

The views expressed in this report are those of the authors and not necessarily those of the NHS R&D HTA Programme or the experts consulted. Any errors are the responsibility of the authors.

Contributions of Authors

Martin Hoyle	Critiqued the model provided by industry. Contributed to the writing of the report.
Gabriel Rogers	Critiqued the effectiveness evaluation and the model provided by industry. Contributed to the writing of the report
Ruth Garside	Provided overall project management. Critiqued the effectiveness evaluation provided by industry. Contributed to the writing of the report.
Tiffany Moxham	Undertook literature searches for report, commented on the searches provided by industry.
Ken Stein	Contributed to the critique of the industry submission and report writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse event
CI	Confidence interval
CR	Complete response
Dex	Dexamethasone
ECOG	Eastern Cooperative Oncology Group
EMEA	European Medicines Agency
EORTC	European Organisation for Research and Treatment of Cancer
G-CSF	Granulocyte-colony stimulating factor
HDT	High dose therapy
HR	Hazard ratio
ITT	Intention to treat
ICER	Incremental cost effectiveness ratio
Len	Lenalidomide
Len/Dex	Lenalidomide plus dexamethasone
MM	Multiple myeloma
nCR	Near complete response
NE	Not evaluable
OS	Overall survival
PD	Progressive disease
PPS	Post-progression survival
PR	Partial response
QALY	Quality adjusted life year
QLQ	Quality of Life Questionnaire
SCT	Stem cell transplant
SD	Stable disease
SE	Standard error
TTP	Time to progression
v.	Versus

1 SUMMARY

1.1 Scope of submission

The use of lenalidomide (in combination with dexamethasone) for multiple myeloma (MM) in people who have received at least one prior therapy: this reflects the licensed indication.

Five subgroups are considered. Those with multiple myeloma who have received:

- One prior therapy
- One prior therapy, and are not suitable for treatment with bortezomib
- At least two prior therapies
- Thalidomide as a single prior therapy
- Thalidomide in addition to at least one other prior therapy.

Comparators considered are:

- High dose dexamethasone
- Bortezomib monotherapy (indirect comparison)

Other comparators outlined in the scope are not considered.

1.2 Summary of submitted clinical effectiveness evidence

The evidence is based on two, identically designed RCTs: MM-010 and MM-009 in people with multiple myeloma who had received at least one prior therapy. Pooled analysis of these trials shows increased time to progression with lenalidomide/dexamethasone [median 48.3 weeks *v.* 20.1 wks. HR 0.35 (95% CI 0.29, 0.43)]. Increased overall survival is also seen with lenalidomide/dexamethasone. Based on updated data, median survival with lenalidomide/ dexamethasone increased from [redacted] compared with dexamethasone alone.

A mixed treatment comparison is undertaken to estimate the effectiveness of lenalidomide/ dexamethasone compared to bortezomib monotherapy.

1.3 Summary of submitted cost-effectiveness of evidence

The manufacturer used discreet event simulation to model the cost-utility of lenalidomide/dexamethasone and dexamethasone alone in the five subgroups. Because of the extensive cross-over in the trials, effectiveness data from MM-009 and MM-010 was supplemented with long term follow up based on database information from MRC trials.

Estimated cost per quality adjusted life-year gained ranged from £22,600 to [redacted] in Celgene's submission. However, taking into account a number of limitations and uncertainties in the model assumptions, the ERG suggest these are likely to be underestimates.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

The searches for clinical- and cost-effectiveness data are appropriate and relevant trial data is included.

The RCTs are good quality.

The approach taken to modelling is reasonable.

1.4.2 Limitations

The main threat to validity for the clinical effectiveness data is the high level of crossover in the trials, leading to a strong lenalidomide effect in the comparator arm. This is likely to underestimate treatment effect, especially for overall survival.

Methods used in the mixed treatment comparison are inappropriate. However, when recalculated using more appropriate methods, lenalidomide/ dexamethasone shows increased time to progression [HR 0.56 (95% CI 0.34, 0.91)]. It is not known what the comparison with bortezomib/ dexamethasone, which is commonly used in clinical practice, would show.

Other comparators, also in use in clinical practice, such as thalidomide and repeat initial therapy are not considered.

Cost-effectiveness is extremely sensitive to the estimate of overall survival with dexamethasone alone. Because of crossover in the trials, survival for dexamethasone is taken from experience in MRC trials. This breaks randomisation and, since survival data is historical, may underestimate with survival with dexamethasone.

The model predicts better overall survival for lenalidomide/ dexamethasone than shown in MM-009 and MM-010 and, if adjusted to better predict trial data, incremental cost-effectiveness ratio increases for all comparators.

1.4.3 Areas of uncertainty

In all cost-effectivenss comparisons, there is a large degree of extrapolation of overall survival with lenalidomide/ dexamethasone. Given that per patient costs and quality adjusted life years are strongly dependent overall survival, all estimates are subject to a large degree of uncertainty.

Celgene assume far lower medical management costs than were assumed in the NICE appraisal of bortezomib. If costs from the bortezomib appraisal are used, the cost-effectiveness of lenalidomide/ dexamethasone worsens in all comparisons.

The cost-effectiveness of lenalidomide/ dexamethasone was assessed against bortezomib monotherapy, however, bortezomib is routinely used in combination with dexamethasone. Bortezomib/ dexamethasone may be more effective than bortezomib monotherapy and dexamethasone is cheap. The cost-effeticvenss of lenalidomide/ dexamethasone compared to bortezomib, therefore, may be underestimated.

Celgene did not model the bortezomib response-rebate scheme. If this scheme is modelled, lenalidomide/ dexamethasone cost-effectiveness estimates for lenalidomide/ dexamethasone increase.

2 BACKGROUND

2.1 Critique of the manufacturer's description of the underlying health problem

In Section 4 of their submission, Celgene outline the incidence of Multiple Myeloma (MM) in England and Wales based on credible sources. A short description of current prognosis and of signs and symptoms is also provided. There is little detail about current understandings of the nature of the disease itself except in relation to suggested mode of action for lenalidomide (Len).

Overall, information provided about the underlying health problem is brief but accurate.

2.2 Critique of the manufacturer's overview of current service provision

The Celgene submission rightly indicates in Section 4.1 that there is no established treatment pathway for MM and that initial and subsequent treatments are likely to vary between individuals based on fitness and prior treatment(s). Although guidelines for diagnosis and treatment exist (from the British Society of Haematology, 2005), these are currently being updated. A range of treatment possibilities are suggested in the submission, however it is not clear that this covers all the possibilities and combinations in use in clinical practice, although some may not be well supported by trial evidence. The sequencing of treatments is not considered and it is currently unclear whether this is important in response.

Celgene suggest that the anticipated place for Len in the treatment pathway is likely to be in those patients who have received at least two prior therapies. The rationale provided is that there are already other treatments likely to be used at earlier treatment stages: thalidomide (in combination with melphalan and prednisone) has recently been licensed for first line use while bortezomib, following recent NICE guidance, is likely to be increasingly used in those with MM having one prior therapy.

Celgene justify the selection of comparator treatments (submission Section 4.4) but it is unclear that 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 CRITIQUE OF THE MANUFACTURER'S DEFINITION OF THE DECISION PROBLEM

3.1 Population

The study population is defined in Celgene submission Section 4.1 as people with multiple myeloma who have received at least one prior therapy. This matches the licensed indication for lenalidomide (in combination with dexamethasone – Len/Dex) from the EMEA. Relevant information is provided about UK incidence and age and gender distribution.

3.1.1 Subgroups

A number of subgroups are also considered in the economic modelling. These are people with multiple myeloma who have received:

- one prior therapy.
- one prior therapy and are unsuitable for treatment with bortezomib (due to existing peripheral neuropathy).
- at least two prior therapies.
- previous treatment with thalidomide (by number of prior therapies).

3.2 Intervention

Lenalidomide (Len) is one of a class of agents known as IMiDs which are structural derivatives of thalidomide (Celgene submission Section 4.3). The exact mode of action is yet to be fully understood but is thought to include anti-neoplastic, anti-angiogenic, pro-erythropoietic and immunomodulatory properties.

Len received EMEA marketing authorisation on 14th June 2007 for use in combination with high dose dexamethasone (Dex) among adults with multiple myeloma who have received at least one prior therapy.

The recommended dose of Len is 25mg orally once daily on days 1-21 of repeated 28-day cycles. Oral administration is noted as an advantage over other chemotherapies, including bortezomib, requiring IV administration.

Dex is delivered as daily, oral 40mg doses on days 1-4, 9-12 and 17-20 for the first four 28-day cycles. Subsequently it is delivered at 40mg once daily on days 1-4 of each 28-day cycle.

Treatment is continued until disease progression (indicated by clinical signs and laboratory tests) or unacceptable toxic effects (for full details of how disease progression was defined in the relevant trials, see Section 4.1.6, below)

3.3 Comparators

There is no current standard treatment for MM at first relapse. Guidelines from 2005 are currently being updated. Celgene's submission outlines the decision problem in Section 2 (p.16). Trial data currently exists comparing Len/Dex with Dex alone.

Comparators included in the submission are:

- High dose Dex (based on direct trial data)
- Bortezomib (monotherapy) (based on indirect comparison methods)

Additional comparators outlined in the NICE scope but *not* included in the submission were:

- Bortezomib in combination with Dex.
- Thalidomide-containing regimens
- Repeat initial therapy (including melphalan, vincristine, cyclophosphamide and doxorubicin)

It is argued that, as the combination of bortezomib and Dex is not currently licensed and only data from phase II studies is currently available on this treatment, it is not a suitable comparator treatment. However, expert opinion suggests that this combination may be standard clinical practice.

Thalidomide containing regimens are not included as comparators because thalidomide is licensed only for first-line use. In response to a request for additional information on this point from NICE, the manufacturer undertook a search for randomised evidence comparing second-line thalidomide with Dex monotherapy. They were unable to identify any relevant evidence (response letter, item A1). Nevertheless, this therapy may be used for further treatments in clinical practice.

Repeat initial therapy is not an included comparator because, it is argued, there is no standard treatment, and these therapies are "non-superior to Dex monotherapy" in terms of either myeloma control or tolerability (Celgene submission p.18). In response to a request for additional information on this point from the ERG, the manufacturer undertook a search for randomised evidence comparing second-line chemotherapy with Dex monotherapy. They were unable to identify any relevant evidence (response letter, item C3). Similarly, they did not find any Phase III evidence on the effectiveness of thalidomide at disease relapse (response letter, item A1).

The submission notes (p.32) that there is no established treatment pathway and that options for initial and subsequent therapy include: melphalan and prednisolone, alkylator based combination therapy, the vincristine, doxorubicin and high dose Dex regimen, high dose Dex, bortezomib and thalidomide.

The submission also suggests that – given recent EMEA authorisation for thalidomide (in combination with melphalan and prednisone) as first line treatment in older patients (aged 65+) unsuitable for high dose chemotherapy, and recent NICE guidelines recommending bortezomib for those with first relapse having failed one prior therapy – the most likely place for uptake of Len will be among those who have received at least two prior therapies (p.32).

3.4 Outcomes specified in the systematic review

Outcomes specified by the manufacturer's definition of the decision problem match those suggested in the NICE scope. These are appropriate and clinically meaningful:

- Time to progression (primary outcome)
- Overall survival
- Response rates

- Health-related quality of life
- Adverse effects of treatment

Overall survival is, arguably, a more robust primary outcome, but time to progression is valid.

3.5 Time frame

A timeframe of 30 years is adopted for analysis with 2% of patient still alive at this point.

4 CLINICAL EFFECTIVENESS

4.1 Critique of the manufacturer's approach

4.1.1 Description of search strategies and comment on whether the search strategies were appropriate

Searches were performed in the following databases:

- EMBASE
- MEDLINE In-Process and Old Medline
- The Cochrane Library
- ISI Science Citation InDex Web of Knowledge
- ISI Biosis Preview
- ISI Proceedings
- National Research Register (NRR)
- Current Controlled Trials
- ClinicalTrials.gov
- American Society of Haematology (ASH)
- American Society of Clinical Oncology (ASCO)
- European Haematology Association (EHA)
- Celgene Company literature

Separate search strategies were provided for each database in the manufacturer's submission except for ISI Proceedings; however, as the other ISI databases were searched with a single strategy and, based on the number of hits recorded, it appears the same strategy was used for this search. The database searches are based on a conjunction of a term identifying a MM population and a term identifying Lenalidomide as an intervention. For each term, a combination of thesaurus headings and free-text search-words is used. No comparators or outcomes are specified to limit the searches. For EMBASE, MEDLINE and ISI databases, an extensive filter was applied to limit the search to randomised trials. An additional meta-analysis limit was applied in EMBASE, and a general clinical trials and a comparative trials filter included in MEDLINE. For EMBASE, MEDLINE, and ISI databases, a human filter was used.

The ASH, ASCO, and EHA, NRR, and CCT searches looked only for the intervention while the ClinicalTrials.gov search combined population and intervention.

All the search strategies and resources used were appropriate, replicable, and the resulting hits appear correct related to the search date and database/interface used.

The time-span column of the search table should have included precise dates instead of "all available years"; however, additional information was not requested from industry as it was clear when the searches were run what the search time-spans were (for example, Cochrane 2008 version 1).

The number of hits in each database, as taken from both the search strategies and the table provided, include four discrepancies. The total on the search strategy for MEDLINE shows 211 hits but on the table records 212. In the QUOROM flow diagram (p.38) MEDLINE hits are shown as 215 (211 in original searches), company submission papers are 19 (17 in Appendix 2), and EHA 25 (21 in Appendix 2 and original searches). The EHA total for included and excluded papers in the QUOROM flow diagram equals 22.

4.1.2 Statement of the inclusion and exclusion criteria and comment on whether they were appropriate

In Celgene's review of clinical effectiveness, studies were included if they were:

- RCTs or systematic reviews
- Subgroup analyses and open label extensions from relevant RCTs
- Studies comparing Len/Dex with another therapy or placebo
- Patients with MM who have received at least one prior therapy
- Full text publications and abstracts

Additional inclusion criteria were:

- Studies that report overall survival, time to progression, progression free survival, clinical response, quality of life or safety outcomes.

Exclusion criteria were:

- Non-systematic reviews, editorials, comments or letters
- Animal, in vitro or pharmacodynamic/ pharmacokinetic studies.
- Patient population is treatment naïve/ newly diagnosed MM patients.

These criteria seem appropriate to identify all relevant evidence with appropriate methods bearing directly on the effectiveness of Len.

In addition, however, the submission relies on indirect comparison between Len and bortezomib, and other possible indirect comparisons – e.g. with second-line thalidomide or repeat chemotherapy – are dismissed in part because of lack of available evidence. In its initial submission, the manufacturer provided no evidence of systematic searches for this kind of evidence. In response to questions from the ERG/NICE, limited systematic searches were subsequently undertaken for randomised evidence comparing (a) second-line thalidomide and (b) repeat chemotherapy with Dex monotherapy (see Response Letter, items A1 and C3, respectively). No additional evidence was found.

4.1.3 Table of identified studies. What studies were included in the submission and what were excluded?

The search results presented by the manufacturer are inconsistent and, although the ERG asked for clarification, the numbers provided do not totally tally. According to the QUOROM flowchart (Celgene submission Section 5.2.3, p.38),

- 1,071 hits were generated by the initial searches (the total number of hits presented sums to 1,075);
- 924 were excluded on the basis of title and/or abstract (exclusions sum correctly);
- 148 studies were reviewed in full (1,071 – 924 = 147, implying one study is double-counted in results);
- 129 references were identified as duplicates or otherwise excluded on perusal of full text (exclusions sum correctly); leaving
- core included evidence-base of 19 references, reporting 2 RCTs

We do not, however, believe that relevant studies have been missed.

All publications relating to two phase III trials – MM010 and MM009 – were included. This includes two full trial reports published in 2007 in the *NEJM*[1;2] and 15 abstracts presented at conferences in 2006 and 2007 which presented preliminary results or *post-hoc* subgroup analyses. Details are presented in Table 1 (adapted from submission Table 7, p. 36).

Table 1: Included Lenalidomide trials

RCT		Reference	
MM-009	Primary publication:	Weber et al. (2007)[1]	full article
	Updated data:	Weber et al. (2007)[3]	abstract
MM-010	Primary publication:	Dimopoulos et al. (2007)[2]	full article
	Updated data:	Dimopoulos et al. (2007)	unpublished conference poster
	Interim analysis:	Dimopoulos et al. (2005)[4]	abstract
Pooled	Interim analysis:	Dimopoulos et al. (2005)[5]	abstract
	Subgroup analyses:	Chanan-Khan et al. (2006)[6]	abstract
		Chanan-Khan et al. (2006)[7]	abstract
		Stadtmauer et al. (2006)[8]	abstract
		Miguel et al. (2007)[9]	abstract
		Harousseau et al. (2007)[10]	abstract
		Foa et al. (2007)[11]	abstract
		Chanan-Khan et al. (2007)[12]	abstract
		Niesvizky et al. (2006)[13]	abstract
		Weber et al. (2006)[14]	abstract
		Wang et al. (2007)[15]	abstract
		Wang et al. (2006)[16]	abstract
		Wang et al. (2006)[17]	abstract
		Bladé et al. (2006)[18]	abstract

The two RCTs were of identical design, but MM-009 was undertaken in the USA and Canada, while MM-010 was undertaken in Europe, Israel and Australia. A description and critique of the design to which both studies conformed is given in the remainder of this chapter.

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

We did not identify any relevant studies that were not included in the submission.

4.1.4.1 Ongoing studies

We did not identify any relevant ongoing studies that were not included in the submission.

4.1.5 Description and critique of the manufacturer's approach to validity assessment

4.1.5.1 Internal validity of evidence

The manufacturer provides a detailed narrative description of the trial methods (Celgene submission Section 5.5 p. 39–64), the key points of which are summarised in a (unnumbered) table (pp. 62-64). This table is reproduced with additional comments in Table 2. The trials seem to be well conducted with little possibility of randomisation or blinding being compromised and adequate samples to detect differences between the groups. Extensive crossover is the biggest threat to validity with the effectiveness of Len/Dex likely to be underestimated, especially as regards overall survival benefit.

Table 2: Validity assessment of available evidence

Question	Celgene response	ERG's comment
How was allocation concealed?	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 Interactive Voice Response System (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.	According to this description, it is unlikely that blinded allocation was compromised in these trials. It is possible that adverse haematological effects with Len may indicate active treatment.

Question	Celgene response	ERG's comment
What randomisation technique was used?	A stratified randomization list was independently generated before the study was initiated, which randomized the subjects in a 1:1 ratio to either the Len/Dex group or the Dex group. Randomization was done centrally using an IVRS. Randomization was centralized and stratified by three factors: baseline serum β2-microglobulin, prior treatment with high-dose chemotherapy or SCT or no prior treatment, and number of prior anti-myeloma regimens.	This is appropriate.
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 are justified by the reported outcomes (TTP HR about 2.8 in both RCTs; OS HR 2.3 [MM-009] & 1.5 [MM-010]) indicating the studies are adequately powered.
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 are contacted every 6 months during the follow-up phase.	It is unclear how participants were followed up after treatment discontinuation for AEs.
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 "adjudication committee" is not mentioned anywhere else in the submission. We note, with reference to the primary outcome (TTP), that disease progression is defined by a mixture of objective (haematological) and subjective (appearance of bone / soft tissue lesions) criteria. It is not clear who assessed this outcome and, especially as far as the subjective dimension is concerned, it would be important to ensure that allocation was concealed from the assessor. 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. Response levels other than PD appear to be defined by objective criteria only.

Question	Celgene response	ERG's comment
Was the design parallel-group or crossover? Indicate for each crossover trial whether a carry-over effect is likely.	It was a parallel-group design. Patients in the Dex group were only allowed to roll over to receive lenalidomide after disease progression, or cross over to receive Len/Dex after the IDMC had declared the studies could be unblinded. Carry-over effect is not applicable in these two trials.	Across both RCTs, 47% of participants randomised to receive Dex alone chose to receive additional Len when it became available at study unblinding (MM-009: 101/176=57%; MM-010: 63/175=36%). It is concluded that OS results, in particular, "for the Dex arm includes a strong Len effect rather than the pure Dex outcome." (Celgene p. 86)
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 is being 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.	The manufacturer notes that the significantly increased risk for thrombosis in the North American trial (MM-009) was found to be associated with concomitant erythropoietin use, which is uncommon in England and Wales (Celgene p. 100)
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 (5;6). The patients enrolled in the trials are slightly younger and have a better status at baseline than those that might be seen in UK clinical practice. However, the trial data shows Len/Dex significantly improves outcomes over Dex regardless of age and performance status [19;20].	The median age at onset in England and Wales is 70, with just 15% aged under 60. The trials contain patients in the treatment arm with a median age of 63 (range 33–84) in MM-010 and 64 (range 36–86) in MM-009. People with an ECOG performance status of three or more were excluded (for details of the ECOG scale, see Table 3). Our expert clinical advice is that MM patients who would be considered for active treatment represent a subset of the whole MM population and, on average, would be younger and have better performance status and so may be similar to those enrolled in trials.

Question	Celgene response	ERG's comment
For pharmaceuticals, what dosage regimens were used in the RCT?	Dosage regimens were the same as those detailed in the Summary of Product Characteristics.	For details of standard Len/Dex regimen, see Section 3.2, above. The control arm in the RCTs received Dex monotherapy in identical dosage to that provided as concomitant therapy in the Len/Dex arm.
Were the study groups comparable?	Yes, the demographic and baseline characteristics of the study groups are comparable.	The selection of baseline characteristics (Celgene Table 8, p. 46-48) and details of previous therapy (Celgene Table 9, p. 49) provided seems reasonably comprehensive and the cohorts well matched.
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.	See Section 4.1.7, below.
Was an intention-to-treat analysis undertaken?	Yes	Intention to treat analysis was undertaken. There are slight anomalies in data presented. According to design summary for the trials in Figure 1, (Celgene p.40), 705 people were recruited to trials MM-009 and MM-010. The pooled safety database as of 31 December 2005 includes 703 patients (Celgene p.96) However, effectiveness data presented in the trials is based on 692 patients (Celgene p.102). We note that, although all time-to-event analyses appear to be based on the full ITT cohort, we note that data were censored for a variety of reasons, including dropout due to intolerable adverse events (Celgene p.65). Arguably, this might be seen as a violation of ITT principles.

Question	Celgene response	ERG's comment
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 Dex group 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 Dex group in general, and is likely to explain the decreasing difference in OS between the study groups over time. TTP in the Dex arms is relatively unaffected by the treatment crossover, because most patients had developed progressive disease (PD) when the studies were unblinded – 75.0% in MM-009 and 81.1% in MM-010.	Agreed, any impact is likely to be in favour of the control arm.

Response source: Celgene submission p. 62-64.

4.1.5.2 External validity of evidence

Of the 705 patients in the two RCTs, 15 were treated in the UK (all in MM-010). The manufacturer accepts that "The strict inclusion and exclusion criteria [of the RCTs] meant that the range of patients were [sic] slightly younger and of higher performance status than might be seen in clinical practice [in the UK]." (Celgene submission p. 110)

The median age at onset in England and Wales is 70, with just 15% aged under 60. The trials contain patients in the treatment arm with a median age of 63 (range 33–84) in MM-010 and 64 (range 36–86) in MM-009. People with an ECOG performance status of three or more were excluded (details of the scale are shown in Table 3).

Table 3: ECOG performance status

Grade	ECOG
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
5	Dead

However, as stated above, although trial data might be based on patients who are fitter than average, active chemotherapy treatment may also only be offered to this group.

4.1.6 Description and critique of the manufacturer's selection of outcomes

The outcomes on which Celgene's submission focuses are a direct reflection of those selected in the two RCTs:

- The primary efficacy outcome in studies MM-009 and MM-010 was time to disease progression (TTP), calculated as time from randomisation to either

  - disease progression according to myeloma response criteria (see below); 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 if the participant in question had either

  - not progressed at the time of analysis; or
  - withdrawn from the treatment phase before documented progression; or
  - died of causes not related to multiple myeloma; or
  - received another antimyeloma therapy without documented progression or intolerable adverse events.

  Additional sensitivity analyses were performed, analysing

  - progression-free survival;
  - 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 outcomes analysed are

  - overall survival (OS);
  - myeloma response rate (criteria based on the International Uniform Response Criteria;[21;22] see Table 4);
  - "functioning and quality of life", consisting of
    - time to first skeletal-related event; and

- time to first worsening of ECOG performance status; and

o adverse events.

Table 4: Myeloma response determination criteria

Outcome	Criteria for Classification[a]
Complete response (CR)	▪ 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)	▪ 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 <200 mg 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 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]. □ 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.

Outcome	Criteria for Classification[a]
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 hypercalcemia (serum calcium level, corrected for albumin concentration, >11.5 mg/dL [2.8 mmol/L]) not attributable to any other cause. 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 ≥500 mg/dL (5 g/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 ≥200 mg/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 hypercalcemia (serum calcium level, corrected for albumin concentration, >11.5 mg/dL [2.8 mmol/L]) not attributable to any other cause.

Source: Celgene submission Table 10, pp. 62-63; based on the International Uniform Response Criteria[21;22]

[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; non-evaluable disease comprises malignant pleural [?additional text omitted in submission]

Critique

This selection of outcomes seems to provide a reasonable range of dimensions in which to assess the clinical effectiveness of Len/Dex.

It is regrettable that no measures directly capturing participants' health-related quality of life were recorded. The manufacturer reports that the European leg of the Phase IIIb expanded access study will provide quality of life data (EORTC QLQ-C30 and EORTC MY-24) (Celgene submission p. 45). This data is not yet available and, as it is

uncontrolled and observational, this study will not provide comparative information on relative health-related quality of life.

The adequacy of outcome measures to predict OS is a recurring problem in MM research. Long-term follow-up in first-line MM therapy suggests that complete response rate is not a valid surrogate for OS[23] and that TTP may also be a poor proxy for OS.[24] If an antimyeloma therapy is shown to prolong time to disease progression, this will doubtless be seen as valuable in its own right; however, it cannot be assumed that an OS benefit will necessarily accrue.

Furthermore, the manufacturer accepts that, in the evidence-base underpinning this appraisal, the measurement of OS is likely to be substantially compromised by post-unblinding crossover, whereby 47% of participants randomised to receive Dex alone chose to receive additional Len. This effect is much more marked in OS data than in TTP results because a smaller proportion of the Dex cohorts had experienced the event in question at the time of unblinding (35% compared to 78% for TTP).

We note that the criteria for interim analyses, and the potential for crossover, were pre-specified, and based on accepted criteria (Celgene submission p.59)

4.1.7 Description and critique of the manufacturer's statistical approach

4.1.7.1 Primary statistical analysis of raw data

Intention to treat calculation are performed.

It is not clear why safety and efficacy data are based on different numbers of patients.

Incomplete information on the methods used to analyse data is reported by the manufacturer. In particular, it is unclear whether hypothesis tests present one- or two-tailed *p*-values. Although it is explicitly stated that "Formal statistical hypothesis tests… were conducted at the 2-sided, 0.05 level of significance" (Celgene submission p.59), there are several juncture in the presented evidence where this is not the case.

Above all, where specified, all log-rank tests are one-tailed (and, therefore, we assume that all such tests follow this method). We note that, in the published reports of the

RCTs, power calculations are reported differently: for MM-009, it is stated that the trial was designed to detect a difference measured by a *one-sided* log-rank test at α=0.025 whereas, in MM-010, the authors report that they powered their trial to detect a treatment effect by a *two-sided* log-rank test at α=0.05. While these calculations are mathematically identical, they carry important implications for the interpretation of results (although the report of MM-009 is explicit that all *post hoc p*-values were two-sided).

The use of one-sided *p*-values is open to criticism. Such a step would only be justifiable on the assumption that the only possible difference in inter-arm efficacy would be one favouring Len/Dex. If it is asserted that such a position is appropriate, it is unclear on what basis that assertion is made. Nevertheless, because, in most instances, highly significant *p*-values are generated by these tests, it seems unlikely that this decision has a substantial impact on findings.

4.1.7.2 Meta-analysis

Although its primary strategy for synthesising outcomes from the two relevant RCTs is to pool data at patient level, the manufacturer also provides a limited conventional meta-analysis, providing trial-level aggregation of data (Celgene submission p. 88-90). Only two outcome measures are considered; both are time-to-event outcomes. For OS, the crude proportion of participants alive in each arm of the two trials at analysis date is compared, to provide an odds ratio for survival. For TTP, a meta-analysis of differences in median TTP times is presented.

Both choices of measure are open to criticism. The use of dichotomised survival proportions is widely recognised as inadequate and potentially misleading, because (a) it relies on an arbitrary cut-off (in this instance, the time of data analysis); (b) by reducing data to a binary status, it discards information about differences in length of survival achieved; and (c) it does not account for differences in follow-up time.[25] Similarly, it has been shown that median survival time is an unreliable summary measure of treatment effect in time-to-event analyses.[26] In both instances, it would be more appropriate to combine hazard ratios (we have performed this analysis; see Section 6.1.1).

4.1.7.3 Mixed treatment comparison

A limited mixed treatment comparison, considering bortezomib monotherapy and Len/Dex, is provided by Celgene in their submission (p.90). Because it is the only population for which bortezomib is approved in England and Wales, the comparison is limited to data reflecting trial participants who received therapy at first relapse (i.e. one prior therapy only).

The bortezomib data is provided by the APEX trial, details of which are outlined (p.91-94). This trial was also subject to considerable post-treatment-phase cross-over with over 62% of the Dex arm receiving bortezomib (Celgene submission p.93).

As in the pairwise meta-analysis, TTP is considered in terms of median value. This analysis is flawed, for a number of reasons:

- As discussed in section 4.1.7.2, median survival is a suboptimal measure of time-to-event data;
- SEs are mostly assumed, by copying values from treatments for which an estimate of dispersion is available to those for which there are no such data. In an analysis of this type, which weights evidence according to its precision, it is inappropriate simply to substitute missing data with that from comparator treatments.
- For one trial, the primary data-point – median TTP – is also assumed.

Taken together, these shortcomings make it difficult to draw any valid inferences from the attempted analysis. As for the pairwise meta-analysis, we suggest that it would be more appropriate to analyse TTP in terms of hazard ratio. This would solve all the problems listed above, as full data (estimate of HR and measure of dispersion) is available for all three trials. We have performed this analysis (see Section 6.1.2).

4.1.8 Summary statement about the review of clinical effectiveness

The searches conducted were appropriate and evidence from two good quality RCTs is identified. Results of a pre-planned interim analysis led to substantial cross over between arms.

Methods used for mixed methods comparison are based on erroneous assumptions.

4.2 Summary of submitted evidence

4.2.1 Summary of results

4.2.1.1 Primary outcome: Time to Progression

Because it is the primary outcome in both underlying RCTs, time to progression (TTP) is given precedence in the manufacturer's appraisal of clinical effectiveness (for details of how the measure was defined, see Section 4.1.6, above).

Results reproduced here are those measured at study unblinding (28 June 2005 [MM-009] / 3 August 2005 [MM-010]). In the manufacturer's submission results are also presented for TTP at time of interim analysis (Celgene submission Table 12, p. 67) but are not reproduced here, as the data is less complete than those collected subsequently.

Whole population

Table 5 shows median TTP in each RCT individually, and in a pooled analysis aggregating both populations.

Table 5: Time to Progression

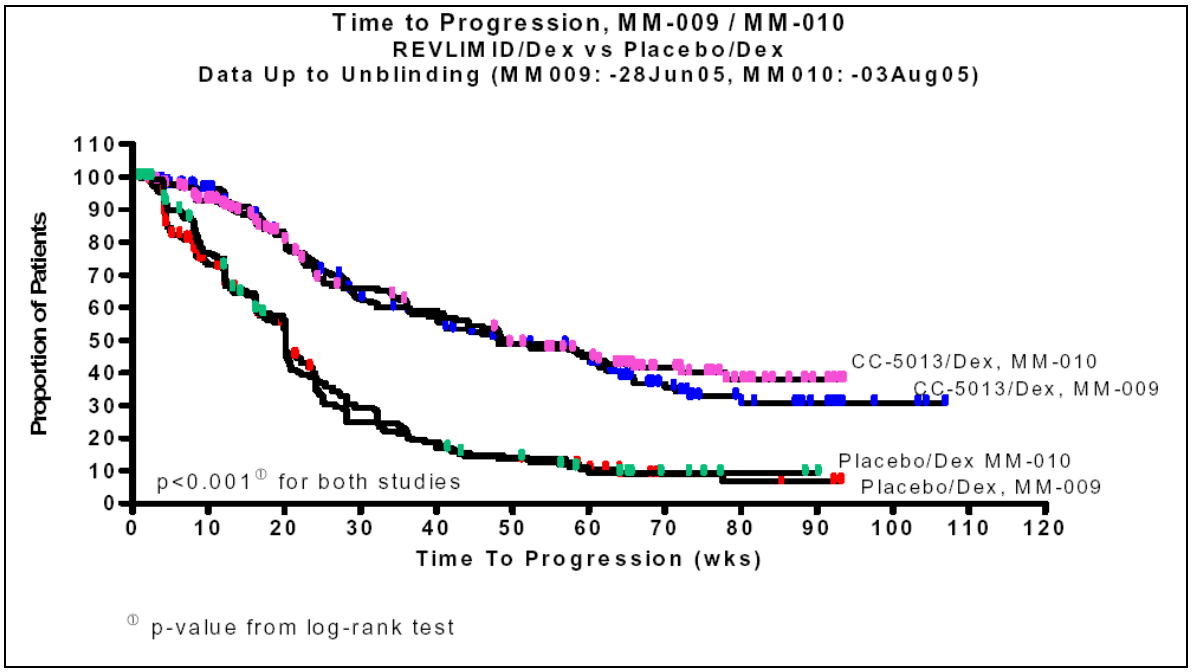
	MM-009[a]		MM-010[a]		pooled[b]	
	Len/Dex	Dex	Len/Dex	Dex	Len/Dex	Dex
N	177	176	176	175	353	351
Median – wk [95% CI]	48.1 [36.9, 61.4]	20.1 [16.7, 23.1]	48.7 [40.9, 72.1]	20.1 [18.1, 20.7]	48.3 [41.1, 60.1]	20.1 [19.9, 20.7]
HR [95% CI]	0.354 [0.270, 0.466][c]		0.351 [0.266, 0.463][c]		0.35 [0.29, 0.43]	
Log-rank *p*-value	< 0.001		< 0.001		< 0.001	

[a] Source: Table 13, p. 68
[b] Source: Table 22, p. 83
[c] The reported data has been inverted, because it is presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

It is noticeable that the data from the two individual trials agree extremely well and, as a result, provide a similar estimate (subject to a lesser degree of uncertainty) when pooled.

Kaplan-Meier curves for TTP at the time of unblinding are presented for each trial in the submission (Celgene submission Figure 4, p. 69 and Figure 5, p. 70, for MM-009 and MM-010, respectively). Figure 1 reproduces the depiction of this data from the EMEA scientific discussion, in which the two pairs of curves are superimposed on each other in the same graph.

Figure 1: Time to progression in both RCTs



source: EMEA scientific discussion, p. 20 Stratified according to relapse phase

TTP results for trial participants who were at first relapse at baseline and those at second/subsequent relapse are presented in Table 6 and Table 7, respectively.

Table 6: Time to Progression (1 prior therapy)

	MM-009[a]		MM-010[a]		pooled[b]	
	Len/Dex	**Dex**	Len/Dex	**Dex**	**Len/Dex**	**Dex**
N	68	67	56	57	124	147[c]
Median – wk [95% CI]	61.4 [redacted]	21.1 [redacted]	NE [redacted]	20.1 [redacted]	62.3 [redacted]	19.9 [redacted]
HR [95% CI][d]	[redacted]		[redacted]		[redacted]	
Log-rank *p*-value	<0.001		<0.001		[redacted]	

[a] source: Table 16, p. 76
[b] source: Table A4, p. A49
[c] we note the reported pooled sample size exceeds the sum of the two trial-specific samples; we are unable to explain this discrepancy
[d] The reported data has been inverted, because is it presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

Table 7: Time to Progression (≥2 prior therapies)

	MM-009[a]		MM-010[a]		pooled[b]	
	Len/Dex	**Dex**	**Len/Dex**	**Dex**	**Len/Dex**	**Dex**
N	109	109	120	118	229	227
Median – wk [95% CI]	40.1 [redacted]	19.9 [redacted]	48.1 [redacted]	20.1 [redacted]	41.3 [redacted]	20.1 [redacted]
HR [95% CI][c]	[redacted]		[redacted]		[redacted]	
Log-rank *p*-value	<0.001		<0.001		[redacted]	

[a] Source: Table 16, p. 76
[b] Source: Table A4, p. A40
[c] The reported data has been inverted, because it is presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

In both strata, there is excellent agreement between the two trials. There is a strongly significant treatment effect in each stratum, though it appears that Len/Dex has a marginally greater relative effect in those with one prior therapy at study recruitment. Median TTP in the Dex arm is similar across strata, but is prolonged to a greater extent in the Len/Dexe arm for those at first relapse (in whom median TTP with Len/Dex is approximately three times as long as observed in the control arm) than in those who have had two or more prior therapies (in whom TTP with Len/Dex is approximately double that observed in the control arm).

Time to progression in various subgroups

Appendix 5 of Celgene's submission provides detail on TTP in subcategories according to a variety of different criteria pertaining to the participants' characteristics, clinical history and therapy during the trial. Table 8 gives a summary of these analyses. The analysis according to prior thalidomide exposure status is especially important, as it forms the basis of a subgroup for which separate analyses are performed in the cost-effectiveness model.

A significant TTP benefit for those taking Len/Dex was observed in each of the subgroups for which data is presented. It is noticeable that median TTP varies relatively little in the Dex arms of each analysis, with most estimates around 20 weeks. In contrast, there is a wide range of estimated medians for those receiving Len/Dex. In comparisons performed within subgroups of participants taking Len/Dex, complete or very good partial response was associated with improved TTP, as was a stable dose of Dex during the trial.

Table 8: Summary of Time to Progression results in various subgroups

	Len/Dex		Dex			
	N	median TTP – wk [95%CI]	N	median TTP – wk [95%CI]	HR[a]	*p*
beta-2M[b]:						
<2.5mg/L	103	[redacted]		[redacted]	[redacted]	<0.001
>2.5mg/L	250					<0.001
prior therapy:						
HDT/SCT[c]	206	[redacted]		[redacted]	[redacted]	<0.001
no HDT/SCT[c]	147					<0.001
thalidomide[d]	**127**	**36.4 [29.1, 48.1]**	**147**	**19.9 [16.1, 20.1]**	**0.376 [0.279, 0.508]**	**<0.001**
no thalidomide[d]	**226**	**60.1 [48.1, 80.0]**	**204**	**20.3 [20.1, 24.0]**	**0.336 [0.260, 0.435]**	**<0.001**
thalidomide-resistant[e]	54	30.3 [24.1, 48.1]	62	16.1 [12.1, 24.1]	0.423 [0.272, 0.659]	<0.001
bortezomib[f]	27	60.1 [29.1, 70.0]	27	14.3 [8.6, 24.0]	0.250 [0.124, 0.504]	<0.001
response to thalidomide[g]:						
[redacted]						
performance status[h]:						
ECOG 0	192	44.3 [36.1,58.6]	196	20.1 [18.7, 23.1]	0.393 [0.295, 0.523]	<0.001
ECOG 1 or 2	152	57.0 [40.1, 71.4]	150	20.1 [16.6, 20.4]	0.333 [0.254, 0.436]	<0.001
type of MM[i]:						
IgA	72	44.3 [34.3,60.1]	82	16.4 [12.1, 20.1]	0.302 [0.199, 0.457]	<0.001
non-IgA	281	57.0 [40.1, 71.4]	269	20.1 [20.0, 22.1]	0.367 [0.295, 0.458]	<0.001
age[j]:						
<65	192	48.1 [40.1, 61.3]	198	20.1 [19.9, 21.6]	0.361 [0.280, 0.466]	NR[k]
65-75	125	57.3 [36.1, 80.0]	121	20.1 [16.1, 22.1]	0.337 [0.238, 0.476]	NR[k]
>75	36	57.0 [24.1, NE]	32	20.1 [16.1, 32.3]	0.357 [0.183, 0.699]	NR[k]
creatinine clearance[l]:						
<30 ml/min	16	33.9 [27.4, 61.3]	12	20.1 [16.6, 24.1]	0.302 [0.097, 0.943]	NR[k]
30-50 ml/min	42	57.3 [36.1, 80.0]	34	12.1 [8.6, 20.1]	0.241 [0.133, 0.438]	NR[k]
50-80 ml/min	125	52.1 [35.1, 72.1]	132	20.1 [18.7, 24.1]	0.384 [0.276, 0.534]	NR[k]
>80 ml/min	158	48.4 [41.0, 64.1]	163	20.1 [20.0, 23.1]	0.344 [0.259, 0.459]	NR[k]
best myeloma response[m]						
CR/VGPR	114	NE			2.43 [NR][n]	<0.001[n]
PR	100	48.1 [NR]				
intra-study dexamethasone[o]						
stable dose	224	44.1 [36.1,60.1]			0.529 [0.337, 0.831][n]	0.005[n]
reduced dose	56	NE [58.1, NR]				
pre-existing neuropathy[p]						
[redacted]					[redacted][n]	[redacted][n]

NE = not evaluable; NR = not reported

[a] The reported data for inter-arm comparisons has been inverted where it is presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)
[b] Source: Celgene submission Table A6, p. A42.
[c] Source: Celgene submission Table A8, p. A44.
[d] Source: Celgene submission Table A10, p. A47.
[e] Source: Celgene submission Table A12, p. A49.
[f] Source: Celgene submission Table A14, p. A1.
[g] Source: Celgene submission Table A13, p. A1.
[h] Source: Celgene submission Table A16, p. A3.
[i] Source: Celgene submission Table A18, p. A5.
[j] Source: Celgene submission Table A24, p. A11.
[k] A *p*-value <0.05 may be inferred, since the 95%CI for the HR does not include 1.
[l] Source: Celgene submission Table A26, p. A13.
[m] Source: Celgene submission Table A20, p. A7.
[n] Inter-stratum comparison.
[o] Participants with stable Len dose only. Source: Celgene submission Table A21, p. A8.
[p] Source: Celgene submission Table A23, p. A10.

4.2.1.2 Secondary outcome: Overall survival

In its submission, the manufacturer explains that OS results should be treated with caution, because follow-up is far from complete (thus data is heavily censored) and results are likely to be substantially influenced by post-unblinding crossover from the control arm (across both trials, 47% of participants who were randomised to receive Dex alone received Len once allocation had been unblinded) (Celgene submission see p. 86). This effect is much more marked in OS data than in TTP results because a smaller proportion of the Dex cohorts had experienced the event in question at the time of unblinding (34.5% had died compared to the 8.1% who had experienced disease progression, in TTP analyses).

Whole population

In both Celgene's submission and the published trial reports, precedence is given to OS data measured in May 2006 (3yr 3mo from study initiation for MM-009; 2yr 8mo for MM-010). This data is reproduced in Table 9.

Table 9: Overall survival

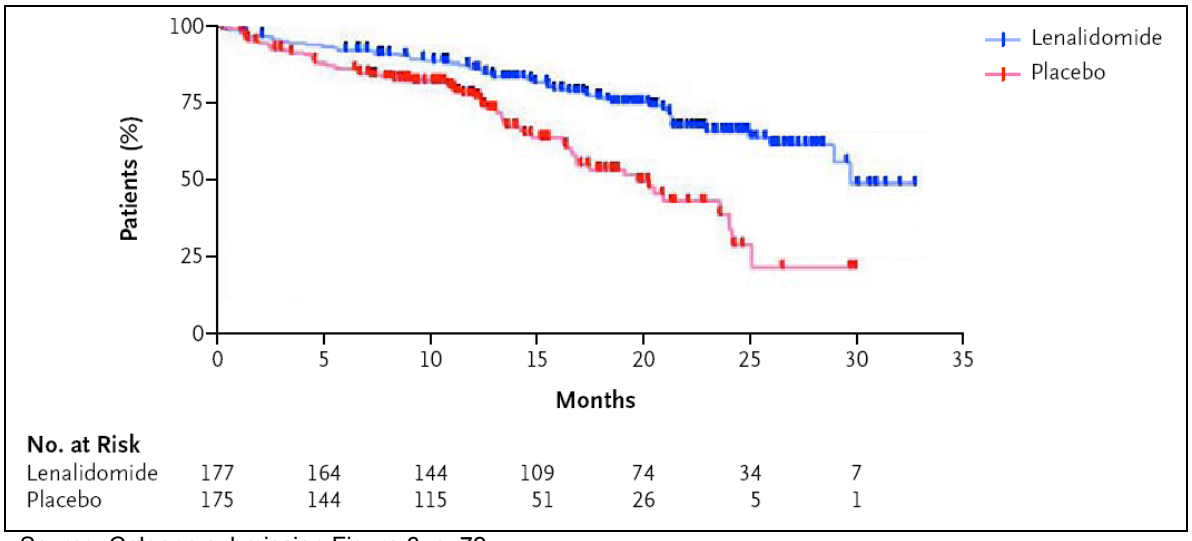
	MM-009		MM-010	
	Len/Dex	**Dex**	**Len/Dex**	**Dex**
N	177	176	176	175
Median – months[a]	29.6	20.2	NE	20.6
Hazard Ratio [95% CI]	0.44 [0.30, 0.65]		0.66 [0.45, 0.96]	
Log-rank Test p-Value	<0.001		0.03	

Source: Celgene submission Table 14, p. 72

[a] no measure of dispersion provided

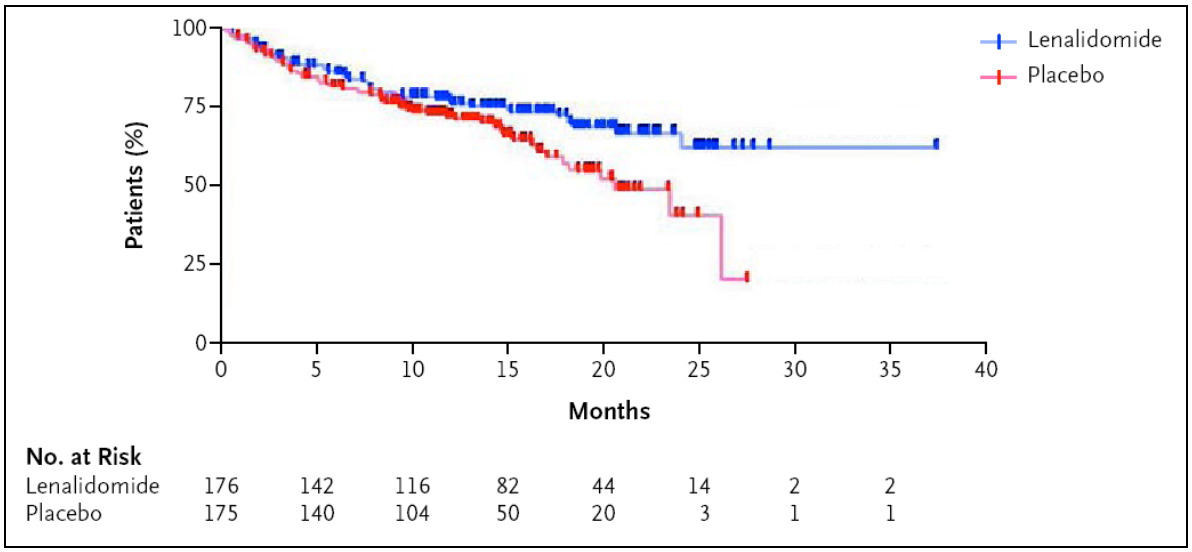
Kaplan-Meier curves depicting observed OS in the two RCTs are reproduced in Figure 2 and Figure 3.

Figure 2: Overall survival in MM-009



Source: Celgene submission Figure 6, p. 72

Figure 3: Overall survival in MM-010



Source: Celgene submission Figure 7, p. 73

The agreement between these two pairs of curves is imperfect: earlier stages of follow-up in the Len/Dex arm of MM-009 show a more positive survival profile but, as the tail of the curve is approached, MM-010 takes over as the more favourable.

A 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).

Whole population – updated results

Celgene also present updated OS results as at January 2007 (both trials pooled). Because it comes from later in the population's progress, this data is more mature (with death having occurred in 47.2% of the whole population), but it is also more susceptible to confounding by post-unblinding crossover in the Dex arm. Results for the whole population are reproduced in Table 10.

Table 10: Updated, pooled OS data

	Len/Dex	Dex
N	353	351
Median – months[a]	34.4	30.7
Hazard Ratio [95% CI][b]	[redacted]	
Log-rank Test p-Value	0.015	

Source: Celgene submission Table 25, p. 85

[a] No measure of dispersion provided; the reported data has been converted from weeks to months to facilitate comparison with Table 9

[b] The reported data has been inverted because it is presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

As would be expected, median survival has increased somewhat from the two immature, trial-specific estimates presented in Table 9, and the effectiveness of Len/Dex is less marked, amounting to a median survival benefit of [redacted] months.

Stratified according to relapse phase

The manufacturer presents separate OS results for first relapse and second/subsequent relapse for the updated (January 2007) analysis only. The relevant data is reproduced in Table 11 and Table 12, respectively.

Table 11: Overall survival (1 prior therapy)

	Len/Dex	Dex
N	124	124
Median – months[a]	38.9	33.4
	[redacted]	[redacted]
Hazard Ratio [95% CI][b]	[redacted]	
Log-rank Test p-Value	[redacted]	

Source: Celgene submission Table 26, p. 87

[a] The reported data has been converted from weeks to months to facilitate comparison with Table 9

[b] The reported data has been inverted because they are presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

Table 12: Overall survival (≥2 prior therapies)

	Len/Dex	Dex
N	229	227
Median – months[a]	33.1	27.1
	[redacted]	[redacted]
Hazard Ratio [95% CI][b]	[redacted]	
Log-rank Test p-Value	[redacted]	

Source: Celgene submission Table 26, p. 87

[a] The reported data has been converted from weeks to months to facilitate comparison with Table 9

[b] The reported data has been inverted because it is presented in the submission as Dex *v*. Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

Notably, neither of these analyses generates a significant result. The manufacturer emphasises that the RCTs were not powered to detect difference in OS amongst subgroups and speculates that the analyses "would likely have demonstrated statistical significant [*sic*] with a larger sample size" (Celgene submission p. 86).

Contamination of the randomised data by post-unblinding crossover may well have reduced the inter-cohort difference in observed OS; however, it is difficult to confirm this hypothesis and impossible to quantify the magnitude of any effect.

4.2.1.3 Secondary outcome: Best myeloma response rates

For each trial participant, response to therapy was categorised – using the International Uniform Response ("Bladé") criteria[27] (see Section 4.1.6) – according to highest assessment of response during the treatment phase of the study. Median follow-up was 17.6 months for MM-009 and 16.4 months for MM-010.

Whole population

Table 13 shows myeloma response rates in each RCT individually, and in a pooled analysis aggregating both populations (this latter data, which is not presented in the manufacturer's submission, has been calculated by the ERG as a simple sum of the counts from the two trials).

Table 13: Best myeloma response rates at unblinding

	MM-009[a]		MM-010[a]		pooled[b]	
	Len/Dex	Dex	Len/Dex	Dex	Len/Dex	Dex
N	177	176	176	175	353	351
Response						
Complete response (CR)	25 (14.1%)	1 (0.6%)	28 (15.9%)	6 (3.4%)	53 (15.0%)	7 (2.0%)
Near-complete response (nCR)	18 (10.2%)	2 (1.1%)	15 (8.5%)	3 (1.7%)	33 (9.3%)	5 (1.4%)
Partial response (PR)	65 (36.7%)	32 (18.2%)	63 (35.8%)	33 (18.9%)	128 (36.3%)	65 (18.5%)
Stable disease (SD)	54 (30.5%)	102 (58.0%)	53 (30.1%)	97 (55.4%)	107 (30.3%)	199 (56.7%)
Progressive disease (PD)	5 (2.8%)	25 (14.2%)	3 (1.7%)	25 (14.3%)	8 (2.3%)	50 (14.2%)
Response not evaluable (NE)[c]	10 (5.6%)	14 (8.0%)	14 (8.0%)	11 (6.3%)	24 (6.8%)	25 (7.1%)
p-value[d]	< 0.001		< 0.001		< 0.001	
Dichotomised Response						
CR, nCR or PR	108 (61.0%)	35 (19.9%)	106 (60.2%)	42 (24.0%)	214 (60.6%)	77 (21.9%)
SD, PD or NE	69 (39.0%)	141 (80.1%)	70 (39.8%)	133 (76.0%)	139 (39.4%)	274 (78.1%)
Odds Ratio [95% CI]	6.31 [3.91, 10.17]		4.80 [3.03, 7.59]		5.48 [3.94, 7.63]	
p-value[e]	< 0.001		< 0.001		< 0.001	

[a] source: Celgene submission Table 15, p. 74
[b] calculated by ERG
[c] Subjects who did not have any response assessment data at the data cutoff point, or whose only assessment was RESPONSE NOT EVALUABLE.
[d] Wilcoxon rank sum test
[e] Continuity-corrected Pearson chi square test

Complete response rate is significantly higher in the Len/Dex group than in the Dex group (p<0.003 by continuity-corrected Pearson chi-square).

Stratified according to relapse phase

Separate response results, stratified according to first relapse *v*. second/subsequent relapse, are reproduced in Table 14 and Table 15, respectively. Significantly better responses are seen in the treatment arms for both these groups.

Table 14: Best myeloma response rates (1 prior therapy)

	Len/Dex	Dex
Number of subjects, n (%)	124	124
Response		
Complete response (CR)	[redacted]	[redacted]
Remission response (RR)	[redacted]	[redacted]
Partial response (PR)	[redacted]	[redacted]
Stable disease (SD)	[redacted]	[redacted]
Plateau phase (PP)	[redacted]	[redacted]
Progressive disease (PD)	[redacted]	[redacted]
Response not evaluable (NE)[a]	[redacted]	[redacted]
p-value[b]	[redacted]	
Dichotomised Response		
CR, RR, or PR	81 (65.3)	32 (25.8)
PP or SD or PD or NE	43 (34.7)	92 (74.2)
Odds Ratio [95% CI]	[redacted]	
p-value[c]	<0.001	

Source: Celgene submission Table A5, p. A41

[a] Subjects who did not have any response assessment data at the data cutoff point, or whose only assessment was RESPONSE NOT EVALUABLE.

[b] Wilcoxon rank sum test

[c] Continuity-corrected Pearson chi square test

Table 15: Best myeloma response rates (≥2 prior therapies)

	Len/Dex	Dex
Number of subjects, n (%)	229	227
Response		
Complete response (CR) [c]	[redacted]	[redacted]
Remission response (RR)	[redacted]	[redacted]
Partial response (PR)	[redacted]	[redacted]
Stable disease (SD)	[redacted]	[redacted]
Plateau phase (PP)	[redacted]	[redacted]
Progressive disease (PD)	[redacted]	[redacted]
Response not evaluable (NE)[a]	[redacted]	[redacted]
p-value[b]	[redacted]	
Dichotomised Response		
CR, RR, or PR	133 (58.1)	45 (19.8)
PP or SD or PD or NE	96 (41.9)	182 (80.2)
Odds Ratio [95% CI]	[redacted]	
p-value[c]	<0.001	

Source: Celgene submission Table A5, p. A41

[a] Subjects who did not have any response assessment data at the data cutoff point, or whose only assessment was RESPONSE NOT EVALUABLE.

[b] Wilcoxon rank sum test

[c] Continuity-corrected Pearson chi square test

Response rates in various subgroups

Appendix 5 of Celgene's submission provides detail on TTP in subgroups categorised according to a variety of different criteria pertaining to the participants' characteristics, clinical history and therapy during the trial. Table 16 gives a summary of these analyses. As previously, we draw particular attention to the analysis according to prior thalidomide exposure status, as separate analyses are performed in the cost-effectiveness model for patients who have previously taken thalidomide.

Table 16: Summary of response rate results in various subgroups

	Len/Dex		Dex			
	N	**Response[a] (%)**	**N**	**Response[a] (%)**	**OR [95%CI]**	***p***
beta-2M[b]:						
<2.5mg/L	103	[redacted]	[redacted]	[redacted]	[redacted]	<0.001
>2.5mg/L	250					<0.001
prior therapy:						
HDT/SCT[c]	206	[redacted]	[redacted]	[redacted]	[redacted]	<0.001
no HDT/SCT[c]	147					<0.001
thalidomide[d]	**127**	**68 (53.5%)**	**147**	**21 (14.3%)**	**6.92 (3.88, 12.34)**	**<0.001**
no thalidomide[d]	**226**	**146 (64.6%)**	**204**	**56 (27.5%)**	**4.82 (3.20, 7.27)**	**<0.001**
bortezomib[e]	27	17 (63.0%)	27	2 (7.4%)	21.25 (4.13, 109.38)	<0.001
performance status[f]:						
ECOG 0	192	103 (53.6%)	196	43 (21.9%)	4.12 (2.65, 6.40)	<0.001
ECOG 1 or 2	152	89 (58.6%)	150	33 (22.0%)	5.01 (3.03, 8.29)	<0.001
type of MM[g]:						
IgA	72	49 (68.1%)	82	15 (18.3%)	9.52 (4.51, 20.10)	<0.001
non-IgA	229	133 (58.1%)	227	45 (19.8%)	5.60 (3.69, 8.52)	<0.001
age[h]:						
<65	192	NR	198	NR	NR	NR
65-75	125	73 (58.4%)	121	26 (21.5%)	5.13 (2.93, 8.99)	<0.001
>75	36	23 (63.9%)	32	7 (21.9%)	6.32 (2.15, 18.59)	<0.001
creatinine clearance[i]:						
<30 ml/min[j]						
30-50 ml/min	42	25 (59.5%)	34	7 (20.6%)	5.67 (2.02, 15.96)	<0.001
50-80 ml/min	125	77 (61.6%)	132	26 (19.7%)	6.54 (3.73, 11.45)	<0.001
>80 ml/min	158	100 (63.3%)	163	41 (25.2%)	5.13 (3.18, 8.29)	<0.001
intra-study dexamethasone[j,k]						

NE = not evaluable; NR = not reported

[a] complete response, remission response or partial response
[b] Source: Celgene submission Table A7, p. A43.
[c] Source: Celgene submission Table A9, p. A45.
[d] Source: Celgene submission Table A11, p. A48.
[e] Source: Celgene submission Table A15, p. A2.
[a] Source: Celgene submission Table A17, p. A4.
[g] Source: Celgene submission Table A19, p. A6.
[h] Source: Celgene submission Table A11, p. A48.
[i] Source: Celgene submission Table A27, p. A14.
[j] Data presented on this outcome does not tally, so has not been reproduced.
[k] Source: Celgene submission Table 22, p. A9.

4.2.1.4 Secondary outcome: Time to first worsening of ECOG performance status

Results for time to first worsening of ECOG performance status were not included in the original Celgene submission. Following a request from the ERG, data for this outcome was provided for the overall population in an updated report, and is reproduced in Table 17.

Table 17: Time to first worsening of ECOG performance status

	MM-009		MM-010	
	Len/Dex	Dex	Len/Dex	Dex
N	177	176	176	175
[redacted][a]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted][b]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Celgene Submission Table 22, p. 82

[redacted]

Celgene do not provide data for this outcome stratified according to relapse phase, or amongst other subgroups.

4.2.1.5 Secondary outcome: Time to first skeletal-related event

Results for time to first skeletal-related event are not available as "there have been too few events for both studies and no analysis can be done" (Celgene submission p. 83).

4.2.1.6 Safety

All grades

According to the EMEA scientific discussion, AEs that were reported significantly more frequently in the Len/Dex group than in the Dex group at any grade were "anaemia, neutropenia, thrombocytopenia, constipation, pneumonia, weight decreased, hypokalaemia, hypocalcaemia, tremor, rash, and deep vein thrombosis" (EMEA scientific discussion, p. 25), and this statement is reproduced in the manufacturer's submission (p. 97; *NB* there appears to be a typographical error, here: hypocalcaemia appears twice but hypokalaemia is not listed).

This statement does not reconcile perfectly with the numerical data on incidence of AEs that is also provided in the EMEA Scientific Discussion (Table 12, p. 25). For example, the incidence of weight loss appears relatively well matched across the two groups (68/353=19.3% *v*. 53/350=15.1%; p=0.178 by χ^2 with Yates's correction). In addition, a number of AEs that are not mentioned were significantly (p<0.05) more common in the Len/Dex arm compared to Dex:

- blurred vision,
- diarrhoea,
- abdominal pain,
- fatigue,
- pyrexia,
- peripheral oedema,
- upper respiratory tract infection,
- anorexia,
- muscle cramp,

- back pain,
- dizziness,
- dysgeusia,
- dyspnoea,
- nasopharyngitis and
- pharyngitis.

It is possible that this discrepancy arises as a result of the methods by which significance was judged in Celgene's analysis (e.g. if analyses were adjusted for multiplicity of testing), but no detail is provided of the approach adopted. We note, however, that magnitude of difference alone does not appear to explain this discrepancy. For example, pneumonia (49/353=13.9% *v*. 30/350=8.6%; p=0.035) is included in the manufacturer's list of AEs with significantly increased incidence, whereas muscle cramp (121/353=34.3% *v*. 76/350=21.7%; p<0.001) is not.

To complicate matters further, the Summary of Product Characteristics for Lenalidomide (Appendix 1 of Celgene's submission) cites a different selection of AEs as significantly more prevalent in the Len/Dex arm: neutropenia, thrombocytopenia, anaemia, leucopenia, lymphopenia, tremor, hypoaesthesia, dyspnoea, rash, pruritus, muscle cramp, pneumonia, deep vein thrombosis, hypotension, fatigue, and asthenia.

The three sources of information on significant inter-arm differences are summarised in Table 18.

Table 18: AEs reported in various sources to be significantly more common in Lenalidomide/Dexamethasone arm

	ERG analysis of raw data[a]	EMEA discussion / manufacturer's submission	Summary of Product Characteristics
abdominal pain	✓		
anaemia	✓	✓	✓
anorexia	✓		
asthenia			✓
back pain	✓		
blurred vision	✓		
constipation	✓	✓	
deep vein thrombosis	✓	✓	✓
diarrhoea	✓		
dizziness	✓		
dysgeusia	✓		
dyspnoea	✓		✓
fatigue	✓		✓
hypoaesthesia			✓
hypocalcaemia	✓	✓	
hypotension			✓
hypokalaemia	✓	✓	
leucopenia			✓
lymphopenia			✓
muscle cramp	✓		✓
nasopharyngitis	✓		
neutropenia	✓	✓	✓
peripheral oedema	✓		
pharyngitis	✓		
pneumonia	✓	✓	✓
pruritus			✓
pyrexia	✓		
rash	✓	✓	✓
thrombocytopenia	✓	✓	✓
tremor	✓	✓	✓
upper respiratory tract infection	✓		
weight decreased		✓	

[a] Raw data from EMEA Scientific Discussion (Table 12, p. 25); significance adjudged by $p<0.05$ by χ^2 with Yates's correction; no correction for multiplicity of testing

The manufacturer notes that the significantly increased risk for thrombosis in the North American trial (MM-009) was found to be associated with concomitant erythropoietin use, which is uncommon in England and Wales (p. 100)

Grade 3/4 adverse events

Concentrating on Grade 3/4 AEs only (Celgene submission Table 33, p.98), the following AEs occurred significantly ($p<0.05$ by χ^2 with Yates's correction) more commonly in the Len/Dex arms of the RCTs:

- **Grade 4** neutropenia, pulmonary embolism and venous thromboembolism
- **Grade 3** neutropenia, anaemia, thrombocytopenia, febrile neutropenia, any infection/other infection, hypocalcaemia, deep-vein thrombosis and venous thromboembolism
- **Grade 3 or 4** neutropenia, anaemia, thrombocytopenia, febrile neutropenia, any infection/other infection, hypocalcaemia, deep-vein thrombosis, pulmonary embolism and venous thromboembolism

In absolute terms, some of these AEs had fairly high incidence among trial participants randomised to Len/Dex, especially those relating to haematological dysfunction and its sequelae: Grade 3/4 neutropenia was experienced by 36.1%, infection 15.9%, thrombocytopenia 13.3% and anaemia 11.0%. In addition, Grade 3/4 deep-vein thrombosis occurred in 13.3% of individuals treated with Len/Dex.

No analysis of Grade 3/4 AEs is provided in the manufacturer's submission or the EMEA scientific discussion. This data accords fairly well with that in the EMEA Scientific Discussion (Table 12, p. 25), although the latter only presents information on AEs with >10% incidence.

Serious adverse events

Although there is no explicit definition in the submission, serious AEs are usually defined as those that result in death, threat to life, hospitalisation (or prolongation of existing hospitalisation), persistent or significant disability, or a birth defect.[28] According to the EMEA Scientific Discussion (Table 13, p. 27), the incidence of serious AEs was fairly high in the trial, with individuals receiving Len/Dex more likely than those on Dex alone to experience at least one (202/353=57.2% *v*. 163/350=46.6%; $p=0.005$ by χ^2 with Yates's correction).

Participants randomised to Len/Dex were more likely to experience serious AEs in the following generic categories: infections and infestations, vascular disorders, cardiac disorders and blood and lymphatic system disorders. The particular serious AEs that occurred significantly more frequently in the Len/Dex group were deep vein thrombosis (25/353=7.1% *v*. 11/350=3.1%; *p*=0.018), pulmonary embolism (13/353=3.7% *v*. 3/350=0.9%; *p*=0.012), atrial fibrillation (11/353=3.1% *v*. 2/350=0.6%; *p*=0.012), congestive cardiac failure (5/353=1.4% *v*. 0/350; *p*=0.041) and febrile neutropenia (6/353=1.7% *v*. 0/350; *p*=0.023).

Interruption / modification / discontinuation of therapy

Neutropenia and thrombocytopenia were the primary reasons for dose reductions in the Len/Dex group.

Discontinuation due to adverse events was relatively uncommon: in Study MM-009, 5 participants discontinued treatment due to neutropenia (2.4%; 4/170) and thrombocytopenia (0.6%; 1/170) and, in Study MM-010, 2 participants dropped out due to neutropenia and thrombocytopenia (0.6% and 0.6%, respectively).

Expanded Access Programme

Before formal regulatory approval, Len/Dex was made available to North American patients under the terms of an Expanded Access Programme (EAP), commencing in 2005. Some early results of this experience were presented at the American Society of Haematology's annual meeting in 2006.[29-31]

According to the published abstract detailing overall experience,[29] at least one Grade 3/4 AE was reported in 261 (35%) of the 746 participants for whom data was available in March, 2006. The most common were neutropenia (7.9%), thrombocytopenia (6.0%), fatigue (3.6%), anaemia (3.5%), pneumonia (3.1%) and hyperglycaemia (2.0%). In the manufacturer's submission, the same publication is cited, but different data is provided, suggesting a higher incidence of these AEs: neutropenia (14.9%), thrombocytopenia (11.1%), fatigue (6.4%), anemia (6.2%), pneumonia (5.4%), hyperglycaemia (3.6%). The source of this data is not clear.

At the same meeting, two posters/presentations detailing experience at a single centre in Canada, treating 69 patients, were presented. In this population, 31 (44.9%) experienced Grade 3/4 neutropenia, 7 (10.1%) febrile neutropenia, 13 (18.8%) infection, and 20 (29%) thrombocytopenia. Because of the high incidence of haematological disorders, 20 (29.0%) individuals required at least one platelet transfusion, and 42 (60.9%) received G-CSF. (*NB* perhaps because they update the data presented in the abstracts, presentation slides[32] detail 70 patients with Grade 3/4 neutropenia in 34 [49%], febrile neutropenia in 11 [16%], infection in 18 [26%], and thrombocytopenia in 27 [39%], necessitating transfusions and G-CSF use in 41% and 63% of patients, respectively.)

Post-marketing experience

Post-marketing experience from the USA includes 5,075 confirmed AEs in 2,275 patients, including 2,087 serious AEs. There were 194 reports with fatal outcome. Of these, the primary cause of death was the progression of the disease in 33% of reports, unknown or unreported in 36% of the reports. For the other reports, the primary causes of death were commonly related to compromised immune function (sepsis [*n* = 11], and pneumonia [*n* = 7]), with other reported causes being leukaemia (*n* = 6), congestive cardiac failure (*n* = 4), renal failure (*n* = 4) and myocardial infarction (*n* = 4). (EMEA Scientific Discussion, p. 29)

Special precautions

The very well known teratogenic properties of thalidomide and its analogues dictate that Len cannot be used by pregnant women. In those of childbearing potential, a pregnancy prevention programme (including an education programme and contraceptive advice) must be followed. The effect of paternal exposure has yet to be determined.

Because of the observed interaction between Len and erythropoietin in the North American RCT, the product characteristics of Len specify that "erythropoietic agents, or other agents that may increase the risk of thrombosis, such as hormone replacement therapy, should be used with caution in multiple myeloma patients receiving lenalidomide with dexamethasone."

4.2.2 Critique of submitted evidence synthesis

An assessment of the manufacturer's submitted evidence synthesis according to the QUOROM checklist[33] is provided in Table 19.

Table 19: Assessment of review of clinical effectiveness according to QUOROM checklist

Heading	Subheading	Descriptor	Reported? (Y/N)	pp.
Title			N/A	
Abstract			N/A	
Introduction		The explicit clinical problem, biological rationale for the intervention, and rationale for review	Y	17-20
Methods	Searching	The information sources, in detail (eg, databases, registers, personal files, expert informants, agencies, hand-searching), and any restrictions (years considered, publication status, language of publication)	Y no details of any non-automated methods, e.g. hand-searching, expert consultation (although arguably not appropriate in this instance)	A19-A28
	Selection	The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design	Y	37, A29
	Validity assessment	The criteria and process used (eg, masked conditions, quality assessment, and their findings)	CONSORT checklist was included in the extraction workbook, but results are not reported	A30
	Data abstraction	The process or processes used (eg, completed independently, in duplicate)	few details about extraction methods (no. of reviewers, data-checking, etc.); extraction workbook appears not to have included some relevant outcomes (QoL and Safety)	A30
	Study characteristics	The type of study design, participants' characteristics, details of intervention, outcome definitions, &c, and how clinical heterogeneity was assessed	Y	36-37; 46-60
	Quantitative data synthesis	The principal measures of effect (eg, relative risk), method of combining results(statistical testing and confidence intervals), handling of missing data; how statistical heterogeneity was assessed; a rationale for any a-priori sensitivity and subgroup analyses; and any assessment of publication bias	For most of the submission, the 2 identified RCTs are effectively treated as a single study; brief meta-analysis and MTC performed	93-101

Heading	Subheading	Descriptor	Reported? (Y/N)	pp.
Results	Trial flow	Provide a meta-analysis profile summarising trial flow (see figure)	Y (although difficult to make all reported figures tally)	38
	Study characteristics	Present descriptive data for each trial (eg, age, sample size, intervention, dose, duration, follow-up period)	Y	39-64
	Quantitative data synthesis	Report agreement on the selection and validity assessment; present simple summary results (for each treatment group in each trial, for each primary outcome); present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses(eg 2×2 tables of counts, means and SDs, proportions)	not reported: agreement on the selection and validity assessment reported: simple summary results data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses	88-96
Discussion		Summarise key findings; discuss clinical inferences based on internal and external validity; interpret the results in light of the totality of available evidence; describe potential biases in the review process (eg, publication bias); and suggest a future research agenda	insufficient studies to assess publication bias; no recommendations for future research (although see p. 180 for recommendations arising from economic analysis)	102-104

4.3 Summary

The searches for clinical effectiveness data are appropriate and relevant trial data is included.

The use of Len (in combination with Dex) for multiple myeloma (MM) in people who have received at least one prior therapy which reflects the licensed indication.

The evidence is based on two, identically designed, good quality RCTs: MM-010 and MM-009 in people with multiple myeloma who had received at least one prior therapy. Pooled analysis of these trials shows increased time to progression with Len/Dex [median 48.3 weeks *v.* 20.1 wks. HR 0.35 (95% CI 0.29, 0.43)]. Increased overall survival is also seen with Len/Dex. Based on updated data, median survival with Len/Dex increased from [redacted] compared with Dex alone.

The main threat to validity for the clinical effectiveness data is the high level of crossover in the trials, leading to a strong lenalidomide effect in the comparator arm. This is likely to underestimate treatment effect, especially for overall survival.

Methods used in the mixed treatment comparison undertaken to estimate the effectiveness of Len/Dex compared to bortezomib monotherapy are inappropriate. However, when recalculated using more appropriate methods, Len/Dex shows increased time to progression [HR 0.56 (95% CI 0.34, 0.91)].

5 ECONOMIC EVALUATION

This chapter provides an assessment of the cost effectiveness analysis submitted by Celgene, plus additional analyses performed by the Evidence Review Group (ERG). We discovered several important logical errors in the economic model first sent to us by Celgene (Table 20). Celgene sent us three versions of the model, each version correcting mistakes in the previous version. In this chapter, we discuss their final model version (version 3 in Table 20).

Table 20: Most important errors in versions of economic model sent to ERG by Celgene.

Model version		**1**	**2**	**3**
Date received by ERG		9th July 2008	7th August 2008	11th August 2008
ICER (Cost / QALY)	Len/Dex *v.* bortezomib	[redacted]	[redacted]	[redacted]
	Len/Dex *v.* Dex (1 prior therapy)	£44,438	£60,000	£46,865
	Len/Dex *v.* Dex (>1 prior therapy)	£28,880	£37,100	£24,584
	Len/Dex *v.* Dex (1 prior therapy) prior thalidomide	£34,734	£45,300	£38,861
	Len/Dex *v.* Dex (>1 prior therapy) prior thalidomide	£29,917	£33,100	£22,589
Errors in model		Base case based on 1 simulation, instead of 300. Values for each modelled patient in the regression equations for time to progression in the 1 prior therapy subgroups were incorrectly coded The regression coefficients in the time to progression equations for the >1 prior therapy were incorrect.	For all subgroups, there were several errors in the calculation of overall survival. One of these was that progression-free survival was frequently double-counted in the estimation of overall survival, i.e. overall survival was calculated as progression-free survival + progression-free survival + post-progression survival.	No logical errors discovered by ERG. This is final Celgene base case.

In the remainder of this chapter, we start with a summary of the systemic review of cost-effectiveness studies presented by Celgene (Section 5.1). Then we present a brief description of the cost effectiveness analysis submitted by Celgene (Section 5.1.1) and its baseline results (Section 5.2). We then explain the workings of the Celgene model in more detail and critique the submission using standard approaches for critical appraisal of economic evaluation (Section 5.3).

5.1 Overview of manufacturer's economic assessment

Celgene performed a systematic review of economic evaluations of Len in combination with Dex for multiple myeloma (MM). The following databases were searched (time span in brackets):

- MEDLINE including In-Process and Old Medline (1951-2008)
- EMBASE (1997-2008)
- CRD (HTA and NHSEED (All years)
- ISPOR (2003-2008)
- Celgene Company Literature (Database of published studies associated with company sponsored trials)

Separate search strategies were provided for each database in the manufacturer's submission (Appendix 3). The EMBASE and MEDLINE searches are based on the same thesaurus and free-text words for the population and intervention used for the clinical effectiveness section (see Section 4.1.1) with an economic evaluation filter added. The EMBASE filter is limited to economic, cost and price terms but does not include terms to identify quality of life or other utility terms. The MEDLINE search includes an additional limited filter of MeSH heading, "Value of Life". The CRD search strategy includes only the population and intervention with no filter and the ISPOR strategy searched for the intervention only. All search strategies reported are appropriate for identifying specific economic evaluations of Len for multiple myeloma but not for additional parameters. The search identified two cost-effectiveness studies, both funded by Celgene, and both in conference abstract form.

Deniz et al (2008)[34] evaluated the long-term health and cost consequences of Lex/Dex versus Dex alone in Scottish patients with multiple myeloma who had received one prior

therapy only. This model was used for the appraisal of lenalidomide by the Scottish Medicines Consortium (SMC)[35]. A discrete event simulation model was used to predict the disease course following second-line treatment. Clinical effectiveness data for the model was derived from the data collected in the MM-009 and MM-010 trials. The median overall survival for Dex was estimated using data from the UK Medical Research Council (MRC) multiple myeloma trials[36]. Disease management costs reflected clinical practice in Scotland. Len/Dex provided substantial clinical benefits compared to Dex alone (modelled median time to progression was 13.5 months with Len/Dex compared to 4.7 months with Dex). This translated to QALY gains of 3.19 against 1.39, with an incremental cost-effectiveness ratio of £28,980 per QALY, which is also quoted in the SMC report. However, the SMC also report that when a lifetime model horizon was used, the ICER increased to £35,673[35]. We suspect that the lower ICER was based on a model time horizon less than lifetime.

In Deniz et al (2008)[37], the same discrete event simulation model was adapted to a Welsh setting. This model was used for the appraisal of lenalidomide by the All Welsh Medicines Strategy Group (AWMSG)[38]. Two subgroups were evaluated: 1) patients who had received one prior therapy only and 2) those with two or more prior therapies. Efficacy data was obtained from the pivotal MM-009 and MM-010 trials. In the one prior therapy group Dex overall survival was estimated using data from the UK Medical Research Council (MRC) Myeloma trials[36], however for those with two or more prior therapies the Mayo Clinic prospective database study[39] was used. The AWMSG state that, using a lifetime model horizon, the ICER for one prior therapy was £34,770 / QALY, and for at least two prior therapies, £30,871 / QALY.

Celgene economic model submitted to NICE

We now turn to the economic evaluation of Len presented to NICE by Celgene. They report cost per QALY estimates for the cases shown in Table 21. The choice of comparisons is discussed in Section 3.3 above.

Table 21: Patient subgroups and treatment comparisons for which cost per QALY estimates are provided.

Patient subgroup	Treatments
One prior therapy only	Len/Dex *v.* bortezomib
One prior therapy only and have pre-existing peripheral neuropathy	Len/Dex *v.* Dex
At least two prior therapies	Len/Dex *v.* Dex
Prior treatment with thalidomide (1 prior therapy only)	Len/Dex *v.* Dex
Prior treatment with thalidomide (2 or more therapies)	Len/Dex *v.*. Dex

Cost-effectiveness is assessed using a discrete event simulation model, with patient-level information. The model was written in Microsoft Excel, with simulation performed using Crystal Ball software. The model is described in detail in Section 6 of the manufacturer's submission; however, we present a summary below.

5.1.1 Model structure

The structure of the model used by Celgene is complex and somewhat unusual in this context. Rather than the Markov approach often used in the assessment of cost-effectiveness of terminal cancer drugs and more frequently seen in NICE submissions, Celgene use a discrete event simulation model. The discrete event approach differs from the Markov approach in that there is no time-based "cycle" in which the model predicts the occurrence of transitions between specific health states. Rather, the occurrence of events is predicted by the model based on patient characteristics and treatments received. At the predicted occurrence of progression, the model then calculates the expected time of death. The key events in the Celgene model are progression and death, which define two corresponding periods – progression-free survival and post-progression survival.

Below, we present our description of the model in several steps.

Establishing a hypothetical population

Firstly, the model establishes a hypothetical population based on patient records from the MM-010 and MM-009 trials. The pooled population from these trials is initially divided according to whether one or more than one prior treatment had been given, and

then whether prior treatments included thalidomide. Within these four sets of trial patient records, the best treatment response achieved is recorded, along with times to progression and death where these had occurred, and a range of personal characteristics.

At this point the treatment allocation in the trials is not used directly, the purpose of this step being to generate a large hypothetical population based on the trial population data. Instead, the simulated patient populations for each treatment group at the start of the simulation are established on the basis of the proportions of best responses achieved by treatment. For example, 19% of patients on Len/Dex and 3% on Dex achieved a complete response as their best response during the trial in the one prior therapy subgroup. Therefore, in the model, 19% of simulated patients on Len/Dex and 3% of simulated patients on Dex were allocated to a patient record that achieved a best response of complete response in the MM RCTs.

Therefore, at baseline for simulating Len/Dex patients, the trial population data is used to construct a total of eight groups, each with expected proportions of patients achieving different best responses to treatment in line with that shown overall in the MM trials. For simulating bortezomib patients, a single group was defined for one prior therapy.

Since patient records for hypothetical patients are allocated at random, and death and progression are predicted in many cases rather than observed, the base case analysis involves averaging over 300 runs of the model.

Predicting progression / progression free survival

Having established the baseline cohorts for the simulation by treatment and population group, the Celgene model goes on to estimate progression (and therefore the progression free survival period) for each patient. This is achieved using one of three approaches:

(a) Where disease progression was observed in the real patient record ascribed to the hypothetical patient, and the treatment group of the real patient is the same as that being modelled, the time to progression is simply noted from the real patient record.

(b) Where progression had not occurred in the patient record being used for simulation, this is predicted by assuming that time to progression follows a Weibull distribution. Time to progression is estimated for that patient based on a regression of patient characteristics and best treatment response achieved.

(c) Where progression occurred in the real patient record, but the actual treatment taken was different from that of the simulated patient, the simulated patient's time to progression is predicted as if they had been included in the treatment group of interest.

For example, if the real patient progressed on Dex at time T1, this corresponds to a point (P1) on the TTP survival curve denoting the probability of progression at T1. It is assumed that, had the patient been taking Len/Dex, progression would have occurred at the same probability (P1), although since expected TTP is greater on len/Dex, this would correspond to a different time of progression (T2), where T2 is greater than T1. This T2 is calculated by solving the regression equation for patient characteristics on TTP for Len/Dex, using the real patient data.

In approaches (b) and (c), and all approaches for bortezomib, the Celgene model uses a regression equation to estimate progression free survival based on best response achieved with individual patient level covariates. In the case of bortezomib, data on the proportions of patients achieving each response is taken from the APEX trial. The progression free survival equation for bortezomib is adjusted to calibrate median progression free survival to the same value as shown in the APEX trial (30.3 weeks). However, the modelling of myeloma with bortezomib therapy does not take into account stopping rules as recommended in NICE Technology Appraisal #129, nor the response-rebate scheme implemented in the NHS.

Predicting death / post progression survival

Broadly, similar approaches to predict post-progression survival are taken by the Celgene model as for time of progression. Post-progression survival is modelled as an exponential function of a range of predictors. Time of death is then calculated as PFS plus PPS.

However, crossovers from Dex to Len/Dex, which occurred in 47% of patients in MM-010 and MM-009, confounds the analysis of post progression survival for Dex. Therefore some adjustment is needed, specifically of the estimated PPS, and hence time to death in the Dex group. This issue is more important in post progression than progression-free survival because most (75%) patients crossed over at or after progression.

In order to correct for the confounding effect of crossovers on Dex survival, the post-progression survival equation for Dex includes an adjustment factor which calibrates the Dex group's overall survival to that shown in the UK Medical Research Council's trials[36]. Inherent in this approach is that Dex confers no less benefit as the range of chemotherapeutic agents included in the MRC trials. Analyses carried out by Celgene suggest that this is the case. The calibration of Dex survival is reported in Appendix 8 of the manufacturer's submission. Briefly, parametric survival analysis was used on the MRC data to derive an equation for time of death, including predictors of age, m-protein, beta-2M and time to progression with first line treatments. The values of these predictors were then set to the corresponding mean values in the MM010 and MM009 trials to estimate median overall survival for MM009 and MM010 under MRC conditions i.e. in the absence of Len treatment. Celgene justify using the MRC data to model post-progression survival of Dex as follows. First, the MRC trials represent the outcomes experienced by a large population (1,372 patients for overall survival) of UK patients on treatment with Dex for multiple myeloma. Second, Celgene state that although the MRC trial data is rather old, with patients enrolling between 1980 and 1997, they believe that the data is still appropriate for the economic evaluation of Len because the MRC data shows no trend for improvement in overall survival over time.

Because the Celgene model aims to predict events at an individual patient level, the preceding step is insufficient and it is necessary to adjust the Celgene PPS equation for Dex so that it might be used to estimate individual times of death. This was achieved by adding a factor to the Dex survival equation and iteratively varying this until the Celgene estimated median OS matched that for the MRC equation, as calculated using mortality predictors from MM009 and MM010.

The use of MRC data is justified by Celgene on the grounds that it is: based on a large population (n=1,372) of UK patients. Although the data is now rather old, Celgene demonstrate no secular trend for improvement in overall survival, suggesting that the

MRC data is still a good representation of overall survival at initiation of second and further lines of therapy.

5.1.2 Treatment effectiveness

The Celgene model does not explicitly model estimates of treatment effectiveness (e.g. hazard ratios) derived from the trials of Len or from indirect comparisons carried out against bortezomib. The modelling of treatment effect is therefore less clear to the reader than would be the case with other modelling approaches.

As stated, the model is based on the prediction of times to events, particularly disease progression and death. Treatment effects are therefore modelled in two main ways:

- Different proportions of patients achieving different best responses to therapy. For example, 19% of patients on Len/Dex achieved complete response, compared to 3% on Dex in the one prior therapy subgroup.

- The regression equation used to calculate time to progression includes a term for lenalidomide treatment, meaning that a complete response on Len/Dex will lead to a longer time to progression than a complete response on Dex or on bortezomib.

For the comparison of Len/Dex with bortezomib, the published median TTP for bortezomib treatment (30.3 weeks) has been used to adjust the TTP Weibull survival curve for bortezomib. Specifically, the median TTP of the adjusted curve matches the published median TTP for bortezomib while keeping the Weibull shape parameter of the curve the same. The model does not take into account the stopping rules and the response-based rebate scheme for bortezomib as recommended by NICE as per guidance[40].

5.1.3 Health Related Quality of Life

No utility or quality of life data was collected in the MM-009 and MM-010 trials. Utility values used in the Celgene model were taken from a cost-utility study carried out by van Agthoven and colleagues (2004)[41] in people with myeloma of intensive chemotherapy alone versus intensive chemotherapy followed by myeloablative therapy with autologous stem cell rescue. Celgene estimate the utility of patients in progression-free survival

(CR/PR/SD) as 0.81, following van Agthoven and colleagues (2004).[41] This value was based on the utility of the general public at an age (median of 54 years) corresponding to that of the patients in the study. In their economic evaluation of Len, Celgene assume a utility of 0.644 in progressive disease, again from van Agthoven and colleagues (2004)[41]. This value was calculated by applying a reduction of 19.5%, specific to multiple myeloma and cited in Mathers and colleagues[42], to the utility for patients in progression-free survival.

Celgene did not model disutility due to adverse events. They give two reasons for this (point C7 in communication from Celgene to ERG, received by ERG on 11th August):

1) The utilities associated with each adverse event that Celgene identified from their literature search were estimated in patient populations different from multiple myeloma patients (e.g. breast, colon, rectal cancer). They state that they would have had to assume that the adverse event of a patient suffering from a different disease would be the same as that of a patient with multiple myeloma.

2) Utility decrements for adverse events are included indirectly through progressive disease since this classification incorporates factors such as new lytic bone lesions or soft tissue plasmacytoma, increase in bone lesion size, and the development of hypercalcaemia.

QALYs for each simulated patient are calculated by weighting duration of the predicted progression-free and post-progression survival periods by the appropriate utilities. For each patient in the RCTs, disease progression status was recorded in the MM trials, the average being approximately 10 assessments per patient. For times beyond those with response recorded in the MM trials, the utility at last recorded response is used for times between the last recorded response and time to progression. A utility of 0.644, corresponding to progressive disease, is used for the period from time of progression to the simulated time of death.

5.1.4 Resources and costs

The perspective of the analysis was the UK NHS and Personal Social Services.

The costs of Len, Dex and bortezomib are taken from the British National Formulary (BNF)[43]. Len is available as 5mg, 10mg, 15mg and 25mg capsules, in 21 capsule packs. The price of a 21 capsule pack of Len (Revlimid) 25mg hard capsules, taken from the BNF[43] at the time of this report, is £4,368. Assuming that patients take the recommended dose of Len of 25 mg orally once daily on days 1-21 of ongoing 28-day cycles, the mean cost of Len per patient per day is £156. This is the cost assumed in the model.

The recommended dose of Dex 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 40mg once daily on days 1-4 of every subsequent 28-day cycle. Dex 40mg costs £2.39[43]. Therefore, the mean cost of Dex per patient per day is £1.02, given that Dex is taken on 12 days in the 28-day cycle. This is the cost assumed in the model.

In the APEX trial[44], and as stated by Celgene, bortezomib was administered at 1.3 mg/m^2 on days 1, 4, 8 and 11 of cycles 1 through 8 (21-day cycles) and on days 1, 8, 15 and 22 of cycles 9 to 11 (35-day cycles), for a maximum treatment period of 273 days. A 3.5mg vial of bortezomib costs £762.38[43].

Patients discontinue drug treatment at disease progression or following unacceptable side effects. For all drugs, dosing is continued or modified based upon clinical and laboratory findings. Dose adjustments are recommended to manage grade 3 or 4 toxicity related to Len.

The costs of disease monitoring and laboratory tests were modelled.

Data on the NHS resources used to treat relapsed/refractory multiple myeloma was obtained by interviews with fifteen haematologists who specialise in the treatment of multiple myeloma[45]. These specialists were selected to provide a broadly representative geographic spread across England and Wales. NHS resources covered in the interview questionnaire included type and frequency of laboratory and disease monitoring, and treatment of disease-specific complications and treatment-related adverse events.

Resource use profiles for management of relapsed/refractory multiple myeloma (including tests to monitor therapy response and disease state) were collected separately by disease status (Table 22, p68 of this report). Specifically, resource use

profiles were collected for patients during relapse and/or on treatment, and for patients in remission/plateau and either on maintenance therapy or off therapy. The "during relapse and/or on treatment" resource use profiles were applied to patients in the model in post-progression, and the resource use profiles associated with "in remission/plateau and on maintenance therapy" were applied to patients in the model in progression-free survival. Resource use profiles associated with "in remission/plateau and off therapy" were not used in the model.

Celgene model costs associated with only Grade 3 and 4 adverse events (AEs), since they assume that these will have the greater impact on resource use, and therefore overall management costs. The following AEs were modelled: anaemia, thrombocytopenia, neutropenia, hypercalcaemia, diarrhoea, constipation, pneumonia, peripheral neuropathy and deep-vein thrombosis. AE rates for Len/Dex and Dex were derived from the MM-009 and MM-010 trial data. AEs are simulated only during the first two years of each simulated patient's course.

Resource use profiles for the management of disease complications and treatment-related adverse events were collected separately for Grade 3 and 4 toxicities as defined by the National Cancer Institute (NCI) Common Toxicity Criteria (CTC) Version 2.0[46], in order to ensure consistency with the MM-009 and MM-010 trials, see Table 44, p141, manufacturer's submission.

For each specific disease-related complication and treatment-related adverse event in the model, the cost per patient was estimated from the following quantities: the probability that a patient suffers the event (taken from MM RCTs), the location where treatment is administered (Table 23) the mean numbers of visits per month by location of treatment (Table 24), and the cost of treatment (Table 25).

The unit cost of inpatient and day-case treatment for disease-related complications and treatment-related adverse events, was calculated from "CHKS" data (abbreviation not defined by Celgene), which contains data from approximately 90% of trusts in the UK and reports in the same structure as Health Episodes Statistics (HES), and data from NHS Reference Costs[47]. For details, see p. 142 and Appendix 11 of the Celgene submission.

Given that drugs are taken only whilst patients are in progression-free survival, undiscounted drug costs are estimated by the simulated TTP multiplied by the cost of the drug per unit time. Undiscounted medical management costs are estimated by the simulated TTP multiplied by the medical management cost per unit time in PFS, plus the simulated PPS multiplied by the medical management cost per unit time in progressive disease. The occurrence of each of several adverse events is modelled for each simulated patient. The cost of treating each adverse event is recorded.

Table 22: Mean number of regular outpatient consultations and disease monitoring tests

	Cost (£)	Frequency (mean # of assessments per year)		
		During relapse / on treatment§	In remission / plateau*	
			Maintenance therapy	Off therapy†
Outpatient	97	12	12	6
Tests to monitor therapy response and disease status				
Routine Blood Counts (FBC)	2.93	20.1	10.7	7.1
Clotting	2.93	3.9	1.1	0.4
INR	2.93	2.6	2.9	0.4
Biochemistry (U&Es)	1.59	17.3	9.7	6.6
Liver function tests (LFTs)	1.59	14.6	7.6	5.1
Erythrocyte sedimentation rate (ESR)	2.93	2.6	1.4	0.9
Plasma Viscosity	1.59	1.6	0.3	0.3
Uric Acid (Urate)	1.59	2.7	1.4	0.9
Immunoglobulin (IGs)	1.59	9.7	6.4	4.9
Paraprotein Measurements (PP)	1.59	11.1	7.6	6.1
Protein Electrophoresis	1.59	9.6	6.7	5.1
Serum β2 microglobulin	1.59	5.0	3.0	2.0
C-reactive protein	1.59	3.3	1.6	1.3
Serum erythropoietin level	1.59	0.5	0.1	0.1
Immunofixation (SIF)	1.59	4.8	3.4	2.9
Creatinine-clearance (CRCL)	1.59	2.3	0.7	0.4
Glomerular filtration rate (GFR)	1.59	7.1	3.3	2.7
Serum Free Light Chains (SFLC)	1.59	4.1	2.9	1.7
Routine urinalysis	1.59	4.4	1.7	1.0
24-hour urine measurement (24hr UR)	1.59	3.0	1.3	1.0
24-hour urine for creatinine (24hr UrCr)	1.59	1.4	0.6	0.1
Total Urine Protein (24hr TUP)	1.59	3.2	1.4	0.4
Urine protein electrophoresis/ light chains	1.59	4.9	2.7	2.1
Urine Immunofixation	18.56	2.1	1.0	1.0
Skeletal Survey by X-Ray (SS)	18.56	1.6	0.1	0.0
Skeletal Survey by X-Ray Individual Sites	2.93	1.6	0.1	0.1
MRI	312.95	0.9	0.0	0.0
Bone Densitometry (BMD)	6.35	0.1	0.0	0.0
Bone Marrow Aspirate (BMA)	1.59	2.1	0.2	0.1
Bone Marrow Trephine Biopsy (BMT)	1.59	2.0	0.2	0.1
Neuropathy (please specify)	2.93	0.1	0.1	0.1
Bacterial investigation	6.35	1.6	0.4	0.3

Source: Celgene submission Table 45, p128

§ Induction or re-induction treatment

* Remission defined as per EBMT criteria. Plateau defined as stable disease following response to induction/reinduction treatment and now on maintenance therapy or off therapy (without maintenance)

† Off therapy would include patients on bisphosphonate treatment alone

Table 23: Average proportion of patients receiving treatment and location of treatment for each of the complications/adverse events

	Grade	% whom receive treatment	Location of care (%)				
			In-patient	Day case	Out-patient	Primary-care	Community care
Disease-related complications							
Anaemia	3	91.86%	5.71%	73.21%	15.36%	0.00%	5.71%
	4	100.00%	19.62%	69.62%	5.38%	0.00%	5.38%
Hypercalcaemia	3	100.00%	50.36%	27.50%	22.14%	0.00%	0.00%
	4	100.00%	77.50%	11.79%	10.71%	0.00%	0.00%
Pneumonia	3	100.00%	98.57%	1.43%	0.00%	0.00%	0.00%
	4	100.00%	100.00%	0.00%	0.00%	0.00%	0.00%
Treatment-related adverse events							
Thrombocytopenia	3	28.85%	6.15%	81.54%	12.31%	0.00%	0.00%
	4	96.43%	17.14%	80.00%	2.14%	0.00%	0.71%
Neutropenia	3	44.11%	5.00%	55.56%	39.44%	0.00%	0.00%
	4	70.71%	12.31%	40.38%	43.46%	0.00%	3.85%
Diarrhoea	3	95.71%	57.50%	12.50%	28.57%	1.43%	0.00%
	4	100.00%	100.00%	0.00%	0.00%	0.00%	0.00%
Constipation	3	100.00%	37.50%	21.43%	35.36%	3.57%	2.14%
	4	100.00%	100.00%	0.00%	0.00%	0.00%	0.00%
Peripheral neuropathy	3	79.29%	0.00%	4.62%	94.62%	0.00%	0.77%
	4	83.85%	9.09%	15.45%	71.82%	3.64%	0.00%
Deep-vein Thrombosis	3	100.00%	12.86%	16.07%	68.93%	2.14%	0.00%
	4	100.00%	81.15%	3.46%	15.38%	0.00%	0.00%

Source: Celgene submission Table 47, p130

Table 24: Average number of visits for treatment of complications/adverse events

	Grade	Visits per Month			
		Day case	Outpatient	Primary care	Community care
Disease-related complications					
Anaemia	3	1	1	NA	4
	4	1	2	NA	4
Hypercalcaemia	3	2	3	NA	NA
	4	4	3	NA	NA
Pneumonia	3	2	1	NA	NA
	4	NA	NA	NA	NA
Treatment-related adverse events					
Thrombocytopenia	3	1	3	NA	NA
	4	2	4	NA	NA
Neutropenia	3	1	3	NA	NA
	4	1	3	NA	2
Diarrhoea	3	2	2	1	NA
	4	NA	NA	NA	NA
Constipation	3	1	2	1	3
	4	NA	NA	NA	NA
Peripheral Neuropathy	3	1	2	NA	2
	4	2	2	1	NA
Deep-vein Thrombosis	3	5	3	3	NA
	4	8	2	NA	NA

Source: Celgene submission Table 48, p131 of Celgene submission

Table 25: Unit costs of complications/adverse events

	Grade	Cost per visit £		
		Inpatient	Day Case	Outpatient
Disease-related complications				
Anaemia	3	1,228.45†	430.53†	97#
	4	1,228.45†	430.53†	-
Hypercalcaemia	3	1,493.06†	420.58†	97#
	4	1,493.06†	420.58†	-
Pneumonia	3	1,670.98†	506.80†	-
	4	1,670.98†	506.80†	-
Treatment-related adverse events				
Thrombocytopenia	3	1,559.56†	547.89†	97#
	4	1,559.56†	547.89†	97#
Neutropenia	3	1,796.67†	470.00†¥	97#
	4	1,796.67†	470.00†¥	97#
Diarrhoea	3	1,302.90†	477.84†	-
	4	1,302.90†	477.84†	-
Constipation	3	1,685.26†	445.77†	-
	4	3,953.50†	445.77†Φ	-
Peripheral Neuropathy	3	1,631.57†	523.80†	97#
	4	1,631.57†	523.80†	97#
Deep-vein Thrombosis	3	1,197.83†	311.28†	199§ / 111ℓ
	4	1,869.50†	282.00†	199§ / 111ℓ

Source: Celgene submission Table 49, p132

\# NHS reference costs 2005 - TOPS FUA - Specialty code: 303 - Clinical Haematology [48]

§ NHS reference costs 2005 - TOPS FAA - Specialty code: 300 - General medicine [49]

ℓ NHS reference costs 2005 - TOPS FUA - Specialty code: 300 - General medicine [50]

† NHS reference costs 2005 [51] combined with CHKS data.

Φ No day case admissions were identified for grade 4 constipation. Therefore, grade 3 constipation average costs per visit were used,

¥ No day case admissions were identified for neutropenia. Therefore, the average of the identified HRG costs was used.

5.1.5 Discounting

Future costs and benefits were discounted at 3.5% as specified in the NICE reference case[52].

5.1.6 Sensitivity analysis

One-way sensitivity analyses and probabilistic sensitivity analyses are reported.

5.1.7 Model validation

There is no evidence that Celgene have tested the model for internal or external validity. We discovered several important logical errors in the economic model first sent to us by Celgene, see Section 5 above.

5.2 Results included in manufacturer's submission

Here, we present a summary of the results of Celgene's corrected model. See Section 6.3.1 of Celgene's report for details.

In Table 26 to Table 29 below, in the Len/Dex and bortezomib treatment arms, drug costs account for the great majority of total costs. Drug costs are far smaller for Dex monotherapy. In all comparisons, life years, QALYs and costs are markedly higher for Len/Dex than for Dex monotherapy. Life years, QALYs and costs are higher, but to a lesser extent, for Len/Dex *v.* bortezomib.

Table 26: Results for patients who have received one prior therapy only.

	Len/dex	**Bortezomib**	Dex
Clinical Outcomes (years)			
Time to progression (median, undiscounted)	1.17	0.56	0.39
QALYs (discounted)	[redacted]	[redacted]	1.53
Life Years (median, undiscounted)	[redacted]	[redacted]	1.65
Life Years (mean, discounted)	[redacted]	[redacted]	2.20
Average Total Cost (£ discounted, per patient)	[redacted]	[redacted]	1,366
Medication	[redacted]	[redacted]	109
Monitoring	[redacted]	[redacted]	1,072
Adverse Event-Complication	[redacted]	[redacted]	185
Incremental cost per QALY of Len/dex versus:		[redacted]	46,865
Incremental cost per Life Year of Len/dex versus:		[redacted]	32,501
Probability Len/Dex cost-effective (willingness to pay £30,000 / QALY)		[redacted]	0%

Source: Celgene submission Table 53, p. 138, Table 54, p.139, Figure 11, p.152 & Figure 13, p.153

Table 27: Results for patients who have received at least two prior therapies.

	Len/dex	Dex
Clinical Outcomes (years)		
Time to progression (median, undiscounted)	0.80	0.39
QALYs (discounted)	3.23	0.77
Life Years (median, undiscounted)	3.39	1.11
Life Years (mean, discounted)	4.76	1.05
Average Total Cost (£ discounted, per patient)	61,171	694
Medication	57,921	109
Monitoring	2,504	404
Adverse Event-Complication	746	181
Incremental cost per QALY	**24,584**	
Incremental cost per Life Year	16,301	
Probability Len/Dex cost-effective (willingness to pay £30,000 / QALY)	approx. 90%	

Source: Celgene submission Table 55 p. 140 & Figure 15 p.155

Table 28: Results for patients who have received one prior therapy of thalidomide only.

	Len/dex	Dex
Clinical Outcomes (years)		
Time to progression (median, undiscounted)	1.57	0.40*
QALYs (discounted)	4.49	1.43
Life Years (median, undiscounted)	5.83	1.56
Life Years (mean, discounted)	6.58	2.10
Average Total Cost (£ discounted, per patient)	119,676	1,311
Medication	115,775	107
Monitoring	3,149	1,017
Adverse Event-Complication	752	187
Incremental cost per QALY	38,861	
Incremental cost per Life Year	26,421	
Probability Len/Dex cost-effective (willingness to pay £30,000 / QALY)	approx. 5%	

Source: Celgene submission Table 56 p. 141 & Figure 17 p.156

* 2.84 months in Celgene report.

Table 29: Results for patients who have received at least two prior therapies, including thalidomide.

	Len/dex	Dex
Clinical Outcomes (years)		
Time to progression (median, undiscounted)	0.66	0.34
QALYs (discounted)	2.96	0.70
Life Years (median, undiscounted)	3.14	1.08
Life Years (mean, discounted)	4.43	1.01
Average Total Cost (£ discounted, per patient)	51,745	694
Medication	48,622	106
Monitoring	2,377	412
Adverse Event-Complication	746	176
Incremental cost per QALY	**22,589**	
Incremental cost per Life Year	14,927	
Probability Len/Dex cost-effective (willingness to pay £30,000 / QALY)	100%	

Source: Celgene submission Table 57 p. 142 & Figure 19 p.158

Celgene performed sensitivity analyses on the following parameters: costs of AEs, costs of disease monitoring, utilities, assuming 5% discount in cost of Len, assuming patients with best response not-evaluable actually had best response of stable disease. They found that ICERs were sensitive to the utilities and very sensitive to the adjustment for post-progression survival for Dex from the MRC data[36].

Celgene also performed probabilistic sensitivity analyses. The probabilities that Len/Dex is cost-effective compared to the relevant comparator, given a willingness to pay of £30,000 / QALY, are shown in the results tables above. Cost-effectiveness acceptability curves and cost/benefit scatter plots are given in Celgene's report.

5.3 Critique of approach used

In this section, we critically appraise the cost-effectiveness model submitted by Celgene. First, we consider the economic model against checklists of good practice. Then we critically appraise the model structure and data.

5.3.1 Critical appraisal frameworks

We considered the economic evaluation submitted by Celgene against the following widely-used checklists: NICE Reference Case[52](Table 30), Drummond et al (1997)[53] (Table 31) and Philips et al (2006)[54] (Table 32).

Table 30: Critical appraisal checklist based on NICE Reference Case requirements

NICE reference case requirement		Critical Appraisal	Reviewer comment
Defining the decision problem	The scope developed by the Institute	✓	
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	X	Comparators are Dex and bortezomib; however, bortezomib plus Dex is not considered as a comparator, despite being routinely used in practice.
Perspective on costs	NHS and PSS	✓	
Perspective on outcomes	All health effects on individuals	X	Disutility of AEs not modelled.
Type of economic evaluation	Cost-effectiveness analysis	✓	
Synthesis of evidence on outcomes	Based on a systematic review	✓	Weber et al (2007) and Dimopoulos et al (2007) RCTs of Len/Dex *v.* Dex and Richardson et al (2005) RCT of bortezomib *v.* Dex
Measure of health benefits	QALYs	✓	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	✓	EQ-5D survey
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	✓	

Table 31: Critical appraisal checklist from Drummond and colleagues[53].

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)?	✓	Len/Dex *v.* Dex Len/Dex *v.* bortezomib
Has the correct patient group / population of interest been clearly stated?	✓	Patient groups are split one prior therapy and more than one prior therapy. Also, subgroup of patients unsuitable for bortezomib. A subgroup of patients with prior treatment with thalidomide is also modelled (split according to one prior therapy and more than one prior therapy).
Is the correct comparator used?	X	Bortezomib is generally administered in combination with Dex, not as monotherapy, as assumed in the analysis. See Section 3.3 for comparators in the scope, but not addressed.
Is the study type reasonable?	✓	Cost-utility model used. However, we are not convinced that it was worth the considerable extra complexity of using an individual patient model compared with a Markov model, see Section 5.3.2.
Is the perspective of the analysis clearly stated?	✓	UK NHS & PSS
Is the perspective employed appropriate?	✓	-
Is effectiveness of the intervention established?	✓	Effectiveness of Len is established from Weber et al (2007) and Dimopoulos et al (2007) RCTs of Len/Dex *v.* Dex. Effectiveness of bortezomib from Richardson et al (2005) RCT of bortezomib *v.* Dex. Quality of RCTs is good. Effectiveness estimates can be confidently adopted for TTP; however, OS data in trials of both active interventions is compromised by substantial post-treatment-phase crossover from control arm to active drug.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	✓	30-year time horizon is used. After 30 years, at least 98% of patients are modelled to have died. Hence the time horizon is effectively life time.
Are the costs and consequences consistent with the perspective employed? *	✓	All costs from UK NHS & PSS perspective.
Is differential timing considered?	✓	-
Is incremental analysis performed?	✓	-
Is sensitivity analysis undertaken and presented clearly?	✓	Univariate and probabilistic sensitivity analysis presented.

Table 32: Critical appraisal checklist of decision analytic modelling practice

Dimension of quality			Comments
Structure			
S1	Statement of decision problem/objective	X	Cost-effectiveness of: ■ Len/Dex *v.* Dex for: □ one prior therapy only and have pre-existing peripheral neuropathy □ at least two prior therapies □ 1 prior therapy with thalidomide □ at least two prior therapies including one therapy with thalidomide, ■ Len/Dex *v.* bortezomib for one prior therapy only. Comparators are correct, but not exhaustive compared to the scope of the appraisal (Section 3.3). Also, we believe that bortezomib is generally administered in combination with Dex, not as monotherapy, as assumed in the analysis, see Section 3.3.
S2	Statement of scope/perspective	✓	NHS and PSS perspective. Cost and benefit inputs are consistent with the perspective. Scope of model stated.
S3	Rationale for structure	?	Although presenting the analysis as a discrete event model is acceptable, we believe that a simpler Markov model would have been sufficient to model the disease, see Section 5.3.2.
S4	Structural assumptions	✓	Model assumptions are explained clearly in the report, see Section 5.1. Overall, we are satisfied with the structural assumptions. Weibull functions were fitted to TTP and exponential functions to PPS of the MM RCTs. Overall survival is defined as TTP plus PPS.
S5	Strategies / comparators	X	See S1 above for comparators.
S6	Model type	?	As stated in Section 5.3.2, we are not convinced that a complex, discrete event model is necessary.
S7	Time horizon	?	The model time horizon is 30 years, which is long enough, since by then at least 98% of patients are modelled to have died. Since OS is immature, it is extrapolated for each treatment. However, we are no convinced by the fit of Len/Dex OS to the RCTs, see Section 5.3.3.2
S8	Disease states / pathways	✓	The disease states: stable disease, partial response, complete response and progressive disease are those generally accepted for this disease.
S9	Cycle length	✓	There is no concept of a model cycle length because this is a continuous time model.
Data			
D1	Data identification	?	Data identification methods are described. The data for time to death and time to progression has been taken from MM-009 and MM-010, but the individual patient level for all patients in these RCTs is not provided.
D2	Pre-model data analysis	✓	TTP and PPS are functions of characteristics of individual patients and, for TTP, treatment with lenalidolmide.
D2a	Baseline data	X	TTP and PPS for the model patients for both Len/Dex and the baseline treatments (Dex and bortezomib) were based on data from MM-009 and MM-010, and the APEX RCT, which is reasonable. The baseline PPS of Dex was based on data from MRC trials. We believe that

Dimension of quality			Comments
			would be better to fit mean OS, not median OS from the MRC data (see Section 5.3.3.2). Baseline PPS of Dex was not taken from the RCTs of Len/Dex *v.* Dex due to the large number of patients originally allocated to Dex who crossed over to Len. We have criticised this approach, see Section 5.3.3.2
D2b	Treatment effects	?	The relative treatment effects are incorporated indirectly by modelling TTP and PPS as a function of treatment response and treatment, see Section 5.1.2. The model incorporates a large degree of extrapolation for estimating PPS given that OS from the RCTs is immature. Note our concerns in point D2a above.
D2c	Quality of life weights (utilities)	X	Utility values were assumed the same as those in a cost-utility study of multiple myeloma[41]. In this model the utility value in CR/PR/SD was 0.81, and was based on the utility value of the general public at an age value corresponding to that of the patients in the study. This value may be too high (see Section 5.3.3.6).
D3	Data incorporation	✓	Data incorporated in the model is referenced and generally well described. Data incorporation is transparent. For the PSA, the choice of distribution for each parameter has been described and justified.
D4	Assessment of uncertainty	X	Not all types of uncertainty have been addressed, see Section 5.3.4.
D4a	Methodological	X	Celgene have used a single type of model
D4b	Structural	X	Not assessed. They could, for example, have modelled Dex patients in the MM trials, with those who crossed over to Len/Dex censored.
D4c	Heterogeneity	✓	The model was applied to different patient subgroups: 1 prior therapy, >1 prior therapy, 1 prior therapy of thalidomide, >1 prior therapy including thalidomide
D4d	Parameter	✓	Probabilistic and univariate sensitivity analyses performed.
Consistency			
C1	Internal consistency	X	No evidence has been presented to indicate that the mathematical logic of the model has been tested. Indeed, previous versions of the model contained several serious logical error (see Table 20).
C2	External consistency	X	The results of the model were not calibrated against independent data. The model has not been reviewed in the context of other models of multiple myeloma.

✓ indicates 'clear', X indicates 'concerns', and ? indicates 'unclear/unknown'

Checklist based on Philips and colleagues[54]

5.3.2 Modelling approach and structure

Cost-effectiveness is assessed using a discrete event simulation model, using patient-level information, rather than using an aggregated cohort approach. Whilst we do not suggest that their approach is wrong, we are not convinced that the extra complexity of a discrete event simulation model is justified. In particular, we believe that a simpler model would be appropriate given that overall survival for Dex is very uncertain, and is a very important influence on the cost-effectiveness of Len/Dex.

Celgene justify their choice of a discrete event model in detail on p.118 of their submission. First, they state that their model captures the clinical history and variation in efficacy among patients. They state that clinical history is very important in multiple myeloma because it affects the course of the disease and its management (e.g. previous treatments affect the choice of future ones). In this way, the model can estimate costs and QALYs for each patient. However, we are not convinced that this flexibility is necessary in this case, especially since the analysis models only one course of drug treatment, i.e. patients are not tracked through a sequence of interventions. Second, Celgene correctly state that the discrete event approach requires no half-cycle model correction. However, such corrections are easily implemented in Markov models, and any advantage of modelling the exact time of events in discrete time is extremely small given the overall uncertainties of the model.

Although bortezomib in combination with Dex may be used in clinical practice in England and Wales, Celgene did not consider this to be an appropriate comparator for this economic evaluation because it is an unlicensed therapy. However, we note that this restriction does not apply to comparators in NICE appraisals and that combination treatment with bortezomib and Dex is widely believed to be more effective than bortezomib monotherapy[55;56]. Given that Dex is very inexpensive, Len is likely to be less cost-effective compared to bortezomib-Dex combination therapy than compared to bortezomib monotherapy.

A 30-year time horizon was used in the model. As survival 30 years after starting treatment is negligible (when less than 2% of patients are still alive) the time horizon is effectively lifetime. Given that data for patients receiving Len/Dex is available for a median combined follow-up of only 31.3 months, by which time over half of patients taking Len/Dex are still alive, the 30-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. Given that the cost-effectiveness of Len is strongly affected by survival times, this introduces considerable uncertainty in the estimates of cost-effectiveness.

The manufacturer's submission states that *"the model does not favour any response level achieved by one treatment over the same response level achieved by another treatment, but focuses on the difference in the proportions of patients achieving*

response" (p.A135). This statement is not entirely supported, however, since the equation used to estimate TTP includes a term reflecting treatment allocation, with a positive coefficient for Len use. As a result, simulated patients receiving Len/Dex achieve longer TTP than those representing comparator (Dex or bortezomib) treatments. For example, amongst patients with one prior therapy who are simulated to achieve a complete response, median TTP is 61.8 months for the Len/Dex cohort, 44.3 months for the Dex cohort and 42.3 months for the Bortezomib (estimates are mean of median values generated by PenTAG in 100 iterations of the model). This asymmetry only applies to the calculation of TTP outcomes, because there are no terms representing treatment allocation in the equations used to calculate overall survival.

The comparison of Len/Dex with bortezomib does not take account of stopping rules and response-based rebate scheme for bortezomib in place in the NHS, where only patients who respond to treatment with bortezomib continue to receive treatment beyond four cycles of therapy. Celgene concede that, therefore, the results of their comparison of Len/Dex *v.* bortezomib should be considered as *exploratory only* (p137 Celgene report). To incorporate the response-rebate scheme in the model, it would be necessary to model the time to response. Celgene state that detailed audit data on both the extent to which the response-based rebate scheme is implemented in clinical practice and its effects on efficacy and safety would be required. They say that they are not aware of such data. Nonetheless, we note that, for all drugs, response to treatment as a function of time is included in Celgene's model.

Furthermore, we believe that when the required data is available to model the scheme, Len/Dex would appear worse value for money compared to bortezomib for the following reason. In their comparison of Len/dex against bortezomib, Celgene currently assume that all patients who start on bortezomib continue treatment until progression or severe adverse events. If the response-based rebate scheme were used, drug costs would cease after four treatment cycles for those patients who do not respond to treatment, i.e. those patients who accumulate only small number of QALYs (since survival time is a function of response to treatment).

5.3.3 Data inputs

In this section, we consider the data used in the Celgene cost-effectiveness model.

5.3.3.1 Patient group

The modelled patient populations are taken directly from the MM-009 and MM-010 RCTs. The extent to which patients from clinical trials are representative of patients in routine practice is always uncertain (see Section 4.1.5.2 in "Clinical Effectiveness" section).

The economic analysis of the patient population with one prior therapy and pre-existing peripheral neuropathy utilises the same efficacy data for the Len/Dex treated patients with one prior therapy only. Celgene state that this is because there were too few patients in the trials with one prior therapy who had pre-existing peripheral neuropathy upon which to base such analyses (p138 Celgene report).

[redacted] to progression in various subgroups [redacted]

5.3.3.2 Clinical effectiveness

In this section, we describe several concerns with the clinical effectiveness data used in the model;

- First, and importantly, we suggest that the modelled OS of Len/Dex is better than experienced in the MM-009 and MM-010 RCTs. We suggest this may be because OS for the individual patient records used in the model for Len/Dex is better than experienced in the MM RCTs.
- We then show that modelled TTP is reasonably close to that experienced in the MM RCTs.
- Finally, we explain why we have concerns about Celgene's method for adjusting post-progression survival for Dex using data from the MRC trials.

For TTP and OS, we now compare the actual survival experience from the MM RCTs to the modelled survival with the aid of several figures. We present the Kaplan-Meier survival curves from the RCTs, where these curves are reproduced from the publications of the RCTs[1;2], against Kaplan-Meier curves produced from the model. We have generated all modelled curves using statistical software.

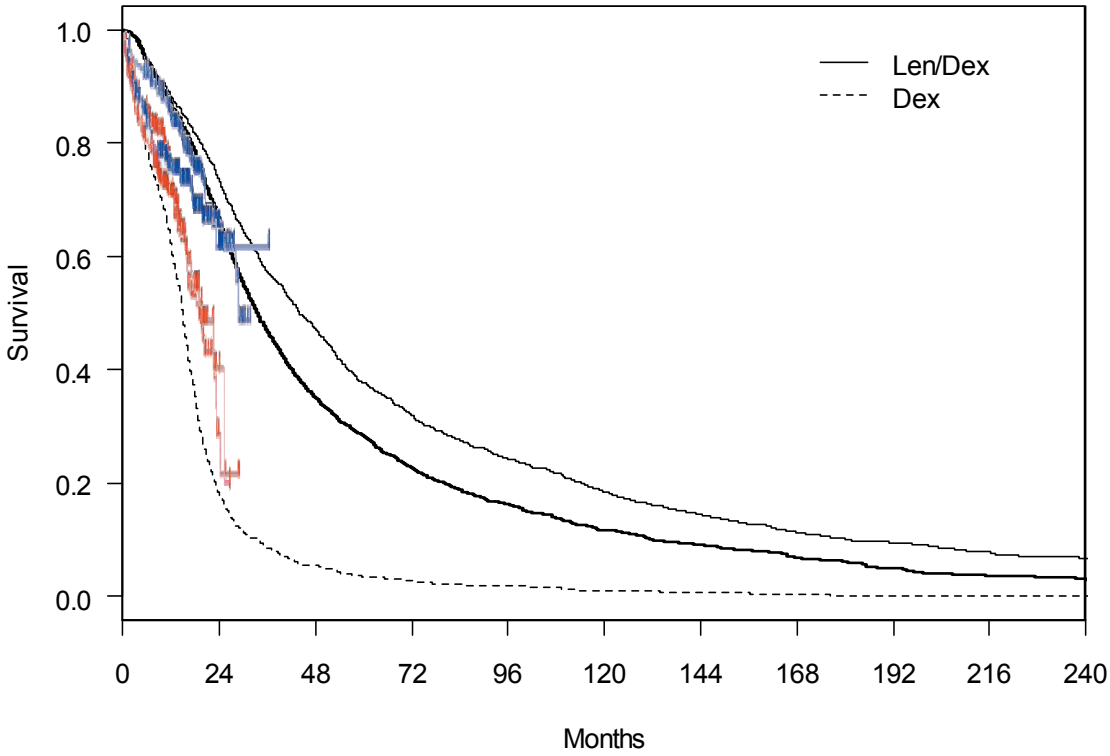
The modelled time to progression for Len/Dex and Dex monotherapy both appear reasonably consistent with the trial results (Appendix Figure 7 p116 below, Figure 8 p117, Figure 9 p117, Figure 10 p118 below). As expected, the modelled overall survival for Dex is far lower than experienced in the Len RCTs due to the adjustment of post-progression survival to reflect experience in the MRC trials (Figure 4 p83, Figure 5 p84 below, Appendix Figure 11 p119 below). However, importantly, the modelled overall survival for Len/Dex is clearly higher than experienced in the Len RCTs, including the thalidomide-exposed subgroup (Figure 4 p83, Figure 5 p84 below, Appendix Figure 11 p119 below).

This is also manifested in the fact that the median OS for Len/Dex is higher in the model compared to the RCTs. In particular, for the 1 prior therapy group, the modelled median OS is approximately 4.27 years, compared to 3.25 years in the RCTs, and for the >1 prior therapy group, the modelled median OS is approximately 3.41 years, compared to 2.77 years in the RCTs (Table 26 Celgene report).

This inconsistency makes Len appear better value for money than if the modelled overall survival had followed the experience of the RCTs. When we adjust OS for Len/Dex in the model so that the model now accurately predicts the median OS from the RCTs, the modelled fit to Len/Dex OS appears closer to the experience from the RCTs (Figure 4 p83, Figure 5 p84 below). Celgene's fit to Len/Dex OS is shown by the thin smooth line, and our improved fit is shown by the thick continuous line in Figure 4 and Figure 5. Technically, this adjustment is achieved by setting the regression intercept coefficient for post-progression survival for Len/Dex from 4.60 to 4.05 for the 1 prior therapy group, and from 4.60 to 4.15 for the >1 prior therapy group.

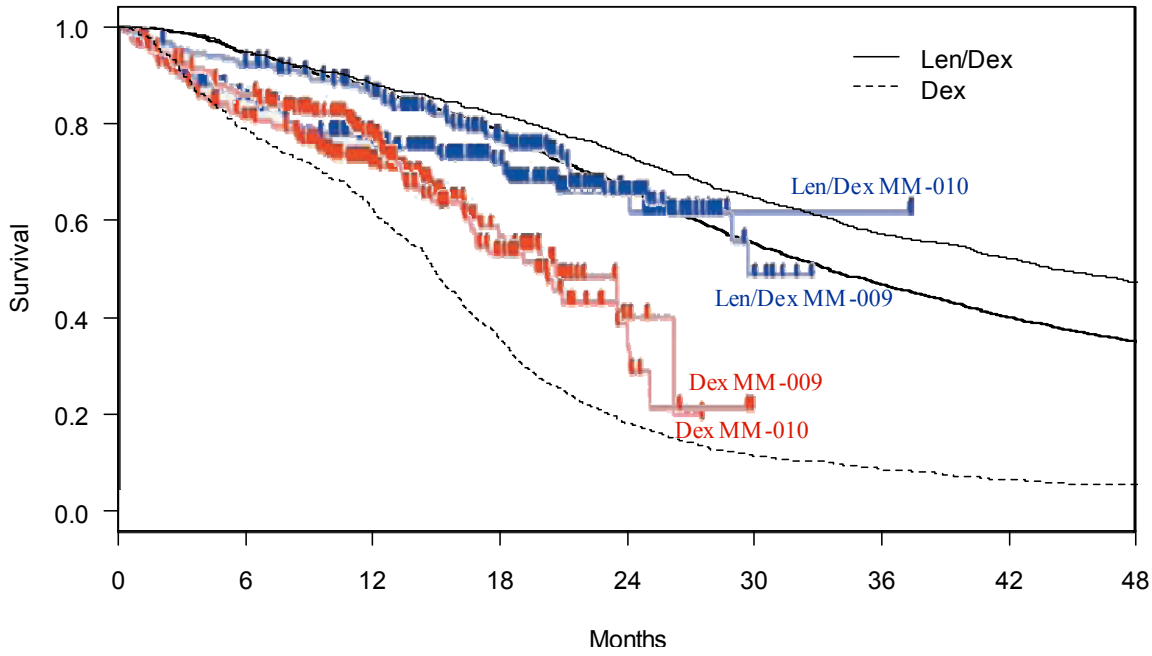
Using these adjustments, the ICERs increase for all comparators. The ICER for 1 prior therapy *v.* bortezomib increases greatly, from [redacted] to approximately [redacted] / QALY, for 1 prior therapy *v.* Dex the ICER increases greatly from £46,900 to £69,500 / QALY, for >1 prior therapy, from £24,600 to £32,900 / QALY. The ICER of Len/Dex *v.* bortezomib is very high because the model now predicts a similar mean overall survival for the two treatments. For subgroups with prior thalidomide treatment, for 1 prior therapy the ICER increases from £38,900 to £56,500 / QALY, and for >1 prior therapy, from £22,600 to £30,800 / QALY.

Figure 4: Overall survival for Len/Dex and Dex for 1 prior and >1 prior therapies pooled.



Pooled population weighted according to reported proportions of trial participants in each stratum (Len/Dex: 1 prior = 35.1%, ≥2 priors 64.9%; Dex: 1 prior = 35.3%, ≥2 priors 64.7%)
Smooth Kaplan-Meier curves represent model output, Kaplan-Meier curves with censorships represent empirical data reproduced from MM-009 and MM-010 papers.
Thin smooth line is Celgene fit to Len/Dex OS.
Thick smooth line is ERG fit to Len/Dex OS.
All Kaplan-Meier graphs of model output were generated by the ERG.

Figure 5: Overall survival for Len/Dex and Dex for 1 prior and >1 prior therapies pooled.



As Figure 4, but expanded x-axis.

As described in Section 5.1.1 above, the modelled cohorts are made up of simulated individuals, each of which is based on a randomly selected patient from the empirical trial records. We have analysed the patient-level data that is included in the model, and found that it differs – in some cases, substantially so – from the trial results presented in the manufacturer's submission and the published RCTs.[1;2]

Most notably, overall survival in the patient data underpinning the model is different, somewhat in Len/Dex's favour, than was observed in the trials. For 1 prior therapy, we have calculated that the hazard ratio for OS is 0.226 (95%CI: 0.112, 0.452) in the patient-level data used in the model, compared to [redacted] in the clinical effectiveness data submitted (see Table 11, above). For >1 prior therapy, the discrepancy is less marked, but also favours Len/Dex: HR=0.657 (0.440, 0.979) in the patient-level data used in the model, compared to [redacted] in the clinical

effectiveness data (see Table 12, above). When both strata are combined, HR=0.468 (0.332, 0.662), compared to [redacted] in the clinical effectiveness data (see Table 10, above). For Kaplan-Meier plots of TTP in the individual patient data in the model, see Appendix 2.

The modelled TTP is slightly lower than experienced in the MM RCTs. In particular, for the >1 prior therapy group, the modelled TTP is 9.5 months, compared to an average of 10.2 months in the MM RCTs (9.3 months and 11.1 months in MM-009 and MM-010 respectively, see Table 55 in Celgene's report). The difference is greater when comparing the modelled median TTP (9.5 months) to the average of 10.7 months as reported in the two MM RCT papers (10.2 months and 11.1 months in MM-009 and MM-010 respectively). Furthermore, there is a slight difference between TTP observed in the individual patient data in the model and that reported in the pooled clinical effectiveness data. For 1 prior therapy, HR=0.286 (0.205, 0.401), compared to [redacted]) in the clinical effectiveness data (see Table 6, above). For >1 prior therapy, the respective values are 0.427 (0.336, 0.544) and [redacted] (see Table 7, above).

When interpreting the significance of any discrepancy, it is important to recognise that, increasing the TTP benefit attributable to Len/Dex tends to *decrease* the estimated value for money offered by the technology. This is because an improved TTP hazard ratio for Len/Dex *v.* Dex means that the TTP for Len/Dex is longer, leading to increased drug costs for Len (as drugs are taken whilst in progression-free survival). Accordingly, basing TTP calculations on a dataset showing less TTP benefit for Len/Dex (as appears to have happened in the >1 prior stratum of the model) would result in a *lower* ICER. In particular, we calculate that if the model is changed so that it predicts the median TTP for Len/Dex as reported by Celgene (Table 55 in Celgene report), the ICER for >1 prior therapy increases by approximately £1,000. If the model is changed so that it predicts the median TTP for Len/Dex as reported by the RCT papers, this ICER increases by approximately £2,000. For Kaplan-Meier plots of OS in the individual patient data in the model, see Appendix 2.

There are three possible explanations for these discrepancies. First, we note that the pool of patient-level trial data incorporated in the model represents only 655 individual patient records from the 704 enrolled in the two RCTs. For the 1 prior therapy stratum, 235 individual patients are represented (119 Len/Dex, 116 Dex); in the >1 prior therapy

stratum, 420 individual patients are represented (210 each for Len/Dex and Dex). As a result, we believe that 49 (7%) of the participants in the two RCTs are not included in the patient files on which the simulation is based. We requested – but did not receive – clarification from the manufacturer regarding the characteristics of these individuals and the process by which they were excluded. It is inevitable that the exclusion of data-points that were used in analysis of the RCTs from model inputs will have impaired agreement between the two datasets.

Second, we note that the patient-level data in the model is less mature than the data presented in the submission and the published RCTs.[1;2] Specifically, the maximum follow-up time for the patient records used in the model is 25.5 months, compared to about 33 months in the published RCTs (with seven months' additional follow-up presented in the submission). If more mature data had been used in the model, this would have improved the accuracy of the estimates of OS in the model. It might be argued that any benefit from using updated data would be attenuated by additional contamination from post-unblinding crossover from control to Len/Dex. However, the manufacturer's methods for modelling OS in the Dex arm provide some compensation for this problem, and the Len/Dex arm (which is not affected by the crossover) would also benefit from enhanced accuracy.

Finally, it is possible that there are inconsistencies between the individual values in the datasets. Without access to the exact patient-level data that was used to plot Kaplan-Meier curves and derive summary estimates of time-to-event outcomes, it is not possible to verify the accuracy of the data in the model.

We emphasise that the inconsistencies discussed here are only of importance if they are reflected in model outputs (that is to say: if the simulated cohort does not agree well with the observed data). However, since we believe that the model's ability to reproduce the trial results is suboptimal (especially as regards Len/Dex OS; see earlier in this section), it is worth noting that biased data input is likely to explain some or all of this discrepancy.

As stated above, in an attempt to correct for the extensive cross-over of patients from Dex to Len/Dex, a factor was added to the post-progression equation for Dex to calibrate the estimated Dex overall survival to that observed in the UK Medical Research Council (MRC) myeloma trials[36]. Their approach is to take the clinical effectiveness of Len/Dex

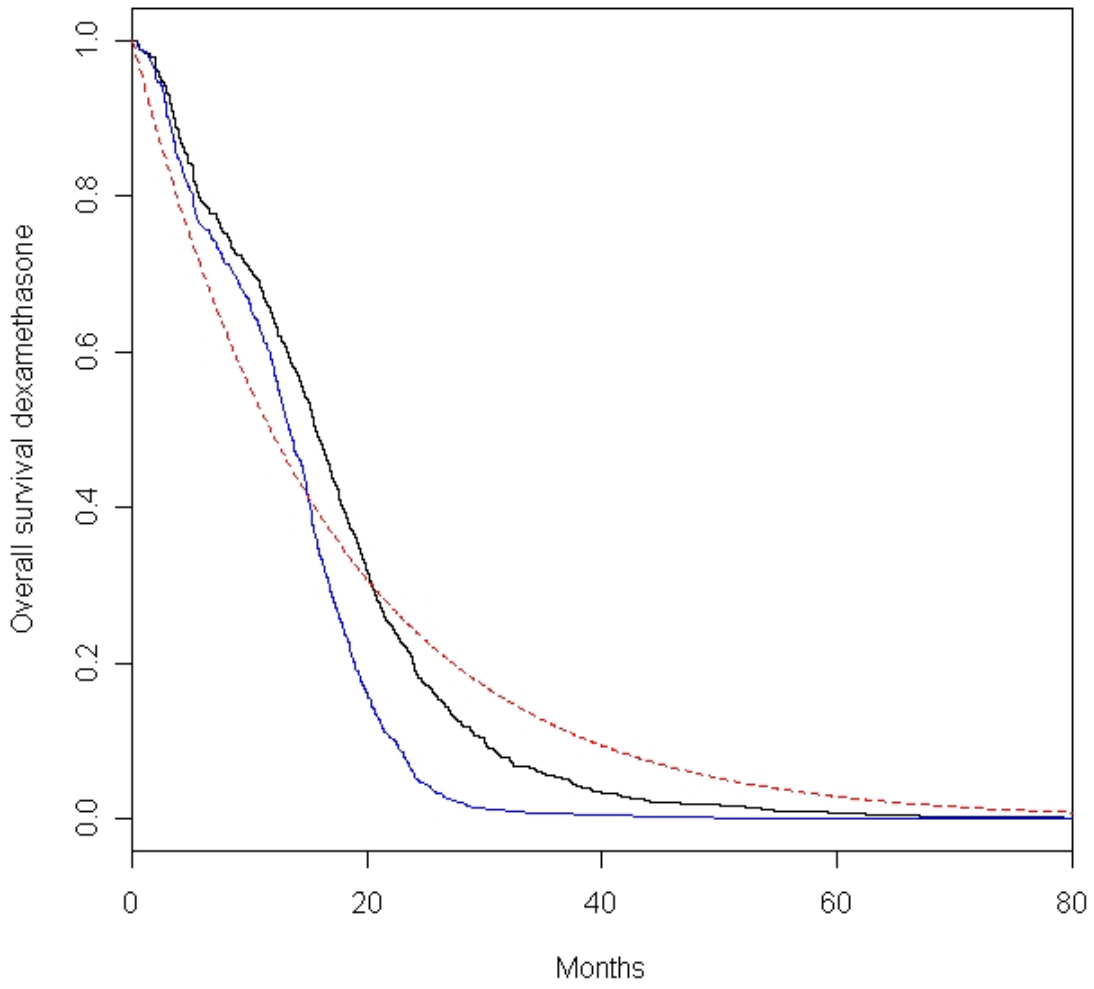
from one group of trials (MM-009 and MM-010) and the clinical effectiveness of Dex from a different group of trials (the MRC trials), with the Dex effectiveness adjusted for the patient characteristics of the patients in the MM09 and MM10 trials. Although Celgene adjusted for differences between the MM trials and MRC trails, there will inevitably be other factors which may be unbalanced between the modelled population of Len/Dex. For example, OS of Dex may have increased from the time of the MRC trials to the time of the MM trials. Celgene state that although the MRC trial data is rather old, with patients enrolling between 1980 and 1997, they believe that the data is still appropriate to the economic evaluation because the MRC data shows no trend for improvement in overall survival over time. However, on p29 of their report, Celgene note that there was a trend in the Mayo clinic data[39] towards improved survival during 1995 to 2000, and a statistically significant improvement in survival from 2000 to 2006. Apparently, the trend to an improvement between 1995 and 2000 coincided with increased use of high dose therapy (with stem cell transplant), which likely contributed to this change[57]. The significant improvement in survival observed between 2000 and 2006 is believed to be due to the introduction of novel therapies[57]. This suggests that the overall survival of patients taking Dex today may be better than calculated from the MRC data. In this case, Len/Dex may actually be substantially worse value for money versus Dex than calculated by Celgene. Given these uncertainties in basing progression-free survival for Dex on the MRC data, it would be useful to populate the cost-effectiveness model with data for Dex taken from MM-009 and MM-010 with patients who crossed over to Len censored.

As Celgene state, their approach results in a higher (more conservative) estimated median survival for the MM-009 and MM-010 Dex patients than was observed in the MRC trials ([redacted], for patients with one or multiple prior therapies respectively). Nonetheless, this higher median survival does not allow for any improvement in OS over time under Dex. Also, their approach results in a lower (less conservative) modelled median survival for Dex patients than was observed in the MM-009 and MM-010 trials ([redacted] versus 33.6 months and [redacted] versus 27.2 months (p87 Celgene report) for patients with one or multiple prior therapies respectively). However, this is of course expected, because Dex survival in the MM trials includes cross-over of some patients to Len/Dex.

The method used to adjust Dex overall survival in the model is extremely important for the assessment of the cost-effectiveness of Len/Dex *v.* Dex. For example, when we remove the adjustment of post-progression survival for Dex, i.e. we model the post-progression survival of Dex as experienced in the MM-009 and MM-010 RCTs, then Len/Dex appears substantially worse value for money, e.g. the ICER for >1 prior therapy for Len/Dex *v.* Dex increases from £24,600 to approx. £79,000 / QALY. We do not advocate using the unadjusted overall survival from the lenalidomide RCTs in the model; however, it is important to emphasise the extent to which the adjustment influences cost-effectiveness.

The adjustment of the post-progression survival equations for Dex was based on matching the **median** overall survival between the MRC data, using patient characteristics of the Len RCTs, with the **median** overall survival of Dex in the model. However, given that the cost per QALY of Len/Dex *v.* Dex equals (**mean** costs in Len/Dex arm - **mean** costs in Dex arm) / (**mean** QALYs in Len/Dex arm - **mean** QALYs in Dex arm), we suggest that it is preferable to match the **mean** Dex overall survivals. We performed the following calculations. Given that overall survival of Dex in the MRC trials was modelled as an exponential distribution, mean OS = median OS / ln(2). Hence the mean OS that we attempt to model for 1 prior therapy = [redacted] / ln(2) = [redacted] months and mean OS that we attempt to model for >1 prior therapy = [redacted] / ln(2) = [redacted] months (Figure 6). By matching to the **median** OS, we are taking no account of the tail of the distribution, or more precisely, the curve beyond the 50th percentile. Indeed, this is evident from Figure 6, where we see that the tail of the curve used by Celgene is a very poor fit to the dotted line representing the exponential distribution. When we run the model with our fit to the **mean** OS, the ICER for 1 prior therapy increases only marginally, but the ICER for >1 prior therapy increases substantially, from £24,600 to £33,200 / QALY. In the thalidomide-exposed subgroup, the ICER for 1 prior therapy increases only marginally, and the ICER for >1 prior therapy increases substantially, from £22,600 to £30,200 / QALY.

Figure 6: Modelled fit to MRC data for overall survival for dexamethasone



The dotted line represents a negative exponential curve which we are aiming to fit to in the model. The lower continuous line represents one simulation of modelled OS by setting the **median** OS approximately equal to the **median** of the exponential curve. This is the method used by Celgene. The upper continuous lines represents one simulation of modelled OS by setting the **mean** OS equal to the **mean** of the exponential curve.

The proportion of simulated patients in each best response rates group is implemented on the basis of the observed trial data, as reproduced above in Table 14 and Table 15 for 1 prior and >1 prior therapy, respectively. In the thalidomide-exposed subgroup, the best response rates used in the model are given in Tables 41 and 42 of Celgene's report.

5.3.3.3 Drug costs and drug administration costs

Drug costs account for most of the costs in the model.

Given that Len and Dex are taken orally, Celgene correctly assume no drug administration costs. Given that bortezomib is administered as an intravenous bolus, they assume an administration cost, namely £1,628 per patient per cycle, i.e. £407 per patient per dose. They therefore calculate the mean cost of bortezomib per patient per day as £762.38 x 4/21 + £1,628 x 4/25 = £210, where they assume 25 days per cycle for the administration of bortezomib. However, we believe that the administration cost of bortezomib is too high.

Celgene cite the cost of £1,628 per cycle as from NHS Reference Costs (2005) - Outpatient Adult Follow Up Attendance (TOPS FUA) - Specialty: Clinical Haematology – Face to Face, Specialty Code: 303. The cost from this reference is actually £108 per outpatient attendance. Given that there are 4 administrations per cycle, this equates to £108 x 4 = £432 per cycle, not £1,628 as assumed by Celgene. When we insert the corrected figure for the cost of administration of bortezomib, Len appears worse value for money versus bortezomib, with the ICER increasing from [redacted] to [redacted] / QALY. Furthermore, in the NICE appraisal of bortezomib, the manufacturer assumed an administration cost of £1,672 per patient (not per cycle). However, the ERG for the bortezomib appraisal suggested that this figure is probably low, suggesting instead a mean administration cost per patient in excess of £2,500. Celgene's current assumption of £1,628 per cycle equates to approximately £8,900 per patient, which is far higher than assumed by the manufacturer of bortezomib.

Although there was a maximum of 11 cycles in the APEX trial, Celgene have assumed a maximum of 8 cycles in their model. When we correct this error alone, Len appears *better* value for money versus bortezomib, with the ICER decreasing from [redacted] to [redacted] / QALY.

In common with the MM-009 and MM-010 trials, treatment with Len/Dex was modelled to continue until the occurrence of either disease progression or unacceptable side effects. The efficacy data from the trials therefore reflects treatment interruptions and dosage reductions. For consistency of costs and clinical effects, the cost of lenalidomide is modelled to be reduced by the dose intensity. This is defined as the average cost of lenalidomide allowing for the actual treatment interruptions and dose reductions experienced in the RCT divided by the cost of lenalidomide given no treatment interruptions or dose reductions. The dose intensity value of 92.4% for Len appears

consistent with the breakdown of use of the different doses of Len and dose interruptions shown in Figure 8, p110 of Celgene's submission. We note that although this factor remains constant in the PSA, it may have been relatively straightforward to assume uncertainty around dose intensity.

The dose intensity of Dex was assumed to be 100%. The modelled dose intensity of Dex has very little impact on the estimated cost-effectiveness of Len/Dex *v.* Dex for two reasons. First, Dex is very cheap compared to lenalidomide, therefore incremental drug costs are dominated by the cost of Len. Second, as Dex is used in both treatment arms, the cost of Dex will cancel to some extent.

However, we are concerned that the dose intensity of bortezomib is assumed as 100% in the model, and that this may bias the analysis in favour of Len. Celgene do not justify this assumption. We suspect that the dose intensity of bortezomib may be less than 100% for the following reasons. Disease progression led to early discontinuation in 29% of patients receiving bortezomib, however only 56% of patients taking bortezomib completed five 3-week cycles of bortezomib[44]. Furthermore, the median length of bortezomib therapy was six 3-week cycles = 4.2 months, whereas the median TTP was 6.2 months[44].

Given Celgene's dose intensities above, the average drug cost per patient is £4,019 per 28-day cycle of Len, £29 per 28-day cycle of Dex, and £3,050 per 21-day cycle of bortezomib.

5.3.3.4 Disease management costs

As noted in Section 5.3.3.3 above, per-patient non-drugs costs are far smaller than drug costs in the model.

In their report (Table 43), Celgene assume that patients in progression-free survival and post-progression would have one outpatient visit per month. However, outpatient appointments are not included in the model. When this is included in the model, all ICERs for Len/Dex increase slightly. Even when we include the costs of outpatient visits, we believe that the costs for medical management assumed by Celgene, £111 per month in progression-free survival and £149 per month in progressed disease, may be too low. For instance, in the assessment for bortezomib for multiple myeloma, the

manufacturer of bortezomib assumed a far higher cost of medical management. The ERG report on the submission of bortezomib[58] states that the manufacturer of bortezomib used Bruce et al (1999) to inform on resource use and unit costs. Bruce et al (1999) is a study of the economic impact of using clodronate in the management of patients with multiple myeloma from the perspective of the NHS. Apparently this study is based on expert opinion, with the experts drawing on experience from the MRC VI Myelomatosis trial. The manufacturer of bortezomib assumed a cost of £443 per month for medical management whilst patients are in progression-free survival and post-progression, whilst the ERG estimated a cost of £470 per month in 2004/5 terms. This included costs of hospitalisation, out-patient visits, tests, hospice care and other factors. When we inflate this amount according to the HCHS indices of PSSRU (2007)[59], this corresponds to £551 in 2008/9 terms, where we assume 4% p.a. inflation in the last two financial years. Assuming £551 / month in progression-free survival and post-progression in the model of Celgene, all ICERs for Len increase substantially, e.g. the ICER for Len/Dex *v.* Dex for >1 prior therapy increases from £24,600 to £33,900 / QALY, and the ICER for the >1 prior therapy with thalidomide increases from £22,600 to £31,500 / QALY.

In the absence of appropriate cost data for a particular disease, costs from other diseases are sometimes used. Therefore we note in passing that in the assessment of four drugs for renal cancer[60], the medical management costs in progression-free survival were assumed to be £159 / month, and in progressive disease, £311 / month. Also a cost of £937 / month (inflated to 2008/9 values) was assumed in progressive disease, in respect of hospital and hospice care, in a study of stage IV breast cancer[61].

All non-drug costs are indexed to 2005. Given that this appraisal is conducted in the 2008/9 financial year, we believe that all costs should be indexed to 2008/9. However, inflating non-drug costs used by Celgene to 2008/9 has negligible impact on the costs per QALY because their non-drug costs are very small compared to drug costs.

5.3.3.5 Costs of adverse events and disease-related complications

The clinical trial data was analysed over short, three-month time intervals to accurately model the timing of the adverse events. No extrapolation of AE is attempted beyond 2

years, that this, the time period of the trials. This approach seems reasonable, since by 2 years, the marginal incidence of AEs is small.

In the MM-009 and MM-010 trials, G-CSF, an expensive therapy, was administered only in response to Grade 3 or 4 myelosuppression. In the Len arm of MM-010, 38 patients (21.6%) received G-CSF, and in MM-009, 60 patients (33.9%) received G-CSF. Whilst G-CSF use is not explicitly included in the model, Celgene state that it is implicitly included in the cost of those inpatient and day case admissions for the treatment of grade 3 or 4 neutropenia. If a higher proportion of patients receiving Len require G-CSF than has been seen in cases of neutropenia generally in the NHS, this would have important implications for the costs of the intervention.

5.3.3.6 Health-Related Quality of Life

Celgene assume no difference in utility between the response levels CR, PR and SD. They suggest that better response may be associated with higher quality of life. They suggest therefore that their assumption of no difference in utility may be conservative for the cost-effectiveness of Len/Dex, since there were more complete and partial responders with Len/dex and a longer duration of response. However, expert opinion suggests that there is probably minimal difference in utility between the response levels CR, PR and SD. Furthermore, as stated below, cost-effectiveness is relatively insensitive to the utility of patients in progression-free survival.

As in the appraisal of bortezomib for multiple myeloma, Celgene sourced utility values from the cost-utility study of intensive chemotherapy followed by myeloablative therapy with autologous stem cell rescue as compared to intensive chemotherapy in multiple myeloma of Agthoven et al (2004)[41]. Celgene assumed a utility value of 0.81 for patients in progression-free survival (CR/PR/SD), based on the utility value of the general public at an age value corresponding to that of the patients in the study. Indeed, given a median age of approximately 63 years from the two RCTs of Len/Dex, the mean utility for members of the general public in the UK, for both males and females is approximately 0.80[62]. Agthoven et al. (2004)[41] also report utilities measured in a study by The Dutch-Belgian Haemato-Oncology Cooperative Study Group (HOVON). In this prospective multi-centre randomised phase III study of multiple myeloma to evaluate the efficacy of intensive chemotherapy followed by myeloablative therapy with autologous

stem cell rescue as compared to intensive chemotherapy, utility scores were obtained from a sample of the general UK population using the EQ-5D. Utilities for patients in response were 0.81 (intensive chemotherapy only) versus 0.65 (myeloablative treatment) at six months from randomisation, 0.80 versus 0.62 (12 months), 0.81 versus 0.69 (18 months) and 0.77 versus 0.75 (24 months). Therefore, Celgene assume a utility of 0.77 after two years for patients who have not progressed at the end of two years. We note that this adjustment has only very marginal impact on cost-effectiveness.

Based on expert opinion, we understand that patients with multiple myeloma in progression-free survival have a lower health-related quality of life than 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. However, cost-effectiveness is insensitive to the utility in progression-free survival. For instance, when the utility in progression-free survival is set equal to that in progressive disease, i.e. reduced from 0.81 to 0.64, the ICER for >1 prior therapies increases only slightly, from £24,600 to £25,500 / QALY.

Cost-effectiveness is far more sensitive to the assumption for the utility in progressive disease than the utility in progression-free survival. This is because patients tend to spend far more time in progressive disease than in progression-free survival.

In addition to Agthoven et al (2004)[41], we are aware of two other studies that quote utilities for patients with multiple myeloma[63;64]. These studies obtained utilities by mapping from disease-specific quality of life instruments. Nord et al (1997)[64] cite a utility of 0.65 at 6 months for patients taking melphalan and prednisone, and Gulbrandsen et al (2001)[63] cite a utility of 0.79 at 6 months for patients taking melphalan and prednisone and 0.73 for patients taking high dose melphalan and stem cell transplant. The disease stages of patients in Gulbrandsen et al (2001)[63], Nord et al (1997)[64] and the two Len/Dex RCTs were similar. Patients in Gulbrandsen et al (2001)[63], with a median age of approximately 52 years, were younger than those in the two Len/Dex RCTs (median 63 years), whereas patients in Nord et al (1997)[64], median age 67.5 years, were older than those in the two Len/Dex RCTs. Gulbrandsen et al (2001)[63], Nord et al (1997)[64] and Agthoven et al (2004)[41] all considered newly-diagnosed patients. Given that patients in the two Len/Dex RCTs had been diagnosed with multiple myeloma for a median of

approximately 3.5 years, and that patients in these RCTs had already received one or more prior therapies, it may be more appropriate to use slightly lower utilities than those cited in these studies. We are unable to suggest exact utility values. However, for illustrative purposes, when we assume a 10% reduction in the utility for progression-free survival, i.e. for CR/PR/PD 0.73 instead of 0.81, and a 10% reduction in the utility for progressive disease, 0.58 instead of 0.64, the ICER for >1 prior therapy increases from £24,600 to £27,400 / QALY.

We note that Celgene do not adjust utility for *treatment-related* adverse events. Given that there are greater frequencies of AEs under Len/Dex compared to Dex, this means that cost-effectiveness is biased in favour of Len/Dex. However, we cannot quantify the magnitude of this bias.

5.3.3.7 Internal consistency

As far as possible, we have thoroughly checked the mathematics, statistics, internal logic, implementation of the model in Excel and the cost-effectiveness results presented in Celgene's report. As stated in Table 20 above, we discovered several important logical errors in the economic model first sent to us by Celgene. Some of these errors were very important because they altered the ICERs by several thousand £ / QALY. The model was corrected for these errors.

5.3.3.8 External consistency

Celgene have provided no evidence that the results of their model have been compared to the results of similar models. However, they do present evidence of the fit of the exponential distribution for overall survival to the MRC trial data in Appendix 8 (p149-150) of their report.

As in Section 5.3.3.2, we find that the modelled overall survival for Len/Dex is clearly better than presented in the two papers of the RCTs of Len/Dex *v.* Dex. This biases cost-effectiveness in favour of Len/Dex. We believe that Celgene should have presented this comparison given that it is important, obvious and easily performed.

5.3.4 Assessment of uncertainty

5.3.4.1 One-way sensitivity analyses

The manufacturer presents extensive sensitivity analyses on the key model parameters (Tables 59-63, p146-150 of their report). They find that cost-effectiveness is most sensitive to the utilities and to the assumed median OS for Dex. We find these results reasonable.

5.3.4.2 Probabilistic sensitivity analysis

The following parameters are varied stochastically in the probabilistic sensitivity analysis: the regression coefficients for time to progression and post-progression survival, utilities and management costs (see Appendix 14 p188 of Celgene's submission). Celgene state on p188 of their report that they run the model for 1,000 sets of input parameters, and for each set of input parameters, they run the model 10 times. This is appropriate. Utilities were modelled as beta distributions, which is appropriate. The standard error of each utility value was assumed as 10% of the mean. Whilst Celgene do not justify this assumption, we appreciate that the required data may not be available. Management costs were modelled by Gamma distributions, which is appropriate. Most of the uncertainty in cost-effectiveness is in the uncertainty in the regression coefficients in the survival equations. These coefficients were assumed to follow a multivariate normal distribution.

Celgene present cost-effectiveness acceptability curves and scatter plots in the cost-effectiveness plane. We have run the probabilistic sensitivity analysis and we are able to reproduce the CEACs shown in Celgene's submission. See Table 26, Table 27, Source: Celgene **submission Table 55 p. 140 & Figure 15 p.155**

Table 28, Table 29 above for the probabilities that Len/Dex is cost-effective at a willingness to pay threshold of £30,000 / QALY.

5.4 Summary of uncertainties and key issues

Whilst Celgene's model is appropriate for modelling the cost-effectiveness of Len/Dex for patients with multiple myeloma, we are not convinced that the complexity of the model

was merited. We believe that the complexity was a major contributor to the very serious logical errors in initial iterations of the model.

Cost-effectiveness of Len/Dex *v.* Dex is extremely sensitive to the estimate of Dex OS. Dex OS is based on experience from the MRC trials. We suspect that Dex OS may have improved over the time since these trials. If so, Len/Dex would appear worse value for money versus Dex than calculated by Celgene in all subgroups.

Furthermore, we have methodological reservations in the use of the MRC data to adjust Dex PPS, because Len/Dex PPS is taken from one group of trials and Dex PPS is taken from a different set of trials (see Section 5.3.3.2), i.e. trial randomisation has been broken for the comparison of Len/Dex with Dex.

Celgene have modelled Dex PPS so that the modelled **median** OS for Dex equals that from the calibrated MRC trials. We suggest that it is preferable to adjust PPS of Dex so that the modelled **mean** OS of Dex equals the **mean** OS in the calibrated trials data. This then makes Len/Dex appear worse value for money than calculated by Celgene in the >1 prior therapy subgroups.

The model predicts better OS for Len/Dex than experienced in the MM RCTs. When we adjust OS for Len/Dex so that the model now accurately predicts the median OS from the RCTs, the modelled fit to Len/Dex OS appears closer to the experience from the RCTs, and Len/Dex appears worse value for money in all comparisons.

In all comparisons, there is a large degree of extrapolation of OS for Len/Dex. Given that the per patient costs and QALYs for Len/Dex are strongly dependent on OS, all ICERs are subject to a large degree of uncertainty.

Celgene have assumed far lower medical management costs than were assumed in the NICE appraisal of bortezomib. If the costs from the bortezomib appraisal had been used, then the cost-effectiveness of Len/Dex would appear worse in all comparisons.

The cost-effectiveness of Len/Dex was assessed against bortezomib monotherapy. However, bortezomib is routinely used in combination with Dex in clinical practice in England and Wales. Given that bortezomib plus Dex may be more effective than bortezomib monotherapy, and that Dex is very cheap, the ICER of Len/Dex *v.* bortezomib may be underestimated.

Celgene did not model the bortezomib response-rebate scheme. We suggest that if this scheme had been modelled, then Len/Dex would have appeared worse value for money versus bortezomib.

In Section 6.2 we quantify the impact on the ICER for each comparison of using alternative inputs for these parameters.

6 ADDITIONAL WORK UNDERTAKE BY THE ERG

6.1 Clinical effectiveness

6.1.1 Meta-analysis

For reasons discussed in Section 4.1.7.2, above, we believe the meta-analysis techniques adopted for time-to-event data in the submission are inappropriate. A more robust approach would be to meta-analyse hazard ratios reflecting the difference between study arms. We have performed such analyses for TTP and OS; results are tabulated in Table 33.

Table 33: ERG's meta-analysis of time-to-event outcomes (all therapy prior strata)

			Fixed-effects model			Random-effects model		
Outcome	**Source data**	***p* (het.)**	**HR**	**(95%CI)**	***p*(HR=0)**	**HR**	**(95%CI)**	***p*(HR=0)**
TTP	Tab. 13, p. 68[a]	0.960	0.353	(0.290, 0.428)	<0.001	0.353[b]	(0.290, 0.428)	<0.001
OS	Tab. 14, p. 72	0.142	0.541	(0.413, 0.709)	<0.001	0.540	(0.363, 0.803)	0.002

[a] For TTP, the raw data has been inverted, because it is presented as Dex *v*. Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

[b] Note that, because inter-study variance (τ^2) is estimated as zero for TTP, the random-effects model produces identical results to the fixed-effects analysis.

Notably, the estimate for TTP is very close to that derived from the pooled individual patient data analysis (HR=0.35 [95%CI: 0.29, 0.43]; Table 23, p. 83).

[redacted]

6.1.2 Mixed treatment comparison

For reasons discussed in Section 4.1.7.3, above, we believe the mixed treatment comparison techniques adopted for TTP data in the submission are inappropriate. A more robust approach would be to use hazard ratios reflecting the difference between study arms as the basis for comparison. We have performed this analysis; results are tabulated in Table 34 (for WinBUGS code, see Appendix 23).

Table 34: ERG's mixed treatment comparison for Time To Progression (1 prior therapy only)

Comparison	Input data		HR (95%CI)
Len/Dex *v*. Bortezomib	[indirect comparison]		0.557 (0.337, 0.912)
Len/Dex *v*. Dex	MM-009: HR 0.311; SE(lnHR) 0.234 MM 010: HR 0.312; SE(lnHR) 0.266	Table 16, p.76[a]	0.312 (0.220, 0.438)
Bortezomib *v*. Dex	HR: 0.56; SE(lnHR): 0.186	APEX trial[b]	0.558 (0.388, 0.804)

[a] The reported data has been inverted, because it is presented in the submission as Dex *v.* Len/Dex (i.e. HR>1 for survival profile favouring Len/Dex)

[b] HR as reported in the main RCT publication[44]; SE calculated on basis of data extracted from published Kaplan-Meier curves using method of Parmar and colleagues,[65] with the aid of spreadsheet available as supplement to the review by Tierney and colleagues[66]

As would be expected, in this analysis, the posterior HRs are very similar to the input HRs for which there is empirical data. The estimated indirect comparison between the two active technologies suggests that TTP benefit is significantly greater for Len/Dex than for Bortezomib.

6.2 Cost-effectiveness analysis

In this section we quantify the impact on the Celgene ICER of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. We distinguish between items where we disagree with inputs used by Celgene, and items which we suggest are more matters of judgement.

For each comparator, we derive a preferred ICER by updating the model for the items where we believe that alternative assumptions are more appropriate. We then derive ICERs based on Celgene's base case and the preferred ICER by updating the model for the items where alternative assumptions are more a matter of judgement.

In all the comparisons below, there is a large degree of extrapolation of overall survival for Len/Dex. Given that the total per-patient costs and QALYs for Len/Dex are strongly dependent on overall survival, all ICERs below are subject to a large degree of uncertainty. For this factor alone, it is not possible to say whether this leads to an overestimate or underestimate of the ICERs.

In all the comparisons of Len/Dex *v.* Dex below, we have methodological reservations in the use of the MRC data to adjust Dex post-progression survival, because Len/Dex post-progression survival is taken from one group of RCTs and Dex post-progression survival

is taken from a different set of trials (see Section 5.3.3.2), i.e. trial randomisation has been broken. As noted above, if the unadjusted post-progression survival, taken from the two RCTs of Len/Dex *v.* Dex were modelled, then all the ICERs comparing Len/Dex *v.* Dex would be far higher (e.g. ICER for >1 prior therapy would increase from £24,600 to £79,000 / QALY). Nonetheless, we appreciate the need to adjust Dex post-progression survival from the Len/Dex RCTs to allow for cross-over of patients from Dex to Len/Dex, and given the available data. However, this does mean that all ICERs of Len/Dex *v.* Dex below are subject to a large degree of uncertainty. For this factor alone, it is not possible to say whether this leads to an overestimate or underestimate of the ICERs.

The combination of the above points means that there is a very large degree of uncertainty in all ICERs comparing Len/Dex with Dex.

In all comparisons, as stated above, cost-effectiveness is relatively insensitive to the utility in progression-free survival, but is sensitive to the utility in progressive disease. Given that there is uncertainty in the utility in progressive disease, this introduces another element of uncertainty in all ICERs below. Because we have no evidence to suggest that Celgene's estimate of utility in progressive disease is biased, we cannot say whether the choice of utility leads to a bias in the ICERs.

6.2.1 Lenalidomide plus dexamethasone *v.* bortezomib

Table 35 and Table 36 assess the impact on the Celgene ICER comparing Len/Dex with bortezomib of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. Where possible, this is quantified, otherwise the direction of impact is indicated.

Table 35: Recalculated ICER for Len/Dex *v.* bortezomib based on ERG preferred assumptions

Item		ICER Len/Dex *v.* bortezomib (£/QALY)
	Celgene base case	[redacted]
1[a]	Bortezomib plus Dex, not bortezomib monotherapy	ICER increases
2[b]	Bortezomib response-rebate scheme modelled	ICER increases
3[b]	Improved fit to Len/Dex OS	[redacted]
4[b]	Maximum 11, not 8 bortezomib cycles	[redacted]
5[b]	Bortezomib administration cost reduced	[redacted]
6[c]	Bortezomib dose intensity <100%, not 100%	ICER increases
	Preferred ICER (combination of all items)	**>** [redacted]

[a] See Section 5.3.2.
[b] See Section 5.3.3.2.
[c] See Section 5.3.3.3.

Table 36: Recalculated ICER for Len/Dex *v.* bortezomib when each item, considered a matter of judgement, is changed in model

Item		Effect on ICER (from Celgene base case)	Effect on ICER (from ERG preferred)
	Base case	[redacted]	[redacted]
1[a]	Medical management costs from bortezomib appraisal	[redacted]	[redacted]

[a] See Section 5.3.3.4.

6.2.2 Lenalidomide plus dexamethasone *v.* dexamethasone (1 prior therapy)

Table 37 and

Table 38 assess the impact on the Celgene ICER comparing Len/Dex with Dex alone (in those receiving one prior therapy) of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. Where possible, this is quantified, otherwise the direction of impact is indicated.

Table 37: Recalculated ICER for Len/Dex *v.* Dex (1 prior therapy) based on ERG preferred assumptions

Item		ICER Len/Dex *v.* Dex (£/QALY)
	Celgene base case	**£46,900**
1[a]	Improved fit to Len/Dex OS	**£69,500**
2[b]	Utilities lower for Len/Dex *v.* Dex due to more AEs with Len/Dex	ICER increases
	Preferred ICER (combination of all items)	**> £69,500**

[a] See Section 5.3.2.
[b] See Section 5.3.3.6.

Table 38: Recalculated ICER for Len/Dex *v.* Dex (1 prior therapy) when each item, considered a matter of judgement, is changed in model

Item		Effect on ICER (from Celgene base case)	Effect on ICER (from preferred ICER)
	Base case	**£46,900**	**> £69,500**
1 [a]	Improvement in OS of Dex over time	> £46,900	> £69,500
2 [b]	Medical management costs from bortezomib appraisal	£55,900	> £79,100

[a] See Section 5.3.2.
[b] See Section 5.3.3.2.

6.2.3 Lenalidomide plus dexamethasone *v.* dexamethasone (>1 prior therapy)

Table 39and

Table 40 and assess the impact on the Celgene ICER comparing Len/Dex with Dex alone (in those receiving more than one prior therapy) of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. Where possible, this is quantified, otherwise the direction of impact is indicated.

Table 39: Recalculated ICER for Len/Dex v. Dex (>1 prior therapy) based on ERG preferred assumptions

Item		ICER Len/Dex *v.* Dex (£/QALY)
	Celgene base case	**£24,600**
1 [a]	Dex mean (not median) OS equal to MRC derived mean	£33,200
2 [b]	Improved fit to Len/Dex OS	£32,900
3 [c]	Utilities lower for Len/Dex *v.* Dex due to more AEs with Len/Dex	ICER increases
	Preferred ICER (combination of all items)	**> £47,100**

[a] See Section 5.3.3.2
[b] See Section 5.3.2
[c] See Section 5.3.3.6

Table 40: Recalculated ICER for Len/Dex *v.* Dex (>1 prior therapy) when each item, considered a matter of judgement, is changed in model

Item		Effect on ICER (from Celgene base case)	Effect on ICER (from preferred ICER)
	Base case	**£24,600**	**> £47,100**
1 [a]	Improvement in OS of Dex over time	> £24,600	> £47,100
2 [b]	Medical management costs from bortezomib appraisal	£33,900	> £56,600

[a] See Section 5.3.2.
[b] See Section 5.3.3.2.

6.2.4 Lenalidomide plus dexamethasone *v.* dexamethasone (1 prior therapy - thalidomide).

Table 41 and

Table 42 assess the impact on the Celgene ICER comparing Len/Dex with Dex alone (in those receiving one prior therapy – that of thalidomide) of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. Where possible, this is quantified, otherwise the direction of impact is indicated.

Table 41: Recalculated ICER for Len/Dex v. Dex (prior thalidomide treatment) based on ERG preferred assumptions

Item		ICER Len/Dex *v.* Dex (£/QALY)
	Celgene base case	**£38,900**
1 [a]	Improved fit to Len/Dex OS	£56,500
2 [b]	Utilities lower for Len/Dex *v.* Dex due to more AEs with Len/Dex	ICER increases
	Preferred ICER(combination of all items)	**> £56,500**

[a] See Section 5.3.2
[b] See Section 5.3.3.6

Table 42: Recalculated ICER for Len/Dex v. Dex (prior thalidomide treatment), considered a matter of judgement, is changed in model

Item		Effect on ICER (from Celgene base case)	Effect on ICER (from preferred ICER)
	Base case	**£38,900**	**> £56,500**
1	Improvement in OS of Dex over time	> **£38,900**	> £56,500
2	Medical management costs from bortezomib appraisal	£47,600	> £65,400

[a] See Section 5.3.2.
[b] See Section 5.3.3.2.

6.2.5 Lenalidomide plus dexamethasone *v.* dexamethasone (>1 prior therapy, treatment with thalidomide).

Table 43and Table 44 assess the impact on the Celgene ICER comparing Len/Dex with Dex alone (in those receiving more than one prior therapy, one of which was thalidomide) of using alternative assumptions for items discussed in Chapter 5, both individually and cumulatively. Where possible, this is quantified, otherwise the direction of impact is indicated.

Table 43: Recalculated ICER for Len/Dex v. Dex (>1 prior treatment, prior thalidomide treatment) based on ERG preferred assumptions.

Item		ICER Len/Dex *v.* Dex (£/QALY)
	Celgene base case	**£22,600**
1 [a]	Dex mean (not median) OS equal to MRC derived mean	£30,200
2 [b]	Improved fit to Len/Dex OS	£30,800
3 [c]	Utilities lower for Len/Dex *v.* Dex due to more AEs with Len/Dex	ICER increases
	Preferred ICER (combination of all items)	**> £43,600**

[a] See Section 5.3.3.2.
[b] See Section 5.3.2
[c] See Section 5.3.3.6

Table 44: Cost / QALY for Len/Dex *v.* Dex when each item, considered a matter of judgement, is changed in model

Item		Effect on ICER (from Celgene base case)	Effect on ICER (from preferred ICER)
	Base case	**£22,600**	**> £43,600**
1	Improvement in OS of Dex over time	> £22,600	> £43,600
2	Medical management costs from bortezomib appraisal	£31,500	> £53,000

[a] See Section 5.3.2.
[b] See Section 5.3.3.2.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

- The searches for clinical effectiveness data are appropriate and relevant trial data is included.

- The use of Len/Dex for multiple myeloma (MM) in people who have received at least one prior therapy reflects the licensed indication.

- The evidence is based on two, identically designed, good quality RCTs: MM-010 and MM-009 in people with multiple myeloma who had received at least one prior therapy. Pooled analysis of these trials shows increased time to progression with len/Dex [median 48.3 weeks *v.* 20.1 wks. HR 0.35 (95% CI 0.29, 0.43)]. Increased overall survival is also seen with Len/Dex. Based on updated data, median survival with Len/Dex increased from [redacted] compared with Dex alone.

- The main threat to validity for the clinical effectiveness data is the high level of crossover in the trials, leading to a strong Len effect in the comparator arm. This is likely to underestimate treatment effect, especially for overall survival. This is a problem in many assessments of new chemotherapy in end stage cancer and it would be unethical to undertake trials that did not allow for such crossover. However, this does introduce uncertainty into the results, particularly for overall survival.

- Methods used in the mixed treatment comparison undertaken to estimate the effectiveness of Len/Dex compared to bortezomib monotherapy are inappropriate. However, when recalculated using more appropriate methods, Len/Dex shows increased time to progression [HR 0.56 (95% CI 0.34, 0.91)].

7.2 Summary of cost-effectiveness issues

We consider that there are a number of incorrect assumptions used in the Celgene economic model, and have calculated the impact of these on the ICERs (Table 45).

Table 45: Comparison of Celgene base case ICERs and ERG preferred ICERs by patient subgroup and treatment comparison

Patient subgroup	Treatments	Celgene base case ICER	ERG preferred ICER
One prior therapy only	Len/Dex *v.* bortezomib	[redacted]	[redacted]
One prior therapy only and have pre-existing peripheral neuropathy	Len/Dex *v.* Dex	£46,900	> £69,500
At least two prior therapies	Len/Dex *v.* Dex	£24,600	> £47,100
Prior treatment with thalidomide (1 prior therapy only)	Len/Dex *v.* Dex	£38,900	> £56,500
Prior treatment with thalidomide (2 or more therapies)	Len/Dex *v.* Dex	£22,600	> £43,600

Key areas of contention and uncertainty are listed below, in order of importance;

- Cost-effectiveness of Len/Dex *v.* Dex is extremely sensitive to the estimate of Dex OS. Dex OS is based on experience from the MRC trials. We suspect that Dex OS may have improved over the time since these trials. If so, Len/Dex would appear worse value for money versus Dex than calculated by Celgene in all subgroups.

- Furthermore, we have methodological reservations in the use of the MRC data to adjust Dex PPS, because Len/Dex PPS is taken from one group of trials and Dex PPS is taken from a different set of trials (see Section 5.3.3.2), i.e. trial randomisation has been broken for the comparison of Len/Dex with Dex.

- Celgene have modelled Dex PPS so that the modelled **median** OS for Dex equals that from the calibrated MRC trials. We suggest that it is preferable to adjust PPS of Dex so that the modelled **mean** OS of Dex equals the **mean** OS in the calibrated trials data. This then makes Len/Dex appear worse value for money than calculated by Celgene in the >1 prior therapy subgroups.

- The model predicts better OS for Len/Dex than experienced in the MM RCTs. When we adjust OS for Len/Dex so that the model now accurately predicts the median OS from the RCTs, the modelled fit to Len/Dex OS appears closer to the experience from the RCTs, and Len/Dex appears worse value for money in all comparisons.

- In all comparisons, there is a large degree of extrapolation of OS for Len/Dex. Given that the per patient costs and QALYs for Len/Dex are strongly dependent on OS, all ICERs are subject to a large degree of uncertainty.

- Celgene have assumed far lower medical management costs than were assumed in the NICE appraisal of bortezomib. If the costs from the bortezomib appraisal had been used, then the cost-effectiveness of Len/Dex would appear worse in all comparisons.

- The cost-effectiveness of Len/Dex was assessed against bortezomib monotherapy. However, bortezomib is routinely used in combination with Dex in clinical practice in England and Wales. Given that bortezomib plus Dex may be more effective than bortezomib monotherapy, and that Dex is very cheap, the ICER of Len/Dex *v.* bortezomib may be underestimated.

- Celgene did not model the bortezomib response-rebate scheme. We suggest that if this scheme had been modelled, then Len/Dex would have appeared worse value for money versus bortezomib.

7.3 Implications for research

- An RCT of continuous Len/Dex until progression or intolerable side effects *v.* shorter regimens is required.

- Longer-term follow-up of overall survival in patients taking Len/Dex would be very useful to inform the cost-effectiveness model.

- Quantification of any improvement in overall survival for Dex patients from the time of the MRC trials (1980 to 1997) to the present day would be useful.

- Accurate estimates of the utilities of patients in progression-free survival and post-progression survival separately for Len/Dex and Dex would be useful to allow more accurate estimates of cost-effectiveness.

- A head-to-head RCT of Len/Dex *v.* bortezomib/Dex would improve the accuracy of the assessment of the relative clinical effectiveness and cost-effectiveness of these treatments.

- We are not convinced that a complex discrete event model is appropriate (see Section 5.3.2). This view could usefully be tested by implementing an alternative model of the cost effectiveness of Len/Dex based on a simpler Markov, cohort-based approach.

8 REFERENCES

(1) Weber DM, Chen C, Niesvizky R, Wang M, Belch A, Stadtmauer EA et al. Lenalidomide plus Dexamethasone for Relapsed Multiple Myeloma in North America. New England Journal of Medicine 2007; 357(21):2133-2142.

(2) Dimopoulos M, Spencer A, Attal M, Prince HM, Harousseau JL, Dmoszynska A et al. Lenalidomide plus Dexamethasone for Relapsed or Refractory Multiple Myeloma. New England Journal of Medicine 2007; 357(21):2123-2132.

(3) Weber D, Knight R, Chen C, Spencer A, Yu Z, Zeldis J et al. Prolonged Overall Survival with Lenalidomide Plus Dexamethasone Compared with Dexamethasone Alone in Patients with Relapsed or Refractory Multiple Myeloma. Blood 2007; 110(11):412.

(4) Dimopoulos MA, Spencer A, Attal M, Prince M, Harousseau JL, Dmoszynska A et al. Study of Lenalidomide Plus Dexamethasone Versus Dexamethasone Alone in Relapsed or Refractory Multiple Myeloma (MM): Results of a Phase 3 Study (MM-010). Blood 2005; 106(11):6.

(5) Dimopoulos M, Weber D, Chen C, Spencer A, Niesvizky R, Attal M. Evaluating oral lenalidomide (Revlimid) and dexamethasone versus placebo and dexamethasone in patients with relapsed or refractory multiple myeloma. Haematologica 2005; 90(suppl 2):160.

(6) Chanan-Khan AA, Weber D, Dimopoulos M, Chen C, Niesvizky R, Wang M et al. Lenalidomide (L) in combination with dexamethasone (D) improves survival and time to progression in elderly patients (pts) with relapsed or refractory (rel/ref) multiple myeloma (MM). Blood 2006; 108:1014a.

(7) Chanan-Khan AA, Yu Z, Weber D, Chen C, Niesvizky R, Spencer A et al. Lenalidomide (L) in Combination with Dexamethasone (D) Significantly Improves Time to Progression (TTP) in Non-Stem Cell Transplant Patients (pts) with Relapsed or Refractory (rel/ref) Multiple Myeloma (MM): Analysis from MM-009 and MM-010 Randomized Phase III Clinical Trials. ASH Annual Meeting Abstracts 2006; 108(11):3554.

(8) Stadtmauer E, Weber D, Dimopolous M, Belch A, Attal M, Prince M et al. Lenalidomide in Combination with Dexamethasone Is More Effective Than Dexamethasone at First Relapse in Relapsed Multiple Myeloma. ASH Annual Meeting Abstracts 2006; 108(11):3552.

(9) Miguel JFS, Dimopoulos M, Weber D, Olesnyckyj M, Yu Z, Zeldis J et al. Dexamethasone Dose Adjustments Seem To Result in Better Efficacy and Improved Tolerability in Patients with Relapsed/Refractory Multiple Myeloma Who Are Treated with Lenalidomide/Dexamethasone (MM009/010 Sub-Analysis). ASH Annual Meeting Abstracts 2007; 110(11):2712.

(10) Harousseau JL, Weber D, Dimopoulos M, Olesnyckyj M, Yu Z, Zeldis J et al. Relapsed/Refractory Multiple Myeloma Patients Treated with Lenalidomide/Dexamethasone Who Achieve a Complete or near Complete Response Have Longer Overall Survival and Time to Progression Compared with Patients Achieving a Partial Response. ASH Annual Meeting Abstracts 2007; 110(11):3598.

(11) Foa R, Weber D, Dimopoulos M, Olesnyckyj M, Yu Z, Zeldis J et al. Lenalidomide/Dexamethasone Improves Response and Prolongs Time to Progression,

Even in Patients with IgA Multiple Myeloma: A Sub-Analysis of the MM-009/010 Studies. ASH Annual Meeting Abstracts 2007; 110(11):4839.

(12) Chanan-Khan AA, Dimopoulos M, Weber D, Olesnyckyj M, Yu Z, Zeldis J et al. ECOG Performance Status Affects Efficacy, but Not Safety, of Lenalidomide/Dexamethasone in Relapsed/Refractory Multiple Myeloma (MM009/010 Sub-Analysis). ASH Annual Meeting Abstracts 2007; 110(11):2721.

(13) Niesvizky R, Spencer A, Wang M, Weber D, Chen C, Dimopoulos MA et al. Increased risk of thrombosis with lenalidomide in combination with dexamethasone and erythropoietin. Journal of Clinical Oncology 2006; 24(18_suppl):7506.

(14) Weber D, Wang M, Chen C, Belch A, Stadtmauer EA, Niesvisky R et al. Lenalidomide Plus High-Dose Dexamethasone Provides Improved Overall Survival Compared to High-Dose Dexamethasone Alone for Relapsed or Refractory Multiple Myeloma (MM): Results of 2 Phase III Studies (MM-009, MM-010) and Subgroup Analysis of Patients with Impaired Renal Function. ASH Annual Meeting Abstracts 2006; 108(11):3547.

(15) Wang M, Knight R, Dimopolous M, Siegel D, Rajkumar SV, Facon T et al. Effect of len/dex in MM despite thal resistance. Haematologica 2007; 92(s2):172.

(16) Wang M, Knight R, Dimopoulos M, Siegel D, Rajkumar SV, Facon T et al. Lenalidomide in combination with dexamethasone was more effective than dexamethasone in patients who have received prior thalidomide for relapsed or refractory multiple myeloma. Blood 2006; 108:3553.

(17) Wang M, Knight R, Dimopoulos M, Siegel D, Rajkumar SV, Facon T et al. Lenalidomide in Combination with Dexamethasone Was More Effective Than Dexamethasone in Patients Who Have Received Prior Thalidomide for Relapsed or Refractory Multiple Myeloma. ASH Annual Meeting Abstracts 2006; 108(11):3553.

(18) Bladé J, Weber D, Zeldis J, Yu Z, Dmoszynska A, Attal M. Lenalidomide (Revlimid®) in combination with dexamethasone (dex) is more effective than dex alone in patients with relapsed or refractory multiple myeloma who have received prior thalidomide therapy. Haematologica 2006; 91(Suppl.1):181.

(19) Chanan-Khan AA, Weber D, Dimopoulos M, Chen C, Niesvizky R, Wang M et al. Lenalidomide (L) in Combination with Dexamethasone (D) Improves Survival and Time to Progression in Elderly Patients (pts) with Relapsed or Refractory (rel/ref) Multiple Myeloma (MM). ASH Annual Meeting Abstracts 2006; 108(11):3551.

(20) Chanan-Khan AA, Dimopoulos M, Weber D, Olesnyckyj M, Yu Z, Zeldis J et al. ECOG Performance Status Affects Efficacy, but Not Safety, of Lenalidomide/Dexamethasone in Relapsed/Refractory Multiple Myeloma (MM009/010 Sub-Analysis). ASH Annual Meeting Abstracts 2007; 110(11):2721.

(21) Durie BGM, Harousseau JL, Miguel JS, Bladé J, Barlogie B, Anderson K et al. International uniform response criteria for multiple myeloma. Leukemia 2006; 20(9):1467-1473.

(22) Bladé J, ., Samson D, Reece D, Apperley J, BJOrkstrand B et al. 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 1998; 102(5):1115-1123.

(23) Durie BGM, Jacobson J, Barlogie B, Crowley J. Magnitude of Response With Myeloma Frontline Therapy Does Not Predict Outcome: Importance of Time to Progression in Southwest Oncology Group Chemotherapy Trials. Journal of Clinical Oncology 2004; 22(10):1857-1863.

(24) Barlogie B, Tricot G, Anaissie E, Shaughnessy J, Rasmussen E, van Rhee F et al. Thalidomide and Hematopoietic-Cell Transplantation for Multiple Myeloma. N Engl J Med 2006; 354(10):1021-1030.

(25) Altman DG. Practical Statistics for Medical Research. Boca Raton: Chapman & Hall/CRC; 1991.

(26) Michiels S, Piedbois P, Burdett S, Syz N, Stewart L, Pignon JP. Meta-analysis when only the median survival times are known: A comparison with individual patient data results. International Journal of Technology Assessment in Health Care 2005; 21(01):119-125.

(27) Durie BG, Harousseau JL, Miguel JS, Blade J, Barlogie B, Anderson K et al. International uniform response criteria for multiple myeloma. Leukemia 2006; 20(9):1467-1473.

(28) FDA Center for Drug Evaluation and Research. Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Guideline for Industry - Clinical safety data management: definitions and standards for expedited reporting. 1995.
Ref Type: Report

(29) Chen C, Reece DE, Siegel D, Niesvizky R, Boccia RV, Stadtmauer EA et al. Expanded Access Program (EAP) for Lenalidomide (Revlimid(R)) Plus Dexamethasone in over 1400 Subjects with Relapsed or Refractory Multiple Myeloma. ASH Annual Meeting Abstracts 2006; 108(11):3556.

(30) Reece DE, Masih-Khan E, Chen C, Wang L, Dean S, Kukreti V et al. Use of Lenalidomide (Revlimid(R) +/- Corticosteroids in Relapsed/Refractory Multiple Myeloma Patients with Elevated Baseline Serum Creatinine Levels. ASH Annual Meeting Abstracts 2006; 108(11):3548.

(31) Reece DE, Masih-Khan E, Chen C, Wang L, Dean S, Trudel S et al. Lenalidomide (Revlimid(R)) +/- Corticosteroids in Elderly Patients with Relapsed/Refractory Multiple Myeloma. ASH Annual Meeting Abstracts 2006; 108(11):3550.

(32) Reece DE, Masih-Khan E, Chen C, Wang L, Dean S, Trudel S et al. Lenalidomide (Revlimid(R)) +/- Corticosteroids in Elderly Patients with Relapsed/Refractory Multiple Myeloma. ASH Presentation Slideshow [2006 Available from: URL:http://myeloma.org/pdfs/ash2006reece1.pdf

(33) Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. The Lancet 1999; 354(9193):1896-1900.

(34) Deniz B. Economic evaluation of lenalidomide use for multiple myeloma in Scotland in patients who have received one prior therapy. 13th Annual Meeting ISPOR . 3-5-0008.
Ref Type: Abstract

(35) Scottish Medicines Consortium (SMC). Assessment of lenalidomide in Scotland. 2008.
Ref Type: Report

(36) Drayson MT, Augustson BM, Begum G, Dunn JA, Barth NJ, Davies F et al. Survival from relapse and the influence of therapy. Haematologica/The Haematology Journal: XIth International Myeloma Workshop Proceedings 92, 173. 2007.
Ref Type: Abstract

(37) Deniz B. Economic evaluation of lenalidomide use for multiple myeloma in Wales in patients who have received one prior therapy. 13th EHA Congress . 2008.
Ref Type: Abstract

(38) All Wales Medicines Strategy Group (AWMSG). Final appraisal report. Lenalidomide (Revlimid) for multiple myeloma. 2008.
Ref Type: Report

(39) Kumar S, Therneau TM, Gertz MA, Lacy MQ, Dispenzieri A, Rajkumar SV et al. Clinical course of patients with relapsed multiple myeloma. Mayo Clin Proc 2004; 79(7):867-874.

(40) NICE. Bortezomib monotherapy for relapsed multiple myeloma: guidance. 2007.
Ref Type: Report

(41) van Agthoven M, Segeren CM, Buijt I, Uyl-de Groot CA, van der Holt B, Lokhorst HM et al. A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study. European Journal of Cancer 2004; 40:1159-1169.

(42) Mathers C, Vos T, Stevenson C. The burden of disease and injury in Australia. 1999. Canberra, Australia.
Ref Type: Report

(43) British Medical Association and Royal Pharmaceutical Society of Great Britain. British National Formulary (BNF) 55. Online edition [2008 [cited 2008 Mar. 19]; Available from: URL:http://www.bnf.org/bnf/extra/current/450035.htm

(44) Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. N Engl J Med 2005; 352(24):2487-2498.

(45) Celgene. A survey of resource utilisation among haemato-oncology clinicians in the UK and Ireland. 2007.
Ref Type: Report

(46) National Cancer Institute (NCI). Common Toxicity Criteria (CTC) Version 2.0. 1999.
Ref Type: Report

(47) Department of Health. NHS Reference Costs. 2005.
Ref Type: Report

(48) Department of Health. NHS Reference Costs 2005 - Outpatient Adult Follow Up Attendance (TOPS FUA) - Specialty: Clinical Haematology - Specialty Code: 303. Department of Health [2005 Available from: URL:http://www.dh.gov.uk/prod_consum_dh/idcplg?IdcService=GET_FILE&dID=7432&Rendition=Web

(49) Department of Health. NHS Reference Costs 2005 - Outpatient Adult First Attendance (TOPS FAA) - Specialty: General Medicine - Specialty Code: 300. Department of Health [

2005 Available from:
URL:http://www.dh.gov.uk/prod_consum_dh/idcplg?IdcService=GET_FILE&dID=7432&Rendition=Web

(50) Department of Health. NHS Reference Costs 2005 - Outpatient Adult Follow Up Attendance (TOPS FUA) - Specialty: General Medicine - Specialty Code: 300. Department of Health [2005 Available from:
URL:http://www.dh.gov.uk/prod_consum_dh/idcplg?IdcService=GET_FILE&dID=7432&Rendition=Web

(51) Department of Health. NHS Reference Costs 2005. Department of Health [2005 Available from:
URL:http://www.dh.gov.uk/prod_consum_dh/idcplg?IdcService=GET_FILE&dID=7432&Rendition=Web

(52) National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. 2008. London, NICE.
Ref Type: Generic

(53) Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. Oxford, UK: Oxford University Press; 1997.

(54) Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment: A Review and Consolidation of Quality Assessment. Pharmacoeconomics 2006; 24(4):355-371.

(55) Davies F, Wu P, Jenner M, Srikanth M, Saso R, Morgan G. The combination of cyclophosphamide, velcadeand dexamethasone (CVD) induces high responserates with comparable toxicity to velcade alone(V) and velcade plus dexamethasone (VD). Haematologica 2007; 92:1149-1150.

(56) Kropff M, Bisping G, et al. Bortezomib in combination with intermediate-dosedexamethasone and continuous low-dose oralcyclophosphamide for relapsed multiple myeloma. British Journal of Haematology 2007; 138:330-337.

(57) Kumar SK, Rajkumar SV, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK et al. Improved survival in multiple myeloma and the impact of novel therapies. Blood 2008; 111:2516-2520.

(58) Green C, Bryant J, Takeda A, Cooper K, Clegg A, Smith A et al. Bortezomib for the treatment of multiple myeloma patients. Evidence review group report commissioned by the NHS R&D HTA programme on behalf of NICE. 2006.
Ref Type: Report

(59) Netten A, Curtis L. Unit costs of Health and Social Care, PSSRU. 2005. University of Kent.
Ref Type: Report

(60) Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T et al. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. 2008. Universities of Exeter and Plymouth.
Ref Type: Report

(61) Remak E, Brazil L. Cost of managing women presenting with stage IV breast cancer in the United Kingdom. British Journal of Cancer 91, 77-83. 2004.
Ref Type: Journal (Full)

(62) Kind P, Hardman G, Macran S. UK population norms for EQ-5D. 1999. University of York, Centre for Health Economics.
Ref Type: Report

(63) Gulbrandsen N, Wisloff F, Nord E, Lenhoff S, Hjorth M, Westin J. Cost-utility analysis of high-dose melphalan with autologous blood stem cell support vs. melphalan plus prednisone in patients younger than 60 years with multiple myeloma. Eur J Haematol 2001; 66:328-336.

(64) Nord E, Wisloff F, Hjorth M, Westin J. Cost-utility analysis of melphalan plus prednisone with or without interferon-alpha2b in newly diagnosed multiple myeloma. Pharmacoeconomics 1997; 12(1):89-103.

(65) Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998; 17(24):2815-2834.

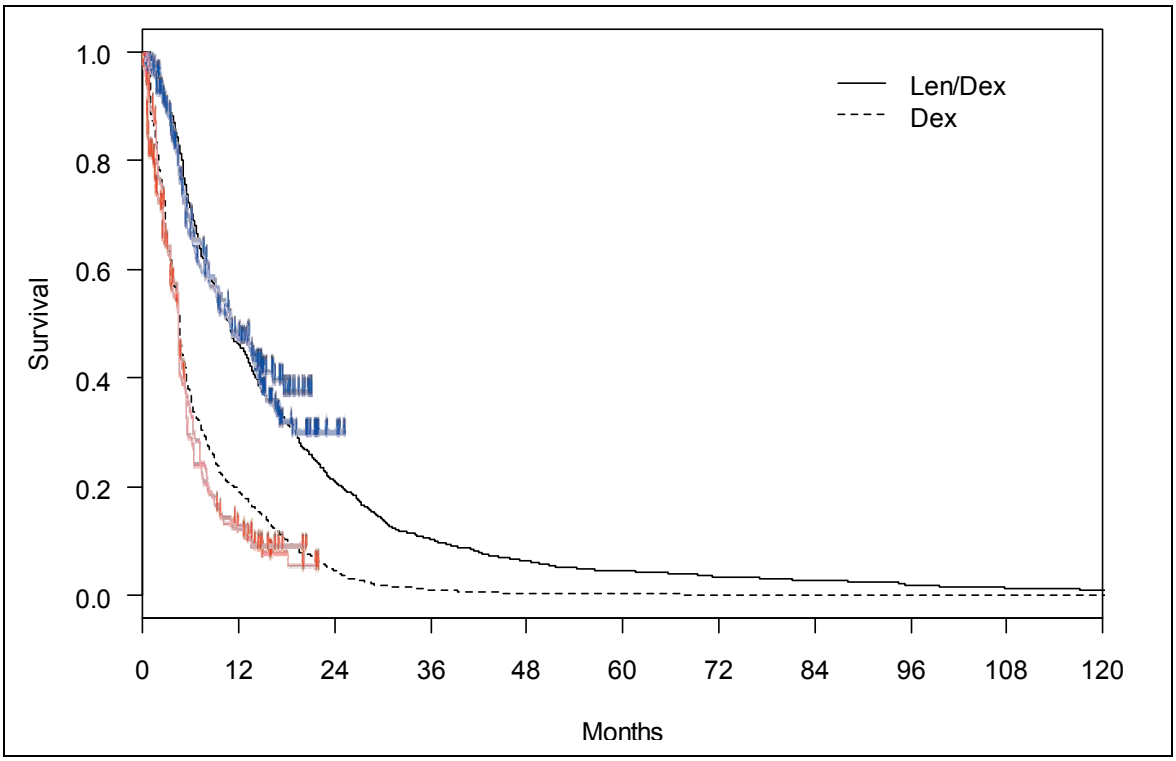
(66) Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007; 8:16.

APPENDICES

Appendix 1: Comparison of model output with empirical data (TTP)

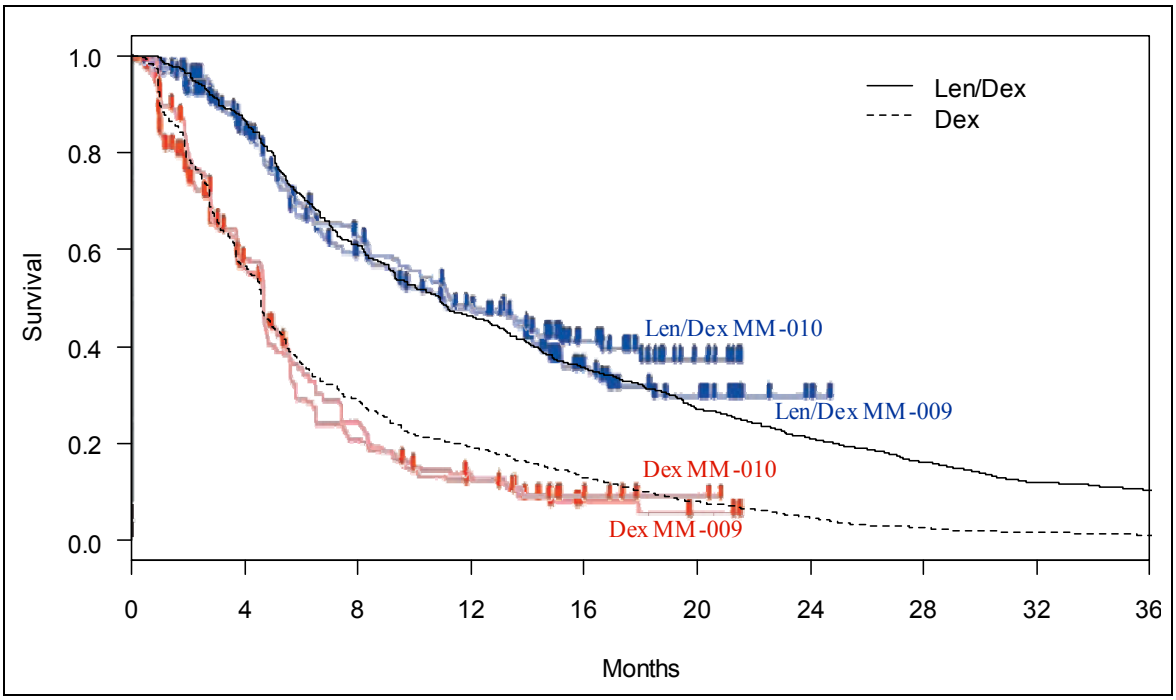
All graphs in the Appendix were produced by the ERG. All empirical Kaplan-Meier curves are reproduced from the two papers of the RCTs of Len/Dex vs. Dex. All Kaplan-Meier graphs of modelled output were generated by the ERG.

Figure 7: Time to progression for Len/Dex and Dex for 1 prior and >1 prior therapies pooled.



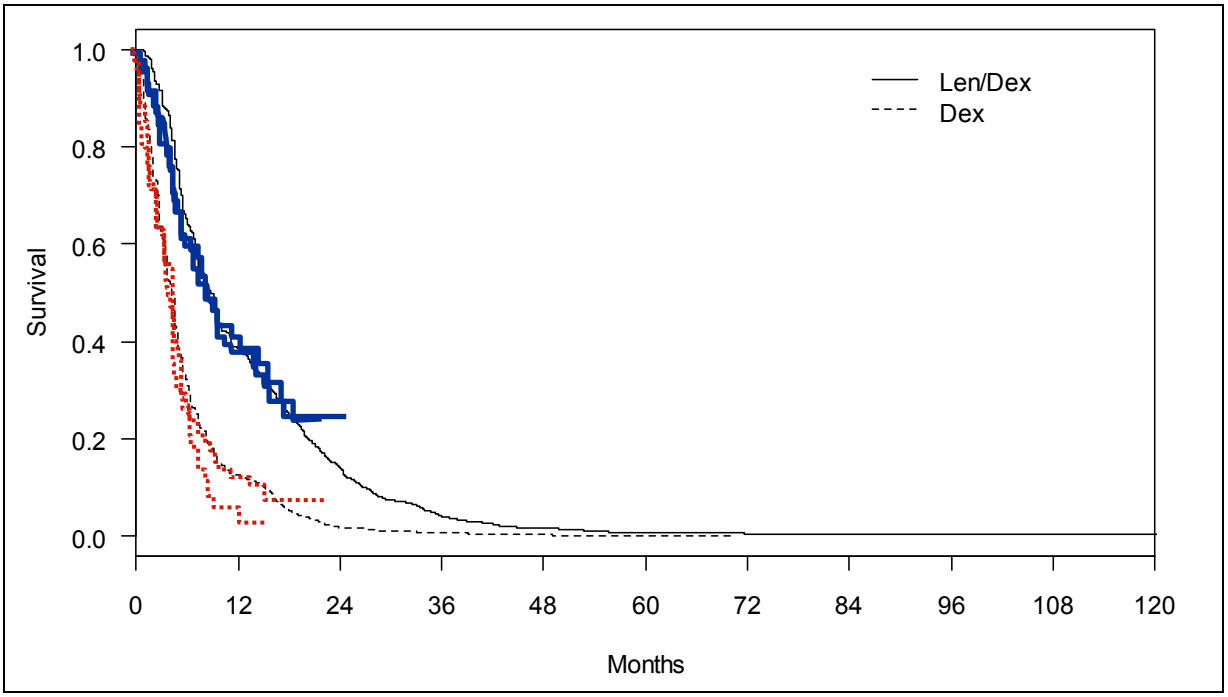
Pooled population weighted according to reported proportions of trial participants in each stratum (Len/Dex: 1 prior = 35.1%, ≥2 priors 64.9%; Dex: 1 prior = 35.3%, ≥2 priors 64.7%)
Smooth Kaplan-Meier curves represent model output, Kaplan-Meier curves with censorships represent empirical data.

Figure 8: Time to progression for Len/Dex and Dex for 1 prior and >1 prior therapies pooled.



As Figure 7, but expanded x-axis.

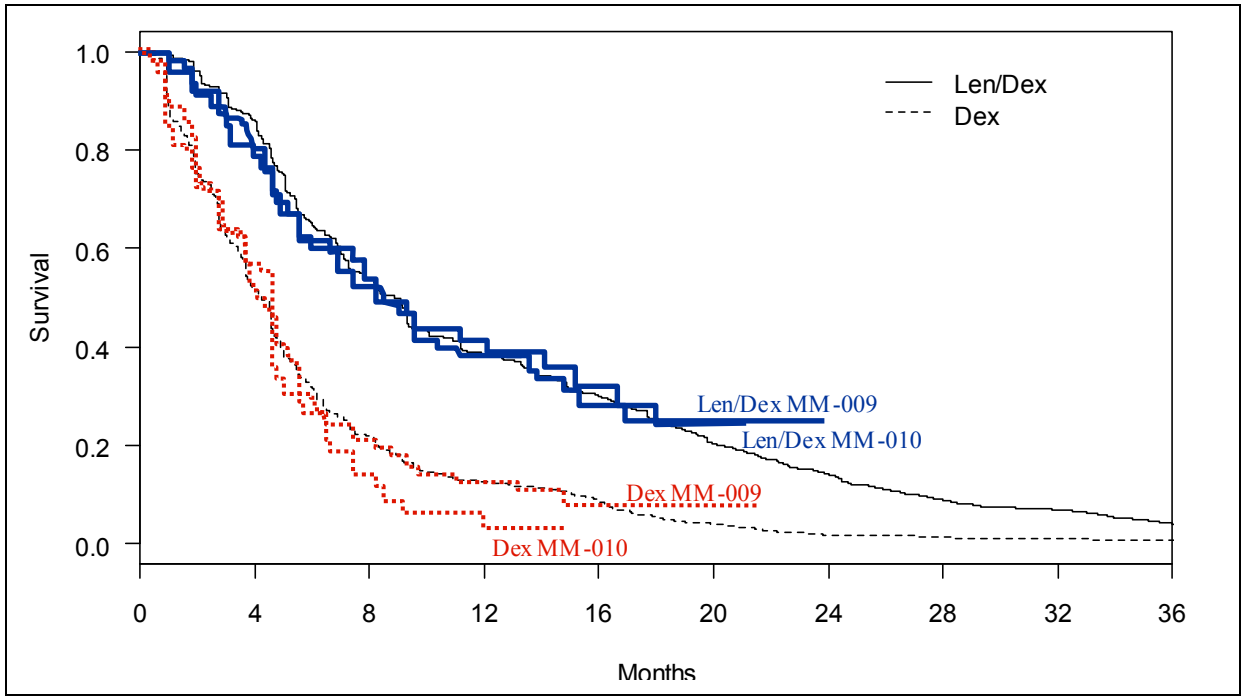
Figure 9: Time to progression for Len/Dex and Dex for 1 prior and >1 prior therapies pooled for thalidomide-exposed patient subgroup.



Pooled population weighted according to proportions of participants for whom IPD is provided in the model in each stratum (Len/Dex: 1 prior = 9.6%, ≥2 priors 90.4%; Dex: 1 prior = 13.2%, ≥2 priors 86.8%)

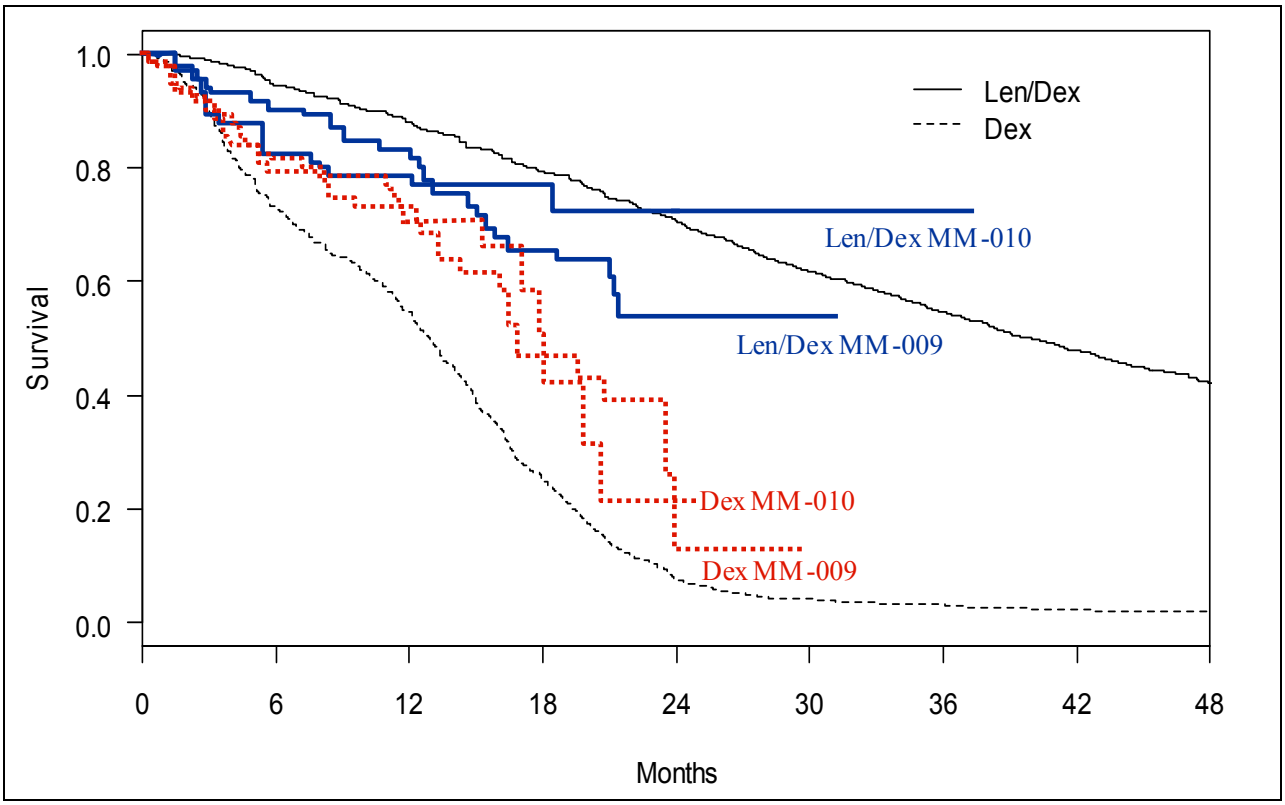
Smooth Kaplan-Meier curves represent model output, Kaplan-Meier curves with censorships represent empirical data.

Figure 10: Time to progression for Len/Dex and Dex for 1 prior and >1 prior therapies pooled for thalidomide-exposed patient subgroup.



As Figure 9, but expanded x-axis.

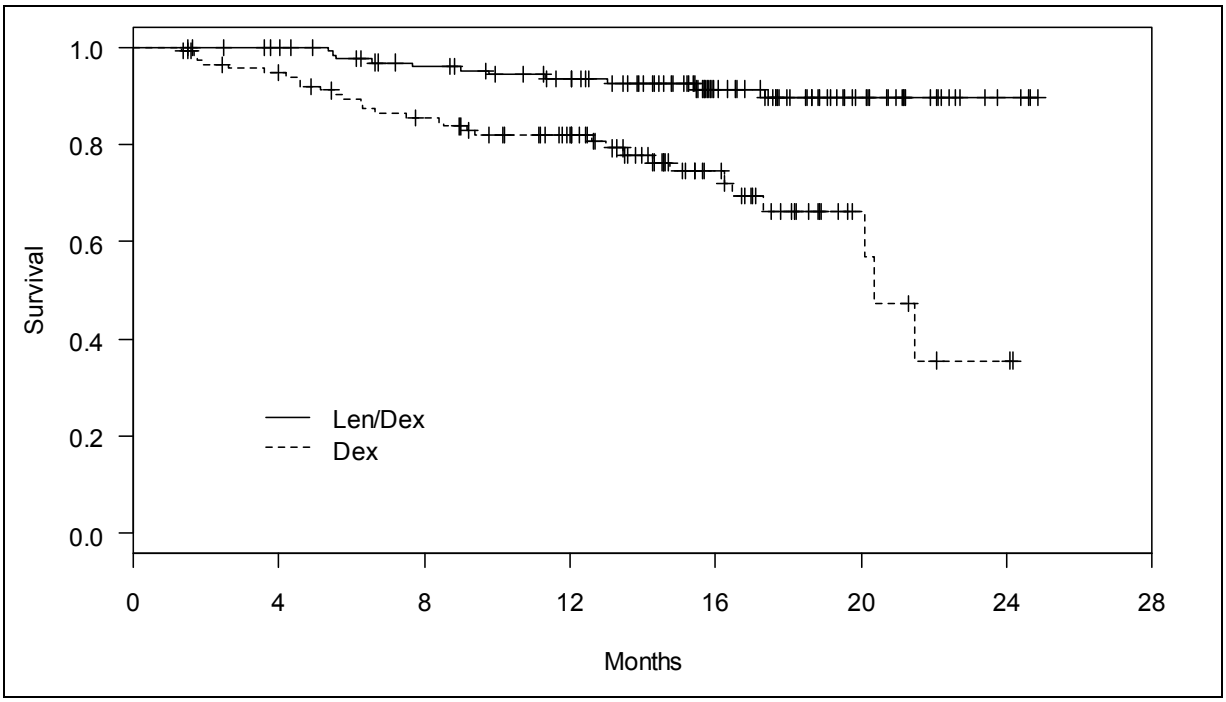
Figure 11: Overall survival for Len/Dex and Dex, thalidomide-exposed subgroup, for 1 prior and >1 prior therapies pooled.



Pooled population weighted according to proportions of participants for whom IPD is provided in the model in each stratum (Len/Dex: 1 prior = 9.6%, ≥2 priors 90.4%; Dex: 1 prior = 13.2%, ≥2 priors 86.8%)
Smooth Kaplan-Meier curves represent model output, other Kaplan-Meier curves represent empirical data.

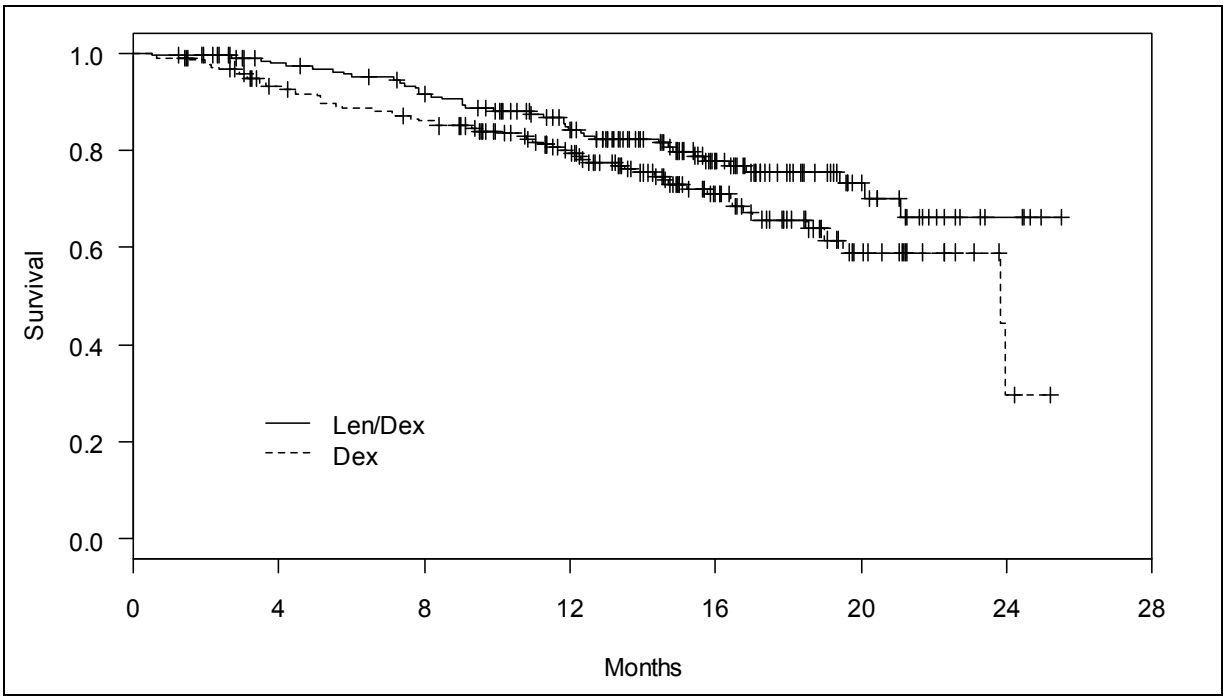
Appendix 2: Kaplan-Meier curves analysing patient-level data underpinning the economic model

Figure 12: Overall survival in patient-level data underpinning economic model – 1 prior therapy only



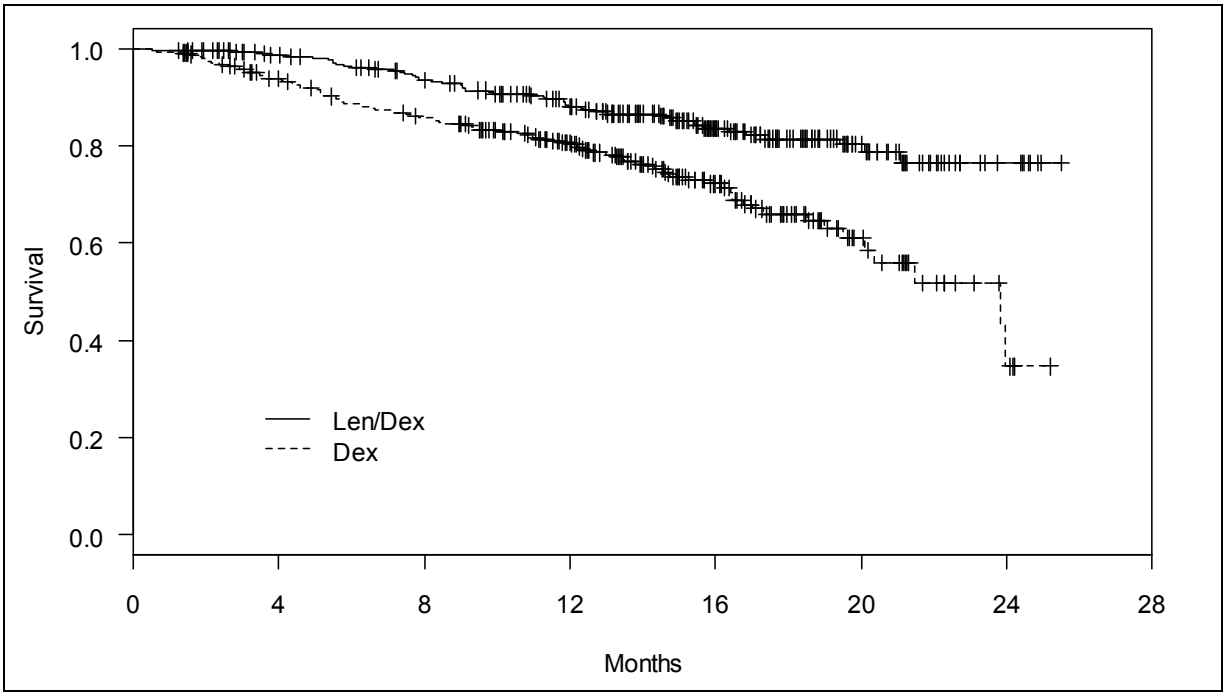
HR = 0.226 (0.112, 0.452)

Figure 13: Overall survival in patient-level data underpinning economic model – ≥2 prior therapies



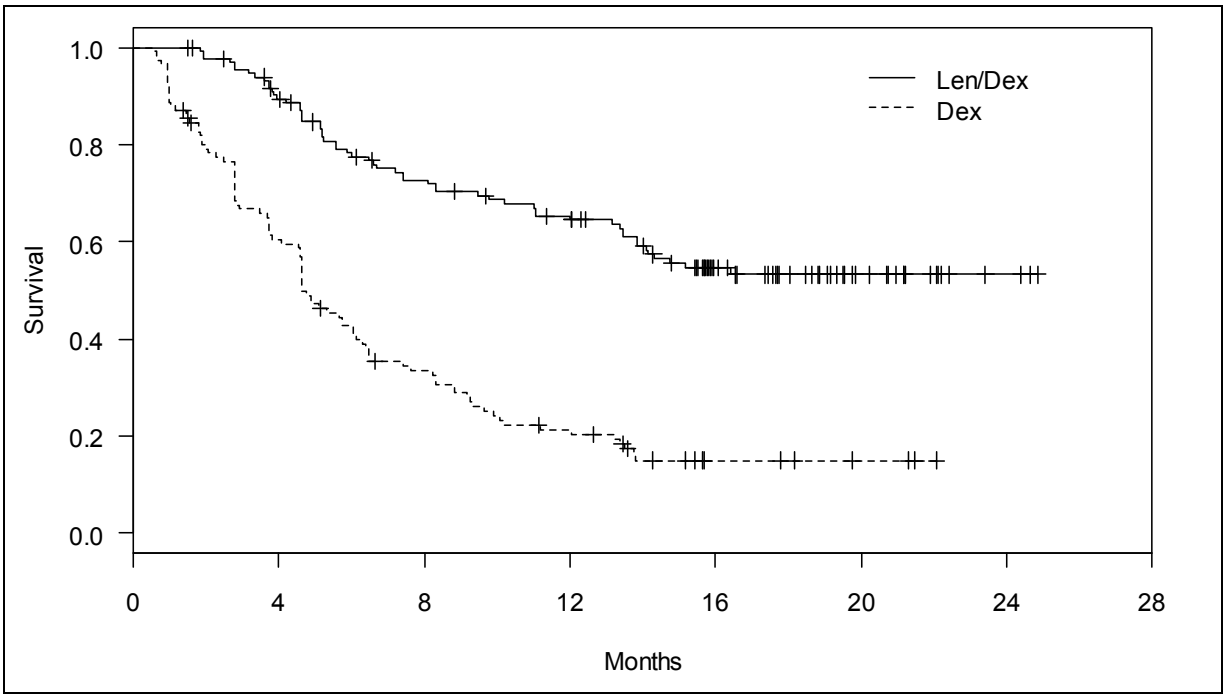
HR = 0.657 (0.440, 0.979)

Figure 14: Overall survival in patient-level data underpinning economic model – both strata pooled



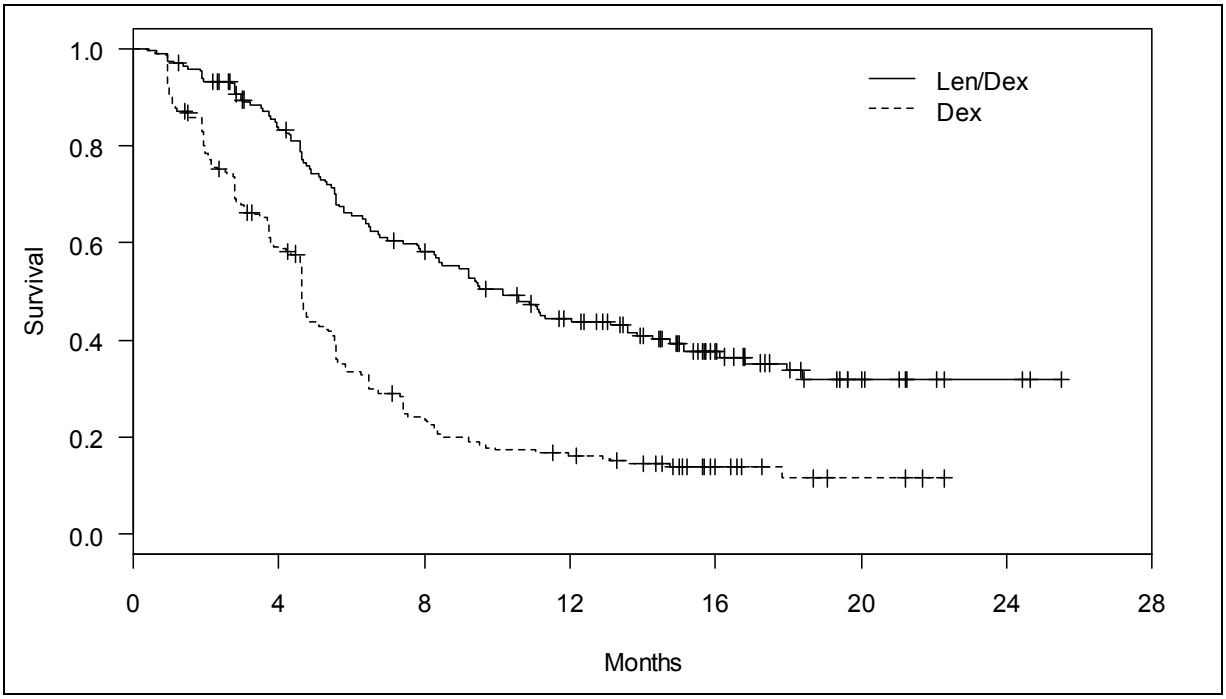
HR = 0.468 (0.332, 0.662)

Figure 15: Time to progression in patient-level data underpinning economic model – 1 prior therapy only



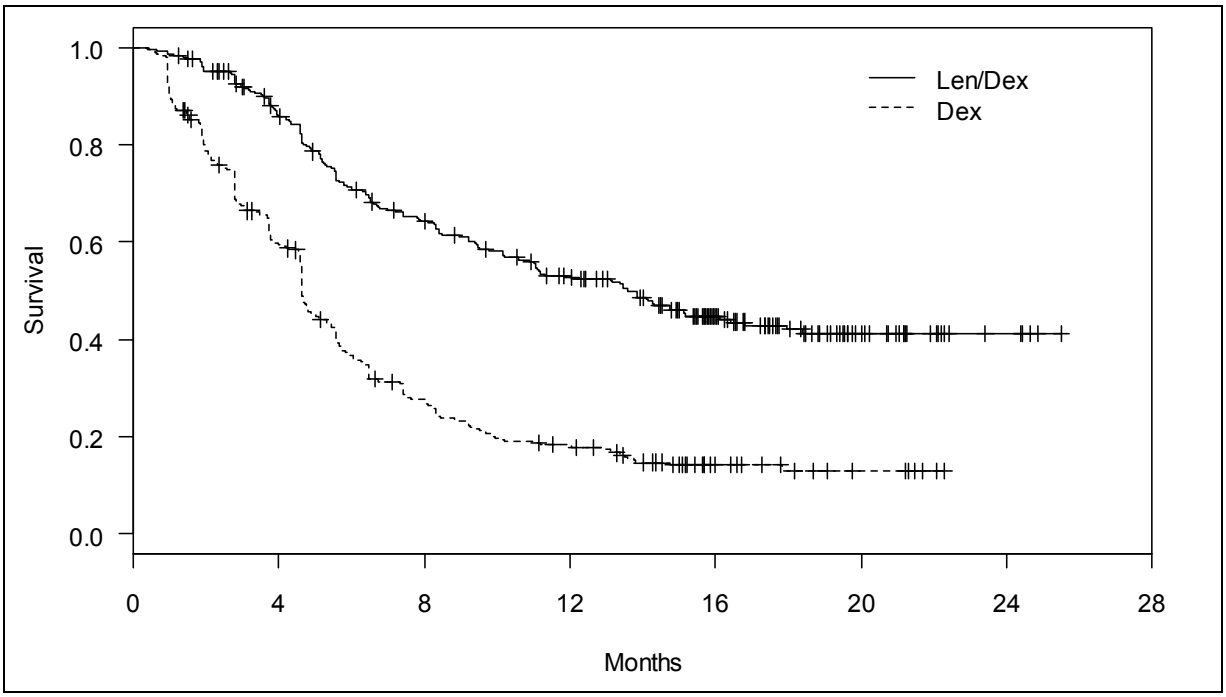
HR = 0.286 (0.205, 0.401)

Figure 16: Time to progression in patient-level data underpinning economic model – ≥2 prior therapies



HR = 0.427 (0.336, 0.544)

Figure 17: Time to progression in patient-level data underpinning economic model – both strata pooled



HR = 0.365 (0.299, 0.443)

Appendix 3: WinBUGS code for ERG's mixed treatment comparison

```
# MODEL
model {
for (i in 1:N) { # indexes arms
  HRPrec[i] <- 1 / (SE[i] * SE[i])
       HR[i] ~  dnorm(mu[i], HRPrec[i]) # Likelihood function
       mu[i] <- (treatmentEffect[t[i]+1] - treatmentEffect[t2[i]+1])
               } # end i loop

# OUTPUTS
for (base in 1:(NT-1)) { # indexes treatments
for (comp in (base+1):NT) { # indexes comparators
   pairHR[base,comp] <- exp(treatmentEffect[base]-treatmentEffect[comp])
                        }} # end base, comp loops

# PRIOR DISTRIBUTIONS
for (k in 1:NT) {
   treatmentEffect[k]~dnorm (0.0,0.001)
                } # end k loop

} # END MODEL

# DATA
list(N=3, NT=3, HR=c(-1.16720508013372, -1.16502405418736, -
0.579818495252942), SE=c(0.233777026273676, 0.265839471144972,
0.186200008791026), t=c(0, 0, 1), t2=c(2, 2, 2))

# INITIAL VALUES
list(treatmentEffect=c(0,0,0))
```