APPENDIX

Response to Lenalidomide ERG Evaluation Report

Herewith are our responses to the summary of uncertainties and key issues (5.4, page 96) raised in the Evaluation Report. We refer to the paragraphs where the issues are fully discussed in the Evaluation Report.

5.3.3.2 Clinical effectiveness

> 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 agree that the tail of the average OS curve resulting from the model is higher than that of a curve produced by averaging the results observed up to a certain point in the trials. Although this does not necessarily mean that the modelling is incorrect, we have implemented the correction to the PPS equation proposed by the ERG to force the model to produce an average Len/dex OS curve that accords better visually with the average of the OS curves displayed in the trial publications.

The reasons we feel that forcing the model to replicate the average observed OS may not be optimal are:

- In the data cut used to produce the KM curves in the trial publications, the median OS had not yet been reached as <50% of patients randomized to Len/dex had died by that point. The curves appear to be getting close to the median because of censoring. As the maximum duration of follow-up is reached for patients who are still alive, they are dropped out of the calculations of the KM curve (NB the ERG report speaks of KM curves generated by the model – this is incorrect as the model does not produce censoring and the curves are not computed using KM techniques). This implies that any subsequent deaths appear to have more impact because they are taken into account with a diminished denominator. As further follow-up accumulates, the KM survival curve will mature, and indeed, in subsequent data cuts the median OS has steadily improved for Len/dex. Thus, calibrating the model to match a published underestimate of the true OS, particularly of its tail end, will lead to inappropriate decreases in the survival gains.
- As can be seen in the trial publications, the number of patients contributing data to the OS in periods beyond 20 months is vanishingly small. Thus, there is a great degree of uncertainty in the estimated tails of the KM curves and so weaker justification for forcing the model to match those points.
- Our modeling technique avoids these problems by weighing the early data more heavily and thus basing the estimated mortality functions on the more robust portions of the observed survival curves. Projections based on these equations are therefore more reliable, and indeed, this is being demonstrated with subsequent data cuts.
- The model is based on the course of the disease according to best response achieved. Thus, the 49 patients whose best response was coded as “non-evaluable” (NE) in the trials were excluded from the model population as explained our original submission (pg 121, 6.2.6.11). The OS Kaplan-Meier curves presented in the trial publications (Weber et al. 2007; Dimopoulos et

al. 2007) included those patients whose best response was NE because they were only displaying OS, unrelated to response or TTP. Inclusion of these 49 patients in the model would be problematic because their response status is unknown (note that NE patients were excluded from TTP calculations even in the trial papers).

- To ensure that identical patients were simulated on each treatment, the model selects individual patients from a population composed of all the evaluable patients from both trials, regardless of treatment. Each patient is then modeled under each treatment option. This variance reduction technique not only reduces the sample size required to achieve stable results, it also removes any residual confounding present in the trial data. Randomization in clinical trials reduces differences across the groups and makes it possible to carry out unbiased comparisons of the average results. The inevitable differences between the groups, however, can become a problem when individuals are simulated over longer periods of time and the full extent of their course is used in computing the consequences of treatment. The pooling removes this problem but means that predictions will differ somewhat from the raw observed data.

5.3.3.2 Clinical effectiveness

Finally, we explain why we have concerns about Celgene's method for adjusting post-progression survival for Dex using data from the MRC trials. … 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 do not agree that it is preferable to calibrate the Dex curve to the mean. That the cost per QALY ratio uses mean values to make the computations does not justify fitting of the mean as the better approach. Fitting has to do with what is most justifiable in terms of reproducing the information as accurately as possible, not with the use of the fits afterwards. Once a reasonable fit is obtained, the resulting function can be used to compute means or whatever other measures are desired.

In the case of the Dex survival function, where we are trying to remove the effect of the cross-over, deciding on the best fit is complicated by several factors:

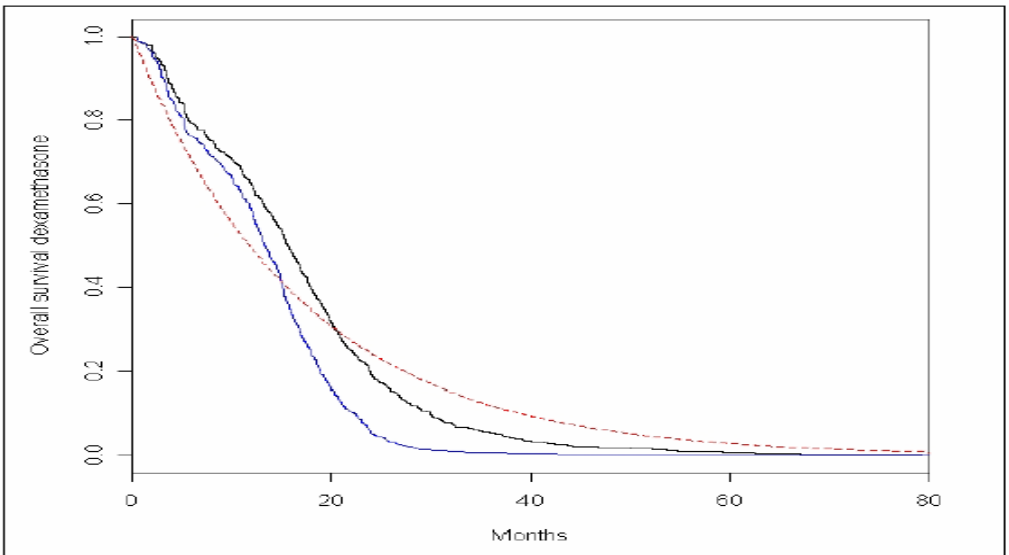
- The MRC population is somewhat different from that of the trial patients receiving Dex. Thus, any fits need to take this into account.
- The MRC data provided OS but the model does not use OS functions as an input. Instead, OS is an output of the combination of TTP and PPS.
- While a simpler exponential distribution fit the available MRC data, the OS results from the model will not be exponential because of their derivation via combining TTP and PPS, with the former fitting a Weibull distribution.

The first point implies that any fitting must take into account the differences between the populations. We did this by recomputing the OS of the patients receiving Dex in the trials using the MRC-derived equations which adjust for patient and disease factors.

The second point means that any calibration and fitting requires that the model be run iteratively to find the correction factor to the PPS equation in the model that produces modeled results that best fit the known information. It is not clear in the ERG report how they generated a different fit or what the resulting calibration factor

was. Based on their Figure 6 (reproduced below) it appears that their calculations are incorrect as the curve representing our submitted model should cross the exponential curve from MRC exactly at the 50% survival point (i.e., the median) and it appears to cross at about 42% instead.

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 third point means that comparing the curves produced by the model with a simple exponential curve from the MRC equation is not very telling. They will necessarily be quite different as is clearly visible in Figure 6 of the ERG report reproduced above. The exponential distribution from the MRC is not likely to be the true shape as it is well known that human mortality accelerates with time, requiring either a Weibull or Gompertz fit (Román et al. 2007; Jucket et al 1993). This was not a concern for our approach as we are not using the MRC-derived shape in the model. The only purpose of the MRC analyses was to provide a calibration point that would allow adjustment of the equations in the model to remove the cross-over effects. By calibrating to the mean produced by the MRC curves, the ERG is taking the exponential shape to be the true function of OS in multiple myeloma. Our analyses of the MM-009/010 trial data revealed a Weibull shape for TTP, however; and, since OS is comprised of TTP and PPS, it suggests that OS in the MM trials would not be exponential. Thus, calibrating to the mean would alter the shape of the survival distribution in the trials, which goes beyond the intended purpose of the calibration.

Apart from this concern, there are other reasons to question that a fit to the mean produces a more accurate reflection of what the actual survival would be. The prediction equations estimated based on the MRC and on the trial data relate mortality (hazards of death) to various patient and disease characteristics. As the hazard itself cannot be directly measured, calibration involves adjusting the intercept until the estimated survival matches a known value. Given that the survival indicates given proportions of the population that are still alive, the calibration involves a centile of the distribution, typically the median. Indeed, the ERG themselves, when adjusting

the Len/dex survival calibrated to the median not the mean (pg 82); and in their discussion of the TTP suggest calibrations to the median from the trials (pg 85). When they try to calibrate to the mean, their curve departs from the target curve significantly more than ours does over the first 20 months or so. Given that the true OS distributions are right-skewed (most of the deaths happen early), calibrating to the mean ignores where most of the known deaths actually occur and over-emphasizes the tail of the distribution where there are fewer patients and much more uncertainty. Thus, the accuracy of predicted survival times in the known earlier parts of the curve would be compromised to gain better fit to the less well known, much more inaccurate, tail. Furthermore, the aim of the calibration was to correct for the impact of cross-over in estimates of PPS. Adjusting to the mean would "delay" the correction, and thus, allow the impact of the cross-overs to remain in the predictions at earlier times.

5.3.3.2 Clinical effectiveness

The modelled TTP is slightly lower than experienced in the MM RCTs. 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.

This critique and suggested adjustments come despite the ERG's statement on page 81 of their report that "**modelled TTP is reasonably close to that experienced in the MM RCTs.**" In any case, we disagree with the critique as it is based on an incorrect averaging of reported medians from the trials: the arithmetic mean of two medians does not equal the median of the pooled trial populations, which is what the model is based on and, therefore, what it ought to be checked against for internal validity. The modelled median TTPs very closely match the observed median TTPs from the pooled trials in our analyses (6.3.1.1 pg 140 in submission):

	Median TTP	
	Predicted by model	Observed in pooled trials
One Prior – Len/Dex	14.1 months	14.3 months
Mult. Prior – Len/Dex	9.5 months	9.5 months

The observed medians do not change when patients with NE as best response are included in the calculation. Thus, the model predicted medians accurately and reflects the observed median for the full pooled population in the data cut used in the model.

5.3.3.2 Clinical effectiveness (pg 87)

the overall survival of patients taking Dex today may be better than calculated from the MRC data. … 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.

The MRC data provide the most accurate means of correcting for the extensive cross-over of patients on Dex in the clinical trials. It is not possible to populate the model with data prior to cross-over with censoring at the time the cross-over occurs because this point occurred very early for most patients. There are too few deaths occurring prior to cross-over and it would be very inaccurate to make projections in the model based on those deaths. As they are the ones that occur very early, they will tend to be the most severe cases and thus the prediction equations would not

only be very uncertain but, worse, inappropriately biased against Dex. Further, the ERG appear to have misunderstood the conclusion of the Mayo Clinic data, which has shown that outcomes have not improved over time for patients receiving conventional therapies such as Dex and that improvements in survival in patients with multiple myeloma have only occurred following the introduction of high dose therapy (with stem cell transplant) since the mid 1990s and the introduction of novel therapies since the turn of the century. We also demonstrated in our original submission that outcomes in the MRC data had not improved over time and, therefore, that the MRC data are indicative of the survival of patients taking Dex today. This point was also made by the independent clinical expert (Dr Jaime Cavenagh) during the appraisal committee meting.

5.3.3.4 Disease management costs (pg 91)

> 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.

Our estimates of the medical management costs before and after progression are based on the expert clinical opinion obtained by interviewing 15 haematologists across England and Wales who specialise in the management of multiple myeloma.

Item	Cost	Source
Outpatient Visit	£107	(76)
Routine Blood Counts (FBC)	£3.23	(77)
Clotting	£3.23	(77)
INR	£3.23	(77)
Biochemistry (U&Es)	£1.75	(78)
Liver function tests (LFTs)	£1.75	(78)
Erythrocyte sedimentation rate (ESR)	£3.23	(77)
Plasma Viscosity	£1.75	(78)
Uric Acid (Urate)	£1.75	(78)
Immunoglobulin (IGs)	£1.75	(78)
Paraprotein Measurements (PP)	£1.75	(78)
Protein Electrophoresis	£1.75	(78)
Serum β2 microglobulin	£1.75	(78)
C-reactive protein	£1.75	(78)
Serum erythropoietin level	£1.75	(78)
Immunofixation (SIF)	£1.75	(78)
Creatinine-clearance (CRCL)	£1.75	(78)
Glomerular filtration rate (GFR)	£1.75	(78)
Serum Free Light Chains (SFLC)	£1.75	(78)
Routine urinalysis	£1.75	(78)
24-hour urine measurement (24hr UR)	£1.75	(78)
24-hour urine for creatinine (24hr UrCr)	£1.75	(78)
Total Urine Protein (24hr TUP)	£1.75	(78)
Urine protein electrophoresis/ light chains	£1.75	(78)
Urine Immunofixation	£20.45	(79)
Skeletal Survey by X-Ray (SS)	£20.45	(79)
Skeletal Survey by X-Ray Individual Sites	£3.23	(77)
MRI	£344.87	(80)
Bone Densitometry (BMD)	£7.00	(81)
Bone Marrow Aspirate (BMA)	£1.75	(77)
Bone Marrow Trephine Biopsy (BMT)	£1.75	(77)
Neuropathy (please specify)	£3.23	(77)
Bacterial investigation	£7.00	(78)
Calcium	£1.75	(77)
Albumin	£1.75	(77)
LDH	£1.75	(77)

Based on that information, and on the cost of a haematology outpatient consultation taken from NHS Reference Costs 2005-06, monthly management cost of multiple myeloma for pre-progression and post-progression states was calculated as £111 and £149, respectively. We agree with the ERG that the costs actually included in the model inadvertently left out the outpatient visits. These have now been included and are part of all the new analyses reported in our response to the ACD.

5.3.3.4 Disease management costs (pg 91)

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 bortezomib submission included a higher estimate of the cost of management but the components of this cost were broader than ours as they included elements that we include separately. Thus, the costs are not truly comparable. The estimate in the bortezomib submission was calculated by taking the mean management costs for

the first 4 years after diagnosis (£16,697) and assuming that these costs were distributed evenly over 48 months (thus, £348 per month), which they then inflated to 2006 costs to obtain an average monthly cost for managing myeloma of £443. This is used for both pre and post progression periods. In the source study (Bruce NJ, McCloskey EV, Kanis JA, Guest JF. Economic impact of using clodronate in the management of patients with multiple myeloma. Br J Haematol 1999;104:358-64), the cost estimate included treatment (multiple chemotherapy sessions) and cost of hospitalizations and other care locations for the treatment of the condition. Therefore, the resulting overall cost accounts for more than routine management only. We consider hospitalizations and other care locations, as well as treatment, separately.

5.3.3.4 Disease management costs (pg 92)

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.

We agree with the ERG's comment and suggestion. All costs have now been inflated to 2008 using the Harmonised Indices of Consumer Prices (HICPs) for Health obtained from the Eurostat website, since UK inflators to 2008 were not found. Available at http://epp.eurostat.ec.europa.eu/portal/page?_pageid=2714,1,2714_61582070&_dad=portal&_schema=PORTAL
These have now been included and are part of all the new analyses reported in our response to the ACD.

5.3.3.5 Costs of adverse events and disease-related complications (pg 92)

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.

Following the question from the reviewers (email communication dated 18 July, 2008 document named "NICE clarification letter – lenalidomide.doc") regarding the consideration of G-CSF in the model for the management of adverse events, the long-term management profile considered in the model was updated in our response in August to the question to include G-CSF and anti-thrombosis as below:

	Medication	% of cases	Dosing	Duration of management
Neutropenia Grade 3	Filgrastim	13%	500mg/day	7 days
	Ciprofloxacin	14%	500mg/bd	7 days
Neutropenia Grade 4	Filgrastim	26%	500mg/day	7 days
	Ciprofloxacin	31%	500mg/bd	7 days
DVT Grade 3	LMW Heparin	94%	10,000iu/week	6 months
DVT Grade 4	LMW Heparin	94%	10,000iu/week	6 months

The cost of G-CSF (£68.41 per pack (Neupogen® (Amgen) filgrastim 30 million-units (300 micrograms)/mL, 1-mL vial British National Formulary Sept 2006 Edition 52 BMJ Publishing Group and RPS Publishing) was included as noted in the costs of managing neutropenia:

	Grade	Cost per visit £		
		Inpatient	**Day Case**	**Outpatient**
Neutropenia	3	£1,980[†]	£518[†¥]	£107[#]
	4	£1,980[†]	£518[†¥]	£107[#]

NHS reference costs 2005 - TOPS FUA - Specialty code: 303 - Clinical Haematology
† NHS reference costs 2005 combined with CHKS data
¥ No day case admissions were identified for neutropenia. Therefore, the average of the identified HRG costs was used.

There is no reason to expect that management of neutropenia in the NHS would change for patients receiving Len/dex.

5.3.3.6 Health-Related Quality of Life (pg 93)

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.

As patients with partial response or stable disease have ongoing disease manifestations, it would be expected that better response improves quality of life. The ERG provides no basis for the contrary opinion.

5.3.3.6 Health-Related Quality of Life (pg 93)

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. … 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.

Pre-progression utility value (0.81) for multiple myeloma patients used in our submission was considered too high, given the age of the trial population (4.12 and 3.12). This comment is surprising for two reasons. First, it is the value suggested by the ERG in the NICE appraisal of bortezomib (6.3.4.3 page 36; Green et al. Bortezomib in treatment of multiple myeloma). The current ERG comments indicate that 'a health state value between 0.644 and 0.789 may be appropriate for patient groups with MM. However, Kind et al (1998) have reported health state values in the UK general population by age group, valued using the EQ-5D, with those aged between 60-69 years ranging between 0.829-0.806.' Thus, a health state value near 0.80 is likely appropriate for the population in pre-progression. The current reviewers sited 3 additional references (page 94 in Evaluation Report) which we discuss below. Also the 0.81 value indicates these patients would accept a 19% chance of death to change from the asymptomatic pre-progression state to normal health – a hefty penalty. Second, the ERG report and the ACD comment implies that a lower utility is more appropriate for patients in the pre-progression state. This is tantamount to saying that keeping them alive for each additional year is less worthwhile than keeping a younger patient population alive. We do not believe that the appraisal committee wished to imply this age specific inequality message in the ACD. Despite our concerns about this utility value, we included sensitivity analyses around the utility values (± 10%) in our original submission and provide these again with the new

base case analysis. We did not adopt the values in the publications suggested by the reviewers (section 5.3.3.6 in the Evaluation Report).

5.3.3.6 Health-Related Quality of Life (pg 94)

> 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), we are aware of two other studies that quote utilities for patients with multiple myeloma.

We had reviewed these two studies in preparation for the submission and did not feel they could be used because they do not report utilities by response level which is required for the model. Nevertheless, we have re-examined them and conclude that they generally support the values we have used.

Gulbrandsen and colleagues (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) used a clinical trial of high-dose melphalan with autologous blood stem cell support (HDM) and a historical trial for MP to collect HRQoL data. The EORTC QLQ-C30 (version 1) was used and was mapped to 15-D. Utility scores at 6 months were 0.7334 for HDM and 0.7896 for MP. At 36 months (presumably when surviving patients are those in remission) the score was 0.81.

Nord and colleagues (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;12:89-103) tried to estimate utilities from a trial comparing MP with and without interferon-alpha2b in newly diagnosed, symptomatic MM patients (age 55-80 years). The utility estimates were obtained via a very subjective and highly uncertain two-step procedure because the trial did not obtain utility data and the authors had to post-hoc map the available instrument to various utility-indexable scores. As the authors stated, " Each of this steps introduced potential sources of error, rendering the reliability of the resulting utilities questionable." Even then, the resulting values reflect a mixture of disease and adverse treatment effects and are not very applicable to our model.

In any case, we have carried out sensitivity analyses with a 10% decrease in the utility estimates and these are reported here.

5.3.3.6 Health-Related Quality of Life (pg 94)

> 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.

We agree that the potential disutility associated with adverse events was not considered in the model. Despite extensive literature searches no suitable estimates were found for the impact of these events on patients suffering from multiple myeloma. The effect on quality of life of a given adverse event depends very much on the underlying condition, the person's understanding of the benefit of treatment and the patient's expectations. Indeed, in a study in multiple myeloma (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;12:89-103) it was found that the difference in utility

between patients in the group with severe adverse events and the one without them was so small that it was not detectable by conventional utility instruments. Thus, utilities obtained in other disease areas are not very transferable.

Nevertheless, in order to explore the potential impact of the adverse events on quality of life, in our response to the ACD we have performed an additional sensitivity analysis in which a disutility has been assigned to each event that may require long term management. These disutilities were expressed as a proportional reduction of the underlying utility score for a given patient to allow application across response levels. The duration of the negative impact of the events is assumed to be the same as the duration of their long term management and shown in the sensitivity analyses.

Adverse Events	Baseline Utility	Disutility of Event	Proportional Reduction	Duration¥ (days)	Reference
Anemia	0.86	0.545	36.6%	180	Ossa et al, 2007
Thrombocytopenia*			36.6%	7	
Neutropenia	0.715	0.565	21.0%	7	Lloyd A et al, 2006
Pneumonia	1.0	0.81	19.0%	7	Cykert et al, 1999
Peripheral neuropathy	0.689	0.624	9.4%	180	Coffey (2002)
Thrombosis or embolism	0.84	0.69	17.9%	180	Mathias et al 1999

* Assumed similar to anemia;
¥ Assumed same as the duration of long term management

References: Coffey JT et al. Valuing health related quality of life in diabetes. Diabetes Care 2002; 25:2238-43; Cykert S et al. Racial differences in patient's perceptions of debilitated health states. J Gen Intern Med 1999; 14:217-22; Lloyd A et al. Health state utilities for metastatic breast cancer. BJ Cancer 2006; 95: 683-90; Mathias SD et al. A health related quality of life measure in patients with deep vein thrombosis: a validation. Drug Info J 1999; 33:1173-87.

6.1.1 Meta-analysis (pg 99 and 32)

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.

A meta-analysis of the two trials was carried out to fulfil requirements for the submission. It has no bearing on the economic analysis as it is not used as input; and in any case, the submission had access to the pooled individual patient data making "conventional meta-analysis" irrelevant.

Use of HRs as the basis for comparison in our analyses is not possible. Our equations for TTP and PPS controlled for best response, which is dependent on treatment. Thus, part of the treatment benefit in our equations is captured through the best response parameters; any benefit beyond best response is reflected in the treatment indicator parameters. This was the case in the TTP equation, but no additional benefit could be detected in the PPS equation.

Thus, the HR for treatment in the in the TTP equation would be different than the HR measured without control for response. As a result, comparisons to other treatments can not be made based on HRs.

6.1.2 Mixed treatment comparison (pg 99 and 33)

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.

The MTC was based on median TTP rather than HR because a meta-analysis of hazard ratios would not be compatible with the way mixed treatment comparisons are made in the economic model. The TTP equation derived from the MM-009/010 trials was used as the basis for alternate comparisons. A calibration term was added to the equation to reflect the TTP distribution for alternate comparator, adjusting for other predictors in our TTP equation. A key predictor in this equation is best response, which is determined by treatment (i.e., it is an intermediate factor between treatment and TTP). Therefore, the hazard ratio for Len/Dex vs. Dex indicator in this equation reflects the benefit of Len beyond best response (rather than the marginal or *total* hazard ratio). Hazard ratios for alternate comparators would have to be similarly adjusted for best response to be valid for use in deriving the calibration term. These are not reported in publications for alternate comparators, however, and applying the published hazard ratios would over-compensate the benefit of the drugs since we adjust for the best response profile of the comparator drugs. Thus, the calibration term was derived based on the median reported TTPs for alternate comparators. Standard errors of the medians were not available in publications, however; thus, some assumptions were made about these to implement the meta-analysis.

References

Bruce NJ, McCloskey EV, Kanis JA, Guest JF. Economic impact of using clodronate in the management of patients with multiple myeloma. Br J Haematol 1999;104:358-64

Coffey JT, Brandle M, Zhou H, et al. Valuing health related quality of life in diabetes. Diabetes Care 2002; 25:2238-43.

Cykert S, Joines JD, Kissling G, Hansen CJ. Racial differences in patient's perceptions of debilitated health states. J Gen Intern Med 1999; 14:217-22;

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

Juckett DA, Rosenberg B. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mech Age Dev* 1993; 69:1-31.

Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. BJ Cancer 2006; 95: 683-90;

Mathias SD, Prebil LA, Putterman CG, et al. A health related quality of life measure in patients with deep vein thrombosis: a validation. Drug Info J 1999; 33:1173-87.

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;12:89-103.

Román R, Comas M, Hoffmeister L, Castellas X. Determining the lifetime density function using a continuous approach. Journal of Epidemiology & Community Health 2007; 61(10):923-925.