Evidence Review Group Report commissioned by the NHS R&D HTA Programme on behalf of NICE

Bortezomib for the treatment of multiple myeloma patients

Produced by Southampton Health Technology Assessments Centre

Authors
C Green, Principal Research Fellow

J Bryant, Senior Research Fellow

A Takeda, Research Fellow

K Cooper, Research Fellow

A Clegg, Director of SHTAC

A Smith, Consultant Haematologist, Southampton General Hospital

M Stephens, Chief Pharmacist, Southampton General Hospital

Correspondence to
J Bryant

Southampton Health Technology Assessments Centre

Wessex Institute for Health Research and Development

University of Southampton

Mailpoint 728, Boldrewood

Southampton SO16 7PX

Date completed
April 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APEX	Assessment of Proteasome Inhibition for Extending Remissions
ASCO	American Society of Clinical Oncology
CEA	Cost effectiveness analysis
CrCl	creatinine clearance
CRD	Centre for reviews and dissemination
CREST	Trial name
CUA	Cost utility analysis
EORTC QLC-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire.
EQ-5D	Euro-Qol (EQ-5D) is a standardised instrument for use as a measure of health outcome.
ERG	Evidence Review Group
FACT/GOG-Ntx	Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity
FDA	Food and Drug Administration
HCHS	Hospital and community health services
HDD	High dose dexamethasone
HR	Hazard ratio
HRQL	Health related quality of life
ICER	Incremental cost effectiveness ratio
ITT	Intention to treat (analysis)
KPS	Karnofsky performance scale
LYG	Life years gained
MM	Multiple myeloma
MRC	Medical Research Council
NE	Not evaluable
NEJM	New England Journal of Medicine
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NRR	National Research Register
OS	Overall Survival
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality adjusted life year
QUOROM	Quality of Reporting of Meta-analyses.
RCT	Randomised controlled trial
RR	Relative risk
SG	Standard gamble
SUMMIT	Trial name
TTO	Time trade off
TTP	Time to (disease) progression
UK	United Kingdom
USA	United States of America
VAD	vincristine, adriamycin and dexamethasone
VAT	Value added tax

1 SUMMARY

1.1 Scope of the submission

- The use of Bortezomib monotherapy for the treatment of multiple myeloma patients at first relapse and beyond; this reflects the licensed indication of Bortezomib.

1.2 Summary of submitted clinical effectiveness evidence

- One RCT, comparing Bortezomib with high dose dexamethasone in multiple myeloma patients who have had a relapse after one to three treatments, was included in the manufacturer's submission.

- Results of the RCT suggest that Bortezomib increases survival and time to disease progression compared with high dose dexamethasone in these patients.

1.3 Summary of submitted cost effectiveness evidence

- The cost-effectiveness analysis uses a decision-analytic model (quasi-Markov) to estimate the treatment effect with Bortezomib compared with high dose dexamethasone.

- The model uses clinical effectiveness data from the RCT supplemented with data from an observational study.

- Primary analysis presents an estimated cost per-life-year-gained of £30,750. Cost per life-year-gained ranges from £27,957 to £36,747 from sensitivity analyses.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

- Searches for clinical and cost effectiveness studies were adequate and all available evidence has been included.

- The RCT is of reasonable quality when assessed according to NICE internal validity criteria.

- The approach taken to model disease progression and cost effectiveness in this patient group seems reasonable.

1.4.2 Weaknesses

- The included RCT is not an absolute reflection of current UK clinical practice so its external validity may be questionable. However, the lack of standardisation in the clinical management of relapsed myeloma suggests that the impact of this on the generalisability of the economic model in terms of patient group and comparator may be minimal.

- The reporting of the trial lacks detail and clarity making interpretation of clinical effectiveness results difficult.

- The manufacturer's submission does not present quality of life issues in the economic model. *Note: The manufacturer has submitted additional information presenting analysis on cost per QALY, see Addendum.*

- The sensitivity analyses undertaken in the economic evaluation are limited.

- Adverse events have not been included in the manufacturer's model, either in terms of loss of quality of life or increased resource use.

1.4.3 Areas of uncertainty

- The RCT is poorly reported, making it difficult to assess the overall treatment effect of Bortezomib on myeloma.

- The economic model in the manufacturer's submission may overestimate the treatment effect from the trial for a UK setting.

- Cost effectiveness results presented may under estimate the cost per life-year-gained for bortezomib compared to HDD.

- There may be greater variability in the cost effectiveness of treatment than presented in the sensitivity analyses in the manufacturer's submission.

1.5   Key issues

- There is no standard treatment for relapsed multiple myeloma patients which makes assessing the effectiveness and cost effectiveness of new treatments problematic in terms of the individuality of treatment protocols and which comparators to use.

- It would be useful for future trials to reflect current practice but this may be difficult as this is a quickly developing area where clinicians are eager to have new treatments options for patients who do not do not easily fit into stereotypical groups.

2 Introduction to ERG Report

This report is a critique of the manufacturer's original submission to NICE received on 1 March 2006. Clarification on some aspects of the submission was requested from the manufacturer by the ERG via NICE on 29 March 2006. A response from the manufacturer via NICE was received by the ERG on 19 April 2006 and this has been included as an Addendum in the ERG report. Annotations referring to the Addendum occur throughout the ERG report where applicable.

3 BACKGROUND

3.1 Critique of manufacturer's description of underlying health problem

The early sections of the manufacturer's submission report do not include much detail on the underlying health problem. The first part of the background chapter focuses on the summary of the decision problem, and a critique of this is presented in section 4. Section 21 of the manufacturer's submission (p.14) provides a brief overview of the disease, including incidence data and a summary of the disease's natural history. This appears to be a clear and accurate overview of the disease.

3.2 Critique of manufacturer's overview of current service provision

The manufacturer's submission correctly states that there is no standard approach to treatment at first relapse and that treatment is likely to vary between individuals according to clinical condition, age, prior treatment regimen and critically, the timing of relapse in relation to previous treatment. A detailed description of possible treatment options at first and subsequent relapse is not provided in the manufacturer's submission.

UK market authorisation is for bortezomib monotherapy only. However, clinical advisors and published guidelines suggest that bortezomib is generally used in conjunction with dexamethasone in the UK, which is acknowledged in the manufacturer's submission; this will be for patients with reasonable life expectancy. The manufacturer's submission mentions that the use of bortezomib varies across the UK due to funding constraints as shown in Appendix 2 of the manufacturer's submission.

The manufacturer's submission justifies the manufacturer's choice of comparators, but it is not clear whether this is representative of UK clinical practice. This is discussed further in Section 4.3 below.

4 Critique of manufacturer's definition of decision problem

4.1 Population

The study population is defined as multiple myeloma (MM) patients at first relapse and beyond. This does not completely match the description of the licence population of 'patients who have received at least one prior therapy (at first relapse) and who have already undergone or are unsuitable for bone marrow transplantation'. The population's characteristics are described briefly in section 2 (p. 6) of the manufacturer's submission. This includes median age at diagnosis, one- and five-year survival rates and very limited epidemiology regarding sex and ethnic risk factors for this disease (with no supporting data).

Median age at diagnosis in the manufacturer's submission differs from that found in the literature and reported by clinicians. The manufacturer's submission states that the median age at presentation is 65 (p. 6, section 2). Recent clinical publications suggest that the average age of UK patients at diagnosis is in the region of 70-72 years. Guidelines on the diagnosis and management of multiple myeloma[1] indicate that median age at presentation is 70 years, and elsewhere[2] a median age at diagnosis of 68 is quoted. The Cancer Research UK website[a] cited in the manufacturer's submission states that 'more than 80% of cases are diagnosed in people over 60 years old'. However, the differences in median ages quoted here are probably not clinically significant.

4.2 Intervention

Bortezomib is a proteasome inhibitor, and works by disrupting normal intra-cellular protein regulation, leading to programmed cell death (apoptosis).

The executive summary states that bortezomib was licensed for the treatment of people with relapsed and refractory multiple myeloma in 2004. The marketing authorisation was extended in April 2005 to allow use as a monotherapy for the treatment of progressive multiple myeloma in patients who had received at least one prior therapy (at first relapse) and who have already undergone (or are unsuitable for) bone marrow transplantation.

The description of the intervention in the definition of the decision problem in the manufacturer's submission does not reflect the use of bortezomib in the UK, as mentioned

---

[a] http://info.cancerresearchuk.org/cancerstats/types/multiplemyeloma/

previously. Therefore there is some concern as to its usefulness in answering a question of relevance to the NHS rather than a question relevant to licensing.

4.3 Comparators

The manufacturer's submission states that high dose dexamethasone (HDD) is the only licensed monotherapy agent with proven efficacy in first relapse patients (p.6), and this is therefore used as the comparator in the manufacturer's submission. HDD, although an accepted, active anti-myeloma regimen used in the management of MM in the UK, it is usually used for second or subsequent relapse or at first relapse where there are concerns about marrow toxicity. Therefore, it appears that the choice of comparator reflects the only phase III RCT of bortezomib (the APEX study) rather than possible relevant comparators in an NHS context. However, the lack of standardised treatment for patients at first relapse means that there is no other obvious comparator, although there are various treatment options for this patient group.

Clinicians and published guidelines suggest that some patients, younger than 65 years, who have had a response of >18 to 24 months after a first stem cell transplant at first relapse, may be treated with a further regimen leading to a second stem cell transplant. Alkylating agents, such as melphalan, with or without corticosteroids, would be regarded as standard, non-transplant-directed therapy at first relapse if the period of first response was >12 to 18 months. Thalidomide, with the addition of dexamethasone if necessary, would also be used, especially in early relapse or refractory patients. As thalidomide is not licensed for use within the UK and the optimum dose and duration of therapy is not clear[1], it was not considered to be a suitable comparator by the manufacturer.

4.4 Outcomes

The outcomes stated in the manufacturer's definition of the decision problem are time to disease progression, response rates, survival and quality of life. They are appropriate and clinically meaningful. No primary outcome is specified. The ERG would have expected survival to be the primary outcome measure, although time to disease progression would also be valid.

5 CLINICAL EFFECTIVENESS

5.1 Critique of manufacturer's approach

5.1.1 Description of manufacturer's search strategy and comment on whether the search strategy was appropriate.

5.1.1.1 Clinical effectiveness searches

The manufacturer ran searches in the expected core databases of Medline, Medline in process, Embase and Cochrane, and checked the websites of ASCO and the American Hematology Society for ongoing trials and conference proceedings. The ERG searches would have included the ISI Science Citation Index, and conference proceedings databases such as ISI Proceedings and Biosis Preview in addition to these. We would also have searched the National Research Register, Current Controlled Trials and www.clinicaltrials.gov for ongoing trials.

The search question is broken down into population and intervention, as would be expected. The population is searched using Medline and Embase descriptors (MeSH and Emtree) and free text terms. The search strategy for both of these facets seems sound and reproducible. The manufacturer did not give detailed search results (by search line) for all the searches, although the results are broken down by database total. *The ERG requested further information on the numbers of studies identified by each line of the search strategy, and these were later provided by the manufacturer (see Addendum).*

ERG searches using the NHS-CRD RCT filter identified an additional 109 references in Medline, and an additional 143 references in Embase. We also found more results in some of the databases when running searches with the manufacturer's RCT filters. Differences could have arisen from the difficulty of translating search terms between databases and the use of long nested search strings in the free text searches. The extra references have been screened for inclusion/exclusion and do not appear to be relevant references that are missing from the manufacturer's searches.

5.1.1.2 Cost effectiveness searches

The search strategy for cost effectiveness (economic evaluations) is reproduced in Appendix 8 (p.127) for all selected databases. The search strategy is only reported in the context of a

review of economic evaluations of bortezomib for MM. The databases searched are listed and seem to be comprehensive. The search was run on the drug intervention alone, which is acceptable for identifying specific economic evaluations (but is regarded as too narrow to search for other references related to quality of life studies).

The manufacturer's submission lists additional strategies for economic studies of multiple myeloma and survival for people with multiple myeloma in Appendix 10. There is no commentary in the manufacturer's submission on the search for economic costing studies. Section 3.4.1 of the manufacturer's submission refers to the epidemiology search. It is not clear how these additional searches were used. The ERG ran two searches for cost effectiveness and quality of life in Medline using our standard filters and found approximately double the number of papers reported in the manufacturer's submission. Again, it is not clear whether these additional references would have been relevant to the review.

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

The manufacturer's submission reports the results of a systematic review using the following inclusion criteria:

1. study design - RCTs, systematic reviews or meta-analyses;
2. interventions - bortezomib monotherapy vs. either placebo or another comparator;
3. population - patients with MM at first relapse, over 12 years old;
4. outcome measures - no outcome measures are specified in the inclusion/exclusion criteria

The inclusion/exclusion criteria seem appropriate and generally reflect the information in the decision problem. However, there is some discrepancy between the listed inclusion criteria (p. 24) and the table of inclusion criteria (p.25). For example, section 34 'inclusion criteria' states 'clinical studies' and 'exclusion criteria' states 'non-systematic reviews or editorials', but the text does not discuss whether or not systematic reviews would be included. However, the table on p. 25 suggests that systematic reviews and meta-analyses were to have been included. This point is relevant because an abstract of a systematic review[3] was listed as a search result, but was not discussed at all in the manufacturer's submission.

There is no discussion of whether or not abstracts would be considered for the review, or whether this would be restricted to fully published papers. Again, this is relevant because an

abstract for a relevant systematic review[3] is listed, but no reason given for its exclusion. It does not appear in the list of excluded studies in Appendix 10.

Inclusion criteria state that MM patients at first relapse will be included in the review, and that trials involving only patients who were at second relapse and beyond will be excluded. The manufacturer's submission does not specify how trials with mixed populations will be handled (for example, whether a trial will be included if at least 50% of patients meet the inclusion criteria). It is possible therefore that a study which does not strictly reflect the stated inclusion criteria could be included in the systematic review. It should also be noted that the population defined in the decision problem is MM patients at first relapse and beyond.

Differences exist between the systematic review and trial inclusion/exclusion criteria and licence indication of bortezomib. Neither the systematic review nor the trial's inclusion/exclusion criteria mention bone marrow transplantation status, yet the licence suggests that bortezomib is only indicated for people who have already undergone or are unsuitable for bone marrow transplantation. This may not be an important point as the use of bortezomib in the UK would be considered in patients with relapsed myeloma on clinical merits irrespective of whether stem cell transplantation had been carried out. The table of inclusion criteria states that only studies with more than 20 patients will be included, but gives no justification for this. *The ERG requested further information from the manufacturer on this point, see Addendum for clarification.*

No QUOROM flow chart of retrieved studies is presented, and there is some confusion around the number of identified studies. Conflicting information was provided on the number of references assessed for inclusion. Text in section 32 on p.24 states a total number of records for evaluation of 283, indicating that there were 245 from searches, plus extras from the Cochrane database and in-house databases. Text in section 35 then states that 'the abstracts of the 257 articles identified were checked for inclusion/exclusion criteria". All excluded references appear to have been rejected on the basis of their titles and abstracts, and no full publications appear to have been retrieved other than the single included RCT. The manufacturer's submission does not state the methods used for screening titles and abstracts for inclusion.

5.1.3 Identified studies

Only one RCT was included in the manufacturer's submission (details are shown in Table 1). This has been published as one primary publication[4] (referred to as 'the NEJM paper' in this report) and a series of conferences presentations and abstracts[5-10]. The manufacturer did not include copies of the relevant papers with the submission, and it has not been possible to find copies of all the cited abstracts.

Table 1 Identified studies

Study: Richardson et al. [4]

Methods	Participants	Outcomes
Design: RCT *Interventions*: Group A: bortezomib (1.3mg/ m^2 BSA) on days 1,4, 8 and 11 for 8 21-day cycles, and on days 1, 8, 15 and 22 for 3 further 35-day cycles, for a maximum treatment period of 273 days. Group B: HDD 40mg on days 1-4, 9-12 and 17-20 for 4 35-day cycles, followed by treatment on days 1-4 for 5 28-day cycles, for a maximum treatment duration of 280 days. Group B patients were permitted to cross over to receive bortezomib in a companion study after disease progression *Number of centres*: 93 *Median duration of treatment*: not stated in NEJM paper. 56% of patients completed 5 3-week cycles of bortezomib, and 56% completed 3 5-week cycles of HDD. 29% of bortezomib group completed 8 twice weekly cycles, and 36% of the HDD group completed 4 cycles. 9% of bortezomib group and 5% of HDD group completed all planned therapy.	*Inclusion criteria*: measurable progressive disease after 1-3 previous treatments; KPS score of ≥ 60; platelet count of ≥ 50,000 per cubic mm; haemoglobin level of ≥ 7.5g per dl; absolute Neutrophil count of ≥750 per cubic ml; creatinine clearance of at least 20ml per min. *Numbers*: 669 participants. Group A: 333 Group B: 336 *Median age (10th and 90th percentiles):* Group A: 62 (48,74) Group B : 61(47,73)	*Primary outcomes*: Time to disease progression *Secondary outcomes*: overall and one-year survival; response rate (complete plus partial response); duration or response; time to 1st evidence of a confirmed response; time to 1st infection of grade 3 or higher, and time to first skeletal event. *Length of follow-up:* 8.3 months (fully published). Updated analysis at 22 months only available as abstract, and contains cross-overs.

BSA= body surface area, KPS = Karnofsky performance scale,

Approximately 67% of patients in the APEX study had received 'prior stem cell transplant/high-dose therapy', but it is not clear from the reporting of the trial how many had either received a transplant or were unsuitable for one. Over 50% of included patients had received at least 2 prior treatments, and almost 50% of the patients had previous thalidomide therapy. ERG requested further information on the timing of previous treatments. *Further information from the manufacturer suggests that two thirds of patients had received*

thalidomide therapy in the three months prior to starting bortezomib treatment (see Addendum).

The manufacturer's searches identified an abstract of a relevant systematic review[3], but this was neither discussed in the text nor specifically excluded from the review.

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

The ERG did not identify any relevant studies that were not included in the submission from searches undertaken and consultation with experts.

5.1.4.1 Ongoing studies

The combination of bortezomib and dexamethasone will be included in a protocol extension to the Myeloma IX study following a patient's first relapse. The study population excludes patients who have previously received chemotherapy. This study is planned to close in May 2008, and has currently recruited 77% of total required participants.

The manufacturer's submission includes a list of ongoing clinical trials (appendix 3). This includes:

1. a large, multicentre phase III trial (protocol number 26866138MMY3005) comparing the efficacy and tolerance of thalidomide + dexamethasone versus thalidomide + dexamethasone + bortezomib for MM patients regressing or relapsing after autologous transplantation (n=452 patients);
2. a small phase III trial (protocol number DOXILMMY3001) comparing doxil and bortezomib with doxil monotherapy for 20 patients with relapsed multiple myeloma;
3. a phase III trial protocol (number 26866138MMY001) with a relevant treatment and comparator (bortezomib-melphalan-prednisone vs. melphalan-prednisone). However, the patient group is newly diagnosed MM patients not suitable for transplant, rather than patients at first relapse, so this may not be relevant.

ERG searches on NRR and the FDA website identified a number of records relating to trials in appendix 3 and to other, recently completed trials. We did not identify any other relevant ongoing trials of bortezomib for patients with multiple myeloma.

5.1.4.2 Additional studies

The manufacturer's submission also describes two phase II studies (SUMMIT and CREST) designed to establish the efficacy of bortezomib and evaluate the safety profile. These trials are not included in the systematic review results.

5.1.5 Description and critique of manufacturer's approach to validity assessment

The manufacturer's submission authors applied the quality assessment criteria developed by NICE, but do not state how many reviewers applied these criteria. The ERG applied the same criteria to the included RCT, and agreed with the manufacturer's submission authors' assessment for all but one criterion (see Table 2). Method of randomisation and concealment of allocation are not described fully in the NEJM paper[4], and the ERG would therefore score it as 'A' on the NICE criteria. The manufacturer's submission authors rate randomisation as 'C'. They cite an interactive voice recognition system that was used for randomisation, yet the NEJM paper does not mention this. The manufacturer's submission authors' assessment must therefore be based on more information than is available in the published trial.

The trial was not blinded, but most of the outcome measures were measurable against standard criteria so would not have been subject to observer bias. However, some outcomes (such as quality of life assessment and reporting of adverse events) could have been affected by the lack of blinding. *The ERG asked the manufacturer to confirm whether or not anyone was blinded in the study, and further information was received as point A8 in the Addendum.*

Allocation to treatment groups was randomised and stratified by number of lines of prior therapy; refractoriness to prior therapy; and baseline serum β_2-microglobulin concentration. Stratification factors are described in the manufacturer's submission as being balanced between treatment arms, but the number of subjects enrolled in each stratum was not controlled. No further data are presented, and it is not possible to check whether adequate numbers were represented in each group.

A consort flowchart is presented. However, it does not include enough detail on the patients that crossed over from HDD to bortezomib to assess how this may have influenced results.

Table 2 ERG quality assessment of RCTs

Quality criteria	Description	Score
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	A – no details in NEJM paper[4]
Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	B
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	B (TTP determined by computer algorithm)

Other questions:	Response
Was the design parallel-group or cross-over? Indicate for each cross-over trial whether a carry-over effect is likely.	The trial was parallel group, with HDD patients who experienced progressive disease being permitted to receive bortezomib at any point during the process in a companion study
Was the trial conducted in the UK (or were one or more centres of the multinational trial located in the UK)? If not, where was the trial conducted and is clinical practice likely to differ from UK practice?	Trial was conducted at 93 centres in US, Canada, Europe and Israel. NEJM paper gives no information on UK centres. Manufacturer's submission states that 7 UK centres were involved, with a total of 51/669 patients.
How do the subjects included in the trial compare with patients who are likely to receive the drug in the UK? Consider factors known to affect outcomes in the main indication such as demographics, epidemiology, disease severity, setting.	Trial participants may have been younger than those in the UK, but otherwise appear comparable.
For pharmaceuticals, what dosage regimens were used in the trial? Are they within those detailed in the Summary of Product Characteristics?	bortezomib (1.3mg/ m^2 BSA) on days 1,4, 8 and 11 for 8 21-day cycles, and on days 1, 8, 15 and 22 for 3 further 35-day cycles, for a maximum treatment period of 273 days. 40mg HDD on days 1-4, 9-12 and 17-20 for 4 35-day cycles, followed by treatment on days 1-4 for 5 28-day cycles, for a max treatment duration of 280 days. Dosage regimens are as specified in SPC.
What was the median (and range) duration of follow-up in the trial?	Not stated in NEJM paper. Manufacturer's submission states median 21.9 months (ref 9-abstract), range 0-35.6 months

5.1.6 Description and critique of manufacturer's outcome selection

The manufacturer's submission did not specify any outcome measures that would be used as part of the inclusion criteria for the selection of studies. ERG would have expected outcomes such as survival and time to disease progression to be inclusion criteria.

5.1.7   Description and critique the statistical approach used

The ITT population was defined as all patients who were randomised to treatment (n=669), and the manufacturer's submission states that patients in this population were analysed according to the treatment to which they were randomised. Safety analyses were based on patients who received at least one dose of the study drug, but according to the treatment they actually received.

Interim results at 8.3 months are presented as ITT for all 669 patients. The manufacturer's submission authors describe the interim analysis as being based on the methods of O'Brien and Fleming (manufacturer's submission ref 33). It was planned to take place when half of the total required number of events had occurred (231 patients with progressive disease), and was anticipated to take place 18.5 months after the first patient was enrolled. Interim analysis took place at only 8.3 months, after a statistically significant difference in time to disease progression was identified.

There is some confusion surrounding timing of cross-overs in the trial. The NEJM paper[4] states that patients were allowed to cross-over *after* interim analysis (median time 8.3 months), and this is also suggested by the flow chart on page 35. However, the manufacturer's submission (p.46) states that "The design of APEX allowed for any HDD patient experiencing disease progression to be offered bortezomib therapy. At interim analysis (median follow-up 8.3 months), 44% of HDD patients had crossed over to receive bortezomib." *The ERG requested further information from the manufacturer regarding the timing of cross-overs, and the number of patients receiving which treatment during the earlier stages of the trial. The manufacturer's response suggests that cross-overs took place before the interim analysis – see Addendum point A1.*

The manufacturer's submission states that: "A total of 60 (9%) of the 669 patients randomised into the study were found to be refractory to dexamethasone, including 32 (10%) of 333 bortezomib patients and 28 (8%) of 336 HHD patients." (p.56). There doesn't seem

to be any mention of how this would have affected power calculations. The NEJM paper[4] states that post-hoc analysis without these patients still found significant benefits for bortezomib. However, the text on the power calculation states that: "a sample size of 310 patients per treatment group provided 80% power to detect a 30% difference in the time to disease progression between the two groups." Therefore, removing the refractory patients would bring the group size to a level (slightly) below that required for statistical power. *The ERG requested further information from the manufacturer regarding the effect on power calculations of removing these patients from time to progression (TTP) analyses. The manufacturer confirmed that statistical power was maintained, as the time to disease progression was so much higher than predicted (>70% compared with the required 30% difference). See Addendum for further details.*

Quality of life data were assessed using a modified intention to treat analysis. T*he ERG requested further information on the statistical methods used for this. The manufacturer confirmed that the modified ITT was only used for this outcome measure, and that 44 patients were excluded from the analysis (see Addendum). .*

Hazard ratios were presented for key outcome measures, but it was not possible to check these as insufficient data were reported. For example, in Table 10 (p.49), HDD results are reported as 'non evaluable', yet a hazard ratio and p value for treatment difference are reported.

Due to the early termination of the APEX trial a high level of censoring was applied to TPP and survival data, and it is not clear what impact this as well as the high rate of attrition will have had on outcomes. Also, it is not known how the high rate of crossovers from HDD to Bortezomib affected results; it may be that the effect of crossovers was to reduce the difference in effectiveness between the two groups but lack of clarity regarding statistical analysis hinders this conclusion.

5.1.8 Summary statement of manufacturer's approach

- The manufacturer's submission appears complete with regard to relevant studies, with one RCT meeting the inclusion criteria.
- Quality assessment of the included study according to the NICE criteria seems to be adequate although details of the process, in terms of whether it was performed by two independent reviewers, are missing.

- The submitted evidence generally reflects the decision problem defined in the submission, although there appear to be some small inconsistencies which probably relate more to reporting than actual differences.

- There is uncertainty about the statistical methods used in the analysis because of lack of clarity in terms of patient numbers and the intervention they received at any time (crossovers, drop-outs, length of follow-up) and because point estimates and confidence intervals are not presented consistently for all outcomes.

5.2 Summary of submitted evidence

5.2.1 Summary of results

Clinical evidence reported in the manufacturer's submission is based on results from the APEX trial which is published as one full publication[4] (NEJM) and a series of abstracts. Clinical effectiveness results are not fully tabulated in the manufacturer's submission for each outcome and time point, and there are a number of differences between the manufacturer's submission and the NEJM paper in reporting of results.

There are differences in the reporting of baseline characteristics, such as the number of patients receiving more than one prior therapy. The percentages of patients receiving 2, 3 or 4 lines of previous therapy in the manufacturer's submission are 98%, 82%, and 34% for the bortezomib group, and 99%, 84% and 35% for the HDD group. However, in the NEJM paper[4], 56% of bortezomib patients and 58% of HDD patients had received 2 or 3 therapies, and 4% of bortezomib patients and 7% of HDD patients had received 4 or more previous therapies.

There are other small differences between the baseline characteristics in the NEJM paper[4] and in the manufacturer's submission (Table 7, p.33). Karnofsky Performance Scale (KPS) scores ≥ 70 are reported in the NEJM paper, whereas KPS scores ≥ 60 and ≥ 80 are reported in the manufacturer's submission. However, the manufacturer's submission's KPS ≥ 60 figures are the same as the NEJM KPS ≥ 70 ones. It is not clear where the manufacturer's submission's KPS ≥ 80 figures come from as these aren't in the NEJM paper. These figures are not reported for the whole ITT population, and use n=322 and n=325 for bortezomib and HDD, respectively. Serum β_2-microglobulin levels are reported as medians in the NEJM paper, but the manufacturer's submission presents these as $\beta_2 M > 2.5$ and $\beta_2 M > 5.5$. It is not clear where this data comes from. CrCl ≤ 20 is presented in the NEJM paper whereas CrCl ≤ 60 is shown in the manufacturer's submission. These figures are quite

different (8/330 and 5/323 in the NEJM paper, 110 and 111 in the manufacturer's submission).

A table of baseline values is presented, but no P values are shown. The two groups are described as: "…comparable with regards to several baseline disease characteristics, including amount of serum and urine M-protein present, proportion of patients with abnormal skeletal survey findings and with plasmacytomas, serum calcium, C-reactive protein and β_2-microglobulin levels, renal function, and refractoriness to prior therapy" (p.33). However, abnormal skeletal survey, plasmacytomas, hypercalcaemia and corrected serum Ca are shown in the manufacturer's submission, but are not in the NEJM paper[4], so it is not clear where some of this information comes from.

The primary outcome measure is time to disease progression (TTP). This is defined as the duration from the date of randomisation until the date of first documented evidence of progressive disease (such as raised M-proteins or new skeletal event) or relapse for patients who experienced a complete response. Evidence of progressive disease included elevation of M-proteins, skeletal events, and other measurable indicators which could be considered to be free from any observation bias. The manufacturer's submission authors state that TTP was chosen as the primary outcome "due to its accepted correlation with survival and because of its clinical relevance". Given the high mortality rate for this condition, the ERG wondered whether survival would have been a more appropriate primary outcome and requested further information from the manufacturer on this point. However, clinical advisors indicate that TTP is an appropriate primary outcome. *Response from the manufacturer's justifying their choice of TTP as the primary outcome is included in the Addendum.*

Response rates (overall, time to achieving and duration of), survival (overall and one year) and quality of life are secondary outcome measures. Response rates were based on the European Group for Blood and Marrow Transplant criteria, and were determined by a computer algorithm developed by the sponsor and validated by the Independent Review Committee. Assessment of this outcome measure should therefore have been free from bias, despite the unblinded nature of the trial. *ERG requested further information on the members of the independent review committee, and this was later provided by the manufacturer (see Addendum).*

Time to new skeletal event (e.g. new fractures, irradiation of or surgery on bone, or spinal cord compression) is listed as an outcome measure but the results are not discussed in detail in the manufacturer's submission. In this population, new skeletal events would not be

a sufficient objective criterion as the extent and nature of the disease would vary according to the individual myeloma patient and to the time frame of the disease. Previous skeletal disease/events would be a pre-existing risk factor for subsequent events.

The incidence of adverse events is listed in the manufacturer's submission. The unblinded nature of the trial may have affected the reporting of these. Grade 3 infections are discussed as a separate measure, assessed using the National Cancer Institute Common Toxicity Criteria.

Quality of life outcome measures used were the EORTC QLC-C30 questionnaire, and a neuropathy-specific measure (FACT/GOG-Ntx). Data are not presented in the manufacturer's submission. *The ERG requested more information on this outcome from the manufacturer, and the manufacturer supplied a conference presentation with further data.*

The NEJM paper[4] discusses key results in the text, but does not tabulate the main outcome measures. These are presented as a series of small graphs, and it is not possible to read accurate data from these. The information in Table 3 is taken from the text of the NEJM, with additional information from the manufacturer's submission which could not be verified shown in italics.

Table 3 Summary of interim results (8.3 months) and one year survival from NEJM publication[4]

Outcome measure	Bortezomib (n=333)	HDD (n=336)
TTP Median TTP (95% CI) Hazard ratio (95% CI) P value	Increase of 78% [a] 6.2 months *(4.6, 6.9)* 0.55 *(0.44, 0.69)* p<0.001	3.5 months *(2,8, 4.2)*
Overall survival [b] median Months (95% CI) Hazard ratio (95% CI) P value	16.6 *(16.6, NE[c])* 0.57 *(0.40, 0.81)* p=0.001*(p=0.0013 in manufacturer's submission)*	Actual figures not reported
One year survival [b] P value	80% p=0.003	66%
	Bortezomib (n=315), not ITT	**HDD (n=312), not ITT**
Overall response rate, n(%) Complete response, n(%) Near complete response, n(%) Partial response (excluding near complete response), n(%) Median time to response Median duration of response	121(38) p<0.001 20(6) p<0.001 21(7) p<0.001 80(25) p<0.001 43 days 8 months	56(18) 2(<1) 3(<1) 51(16) 43 days 5.6 months

[a] stated to be 58% on p.45 of the manufacturer's submission and 78% elsewhere in the report and in the NEJM publication[4]
[b] includes 44% of HDD patients who crossed over to companion study after experiencing progressive disease on HDD prior to the interim analysis.
[c] NE was not defined in the manufacturer's submission report. ERG requested clarification of this, and it was confirmed to mean 'not evaluable'

There is very little discussion of the primary outcome measure (TTP) in the manufacturer's submission report, and the majority of the results section is based on analysis of response rates. A significant increase in TTP (primary outcome) is reported for the bortezomib arm, but this is stated to be 58% in one part of the manufacturer's submission (p. 45) and 78% elsewhere in the report (p.47). The latter figure corresponds to results in the NEJM. *ERG requested clarification of this point from the manufacturer, who confirmed that the figure 78% was correct.*

Overall survival was stated to be significantly prolonged for patients in the bortezomib treatment group compared to patients in the HDD group, with a hazard ratio of 0.57 (p=0.0013) (p.47). It is not clear where this information comes from. The NEJM paper states that, at interim analysis, bortezomib patients had a significantly improved overall survival (p=0.04) compared with HDD patients, but does not give any further information other than in graphs which are too small to read accurately.

Both the NEJM paper and the manufacturer's submission (p.47) state that one-year survival was significantly prolonged in the bortezomib arm, compared to the HDD arm (p=0.003). The manufacturer's submission states that patients in the bortezomib group experienced a 41% decreased risk of death in the first year of treatment (HR 0.57, p=0.001). This analysis of overall survival includes data from 147 patients in the HD group who had disease progression and subsequently crossed over to receive bortezomib in a companion study (44%).

Statements in the manufacturer's submission regarding response to therapy (n=627, not ITT) reflect data in the NEJM paper (and in abstracts for later results) although no information was found in the NEJM paper[4] regarding durability of responses.

5.2.1.1 Updated analyses

Updated analyses in the manufacturer's submission generally reflect data in published abstracts, but confidence intervals included in the manufacturer's submission are from extra information not available in these. The 22 month data shown in the manufacturer's submission's Table 10 are referenced as coming from reference 9, but do not seem to be in this abstract. The partial response rate at 15.8 months in table 11 of the manufacturer's submission is 87 (28%), whereas the figures are higher in the cited abstract (manufacturer's submission ref 9) at 108 (34%). Table 12 in the manufacturer's submission shows

percentage of responders by treatment cycle. The data for this come from 'Data on file, Ortho Biotech' so we cannot check this.

5.2.1.2 Health related QoL

The manufacturer's submission text reflects information in the cited reference. Table 14 data is not from the referenced abstract, so cannot be cross-checked.

5.2.1.3 Subgroup analyses

Subgroup analyses of 'earlier vs. later bortezomib' are presented on p.54 of the manufacturer's submission, and these data generally reflect data in the cited abstract. No supporting data were found in the cited abstract for extended overall and one year survival. Information in Appendix 5 of the manufacturer's submission discusses survival in relation to number of previous therapies, which relates to earlier versus later treatment with bortezomib.

Subgroup analyses of high-risk patients reflect data in the cited reference. It is not clear where the data for low risk patients come from, and no supporting data were found for the stated 47% higher risk of death for the HDD arm for patients <65 yrs old (p.56).

5.2.1.4 Adverse events

All adverse events in the NEJM paper are presented in the manufacturer's submission. Clinical advisors consulted by the ERG confirmed that the adverse events observed in the trial were to be expected in the patient group, and that they were manageable and had relatively mild effects.

P values are not presented for adverse events, and there are some large differences between the groups (e.g. 57% vs. 21% for diarrhoea, 57% vs. 14% for nausea, peripheral neuropathy 36% vs. 9%, thrombocytopenia 35% vs. 11 % for bortezomib and HDD respectively). P values are not presented for differences in withdrawals due to adverse events, and these seem quite high in the bortezomib group (37% vs. 29%).

Grade 3 adverse events were reported in 61% of patients receiving bortezomib and in 44% of patients receiving dexamethasone. Both the manufacturer's submission and the NEJM state this, but the manufacturer's submission does not include the significant p values stated in the NEJM paper (p<0.01). The manufacturer's submission discusses the rate of grade 3

infections rather than grade 3 events, stating that this is 13% in the bortezomib group and 16% in the HDD group (p=0.19). The NEJM paper is referenced for this information, but does not seem to contain these data. *The ERG requested further information from the manufacturer on this point- see Addendum for details.*

5.2.2 Critique of submitted evidence syntheses

No evidence synthesis in the form of a meta-analysis was possible as there was only one RCT, and this was reported by narrative means.

5.2.3 Summary

- The manufacturer's submission probably contains an unbiased estimate of the treatment effect of bortezomib within the stated scope of the decision problem. This is based on the results of a single RCT which is of reasonable quality when judged using the NICE quality assessment criteria, but the reporting of the trial results is not totally transparent and not all results are fully tabulated for each outcome. It is difficult to interpret the data with full confidence due to the lack of details and unclear reporting of cross-overs from the HDD group to the bortezomib group, censoring of data and the reporting of relative rather than absolute values for some outcomes, and the discrepancies between the manufacturer's submission and the paper published in the NEJM. Whilst these factors make it difficult to assess the size of the treatment effect, these limitations probably do not significantly affect the overall results especially in the light of clarifications received from the manufacturer.

- Results suggest that bortezomib treatment in MM patients who have had a relapse after one to three treatments is superior to HDD in terms of survival, time to disease progression and response rates.

6 ECONOMIC EVALUATION

6.1 Overview of manufacturer's economic evaluation

The manufacturer submission (received 1st March 2006) to NICE includes (i) a short report on the cost effectiveness literature available to inform on economic evaluations of bortezomib in patients with MM (4 references, 2 of which are conference abstracts, Table 22 p71), (ii) reference to an early economic analysis undertaken using individual patient-level trial data (from Phase II RCT), and (iii) a report on an economic evaluation undertaken and presented specifically for the NICE STA process.

Note: See Addendum to this report, where the manufacturer reports additional information in response to some specific early questions from the ERG.

The results from the economic evaluation are presented for the base case assumptions (i.e. patients with 1 relapse, up to 6 treatment regimens), and thereafter for a further three scenarios; (1) limiting the number of cycles of bortezomib in non-responding patients, (2) using the overall population of the APEX trial (i.e. not limited to 1 relapse), (3) looking at treatment using bortezomib and HDD in combination versus HDD only.

CEA Methods
The CEA uses a decision-analytic model (quasi-Markov) to estimate the effect of treatment with bortezomib compared to HDD, summarised below:

6.1.1 Natural history

The model uses (non-trial) observational data to predict the treatment experience of a cohort of patients treated with HDD. The patient group are defined as MM patients who have experienced a first relapse of MM treatment. Data from the RCT on HDD are deemed to be unavailable/inappropriate because of the early termination of the trial and subsequent inability to predict long term outcomes and mortality data with HDD. The model uses data from the Mayo Observational Study[11] (Kumar et al 2004) to model this baseline/comparator cohort. Data from the Mayo Study are from patients who have been treated with a range of different drugs, although few of these were treated with HDD. The submission states that 188 patients (32.5%) were treated at some point in their follow-up with VAD (combination of vincristine, adriamycin and dexamethasone), with 74 of these receiving VAD after their first relapse. The model predicts treatment experience, rather than disease progression, modelling the flow of patients through a series of treatment regimens (from regimen 2 to 6). There are two general health states used in the model; these are 'on treatment regimen i',

and 'death whilst on treatment regimen i' (i=1 to 6). These health states apply to each of the potential treatment regimens (regimens 2 to 6), therefore there are 10 health states in which patients can arrive (5 regimens x 2 states). In any cycle patients may remain on that treatment regimen, progress to a new regimen or die on that regimen. The first and second cycle times are 3-months each, the third cycle time is 6 months, and all subsequent cycles are 1-year. There are 18 time cycles in the model, with all patients starting in the second regimen, giving the model a time horizon of 15-years (a lifetime horizon in this patient group).

6.1.2 Treatment effectiveness

The clinical effectiveness data from the APEX RCT, showing a relative benefit in time to treatment progression (HR=0.56) and a relative benefit in overall survival (HR=0.42), are applied to the baseline prediction for HDD patients. The model adjusts the baseline transition rates (between health states) according to the hazard rates estimated in the APEX RCT (bortezomib vs. HDD). The model uses the comparative data to simulate the treatment effect from bortezomib, its impact on survival, and the subsequent cost effectiveness of treatment. Treatment effect on adverse events is not included in the CEA. Bortezomib is assumed to have a treatment effect lasting for up to three years.

6.1.3 Health related quality-of-life

The manufacturer's submission presents results from CEA as cost per life-year-gained; it does not use a single index measure of HRQL to calculate QALYs. The submission presents some discussion to justify this approach. The justification is based on the manufacturer's opinion that "increased survival is the single most important outcome for clinicians and patients" (p76), that meaningful interpretation of data collected on EQ-5D within the APEX trial was not possible, and a stated view from the manufacturer that there is uncertainty over the validity of the EQ-5D in MM patients. On this latter point the submission includes an outline of views from a seven person focus group (MM patients), which reports that the main finding was "the EQ-5D is unlikely to be an appropriate utility measure in this patient group" (p77). Whilst the submission does highlight, from focus group findings, that MM would be expected to have a considerable impact on health related utility, the analysis uses survival (life years gained) only, with no consideration of HRQL.

Note: See Addendum to this ERG Report, where the manufacturer provides additional information including estimates of cost per quality-adjusted life-year (QALY).

6.1.4 Resources and costs

The dose data for bortezomib is from the APEX RCT, and other resource use data are from a published study[12] (Bruce et al 1999) reporting on a survey of clinicians (expert opinion) with experience of treating patients in the MRC VI myelomatosis trial[13] (Table 26,27 p80,81).

6.1.5 Discounting

A discount rate of 3.5% is used for future costs and life years.

6.1.6 Sensitivity analyses

Sensitivity analysis is reported: simple one-way sensitivity analyses, probabilistic sensitivity analyses, and scenario analyses.

6.1.7 Model validation

The submission reports model validation for survival against the Mayo clinic observation arm for the HDD baseline cohort (submission, Appendix 12). Survival is then calibrated to the control arm of the APEX trial (submission, Appendix 12).

6.2 Results

Results are presented as cost per life-year-gained, with incremental costs and life-years-gained also presented separately (in most cases). The primary analysis presents an estimated cost per life-year-gained of £30,750. One-way sensitivity analysis is reported for the variables of time to progression (hazard rate), overall survival (hazard rate), duration of treatment effect, treatment cost, and other cost. A limited range of analyses is reported, with the cost per life-year-gained ranging from £27,957 to £36,747 in these analyses. Table 4 below summarises the results presented in the submission for the primary analysis (deterministic methods), probabilistic analyses (range) and for the scenario analyses undertaken.

Table 4 Cost effectiveness results presented in manufacturer's analyses

Analysis	Difference in mean discounted overall survival (months)	Difference in mean total costs	Estimated cost per discounted life-year-gained
Primary analysis (patients with 1 relapse)	9.9	£25,327	£30,750
Primary analysis with probabilistic sensitivity analysis			£26,855 (5th percentile) £38,564 (95th percentile)
Scenario: only 3 cycles for non-responders	8.2	£19,172	£27,926
Scenario: overall Apex data (i.e. patients with 1 or more relapse)	7.3	£24,165	£39,954
Scenario: bortezomib + HDD versus HDD alone	11.0	£25,898	£28,281

Note: See Addendum to this ERG Report, where the manufacturer provides additional information, including cost per QALY estimates.

6.3 Critical appraisal of the manufacturer's submitted economic evaluation

6.3.1 Critical appraisal of economic evaluation methods

The ERG have considered the methods applied in the economic evaluation in the context of the critical appraisal questions listed in Table 5 below, drawn from common checklists for economic evaluation methods (e.g. Drummond et al 1997).[14]

Table 5 Critical appraisal checklist of economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	✓ - Yes	None
Is there a clear description of alternatives?	✓ - Yes	Bortezomib versus HDD (primary analysis)
Has the correct patient group / population of interest been clearly stated?	? – Yes/No	The CEA uses MM patients after 1st relapse for the primary analysis. This is a subgroup of the eligible patient group from the licence indication. Scenario analysis presents results using the overall trial group (eligible group) of patients with at least one relapse.
Is the correct comparator used?	?	Whilst the analysis justifies the use of HDD as the comparator, clinical opinion suggests that HDD is not a commonly used comparator strategy in the UK. Although HDD is one of a variety of approaches that can be used for treating MM at 1st relapse. The manufacturer's submission justifies the used of HDD as the comparator on the grounds that there is no UK consensus on best practice for the treatment of MM at first relapse and that HDD is commonly used. Furthermore they comment that thalidomide is not an appropriate comparator.
Is the study type reasonable?	✓ - Yes	CEA model used. But no CUA results presented.
Is the perspective of the analysis clearly stated?	✓ - Yes	Perspective stated as UK NHS
Is the perspective employed appropriate?	Cost: ✓ - Yes Outcomes: ✓/?	Submission appears to adopt a UK NHS and PSS perspective for costs (consistent with NICE reference case). Perspective on outcomes is that of the patient, but not all effects considered (i.e. HRQL not considered)
Is effectiveness of the intervention established?	?	The CEA is based on clinical effectiveness data from the APEX RCT. Hazard rates calculated as part of the APEX analysis are applied in the CEA. Detail is presented on the RCT, but interpretation of trial findings, and consideration of methods employed are required by NICE.
Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	✓ - Yes	CEA uses 15-year time horizon, which is lifetime in this patient group.
Are the costs and consequences consistent with the perspective employed? *	? – Yes/No	Costs appear to be consistent with perspective employed, but limited info/justification provided. Consequences limited to patient life-years-gained.
Is differential timing considered?	✓ - Yes	None
Is incremental analysis performed?	✓ - Yes	None
Is sensitivity analysis undertaken and presented clearly?	? – Yes/No	Yes sensitivity analysis is undertaken, but it may be regarded as limited.

* More on data inputs for costs and consequences in the review of modelling methods below

Note: See Addendum to this ERG Report, where the manufacturer provides additional information, including cost per QALY estimates.

6.3.2 NICE reference case

Table 6 reports the manufacturer's submission against the requirements of the NICE reference case. There is some overlap with the items presented above on the general approach for the economic evaluation. There is some interpretation required of the

comparator used in the analysis (HDD) in the context of routine practice in the UK. The fact that the submission does not include consideration of HRQL, and the subsequent presentation of cost per QALY, is highlighted heavily in the cross referencing of the submission with the NICE requirements for technology appraisals.

Table 6 NICE reference case requirements

NICE reference case requirements (see detail in NICE report):	Included in Submission
Decision problem: As per the scope developed by NICE Comparator:	N/A
Alternative therapies routinely used in the UK NHS	?
Perspective on costs: NHS and PSS	✓
Perspective on outcomes: All health effects on individuals	X
Type of economic evaluation: Cost effectiveness analysis	✓
Synthesis of evidence on outcomes: Based on a systematic review	X
Measure of health benefits: QALYs	X
Description of health states for QALY calculations: Use of a standardised and validated generic instrument	X
Method of preference elicitation for health state values: Choice based method (e.g. TTO, SG, not rating scale)	X
Source of preference data: Representative sample of the public	X
Discount rate: 3.5% pa for costs and health effects	✓

N/A=not applicable

Note: See Addendum to this ERG Report, where the manufacturer provides additional information, including cost per QALY estimates.

6.3.3 Modelling methods

An outline critical review of modelling methods has been undertaken. The review has used the framework for good practice in modelling presented by Philips et al (2004)[15] as a guide, addressing issues of model structure, structural assumptions, data inputs, consistency, and assessment of uncertainty.

6.3.3.1 Modelling approach / Model Structure

The manufacturer's submission provides very little commentary on the development of the model structure, other than stating that clinical experts were involved in all stages of the modelling approach. The model structure covers the treatment of MM, rather than presenting a framework depicting the natural history of disease for MM (e.g. by type of response, remission, plateau, for disease progression). Therefore the health states used predict the progress of patients to different treatment regimens and to death, rather than the underlying biological process of the disease. Competing theories of model structure have not been discussed in the report.

The model is described as a semi-Markov approach, and whilst noting the points above, it would appear to be an appropriate modelling approach given the decision problem, the data available, and the specified causal relationships (e.g. impact of relapse on overall survival) within the model.

The cycle length used in the model is variable, with the first two cycles at 3-months and the third at 6-months, with a 1-year cycle length used thereafter. This approach fits the model around the trial data available, in the first 12 months, and applies a 1-year cycle thereafter. Given that the time to relapse (between different regimens) may be an important factor in the prognosis of disease a 1-year cycle may not be appropriate, and the use of a 3-month or 6-month cycle throughout the time horizon may reflect a better profile of disease over time. A half-cycle correction has not been used.

The time horizon of the model is 15-years, and reflects a lifetime horizon (in patients with MM at mean age of 65-years). Mortality in the patient group is high and a shorter time horizon may be justified. No sensitivity analysis is reported against the issue of time horizon.

Sources of data used to develop/populate the model structure are specified (e.g. APEX RCT data, observational data from the Mayo clinic).

6.3.3.2 Structural Assumptions

The submitted economic evaluation assumes that modelling 'treatment' of MM in the specified patient group captures all of the important factors related to the natural history of the condition. However, as above, the model may not capture important issues such as the relevance of 'time to relapse' (e.g. 1-month treated the same as 10-months to relapse), or the importance of the adverse events profiles.

The economic evaluation assumes that costs for 'other resource use', as estimated in the study by Bruce et al (1999),[12] (non bortezomib/HDD costs) are evenly spread over time. The ERG considers this to be a pragmatic assumption given the limited evidence presented in the manufacturer's submission on resource use and costs for 'other' resources. Whilst Appendix 10 of the manufacturer's submission indicates a literature search has been undertaken to identify UK cost studies no comment or detail is provided on such a literature search in the submission.

The model uses data from the study by Kumar et al (2004)[11] (Mayo clinic data, Rochester USA), to predict the baseline disease progression for the comparator group, i.e. HDD treated patients. This assumes that the data from Kumar et al[11] are able to reflect disease progression in the specified patients (controls). Whilst the Kumar et al study seems a good quality observational study, and there is an absence of alternative data sources available (see below), when applying this data in the context of the CEA presented there may be some areas of uncertainty. For example, the patients used from the Kumar et al[11] study are (i) a subset of the Mayo patient data presented, (ii) this observational study reports data collected over a 13-year period (in a USA context), and patients may not have benefited from the latest treatment protocols, (iii) HDD was not one of the reported treatment regimens for the observational study, (iv) the observational data are not specific on which patients had what treatment and when, (v) that there are some differences in the APEX RCT and Mayo patient profiles, e.g. patients in APEX RCT are diagnosed approximately 5 years earlier than the Mayo patients. The Mayo study data show 355 persons receiving a 2nd regimen, the biggest group (n=160) getting combination alkylating agents, with 33 patients received VAD (where dexamethasone is expected to be the most active ingredient). In the dataset presented by Kumar et al[11] 114 persons received VAD as 1st regimen. The suitability of this data is open to some judgement and interpretation. The manufacturer's submission indicates a literature search was undertaken to identify epidemiological data to model longer-term survival for both bortezomib and HDD patients (Appendix 10), but the study by Kumar et al, Mayo Study, was the only one identified. The manufacturer's submission does not discuss any alternative data sources.

The ERG suggest that the data used may predict a more severe disease progression/profile (e.g. time to progression may be shorter than expected, and mortality may be higher) than may be expected in a hypothetical cohort of patients treated with HDD in a RCT context (i.e. a direct comparison to the data applied for bortezomib from the APEX RCT). This suggestion is based on the issues discussed above. These issues could bias the estimates of treatment effect, given the model uses transit probabilities for the base case and adjusts these for treatment effect using hazard rates from the APEX RCT data. The submission does make an adjustment to the observational data to reflect the survival rate of HDD patients calculated from the APEX RCT (patients who had received only one prior therapy and then received HDD). The impact of alternative estimates for the transit probabilities used is uncertain, and it has not been addressed as part of the sensitivity/scenario analyses reported.

The submitted model assumes that patients may switch to more than one regimen when in later cycles of a 1-year duration. For example patients may progress from regimen 2 to regimen 6 in the course of one (1-year) time-cycle. The model also assumes that patients may die in any of these cycles (p240). There may be an overestimation of the death rates but it is not clear what effect this has on results. The validation of the model outcomes is discussed in more detail below. A shorter cycle length may overcome/weaken these assumptions.

The model assumes that there is an overall survival hazard ratio of 0.42 (from APEX RCT) in the first year and 0.83 in years 2 and 3. The latter hazard ratio has not been justified.

The manufacturer's submission assumes that there are independent benefits for TTP and OS. Given the workings of the model these benefits may not be independent, and it may be that the group of patients who have OS benefits will also have TTP benefits. Thus there may be some double counting for the effect of bortezomib. The submission states this not to be the case, but the ERG would like further clarification on this.

6.3.4 Data Inputs

6.3.4.1 Patient Group

The analysis uses MM patients who have experienced a first relapse of treatment for MM. The model does not have patient characteristics as specific model/data inputs. The primary CEA uses a patient group with only one prior treatment (at first relapse) i.e. a subgroup of the APEX trial (approx. 40% of patients). Scenario analyses are presented for a patient group consistent with the APEX trial participants where the inclusion criteria is for at least one relapse (in APEX 60% of patients started the trial at second relapse and beyond). The APEX trial data is from patients with a median age of 61-62 years, with median time from diagnosis between 3.1 and 3.5 years.

UK practice up to now would seem to be that bortezomib has been used in patients after second relapse, and whilst it is likely to be more readily used in earlier stages of disease (after first relapse) where recommended by NICE, it may still be used in a broader patient group than those with only one relapse. The cost effectiveness of treatment differs quite widely between the specified patient groups.

6.3.4.2 Clinical Effectiveness

Treatment effect is modelled over time. It is based on hazard ratios reported in the APEX RCT for (a) time to progression and (b) overall survival, with the effect from treatment modelled on the basis of an adjustment (via hazard ratios) to a baseline prediction of treatment pathway and disease progression. Actual treatment effect, i.e. the difference in the numbers of patients in the treatment and control group, is relative to the underlying transit probabilities used to model baseline treatment and progression. The earlier sections of the current report have presented a review of the APEX RCT. An equally important aspect of the CEA/model is the baseline cohort predictions. The baseline cohort are modelled on the basis of data from an observational study in MM patients at the Mayo clinic (USA), (Kumar et al 2004[11]). Whilst the approach taken appears reasonable to the ERG, the submission does not discuss quality assessment / limitations in the data used for the HDD group.

In the model the treatment effect is assumed to last for three years in the model, and the ERG consider that this has not been justified sufficiently in the report.

One observation of the ERG is that the modelling undertaken has used the rates seen in the Mayo clinic study without aggregating data from different groups. Hence there may be concern that for some transition rates (probabilities) the small numbers in the subgroups (see Table 7 of ERG report) have led to unrealistic transition rates for the 5th and 6th regimens. A more appropriate approach may have been to aggregate these subgroups.
Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.

6.3.4.3 Patient outcomes

Trial outcomes of time to disease progression, and overall survival have been used in the CEA. No patient related measures of HRQL have been discussed, or applied. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information including discussion of QALYs.* The submission refers to the collection of EQ-5D data as part of the APEX RCT but no data are presented on this outcome. The submission refers to data collected in the APEX RCT on the EORTC-QLQ30 but data for the individual dimensions of this health status instrument have not been presented in the submission (or in the published study).

As above (5.1.3) the submission presents an argument for not using the EQ-5D in this patient group, however, this argument is not generalisable to use of HRQL *per se*, especially as the submission itself highlights the negative impact of MM on the quality-of-life of this patient group. Although the ERG have not undertaken systematic reviews in this area we have identified a number of studies that have reported health state values for the MM patient group. Gulbrandsen et al (2001)[16] use EORTC-QLQ30 data from trial participants and 'map' to provide a single index utility score (using the 15-D generic health state classification system). They present estimated societal preferences from this mapping process (map to 15-D general population values), from the EORTC-QLQ30 health state descriptions (which they admit involves some subjective judgement and uncertainty), with a value of 0.7334 estimated at 6-months in the treatment arm (high dose melphalan) of the trial data used, and 0.7896 in the control group (MP). Further detail is provided on HRQL domains and the QLQ30 scores by the Gulbrandsen et al (2001), (see original study for further detail).

Van Agthoven et al (2004)[17] report an estimated health state value for MM at 0.644 for patients with a mean age of 54-years with previously untreated MM, stage II or III A/B disease. The authors used an age-specific EQ-5D weight (0.8), and adjusted this by -19.5% to correct for MM. This analysis was reported in the context of a RCT (the HOVON 24 Study).[18] The health state value is stated to be for patients who "were in an undefined state following intentionally curative primary therapy" (van Agthoven et al, p1162).

The above studies indicate a health state value between 0.644 and 0.789 may be appropriate for patient groups with MM. However, Kind et al (1998)[19] 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. Therefore health state values for those with MM may be expected to be somewhat lower. The use of a health state value of 0.80 (simply applied to the estimated survival difference) in patients treated with bortezomib/HDD, would lead to an estimated cost per QALY of £38,374 in the primary cost effectiveness analysis presented (compared to £30,750 per LYG presented in the submission).

Health state values could be simply incorporated into the model. This could be done by assuming a constant health state value for all patients in all different regimens or by varying health state values for different regimens. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.*

6.3.4.4 Resource use

Treatment cost is calculated from therapy given (doses administered) in the APEX trial, and UK unit cost data from the British National Formulary (no. 48).[20]   The submission reports a calculation of doses across a range of patients (responders, non-responders, progressive disease) to estimate an 'all patients' ITT cost for both bortezomib and HDD therapy. The estimated mean cost per patient for bortezomib in the model is £21,035 (based on an estimate mean of 25 doses of bortezomib per patient), and for HDD it is £82. These data differ slightly from those presented in Appendix 14 of the submission. The estimated mean cost for bortezomib therapy for responders is circa. £27,000 per patient (based on a mean 33 doses of bortezomib) in the 'ITT' cost calculation.

The ERG consider the estimate of cost for bortezomib to be a reasonable estimate. Each treatment cycle comprises bortezomib therapy at $1.3mg/m^2$, with 4 doses per cycle. Each vial is 3.5mg and the price excluding VAT is £762.38, giving a bortezomib cost of £3,050 per cycle, assuming 1 vial per dose. Therefore a patient receiving 6 cycles of treatment will have a bortezomib cost of £18,297, plus an additional administration cost per dose.

The manufacturer's submission estimates an administration cost per dose of £79, with a mean total administration cost of £1,672 per patient. The ERG considers this may be an underestimate of the actual resource use and cost for the administration of each dose of bortezomib. Assuming treatment via an outpatient clinic, each dose will involve an outpatient visit (estimated at between £86 and £93 each, for clinical haematology or clinical oncology, NHS Reference Costs[21]), plus additional pharmacy preparation time (estimated at £20-£25 each dose, from clinical advisors). Therefore the administration cost per dose may be in excess of £100, with a total mean administration cost per patient in excess of £2,500.

Of note is the issue of waste in the dosing / administration of bortezomib. Given that the average person's body surface area is much less than $2.7m^2$ (the surface area relevant for a 3.5mg vial), the packaging for bortezomib leads to significant waste where one single vial is used per dose. In the Phase II RCT[22] for bortezomib the mean patient body surface was $1.98m^2$ (Richardson et al 2003). With the assumption of a single use per vial there will be waste in the provision of bortezomib therapy. The ERG estimates that almost 1mg per vial will be wasted, where patients have a surface area of $1.98m^2$. The ERG estimates that where patients can be treated in a group / cohort (i.e. in scheduled clinic approach) there may be cost savings where the product can be used without waste. This saving could be as much as £5,000 per patient over a course of 6 treatment cycles (24 doses), assuming $1.98m^2$ surface area and no waste. Assuming that 0.5mg of a 3.5mg vial is wasted (unused)

the estimated cost saving is circa. £2,350 per patient (over 24 doses). These estimated cost savings where the assumption of 'single use' per vial is dropped are illustrative, and it is not known how feasible it is in practice to deliver such savings in a hospital clinic or inpatient setting. The manufacturer's submission does not discuss/refer to these potential cost savings.

The manufacturer's submission does not discuss any systematic methods used to search for and identify data on resource use. Although Appendix 10 does indicate that a systematic search was undertaken to identify UK studies on resource use/cost for MM. The submission uses data from one published study (Bruce et al 1999)[12] to inform on resource use, and unit costs associated with, "other" cost items i.e. costs other than the intervention cost. This study is based on expert opinion, with the experts drawing on experiences generally and from involvement in the MRC VI Myelomatosis trial.[13]) The MRC VI Myelomatosis trial included patients at diagnosis of MM treated with initial chemotherapy (using non-stem-cell-directed therapy) i.e. patient group differed from that used in the APEX RCT and the economic analysis presented in the manufacturers submission. That the study data are based predominantly on expert opinion is not discussed clearly in the submission. Quality assessment / limitations in the data used are not discussed.

The data from the cited study were in 1997 UK pounds, and the methods used in the submission to up-rate these costs were incorrect (from the Excel model, the submission used 1999 as a base year, and did not apply accepted cost/year indices). Where the 1997 estimate of cost per month is taken from the Bruce et al study[12] and up-rated to 2004/5 costs (latest available via PSSRU report[23]) using HCHS indices (173.5 for 1997/8, 234.2 for 2004/05),[23] the ERG estimate a cost per month of £470 (as opposed to £443 in the manufacturer's submission). This issue does not have a substantive impact on the cost effectiveness estimates presented. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.*

The manufacturer's submission does not address the issue of additional resource use attributable to adverse events which may be more frequent with the use of bortezomib, compared to HDD. The clinical evidence has indicated that there are a greater number of adverse events in the bortezomib group than the HDD group (APEX RCT[4]). The economic evaluation presented only allows for some additional treatment costs associated with bortezomib in a sensitivity analysis (increasing the monthly cost for 'other costs' by 25% in the bortezomib patient group i.e. from £443 per month to £554 per month). This sensitivity analysis, based on an extra £1,329 per year per patient leads to an increase of £2,646 in the

cost per life-year-gained (from £30,750 to £33,406 per life-year-gained). In practice patients may be prescribed bisophosphates to prevent/treat adverse events, with an annual cost of approx. £360 (estimate of £30 per month from clinical advisor). Where serious adverse events occur (e.g. thrombocytopenia, peripheral neuropathy) they may involve significant resource use / cost for treatment of the events themselves and their consequences. This issue has not been discussed in the manufacturer's analysis.

6.3.4.5 Costs

As above, unit cost data for bortezomib and HDD are from the UK BNF (no. 48). The cost used for bortezomib is still current at March 2006. Other cost data are used directly from one published study (Bruce et al 1999)[12]. Cost data are presented in the submission as 2005/6 UK pounds (see comments above). Costs associated with adverse events, and treatment to prevent adverse events, have not been discussed.

6.4 Consistency

The ERG has examined the submitted Excel model for internal and external consistency and accuracy. This was made more difficult by a lack of documentation on how the model works, and the ERG provide a brief overview in this section.

6.4.1 Internal consistency

Random checking has been done for some of the key equations of the model, for example on Excel worksheets titled *Vprobs*, *Velcade only* and *V mn 1-3*, *Vmn 4-6*. The ERG has not undertaken a comprehensive 'checking' process against all cells in the model. The model is fully executable and inputs changed in the *Summary* sheet (cells D4:F25) produce immediate changes to the deterministic results in this sheet (cells J5:N17). These parameter inputs and model outcomes are shown in Table 28 and Table 31 respectively of the manufacturer's submission. The *'Summary'* worksheet in the Excel model is clear and user friendly. The model is shown using the baseline scenarios described and the user is able to choose the other scenarios described in the submission (on pages 85-88) by clicking on a drop down menu/button (worksheet cell H20).

Excel worksheets titled *Mayo* and *Dex* show the transition parameters/probabilities used from the Mayo observational study which are used for the baseline worksheet, *Dex*. These data are shown in Table 7 below. The transition probabilities used are taken directly from these data although they have been adjusted by calibration to the APEX trial (see

submission, p242). Furthermore the timing of cycle lengths in the model correspond with the data presented in this Table. The data shown in the Table correspond to the percentages of patients who die, switch to another regimen or remain on that regimen by the end of specified time periods. As mentioned above there are concerns that the direct use of these data leads to unrealistic transition rates in some regimens and time cycles and that the cycle length used is too long.

Table 7 Patient experience with each regimen*

Time from diagnosis	Regimen					
	First	Second	Third	Fourth	Fifth	Sixth
3 mo	11/10/79	10/16/74	16/16/68	19/20/61	22/22/56	18/27/55
6 mo	17/17/66	18/27/55	22/28/50	26/30/44	39/31/30	25/37/38
1 y	21/33/46	25/40/35	28/40/32	32/44/24	44/44/12	37/41/22
2 y	27/42/31	29/54/17	34/51/15	36/53/11	45/50/5	55/45/0
3 y	31/50/19	31/62/7	35/54/11	40/54/6	45/52/3	55/45/0
Median EFS† (d)	302	221	182	137	122	99

*Values represent percentages of patients receiving regimen who died during treatment/received another treatment/are still receiving this regimen.
†EFS = event-free survival from initiation of regimen.

Source: Kumar et al[11]. Mayo Clin Proc 2004.

Excel worksheets *Dex* and *Velcade* show the transition probabilities. The proportion of people in each of the regimens over time is shown in the worksheets *D probs*, *V probs* and *D + V probs* for each of the comparators. The model has incident cohorts who start in each regimen in each cycle. These values are worked out by matrix multiplication using values from the worksheets titled in the format *V Mn 1-3,* which themselves are from the input parameters shown in the related worksheets, e.g. *Velcade only*. The transition probabilities are defined according to the time that a patient has spent in a regimen so the model is subdivided into the time each of the cohorts have been in this regimen. The effect of this approach is a cumbersome and difficult to interpret series of spreadsheets.

The ERG has discovered several potential errors:

- The model uses a survival benefit for other regimens when the patients are not treated with Bortezomib, ie regimens 3-6, see sheet *Velcade only*.

  Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.

- There seem to be mistakes on some of the sheets with the transition probabilities. For example, cell AP43 on the *V mn 1-3* sheet uses HDD probabilities rather than

velcade probabilities (='Dex only'!$Q$15). This occurs countless times on this sheet and the other transition probability sheets.

Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.

- There may be a problem with double counting. The group who are saved from dying may also have a longer time to disease progression.

However, notwithstanding the comments above, the ERG view the model as a reasonable approach to predicting treatment profiles in this patient. From the random checking undertaken the 'wiring' of the model appears accurate, with the exception of the points raised above. Parameter inputs were checked for correct predictive validity i.e. ERG undertook some independent sensitivity analyses and results were consistent with those expected (results were in the right direction).

6.4.2 External consistency

The model used the Mayo observational study data and was validated against this data for overall mortality (submission, p241). The authors of the manufacturer's submission claim that the model has survival of 72% at 1-year, 61% at 2-years and 25% at 5-years compared to Mayo study findings of 72% at 1-year, 55% at 2-years and 22 % at 5-years. However the ERG were not able to replicate these results. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.*
Where the ERG has explored the model, excluding the adjustment of a reduced mortality risk in the first year of 21% (see below), the ERG found the results in the model to be 64% at 1-year, 39% at 2-years and 10% at 5-years. It may be that the model uses different parameter values, for example for the hazard ratios, but this has not been made clear.

The manufacturer's submission states that the model has been calibrated against the trial arm of the APEX trial (submission, p242) by reducing the mortality risk while on a 2nd regimen in the first year by 21%. The ERG were unable to replicate these results. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.* The model shows 87% and 72% 1-year survival for bortezomib and HDD, compared to 80% and 66% as reported in the published APEX RCT findings (Richardson et al).[4]

Figure 1 Patient survival for HDD and bortezomib for the APEX trial and model results
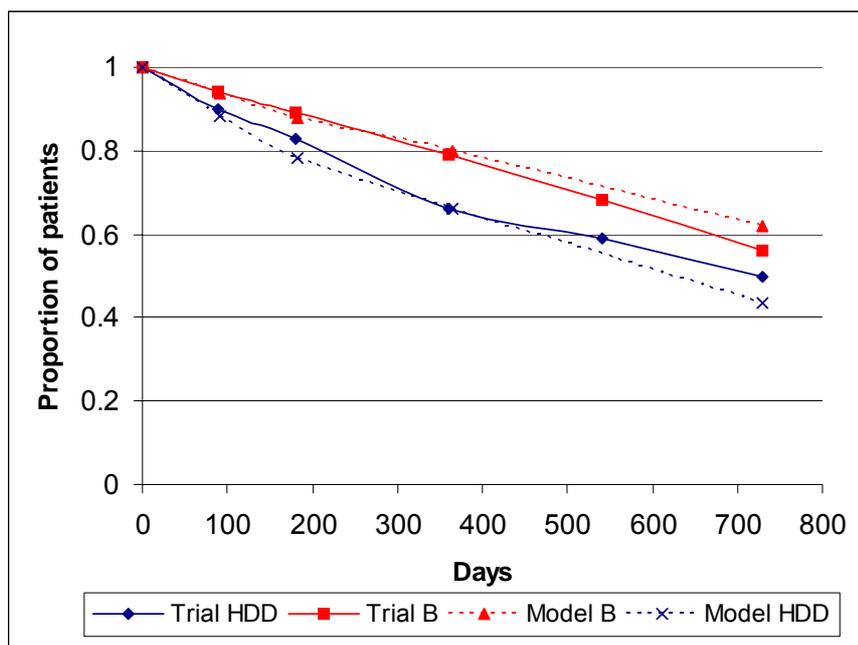
The ERG has compared model results to the Kaplan Meier Survival Curve presented in the manufacturer's submission (p49, Figure 10). The Figure 10 presented in the manufacturer's submission is for all patients in the APEX trial, rather than only those after first relapse (OS RR = 0.57). Therefore the ERG has compared results from the model scenario using the all APEX patient group. The ERG has used hazard ratios of 0.95 and 1 in the worksheet (worksheet cell reference F85 and E85), rather than 1 and 0.79 respectively in the manufacturer's submission analysis. This was done to calibrate the model for first year survival against the APEX trial. The ERG Figure 1 above shows that using these hazard ratios, the model shows good agreement with the control arm of the trial and for the treatment arm for the first year. But, after the first year, the model overestimates patient survival in the treatment arm. Furthermore the treatment effect is also higher than shown in the trial by roughly 40% over the two years. The ERG considers that the model may overestimate the treatment effect of bortezomib.

6.5 Assessment of Uncertainty

6.5.1 One-way sensitivity analyses

The manufacturer's report presents sensitivity analyses on the key parameters in the model (Table 8). It presents the influence of the parameters in a Tornado diagram (Figure 37, p89). According to results presented, the most influential variable is the duration of the treatment

effect. The sensitivity analyses can be changed in the Excel model on the *Sens sheet* and produces results in the *Figures* worksheet. Sensitivity analysis inputs can be changed easily and the results are updated quickly. The ranges chosen for the sensitivity analyses are not representative of the variability of the data. For example the 95% confidence interval for the TTP hazard ratio was 0.44-0.69 (p48) whereas the range used in the report was 0.51-0.61. In general the ERG consider the parameter ranges used were not wide enough.

Table 8 One way sensitivity analyses (from manufacturer's submission, p.88)

Variable	Basecase	Inputs		CE ratios		Range
		Left	Right	Left	Right	
Duration of treatment effect, years	3	4	2	£27,957	£36,747	£8,790
Hazard ratio - TTP	0.56	0.51	0.61	£28,465	£33,605	£5,141
Cost of other care - Bortezomib pre-progression	£443	£443	£554	£30,760	£33,407	£2,646
Cost of Bortezomib per course	£21,035	£20,033	£22,086	£29,534	£32,027	£2,493
Cost of other care - pre- and post-progression	£443	£354	£554	£29,431	£31,804	£2,374
Hazard ratio - OS	0.42	0.38	0.47	£30,200	£31,389	£1,189
Cost of HDD per course	£82	£86	£78	£30,745	£30,755	£9

Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.

Further sensitivity analysis has been undertaken by the ERG, see below.

6.5.2 ERG sensitivity analysis

More appropriate ranges have been used for the one way sensitivity analysis. The ERG has used the 95% confidence intervals for the hazard ratios and have estimated a range of +/- 25% for the costs. A cost of £470 has been used for the other care costs not including bortezomib. The amended one-way sensitivity analyses are shown in Table 9. According to these analyses, the most influential variables were the TTP hazard ratio and the cost of bortezomib.

Table 9 Amended one way sensitivity analyses

Variable	Basecase	Inputs		CE ratios		Range
		Left	Right	Left	Right	
Hazard ratio - TTP	0.56	0.44	0.69	£25,339	£39,141	£13,802
Cost of Bortezomib per course	£21,035	£15,776	£26,294	£24,365	£37,136	£12,770
Duration of treatment effect, years	3	4	2	£27,957	£36,747	£8,790
Cost of other care - Bortezomib pre-progression	£470	£352	£588	£28,266	£33,892	£5,627
Hazard ratio – OS (year 1)	0.42	0.30	0.60	£29,317	£33,175	£3,858
Cost of other care - pre- and post-progression	£470	£352	£588	£29,682	£32,476	£2,795
Cost of HDD per course	£82	£103	£62	£30,725	£30,774	£50

The sensitivity analysis only considers OS hazard ratio for the first year. A sensitivity analysis was run for the OS hazard ratio for year 2 and 3 using an estimate of its 95% confidence intervals. The cost effectiveness ratios ranged from £28,782 to £32,202. A sensitivity analysis was run where each of the hazard ratios (TTP and OS) were varied in the same direction at the same time (low and high scenarios) and the cost effectiveness ratios ranged from £23,287 - £46,814. A sensitivity analysis where the cost of bortezomib varies by +/- 50% gave cost effectiveness ratio from £18,311 - £43,850.

6.5.3 Scenario Analysis

Following discussion of the base case scenario, three further scenarios analyses are presented in the manufacturer's submission:

(i) Limiting the number of cycles of treatment in non-responding patients

The number of cycles is reduced from 8 cycles to 3 cycles for non-responding patients. However, almost 15% of patients achieved a response in the trials after 3 cycles and these would have their treatment stopped. They assume that overall survival falls from 11.2 months to 9.4 months and the cost of treatment drops by £5,427. The ICER falls to £27,926 per life-year-gained.

From discussion with clinicians the ERG understands that this scenario is reflective of UK current practice.

(ii) Using the overall population of the APEX Trial

In this scenario, 40% of patients started at first relapse with the remaining 60% at second relapse and beyond. It is also possible in the model to start with any desired mix of patients.

For patients in second relapse and beyond there is a lesser benefit from bortezomib treatment. In this scenario the cost per life year gained increases to £39,954. However, this scenario uses a hazard ratio of 0.43 for the 2nd and 3rd years of treatment for overall survival (in comparison to 0.83 in the baseline scenario). If a hazard ratio of 0.83 is used in the later years, as in the base case scenario, the ICER increases to over £46,000 per life-year-gained.

Treating patients with bortezomib earlier in the treatment pathway is shown to be more cost-effective compared to later intervention. Where patients are treated at a later stage the cost per life-year-gained increases significantly. For example, where the ERG has run analyses for a scenario where all patients are treated after their second relapse the ICER is over £77,000 per life-year-gained, with an ICER of £107,000 where all patients are treated after the third relapse. This finding may be in conflict with a situation where service provision is considered in the context of budget impact (i.e. providers only being able to treat a certain number of patients, and therefore waiting until later in the treatment pathway), such as may have been the case in current practice with bortezomib used as third-line therapy.

From discussions with clinical advisors the ERG understands that current practice has been to use bortezomib for third line treatment. Should NICE recommend bortezomib for the treatment of patients with MM in line with the licence indication it may be that bortezomib is used across a patient group that varies by staging of disease/treatment (e.g. 1st 2nd, 3rd, 4th relapse), and the cost effectiveness considerations of this will be important.

(iii) Bortezomib and HDD combination therapy Vs HDD only

The scenario presented in the submission using a combination of HDD and bortezomib assumes that the response rate is increased by a factor of 1.24. The manufacturer's submission makes several assumptions in estimating the new parameters for the model and these are unclear (Appendix 15). This scenario analysis produces a cost per life-year-gained of £28,281.

The Excel spreadsheets *Velcade + Dex* and *Velcade* used different methodology to calculate transition probabilities after one year, e.g. cell S57. In particular, patients switch to other treatments regimens in *Velcade* but do not in *Velcade + Dex*.

From discussions with clinical advisors the ERG believe that the combination of bortezomib and HDD is likely to be the treatment option used in practice.

Further scenario analyses have been undertaken by the ERG, see below.

6.5.4 ERG scenario analysis

From discussions with clinical advisors the ERG considered that a combination of the three scenarios is likely to be used in practice. The ERG considered an analysis with bortezomib and HDD combination therapy vs HDD across a patient group that varies by staging of disease and treatment where the number of cycles of treatment is limited in non-responding patients. Table 10 shows the estimated cost effectiveness ratios with different patient groups.

Table 10 Cost effectiveness results for additional scenario analysis (combining scenarios 1-3)

Patient group*	Cost per life-year-gained
All patients treated at 1st relapse	£27,334
80% patients treated at 1st , 20% at 2nd relapse	£30,219
60% patients treated at 1st , 30% at 2nd , 10% at 3rd relapse	£35,783
40% patients treated at 1st , 40% at 2nd , 20% at 3rd relapse	£44,602

*Note: Intervention is bortezomib plus HDD vs. HDD alone; the number of cycles of treatment is limited in non-responding patients; mix of patients by stage of treatment.

6.5.5 Probabilistic Sensitivity Analysis

The manufacturer's model has a probabilistic sensitivity analysis (PSA) on the *Sim parameter* Excel worksheet. The PSA can be run by clicking on the 'Velcade vs HDD button' and takes about 3 minutes to run (on a computer with 2.8 GHz processor). Results are updated on the '*V v D - Sim Results'* sheet. The PSA is based around 300 simulations.

The results of the PSA are presented as an acceptability curve and a scatter-plot (cost-effectiveness plane) of the cost effectiveness results (submission, p89-90, Figures 16a & 17a). The submission reports results from PSA (for base case scenario) between £28,855 and £38,654 per life-year-gained for the 5th and 95th percentile. The manufacturer's submission provides the following information (in the Excel worksheet only) on the assumptions for selection of parameter values:

Summary of assumptions for manufacturer's PSA:

1. Hazard ratios and costs assumed to be distributed lognormal.

2. Duration of treatment effect set to probability of 35% for 2 years, 50% for 3 years, and 15% for 4 years.

3. Standard deviation of the lognormal distributions of OS and TTP hazard ratio were set so that 95th percentile equalled +0.05 from their baseline mean values.

4. Standard deviation of the lognormal distribution of costs of other care based on Bruce el al.[12]

5. Range of costs of bortezomib and HDD varied by +/- 5%.

6. The PSA did not vary policy variables (discount rate, time horizon, unit cost of bortezomib), but did vary random variables.

7. Assumed that hazard ratios on TTP and OS are correlated (70%); based on expert opinion.

Similar to the one way sensitivity analysis, the ERG considers the ranges used for the PSA to be inappropriate (e.g. where possible ranges should be related to the confidence intervals of the data). In general the ranges used should be wider. The distributions used (e.g. lognormal distribution for hazard ratios) are generally appropriate for these data, as there are no hard and fast rules. However it is common practice for the Gamma distribution to be used for cost data parameter inputs (where a lognormal distribution is used in the submission).

The ERG has undertaken further probabilistic sensitivity analysis, see below.

6.5.5.1 ERG probabilistic sensitivity analysis

The ERG has used the 95% confidence intervals for the hazard ratios and has estimated a range of +/- 25% for the costs. A cost of £470 has been used for the 'other care costs'. The baseline scenario is shown in Figure 2 and Figure 3 with more appropriate ranges for the probabilistic sensitivity analysis. The result of the probabilistic sensitivity analysis shows that the 5th percentile is £22,693 and the 95 percentile is £46,751 (cost per life-year-gained). A probabilistic sensitivity analysis where the cost of bortezomib varies by +/- 50% had a 5th percentile of £20,364 and 95th percentile of £49,876.

Figure 2. Cost-effectiveness acceptability curve from ERG Probabilistic sensitivity analyses
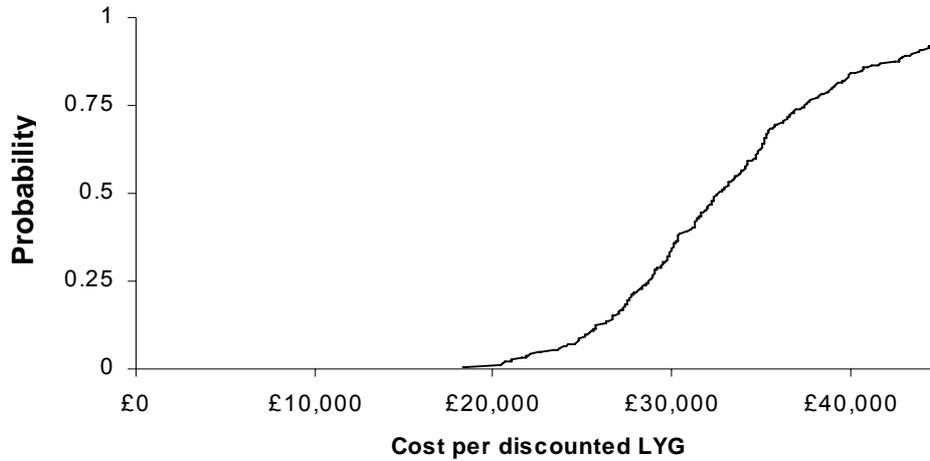
Figure 3. Cost Effectiveness Plane
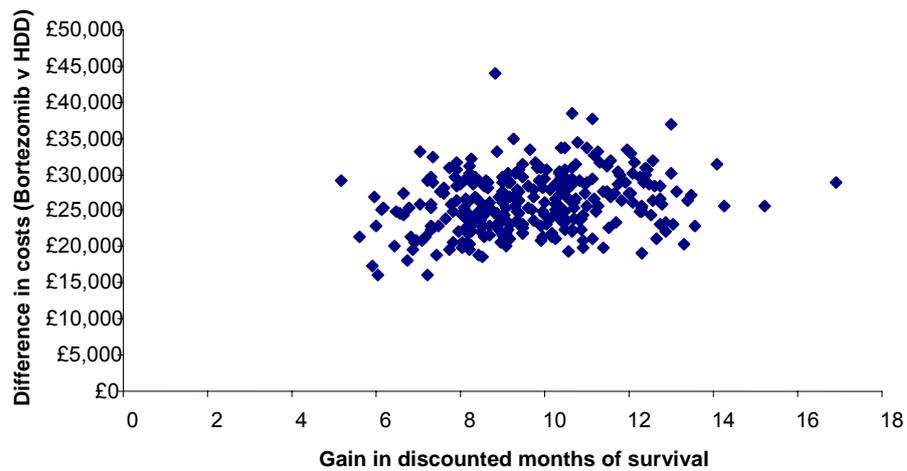
6.6 Comment on validity of results presented with reference to methodology used

In general the approach taken to model disease progression and cost effectiveness in this patient group seems reasonable. A number of concerns have been raised above by the ERG, some general and some specific. The ERG has not undertaken systematic searches in this area therefore the comments made are for broader consideration by the NICE Appraisal Committee in the context of this STA.

6.7 Summary of uncertainties and issues

In general the approach taken to model disease progression and cost effectiveness in this patient group seems reasonable. A number of concerns have been raised above by the ERG, some general and some specific:

– The manufacturer's submission uses data from the APEX RCT and Mayo observational study.[11] The ERG has concerns on the generalisability of these data to UK treatment practice.

– No patient related measures of health-related quality-of-life have been discussed or applied in the manufacturer's submission and the model outcomes are reported in terms of cost per life-year-gained. Mulitple myeloma patients generally have a relatively poor quality of life (compared to age matched controls) and the ERG suggest that this should be incorporated in the manufacturer's model. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.*

– The ERG has identified several mistakes in the Excel spreadsheet submitted. Furthermore the results from the model may overestimate the treatment effect shown in the APEX trial. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.*

– Treatment cost is a key issue in the cost effectiveness estimates, and the ERG note that (i) the cost for bortezomib will vary by responder status (number of cycles), (ii) cost savings may be possible where group administration of bortezomib is an option.

– The sensitivity analyses undertaken are limited, and there may be a greater variability in the cost effectiveness of treatment than presented in the manufacturer's submission

– Adverse events have not been included in the manufacturer's model, either in terms of loss of quality of life or increased resource use.

In view of the points raised above, the ERG suggest that the cost effectiveness results presented may underestimate the cost per life-year-gained for bortezomib therapy (vs. HDD), where potential UK practice and scenarios are considered.

7 Discussion

7.1 Summary of clinical effectiveness issues

The manufacturer's submission to NICE includes a systematic review of the clinical effectiveness literature and narrative reporting of the RCT that met the inclusion criteria of the review (APEX trial).

The ERG has two main areas of concern relating to clinical effectiveness issues in the manufacturer's submission; firstly, the limited evidence base and its relevance to the NHS, and secondly, the interpretation of the included RCT. Whilst the submitted evidence generally reflects the decision problem as defined in the manufacturer's submission, it is not totally representative of current clinical practice in the UK in either the use of bortezomib or the comparator. Due to the nature of the disease and patient population no standard treatment exists for MM patients at first relapse and beyond, which makes comparisons difficult and limits available evidence. The submitted evidence consists of the only published RCT concerning bortezomib and as such may be helpful for answering some questions concerning myeloma treatment that will impact on the NHS.

The manufacturer's submission states that bortezomib increases survival and time to disease progression compared with HDD. Although this conclusion is probably valid, the size of the treatment effect of bortezomib is not known. There is uncertainty about the results and the statistical methods used in the analysis because of lack of details and clarity in reporting patient numbers and the intervention they received at any time, and because point estimates and confidence intervals are not presented consistently.

7.2 Summary of cost effectiveness issues

The manufacturer's submission to NICE includes a report on the cost effectiveness literature, reference to an early economic analysis undertaken using individual patient-level trial data and an economic evaluation using a decision-analytic model.

The state transition model predicts treatment experience, modelling the flow of patients through a series of treatment regimens. The clinical effectiveness data from the APEX RCT,

showing relative benefits in time to treatment progression and overall survival, are applied to the baseline prediction for HDD patients to simulate the treatment effect from bortezomib.

In general the approach taken to model disease progression and cost effectiveness in this patient group seems reasonable. A number of concerns have been raised above by the ERG, some general and some specific. The ERG has concerns on the generalisability of the data used to UK treatment practice. No patient related measures of health-related quality-of-life have been discussed or applied in the manufacturer's submission and the model outcomes are reported in terms of cost per life-year-gained. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.* The ERG has identified several mistakes in the Excel spreadsheet submitted. *Note: See Addendum to this ERG Report, where the manufacturer provides additional information on this point.* Furthermore the results from the model may overestimate the treatment effect shown in the APEX trial. Treatment cost is a key issue in the cost effectiveness estimates and the cost for bortezomib will vary by responder status and method of administration. The sensitivity analyses undertaken are limited, and there may be a greater variability in the cost effectiveness of treatment than presented in the manufacturer's submission. Adverse events have not been included in the manufacturer's model, either in terms of loss of quality of life or increased resource use.

In view of the points raised above, the ERG suggest that the cost effectiveness results presented may underestimated the cost per life-year-gained for treatment with bortezomib compared to HDD, where potential UK practice and scenarios are considered.

Reference List

(1) Smith A, Wisloff F, Samson D, on behalf of the UK Myeloma Forum NMSGaBCfSiH. Guidelines on the diagnosis and management of multiple myeloma 2005. British Journal of Haematology 2005; 132:410-451.

(2) Sirohi B, Powles R. Multiple myeloma. The Lancet 2004; 363(9412):875-887.

(3) Prince M, Adena M, Smith DK, Hertel J. Efficacy of single-agent bortezomib versus thalidomide in patients with relapsed or refractory multiple myeloma: A systematic review. Blood 2005; 106(11):5160.

(4) Richardson PG, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. New England Journal of Medicine 2005; 352(24):2487-2498.

(5) Lee SJ, Richardson PG, Sonneveld P, Schuster M, Irwin D, Massaro J et al. Health-related quality of life (HRQL) associated with bortezomib compared with high-dose dexamethasone in relapsed multiple myeloma (MM): Results from APEX study. Journal of Clinical Oncology 2005; 23(16):568S.

(6) Richardson P, Sonneveld P, Schuster M, Irwin D, Stadtmauer E, Facon T et al. Bortezomib demonstrates superior efficacy to high-dose dexamethasone in relapsed multiple myeloma: Final report of the APEX study. Blood 2004; 104(11):366.

(7) Richardson P, Sonneveld P, Schuster MW, Irwin D, Stadtmauer EA, Facon T et al. Bortezomib vs. dexamethasone in relapsed multiple myeloma: A phase 3 randomized study. Journal of Clinical Oncology 2004; 22(14):6511.

(8) Richardson P, Sonneveld P, Schuster M, Irwin D, Stadtmauer E, Facon T et al. Bortezomib continues demonstrates superior efficacy compared with high-dose dexamethasone in relapsed multiple myeloma: Updated results of the APEX trail. Blood 2005; 106(11):2547.

(9) Sonneveld P, van Oostrum NHM, Schoester M, van Oosterhoud S. PS341 (Bortezomib) and the histone deacetylase inhibitor LAQ824 sensitize Melphalan resistant Multiple Myeloma cells to Melphalan. Blood 2003; 102(11):660.

(10) Zangari M, Esseltine D, Lee CK, Barlogie B, Elice F, Burns MJ et al. Response to bortezomib is associated to osteoblastic activation in patients with multiple myeloma. British Journal of Haematology 2005; 131(1):71-73.

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

(12) Bruce N, McCloskey E, Kanis J, Guest J. Economic impact of using clodronate in the management of patients with multiple myelmoa. British Journal of Haematology 1999; 104(2):358-364.

(13) McCloskey EV, MacLennan ICM, Drayson MT, Chapman C, Dunn J, Kanis JA. A randomised trial of the effect of clodronate on skeletal morbidity in multiple myeloma. British Journal of Haematology 1998; 100:317-325.

(14)   Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 1997.

(15)   Philips Z, Ginelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Asses 2004; 9(36).

(16)   Gulbrandsen N, Wisloff F, Nord E, Lenhoff S, Hjorth M, Westin J et al. 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. European Journal of Haematology 2001; 66(5):328-336.

(17)   van Agthoven M, Segeren CM, Buijt I, Uyl-de Groot CA, van der HB, 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(8):1159-1169.

(18)   Segeren CM, Sonneveld P, van der Holt B, et al. Overall and event-free survival are not improved by the use of myeloablative therapy following intensified chemotherapy in previously untreated patients with multiple myeloma: a prospective randomised phase 3 study. Blood 2003; 101:2144-2151.

(19)   Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. British Medical Journal 1998; 316:736-741.

(20)   Joint Formulary Committee. British National Formulary. 48. 2004. London, British Medical Association and Royal Pharmaceutical Society of Great Britain.

(21)   Department of Health E. NHS reference costs. Department of Health, England, UK [2005 [cited 2006 Mar. 6]; Available from: URL:http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/FinanceAndPlanning/NHSReferenceCosts/fs/en

(22)   Richardson PG, Barlogie B, Berenson J, Singhal S, Jagannath S, Irwin D et al. A phase 2 study of bortezomib in relapsed, refractory myeloma. New England Journal of Medicine 2003; 348(26):2609-2617.

(23)   Netten A, Curtis L. Unit Costs of Health and Social Care, PSSRU, University of Kent. 1-213. 2005. University of Kent, PSSRU.

Addendum: Manufacturer's response to questions raised by the ERG on the manufacturer's submission to NICE

JANSSEN-CILAG Ltd

Mr Meindert Boysen
Associate Director – Single Technology Appraisals
Centre for Health Technology Evaluation
National Institute for Health and Clinical Excellence
MidCity Place
71 High Holburn
London
WC1V 6NA

12th April 2006

Dear Mr Boysen,

RE: SINGLE TECHNOLOGY APPRAISAL – VELCADE® (BORTEZOMIB) FOR MULTIPLE MYELOMA

Thank you for providing Ortho Biotech, a Biopharmaceuticals Division of Janssen-Cilag with an opportunity to respond to the issues and queries that were raised by the SHTAC Review Group. The structured critique provided by SHTAC was very helpful in laying out those areas of our submission that required further clarification. We have addressed each of these comments in turn in the attached document.

Ortho Biotech particularly wishes to draw to the Appraisal Committee's attention our response to point B1 in which we were requested to undertake a cost per QALY analysis using our model. Ortho Biotech did not originally provide a QALY analysis because of concerns over its validity and relevance to this appraisal. However, in response to this request, we have now provided the analysis, but our concerns with its inherent validity remain. We would therefore request that the issues we raise in our response to point B1 are given due consideration in the interpretation of this cost per QALY value at the Appraisal Committee meeting.

We acknowledge that health-related quality of life (HRQL) is an important consideration in the evaluation of a new technology for the treatment of multiple myeloma (MM) and recognise the desire of SHTAC and NICE to derive a cost per QALY figure. Indeed, in the APEX trial, the beneficial impact of Velcade on disease-specific quality of life was clearly demonstrated despite the severe limitations on the available data imposed by early termination of the study.

EQ5D was a secondary outcome measure in the APEX trial; however, the completion of this instrument was poor. Coupled with problems caused by early termination of the study, this means that the utility estimates are highly unreliable and could not be used in our economic model.

As well as the methodological issues associated with the APEX data, we do not believe that the cost-effectiveness of VELCADE within the treatment of MM should be dependent solely upon a cost per QALY analysis. Our concerns are outlined as follows:

1. *Survival is the single most important outcome for this aging patient group.* The results of the APEX trial have demonstrated a significant survival advantage

1

compared to HDD. Without treatment 40% of MM patients would be dead within one year and approximately 75% within five years, this incremental survival benefit is highly clinically significant.

2. *There is a lack of robust utility data in relapsed MM patients.* Estimates obtained reflect the utility of patients at an earlier stage of treatment (newly diagnosed) and are unable to capture the treatment specific utility associated with treatment as well as accurately reflect the utility at first relapse and beyond.

3. *The EQ5D is not sensitive to some important facets of MM.* Patient focus group research indicates that the EQ-5D fails to consider the primary drivers of quality of life in this patient population. The over-riding predictor of QoL is their level of anxiety over their paraprotein levels, which patients know predicts future prognosis. This subtle issue will not be captured in generic multi-attribute utility scales such as the EQ-5D. This must be considered when reviewing the estimates of utility from the published literature as these are based on utility values obtained from the EQ5D.

However, despite these reservations discussed above we have responded to the request by SHTAC and NICE's and derived an estimate of the cost per QALY estimate. We have tried to base this estimate on the least flawed of the published utility data and based this on data from van Agthoven et al 2004. These results are presented in the attached document.

Furthermore, VELCADE is the first new treatment to have been licensed for use in this patient group in 10 years. The lack of new product introductions in this cancer are striking when contrasted with extensively researched and resourced cancers such as breast. This lack of innovation has created a high level of unmet medical need and yet, perversely, lack of innovation also means that new innovative treatments are referenced against extremely low, generic cost alternatives. The challenges of meeting cost-effectiveness thresholds in areas of high unmet need with cheap generic standard agents are therefore greater than in areas where modern technologies are accepted standards of care.

Please do not hesitate to contact me should you require further information.

Yours Sincerely

Angela Christie

Miss Angela Christie
Senior Outcomes Research Manager
Ortho Biotech, a Biopharmaceuticals Division of Janssen-Cilag

2

JANSSEN-CILAG'S RESPONSE TO SHTAC REVIEW GROUP COMMENTS

<u>Section A. Clarification on effectiveness data</u>

A1. Please provide for the APEX trial confirmation of:
- **Group size at 8.3 months;**
- **How many patients crossed over;**
- **When patients crossed over;**
- **How this was accounted for in the analysis.**
- A flowchart specifying numbers of patients receiving which treatment at particular time points would be helpful.

When did HDD patients cross over to VELCADE? The NEJM publication states that patients were allowed to cross-over AFTER interim analysis (median time 8.3 months), and this is also suggested by the flow chart on page 35. However, the industry submission (p.46) states that "The design of APEX allowed for any HDD patient experiencing disease progression to be offered VELCADE therapy. At interim analysis (median follow-up 8.3 months), 44% of HDD patients had crossed over to receive VELCADE."

There was a companion non-comparative trial M34101-040 in which patients randomised to the high-dose dexamethasone (HDD) arm in the APEX trial (M34101-039) could receive VELCADE as a subsequent therapy line after progression. In the patient flow chart presented on P35 of our submission it states that 158 patients from the HDD arm crossed over to VELCADE within the companion study. (152 patients after obtaining progression and before the interim analysis (8.3 months) (45%) and 6 patients had crossed over before the cutoff date (January 2004). After the interim analysis the IDMC recommended to offer all patients on HDD (whether they had progressed or not) the possibility to switch over to 040.

A2. Please supply further information on your statistical analyses of the difference in time to disease progression between treatment groups in the APEX trial.
A total of 60 (9%) of the 669 patients randomized into the study were determined to be refractory to dexamethasone, including 32 (10%) of 333 VELCADE patients and 28 (8%) of 336 HHD patients. The NEJM paper states that post-hoc analysis without these patients still found significant benefits for VELCADE. However, power calculation states that: "a sample size of 310 patients per treatment group provided 80% power to detect a 30% difference in the time to disease progression between the two groups." Therefore, removing the refractory patients would bring the groups size to a level (slightly) below that required for statistical power.

For a given significance level, statistical power is the probability to detect the significant treatment difference of a trial. At the design stage we need to plan enough sample size to make sure of having adequate probability (power) for the detection of a positive finding. The power calculation at the design stage is dependent on the projected treatment effect size. With a given sample size and a fixed significance level (e.g., alpha=0.05), a higher projected treatment effect corresponds to a higher power. However, when the study is done, the actual treatment effect size is usually different

3

from what was projected at the design stage, it could be either higher or lower. In this study, 80% power is calculated based on the projected effect size of 30% improvement in TTP. However, the actual effect size of the VELCADE in this trial is much higher, > 70% improvement in TTP. With this higher treatment effect observed in the trial, the actual study power is still > 80% even the statistical test is just based on only 90% of the planned number of patients.

A3. Please state which is correct: Increase in TPP for patients in the VELCADE arm of the APEX study at 8.3 months is reported as 58% on p45 and 78% on p47 (and in the NEJM paper).

Increase in TTP for patients in the VELCADE arm is 78% at 8.3 months.

A4. Please provide the washout period from the previous treatments in the APEX trial? Nearly 50% of patients had previous thalidomide therapy, which might be considered a comparator

In answer to your question we have undertaken an additional analysis. Please find the results in the table.

Months from end of last line (Thalidomide-based) to randomisation	VELCADE	HDD
Number of patients	131	135
Mean (SD)	4.5 (5.59)	4.7 (5.00)
Median	3	3
Range	1-37	1-25
Category, n (%)		
0-3 months	91 (69)	88 (65)
4-6 months	16 (12)	14 (10)
6-12 months	14 (11)	18 (13)
>12 months	10 (8)	15 (11)

A5. Please provide more information on the EORTC QLQ-C30 questionnaire data was collected in the APEX study, but not presented in detail in the industry submission. Table 14 lists P values but does not give actual data.

Please find attached the abstract and presentation for this analysis, which was presented at the annual meeting of the American Society of Clinical Oncology 2005. This contains all the information on the specifics of the analysis.

A6. Please provide a clearer indication of what is mean by different definitions used for adverse events: the submission discusses grade 3 infections rather than grade 3 events in the APEX study. The table reports grade 3 events, not grade 3 infections.

Within our submission document in p62 we report the adverse events during treatment reported by 15% or more of patients receiving VELCADE or HDD, including grade 3 and 4 adverse events at 8.3 months follow-up. On p63 we then report the results of the

4

planned secondary endpoints namely ≥Grade 3 infection and time to first skeletal event. These disease-related parameters were analyzed for the purposes of determining clinical benefit beyond the potential prolongation of TTP:

The following definition of ≥Grade 3 infection was applied within the APEX trial:

• ≥Grade 3 infection: any Grade 3 or 4 infection from randomisation to the End of Treatment visit (a minimum of 30 days after the last dose).

The rate of Grade 3 or 4 infection was analysed using Fisher's Exact Test, with the test performed on treatment group differences in the rates of patients with 1 or more infections. In addition, supportive analyses on the time to first ≥Grade 3 infection was performed using survival analysis methods (specifically, the log-rank test with an anticipated time of analysis approximately 14 months after the study start).

A7. Please clarify: what does NE refer to in table 10?

Not evaluable.

A8.Please clarify the following: the response to question 49 of the industry submission states that observers were kept fully blinded, but also refers to the response to question 41 that the trial was not blinded. Many outcomes were subjective measures and may be open to bias. Please clarify whether or not anone was actually blinded in the APEX study

APEX was an open label trial but it must be noted that all efficacy assessments are based on objective laboratory data (for example SPEP, UPEP, bone marrow measurements, serum calcium) or radiology data. Furthermore, analysis of progressive disease for each patient was undertaken using a computer algorithm based on the independent EBMT objective response criteria.

Also an independent data monitoring committee (IDMC) was set up which included physicians with expertise and experience in the diagnosis and management of multiple myeloma but without direct involvement in the conduct of the study. The IDC members performed an independent majority review of selected data collected in this study in order to assess each patient's disease response, including PD, based on the SWOG+ criteria. The review was independent of the investigators' disease response assessments performed during the conduct of the study to appropriately manage patients.

The following data was provided to the IDMC for disease response assessment:
• Serum protein electrophoresis results (with quantitation of M-protein and immunoglobulins, and immunofixation).
• Urine protein electrophoresis results (with quantitation of M-protein and immunofixation).
• Selected hematology and clinical chemistry results, as needed.

Efficacy analyses were based primarily on the results of the IDMC's independent majority disease response assessment. The IDMC members recorded assessments by cycle and an overall assessment on study, but did not evaluate follow-up information; they did review assessments through development of PD or the start of alternative

5

therapy. Of most importance, the IDMC did not be provided with the identification of the patient or the patient's study drug assignment (VELCADE or HDD).

Efficacy analysis of the potential superiority of VELCADE relative to HDD in disease response, defined as a complete plus partial response, were based on the results of the IDMC's independent majority disease response assessment.

A10. **Please provide some details** of the Independent Review Committee, e.g., description, qualifications, membership.

An independent data monitoring committee (IDMC) was formed in order to evaluate safety data during the course of the APEX study and both efficacy and safety at the pre-specified interim analysis (8.3 months follow-up).

The members of the IDMC were as follows:
Professor Gösta Gahrton - DMC chairperson, Oncology
Michael Grossbard, MD - DMC member, Oncology
Giuseppe Giaccone, MD, PhD - DMC member, Oncology
David Harrington, PhD - DMC member, Statistics
Michael Seiden, MD, PhD - DMC member, Oncology

In addition, Burt Seibert, PhD served as the reporting statistician. Dr. Seibert is an independent consultant who is not affiliated with the APEX study. His role as a reporting statistician was to be responsible for preparing interim results prior to each meeting and storing minutes from closed meeting sessions (until after the final database lock or in the event that the IDMC recommends the termination or significant modification to the conduct of the study).

IDMC members had were free of conflicts of interest to ensure that they were not substantially affected by the outcome of the trial. This included professional conflicts of interest, financial conflicts of interest (in accordance with the Code of Federal Regulations (CFR), financial disclosure forms was obtained by the IDMC members and will be sent to the Food and Drug Administration), and intellectual conflicts of interest, which might influence the objective interpretations of the data. Investigators participating in the trial were not IDMC members to avoid any possible influence of knowledge of interim results on their conduct of the trial. During the trial, IDMC members were responsible for disclosing to the IDMC Chairperson and to the Company potential conflicts of interest that arise relating to the study. If an IDMC members developed real or perceived conflicts of interest that impact objectivity they were removed from the IDMC and replaced.

IDMC SCOPE OF RESPONSIBILITIES

The IDMC was responsible for reviewing accumulating data from the M34101-039 study on an intermittent basis. Based on this review, the IDMC advised the company regarding the continuing safety of current participants in the study and those yet to be recruited, as well as the continuing validity and scientific merit of the trial. The IDMC advised the Company whether the trial should continue unchanged, be modified, or be terminated due to safety concerns or observing superior efficacy in either the VELCADE™ arm or

6

the HDD arm. In the interest of providing timely information, data that were reviewed included data that had not yet been fully processed and validated by data management.

The duration of the IDMC membership covered the duration of the clinical trial including the production of the final clinical study report. The IDMC was formally dissolved upon completion of the final clinical study report.

A11. **Please supply PDFs**: Appendix 7: SUMMIT and CREST publications-these do not open.

Please find these attached.

A12. Outcomes were analysed using a 'modified ITT analysis', **can more information on this be provided?**

We would like to clarify that a 'Modified ITT analysis' applies only to the Quality of life analysis undertaken within the trial where 44 of 642 patients were excluded due either to absent baseline data or lack of follow-up QoL data. All other efficacy analysis was conducted using a complete ITT analysis.

A13. It would be useful to have an indication of the number of studies identified by each line of the search strategy.

Tables 1-3 below present these results.

Table 1: MedLine Search Strategy

No	SEARCH STRATEGY	NUMBERS RETRIEVED
1	Multiple-myeloma.DE. OR plasmacytoma#.DE. OR paraproteinemias#.DE.	34770
2	Myelom$ OR multiple ADJ myeloma OR plasmacytom$ OR plasmocytom$ OR MGUS OR monoclonal ADJ gammopathy ADJ of ADJ undetermined ADJ significance	47697
3	1 OR 2	54561
4	Bortezomib OR VELCADE OR ps341 OR ps-341 OR ps ADJ '341' OR proteasome ADJ inhibit$5	2617
5	Randomized-controlled-trial.PT. OR comparative-study.PT. OR meta-analysis.PT. OR controlled-clinical-trial.PT.	288864
6	Controlled-clinical-trials#.DE. OR randomized-controlled-trials#.DE. OR double-blind-method.DE. OR control-groups.DE. OR cross-over-studies.DE. OR meta-analysis.DE. OR random-allocation.DE. OR single-blind-method.DE.	191101
7	(intervention OR clinical$) NEAR (trial$ OR study OR studies) AND (random$ OR placebo$ OR rct$ OR (control OR controls OR controlled) WITH (trial$ OR study OR studies) OR (cross ADJ over OR crossover OR parallel OR prospectiv$) WITH (trial$ OR study OR studies) OR (singl$ OR doubl$ OR trebl$ OR tripl$) WITH (blind$ OR Mask$))	129066

7

8	5 OR 6 OR 7	423002
9	3 AND 4 AND 8	22
10	9 LG=EN AND HUMAN=YES	16

Table 2: MedLine-In-Process Search Results

No	SEARCH STRATEGY	NUMBERS RETRIEVED
1	(Multiple-myeloma.DE. OR plasmacytoma#.DE. OR paraproteinemias#.DE.) OR Myelom$ OR multiple ADJ myeloma OR plasmacytom$ OR plasmocytom$ OR MGUS OR monoclonal ADJ gammopathy ADJ of ADJ undetermined ADJ significance	793
2	Bortezomib OR VELCADE OR ps341 OR ps-341 OR ps ADJ '341' OR proteasome ADJ inhibit$5	248
3	Randomized-controlled-trial.PT. OR comparative-study.PT. OR meta-analysis.PT. OR controlled-clinical-trial.PT.	1322
4	Controlled-clinical-trials#.DE. OR randomized-controlled-trials#.DE. OR double-blind-method.DE. OR control-groups.DE. OR cross-over-studies.DE. OR meta-analysis.DE. OR random-allocation.DE. OR single-blind-method.DE.	0
5	(intervention OR clinical$) NEAR (trial$ OR study OR studies) AND (random$ OR placebo$ OR rct$ OR (control OR controls OR controlled) WITH (trial$ OR study OR studies) OR (cross ADJ over OR crossover OR parallel OR prospectiv$) WITH (trial$ OR study OR studies) OR (singl$ OR doubl$ OR trebl$ OR tripl$) WITH (blind$ OR Mask$))	3872
6	3 OR 4 OR 5	4979
7	1 AND 2 AND 6 AND LG=EN AND HUMAN=YES	4

Table 3: EmBase Search Strategy

ID	SEARCH STRATEGY	NUMBERS RETRIEVED
1	Multiple-myeloma.DE. OR malignant-plasmacytoma#.DE. OR plasmacytoma.DE. OR myeloma.DE. OR myeloma-cell.DE. OR monoclonal-immunoglobulinemia.DE. OR paraproteinemia#.DE.	39167
2	Myelom$ OR multiple ADJ myeloma OR plasmacytom$ OR plasmocytom$ OR MGUS OR monoclonal ADJ gammopathy ADJ of ADJ undetermined ADJ significance	34204
3	1 OR 2	51006
4	Bortezomib.de.	951
5	Bortezomib OR VELCADE OR ps341 OR ps-341 OR ps ADJ '341' OR proteasome ADJ inhibit$5	1550
6	4 OR 5	1550

8

7	Randomized-controlled-trial.DE. OR comparative-study#.DE. OR drug-comparison#.DE. OR controlled-study#.DE. OR randomization#.DE. OR double-blind-procedure.DE. OR single-blind-procedure.DE. OR triple-blind-procedure.DE. OR parallel-design.DE. OR crossover-procedure.DE. OR meta-analysis#.DE.	2449064
8	(intervention OR clinical$) NEAR (trial$ OR study OR studies) AND (random$ OR placebo$ OR rct$ OR (control OR controls OR controlled) WITH (trial$ OR study OR studies) OR (cross ADJ over OR crossover OR parallel OR prospectiv$) WITH (trial$ OR study OR studies) OR (singl$ OR doubl$ OR trebl$ OR tripl$) WITH (blind$ OR Mask$))	663130
11	7 OR 8	2565269
12	3 AND 6 AND 11 AND LG=EN AND HUMAN=YES	238

A14. The table of criteria includes criteria that the study should include more than 20 patients, why? This is not in other stated criteria.

We have repeated the literature search strategy with this criterion removed and can confirm that this criterion does not change the results included as part of our submission. There were no RCTs retrieved that met the scope of the submission that enrolled less than 20 patients.

Section B: Economic Analysis

B1. Please incorporate a HRQL (QALY analysis) into the model to estimate cost per QALY.

Health state values have not been incorporated into the CEA. The submission states that the manufacturer believes it inappropriate to use the EQ5D in the analysis. However, the case is not made for ignoring the issue of HRQL (QALY values). In fact the submission highlights that HRQL is important in this patient group and that multiple myeloma is likely to have a big impact on HRQL.

We believe that life-years gained rather than QALYs are the most appropriate measure of effectiveness to use in the model. There are a number of important reasons to justify our deviation from the NICE reference case and these are discussed below.

The incremental survival advantage with VELCADE in the model was 9.9 months

The results of the APEX trial have demonstrated a significant survival advantage (mean overall survival of 9.9 months) compared to HDD. Given that without treatment 40% of myeloma patients would be dead within 1 year and 75% within 5 years, this incremental survival benefit in an aged population is highly clinically significant.

The Robustness of Existing Utility Measures Is Questionable, Which Introduces Considerable Uncertainty into the Precision of Cost per QALY Estimates.

9

As well as these limitations, there are a number of methodological issues that informed our decision not to calculate a cost per QALY in this submission.

The APEX trial included the EQ-5D as a secondary outcome measure. Early termination of this trial, crossover of patients from HDD to VELCADE and poor rates of questionnaire completion at some sites meant that these data were highly unreliable and could not be used to estimate utility gain within the model.

Although these significant problems existed with respect to collection of robust utility estimates, we recognise the desire of SHTAC and NICE to derive a cost per QALY figure. We have therefore undertaken an additional scenario analysis based on published utility data, which is outlined below. We must emphasise that these values should be considered as an approximate estimate, rather than a definitive cost per QALY value.

Literature Review

In developing our original model, we undertook a literature review of utility values in myeloma. We identified three published studies that provided possible utility estimates.

The three published studies were undertaken in newly diagnosed myeloma patients and were conducted alongside clinical trials. In two of the studies, utilities were obtained by a mapping exercise from disease-specific quality of life instruments (Nord et al 1997, Gulbrandson et al 2003) while the other (van Agthoven, et al 2004) used the EQ-5D.

A summary of three studies is provided in Table 1. The range of utility values obtained using the EuroQol EQ-5D for other chemotherapy treatments ranged from 0.81 (van Agthoven et al 2004 at 6 and 18 months for intensive chemotherapy) to 0.65 (Nord et al 1997 at 6 months for melphalan and prednisone). The van Agthoven study also confirmed that utility values show little decline over time from 6 to 24 months.

Table 1: Summary of the identified utility studies in multiple myeloma

Study feature	Study		
	Nord et al 1997	Gulbrandsen et al 2001	van Agthoven et al 2004
Country, N	Scandinavia N = 583	Scandinavia N = 344	Netherlands N=261
Disease status	Newly diagnosed, symptomatic	Newly diagnosed, symptomatic	Newly diagnosed, symptomatic
Patient demographics	Age: median 67/68 years Gender: 56%/60% male	Age: median 51/54 years Gender: 65%/56% male	Age: median 55/56 years Gender: 57%/61%
Study design	Multicentre RCT	Multicentre non-randomised trial with matched control group	Multicentre RCT
Treatments, n	MP, n = NR MP + αINF, n = NR	HDT/SCT, n = 274 Oral MP, n = 70 [a]	Intensive chemo, n =129 HDT/SCT, n = 132
Health value methodology(s)	EORTC QLQ-C30 mapped to EuroQol Index, IHQL and 15-D	EORTC QLQ-C30 mapped to 15-D	EuroQol Index

10

Reporting of utilities	6 months	6 months	6, 12, 18 and 24 months
Utilities available for multiple myeloma disease stages	No	No	Yes, for responders. Assumptions made to obtain utility of 0.644 for non-responders

Utility values	EuroQol
	MP 0.65
	MP+ αINF 0.65
	IHQL
	MP 0.70-0.80
	MP+ αINF 0.60-0.70
	15-D
	MP 0.60-0.70
	MP+ αINF 0.60-0.65
		15-D
		HDM/SCT 0.7334
		MP 0.7896
			EuroQol for responders
			6m Chemo 0.81
			     HDT/SCT 0.65
			12m Chemo 0.80
			     HDT/SCT 0.62
			18m Chemo 0.81
			     HDT/SCT 0.69
			24m Chemo 0.77
			     HDT/SCT 0.75

RCT = randomised controlled trial, MP = oral melphalan and prednisone, HDM/SCT = high dose melphalan and stem cell transplant, αINF = subcutaneous interferon α2b, EuroQol Index = time trade off values, IHQL = Index of Health-Related Quality of Life (uses standard gamble values), 15-D = 15 dimensional scale using rating scale values, m = months, NR = not reported.
[a] Subgroup of patients from Nord et al 1997.

Of these studies, the van Agthoven study is most relevant to this appraisal because it is the only one of the three studies that reports utility values according to responder rate (an important consideration in our modeling work) and because it uses a direct valuation technique rather than a less well accepted, indirect mapping exercise.

To provide NICE and SHTAC with an estimate of the cost per QALY, we have incorporated the values for responders and non-responders into our model. The results are presented in the table below.

Outcomes	VELCADE	HDD
Discounted survival gain (months)	32.6	22.7
QALY gain (months)	26.35	18.33
Total treatment cost (£)	36,566	10,890
Cost per QALY (£)	38, 452	-
PSA Range (£)	33,992 – 49,894	

Interpretation of the Cost per QALY Analysis

This additional analysis suggests that the Incremental Cost per QALY with VELCADE compared to HDD is in the range of £33,992 to £49,894. In interpreting this value, we would like to draw attention to two other critically important points.

11

Current methods for deriving QALY weightings fail to reflect the attributes of importance to multiple myeloma patients

The problem with quality adjusting the additional survival gained with VELCADE using a generalized multi-attribute utility scale (MAUS) such as the EQ-5D is that it fundamentally fails to capture and quantify the key issues that impact myeloma patients' quality of life.

In preparing for this NICE submission, and before developing the economic model, we commissioned a Professor of Psychology to run a focus group with myeloma patients. The objective of this focus group was to understand the issues affecting the quality of life of myeloma patients and to evaluate whether the EQ-5D instrument (which was used in the APEX trial) adequately represented these issues.

This focus group raised some important issues, which seriously calls into question the construct validity and responsiveness of the EQ-5D in this patient population. The most important were that:

- The single most important determinant of QoL is the paraprotein level because patients understand this is a prognostic factor for relapse and survival.
- Myeloma patients show a high degree of psychological adaptation to the co-morbidities associated with the condition, showing that they learn to live with the physical limitations of the condition and see them as being less important than the laboratory results, which predict survival.
- Myeloma patients did not consider that the EQ-5D adequately described the key drivers of their QoL.

In summary, we believe that the incremental cost per QALY number presented here is of limited importance to this appraisal. The key result is an incremental survival advantage of 9.9 months at a cost per life-year gained of £31k.

B2. Please provide further clarification on the stated validation and calibration of the model.

The model used the Mayo clinic data and was validated against this for overall mortality (p241). The submission claims that the model has survival of 72% at 1yr, 61% at 2yr and 25% at 5 yr compared to Mayo study of 72% at 1yr, 55% at 2 yr and 22% at 5 year. However, SHTAC was not able to replicate these results. *By not using a reduced mortality risk in the first year of 21% we found the results in the model to be 64% in the 1st year, 39% in the 2nd year and 10% in the 3rd year.* The model was calibrated against the trial arm of the APEX trial (p242) by reducing the mortality risk while on a 2nd regimen in the first year by 21%. However, again this is the case – model results show 87% and 72% 1yr survival for VELCADE and HDD compared to 80% and 66% as reported in Richardson et al.

A more detailed explanation of the validation can be found in the following steps described below:

12

Step 1: Check goodness of fit with Kumar et al data. We applied this technique across all intervals and then compared the goodness of fit with data presented in Kumar et al on overall survival from the time of diagnosis.
set cell d16 = e16 = f16 in 'Summary sheet' to 100% and cells d17-f20 to 0%; that is, start with all subjects at 1st regimen.
Set all cells K57-K61 and K71-K75 to empirically derived estimates (eg K61=f61/g61).

FIGURE 1. Overall survival from diagnosis of multiple myeloma for the study group of 579 patients. The overall survival at 1 year was 72%, at 2 years was 55%, and at 5 years was 22%. The median survival for the study group is 28.4 months (95% confidence interval. 25.0-30.6 months).
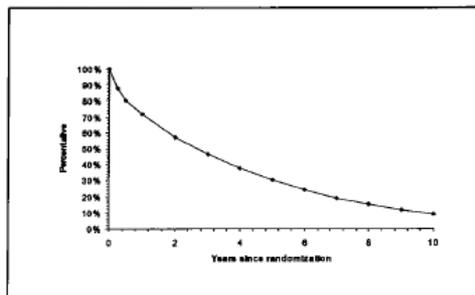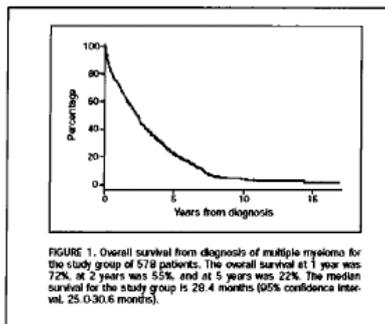
Under these assumptions, the model over-estimates survival at 2 and 5 years – 57% and 30% respectively (See L9 and L10 of 'Summary spreadsheet').

Step 2: Recalibrate transition probabilities to improve goodness of fit. We believe that this over-estimation stems from the limited numbers of patients alive over time in the Mayo study, and especially in later regimens. Therefore, we recalibrated the model to provide a more precise match of the Kumar et al data. Specifically, we reduced the rates of staying on regimens 3-6 in the intervals from 13-24 months and 26-36 months. The resulting model then identically replicated the results from Kumar et al (55% at year 2 and 22% at year5).

Step 3: Contrast results with HDD arm of APEX trial. We compared the survival results from this literature-based model with results reported in the APEX study. In Apex, 40% of patients had only one prior therapy and 60% had more than one prior therapy. Applying this distribution of prior therapies, we predict 1- and 2- survival equal to 60% and 31% respectively. By contrast, APEX reported 1- and 2- year survival in the comparator arm (HDD followed by VELCADE in >60% of subjects) equal to 66% and 47%. The following tables compares 1-, 2- and 5-year survival for the 'calibrated' Mayo model versus findings reported for APEX in the HDD arm, >60% of whom were able to switch to VELCADE after progression.

Year	Percent alive	
	HDD only	HDD - > VELCADE **
1	60%	66%
2	33%	47%
5	5%	-

There are several possible explanations for why survival was higher in APEX than in the data derived from a community-based study. First, the data from Kumar et al was

13

obtained from patients diagnosed between 1985 and 1998. Thus survival may have improved over time with the advent of other treatments being available. However, it is interesting that Kumar et al do not report a steady improvement over time in survival and we were unable to find a systematic literature search evidence from other longitudinal studies that this was the case. Secondly, as we have stated previously, patients enrolled into trials tend to have better outcomes than patients within a community-based study (eg selection bias towards patients who are healthier and more likely to comply with treatment, enhanced monitoring and response to prevent and manage adverse events). Thirdly, the higher survival in APEX may reflect the net health effect of VELCADE when administered after HDD failure. In the next step we adjust the transition probabilities to replicate the 1-year survival reported in APEX (internal validation).

Step 4: Adjust model to obtain accurate internal validation of the APEX study. We believe that the estimates of 1-year survival for the HDD in APEX are unbiased because it reflects survival prior to subjects having the opportunity to switch. However, the 2-year survival in APEX is probably an over-estimate if VELCADE is effective after initial treatment failure. We therefore, adjusted the APEX model to obtain survival as reported in APEX, using cell e85 in 'Summary' worksheet. We lowered the transition probability by 14% (hazard ratio of 0.86) and obtained results shown in scenario 4 on survival at one year that are identical to that in APEX. Survival at year two is lower for HDD because this model does not assume patients may use VELCADE after initial failure. We used GoalSeek in Excel to identify the estimate for cell e85 that would replicate the trial findings. We conducted the same exercise to identify the estimate if all patients started in regimen 2 (21% lower).

It is important to note the scenarios that reflect the validation of the trial findings and these are:
Scenario 4 reflects the comparison of VELCADE with HDD only with the same distribution of patients starting regimens as in the trial (40% - 1st regimen, 58% 2nd regimen and 2% - 3rd regimen).
Scenario 2 reflects the comparison of VELCADE with HDD only with patients starting in 2nd regimen (only 1 prior regimen).

B3. Please provide sensitivity analyses run with more appropriate ranges, or clarify the rationale for the ranges that have been used.
For the sensitivity analysis and PSA, the ranges used have been chosen arbitrarily and are not related to the confidence intervals of the parameters, in particular the ranges chosen would not appear to be wide enough.

For continuous variables, we assigned ranges based on 5th and 95th percentiles of the variable distribution, using the study's published standard deviations (eg mean ± 1.96*SD). We obtained estimates of the mean and the SD for costs directly from Bruce et al and increased it to 2006 prices using the national price index of 2% per annum.

For rates and proportions, the approach used for continuous variables would not be appropriate. For example if the true rate of in the population equals 0.4 then the variance equals 0.4* (1-0.4) = 0.24 and the SD = sqrt(0.24) = 0.49. Hence the range lies from 0-1. Instead, we examined the range of rates among key subgroups (by age, gender and extent of disease to determine ranges of rates.

14

B4. Please investigate and report on the following potential errors:

- There seems to be mistakes on some of the sheets with the transition probabilities. For example, cell AP43 on the V mn 1-3 sheet uses HDD probabilities (='Dex only'!$Q$15). This occurs many times on this sheet and the other transition probability sheets.
- There seems to be (may be) a mistake with use of clinical effectiveness data on overall survival (OS hazard rate). The model has used an improved survival benefit (treatment effect) in regimens 3-6 (eg cells i15-19 on VELCADE only worksheet) where patients are not treated with VELCADE. Was this your intention, or is it an error?

We developed the model to be flexible and permit explorations of various scenarios that may be useful for other submission outside of the UK environment.

However, for this submission we focused only on use of VELCADE used for 2nd or 3rd line regimen. The spreadsheets (V mn 1-3, V mn 4-6 etc) refer to the VELCADE transition probabilities only during use of the initial regimen. If a patient fails this regimen, then the transition probabilities are the same as for the HDD arm (labeled Dex only in the spreadsheet). This is deliberate and reflects no error in the model framework or its implementation.

The same issue applies in answer to the question raised in the second bullet point. The model contains estimates of VELCADE relative treatment effect similar for regimen 4-6 similar to that observed in APEX for regimen 2-3. However, these estimates are not applied (using model auditor in Excel, please observe they point to no cell) in this model; they are useful only to explore as other audiences wish the implication of using VELCADE in later regimens. Again there is no error in the model.

B5. Please send us a copy of reference 52 (ISPOR conference abstract – Yoong et al 2005).

Please find this attached.

B6. The following comments are listed below for your information:

- In general, the ERG suggests that the model seems a little over complicated. There are small numbers in the later regimens, especially for the time periods >2 years which results in strange values for some of the probabilities (see table for 3rd year). For example in the third year 82% die on regimen 6 but only 8% on regimen 5. It would appear to be down to data, and a more general approach would probably see (or assume) equal long term death/switch rates for regimen these later regimens.

The question pertains to imputation of estimates when the sample size is small enough that published estimates are imprecise and give strange values.

We presented only the analyses that we undertook using the empirically derived estimates. However, we conducted other imputations and they revealed minor effects on the model. This is not included in the submission.

15

Specifically, we conducted the following three imputations for intervals 13-24 months and 25-36 months:

- Assume mortality rates are average rates in regimen 5 and 6.
- Assume mortality rate for 6 is the same as 5.
- Assume mortality rate for 5 is same as for 6.

The following table shows the effect of the 4 approaches on the base case OS and ICER. Only in the highly unlikely situation of 0% mortality with regimen 5 and 6 in interval 25-36 months (VELCADE+HDD versus HDD) does the ICER increase significantly. Use of empirically based estimates therefore reveals no bias in the analyses.

VELCADE versus HDD					
	13-24 months	25-36 months	Discounted OS	Total Costs	Cost per LYG
Empirical (no imputation)* Imputation of mortality rates	8%, 80%	0%, 50%	9.9	£25,677	£31,1146
Average of regimens 5 & 6	45%	25%	9.9	£25,690	£31,074
Set rates of regimen 6 to 5	8%	0%	10.2	£25,815	£30,421
Set rates of regimens 5 to 6	80%	50%	9.7	£25,588	£31,634

VELCADE + HDD versus HDD					
	13-24 months	25-36 months	Discounted OS	Total Costs	Cost per LYG
Empirical (no imputation)* Imputation of mortality rates	8%, 80%	0%, 50%	11.0	£26,286	£28,682
Average of regimens 5 & 6	45%	25%	10.9	£26,220	£28,977
Set rates of regimen 6 to 5	8%	0%	8.8	£25,224	£34,500
Set rates of regimens 5 to 6	80%	50%	10.9	£26,250	£28,841

As a minor point the ERG suggests that indexing of 'other costs' seems to be incorrect.

16

We have reviewed the model and accept that we have under-inflated the 'other costs' by two years. In the model we converted costs from 199 instead of 1997. We have now revised this and the cost of 'other costs' is £478 per month. The implications of this change to the estimate of the cost per life year gained is marginal and the revised bas case estimate is now £31,146.

Section C: Additional discussion and rationale

C1. Please clarify the patient group under consideration. Various different descriptions have been used throughout the submission. For example, the trial's inclusion/exclusion criteria don't mention bone marrow transplantation status, yet the statement on license (p.3) suggests VELCADE is only licensed for people who have already undergone or are unsuitable for bone marrow transplantation; see also pages 1,6,9,12,24,30,32.

It is absolutely correct that the in/exclusion criteria did not require prior BMT nor unsuitability for BMT. However, retrospective analysis of the patient population indicated that this description applies to all but 18 second-line patients. Therefore the CPMP requested that we include this in our indication.

It is important to note that for early or first line management of Multiple Myeloma, dependant upon the patient's physiological status, treatment options are principally either intensive chemotherapy combined with bone marrow transplant, or a less intensive therapy usually melphalan combined with prednisolone. Following this treatment once patients have entered a period of stability or remission, there will come a time when all patients relapse. It is at this point, or second line therapy that patients become (according to license) eligible for treatment with Velcade: at this point all patients would have either had a bone marrow transplant, or have been deemed unsuitable to receive one.

C2. Primary outcome was time to progression due to association with survival. Why not use survival, given the high mortality rate?

Survival was not chosen as primary endpoint for several reasons:

Time to progression (TTP) looks exclusively at an effect of the experimental drug/regimen on the malignant disease, where survival information can be confounded by other factors than the malignant disease (unrelated deaths due to other diseases, accidents etc). It is important to note here that following treatment with Velcade, even those who accomplish a complete remission, will eventually (due to the progressive nature of the disease), relapse. At this subsequent point they may receive different treatments, and this will also confound the overall survival.

Within the APEX trial patients were allowed to crossover from the HDD arm to VELCADE in the 040 study (Companion study) after the progression event was documented. Such systematic crossover carries the potential risk of masking a survival improvement provided by an experimental drug/regimen because it may result in a comparison of treatment effect of the experimental drug/regimen in one line versus the subsequent line instead of a comparison versus standard-of-care.

17

Despite this study design, statistically significant improvements in both TTP and OS were observed in the APEX trial.

18