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

A critique of the submission from Celgene

Erratum

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

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

Confidentiality This document contains information which is proprietary to Matrix and may not be disclosed to third parties without prior agreement.
Except where permitted under the provisions of confidentiality above, this document may not be reproduced, retained or stored beyond the period of validity, or transmitted in whole, or in part, without Matrix's prior, written permission.

© TMKG Ltd, 2012

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

Report commissioned by:

NHS R&D HTA Programme

On behalf of:

NICE

Produced by:

Matrix Knowledge
Peninsula Technology Assessment Group (PenTAG)

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

Authors:

Mariana Bacelar, Economist, Matrix Knowledge
Adeline Durand, Senior Economist, Matrix Knowledge
Chris Cooper, Senior Information Specialist, PenTAG
Chris Hyde, Professor of Public Health and Clinical Epidemiology, PenTAG
David Murray, Senior Associate, Public Health & Economics, Matrix Knowledge

Acknowledgements:

We particularly acknowledge the help of Claudius Rudin, Royal Devon and Exeter Hospital who advised us and Nick Latimer, Research Fellow in Health Economics, University of Sheffield.

Declaration of competing interest of the authors

None

Rider of responsibility for report

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

This report should be referenced as follows

Bacelar M, Durand A, Cooper C, Hyde C, Latimer N, Murray, D. The clinical and cost-effectiveness of lenalidomide for people who have received at least one prior therapy with

bortezomib (partial review of TA171).Erratum. Single Technology Appraisal NIHR HTA Programme (13/07/01), Matrix and Peninsula Technology Assessment Group, 2014.

Contribution of authors

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

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

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

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

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

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

Correspondence to:

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

Date completed:

February 2014

Contents

This document contains errata in respect of the ERG report in response to the manufacturer's factual inaccuracy check.

The table below lists the location of the change in the original ERG report and the nature of the change.

Page no.	Change
81	The manufacturer requested that the ERG clarifies that the consistency in the excel flow sheets and the model structure are two different issues.
91	The ERG have overlooked the fact that where it is mentioned *"Celgene clarified that the **second** approach had been taken. However, Celgene's explanation for excluding the number of prior therapies as a covariate would only make sense if the **first** approach had been taken."* this should in fact read *"Celgene clarified that the **first** approach had been taken. However, Celgene's explanation for excluding the number of prior therapies as a covariate would only make sense if the **second** approach had been taken."*
92 - 95	The ERG agreed that it would be most appropriate to compare the extrapolated PFS and TTF curves against the KM curves for second-line patients.
104 - 106	The ERG agreed that it would be most appropriate to compare the extrapolated OS curves against the KM curves for second-line patients.
107	In light of the ERG sentence "However this is still an implausible scenario and a not acceptable one, for the reasons explained before." the manufacturer have requested that the ERG remove " and a not acceptable one" from the sentence aforementioned.

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

Furthermore Celgene claimed to have made the excel flow sheets consistent across intervention and comparator arms of the model. However, the issue initially raised by the ERG was concerning the inconsistency between calculations in the intervention and the comparator arms of the model and not to in the excel sheets layout.

ERG critique of the updated model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

It is therefore crucial how PFS and TTF were extrapolated:

- Progression-free survival

A log-logistic distribution was used to fit the MM-010 PFS data in order to extrapolate the study results to a 25 year horizon.

Celgene report undertaking visual inspections of the fitted curves and using Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) to assess the best model fit. Although these are common steps in the assessment of fit process, they should not be the only ones used (for example, to ensure external validity, the plausibility of the extrapolated portion of the curves should also be assessed).

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

Furthermore, the ERG have the following concerns with Figure 13, presented in the previous section (reported again below) and taken from the original submission (Figure 25) which shows the KM PFS curve for Len/Dex as well as the fitted PFS curve:

1. It is not very informative to show the curves only to the point where the KM curve ends. The time period of the graph should be wide enough so the shape of the fitted curve is observed in the longer term and a judgment can be made of the appropriateness of the fitted curve in estimating PFS. Figure 19 shows the graph produced by the ERG, with a time horizon of 25 years (1300 weeks). The fitted curve is presented alongside the KM for the MM-010 second-line treatment subgroup.
2. The ERG could not replicate Figure 13. In the graph produced by the ERG (Figure 19) the fitted curve does not seem to overlap the KM curve as perfectly as in the graph produced by Celgene.

Reproduction of Figure 13. KM plot and fitted log-logistic model for PFS (Figure 25 in Celgene submission)

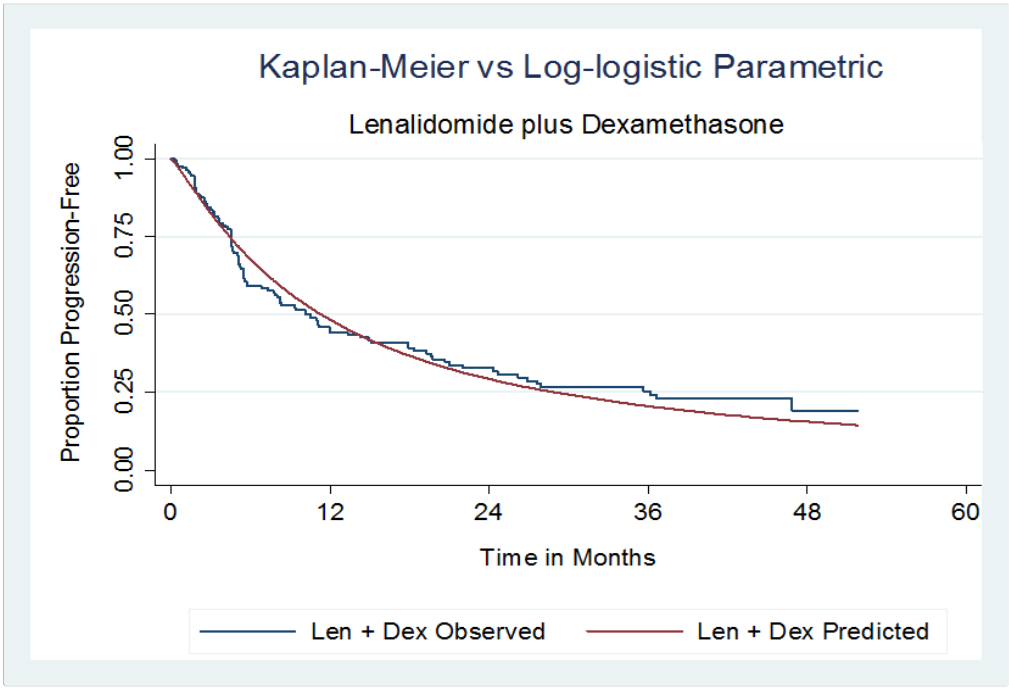
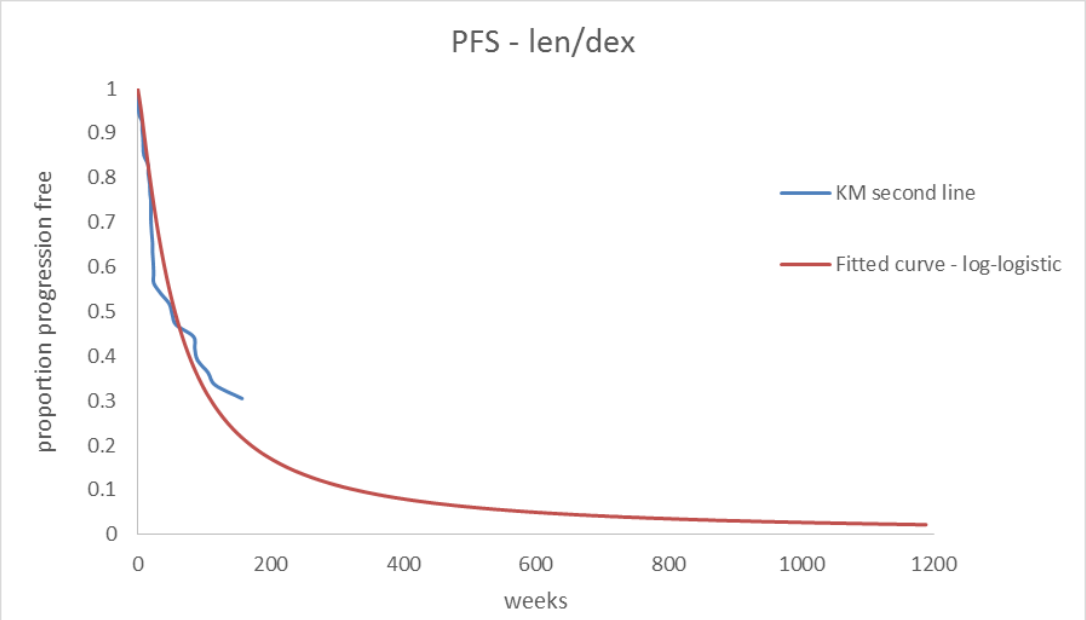


Figure 1. KM plot and fitted log-logistic curve for PFS over 25 years produced by the ERG



Source: produced by the ERG

- Time to treatment failure

Similarly to PFS, a log-logistic distribution was used to fit the MM-010 TTF data in order to extrapolate the study results to a 25 year horizon.

Celgene report undertaking visual inspections of the fitted curves and using Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) to assess the best model fit. Again, other steps could have been taken to assess the appropriateness of the distribution used.

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

The ERG also identified problems for Figure 14 (reported again below) which was taken directly from the submission (Figure 26) and presents the KM TTF curve for Len/Dex as well as the fitted TTF curve:

1. It is not very informative to show the curves only to the point where the KM curve ends. The time period of the graph should be wide enough so the shape of the fitted curve is observed in the longer term and a judgment can be made of the appropriateness of the fitted curve in estimating TTF. Figure 20 shows the graph produced by the ERG, with a time horizon of 25 years (1300 weeks). The fitted curve is presented alongside the KM for the MM-010 second-line treatment subgroup.
2. The ERG could not replicate Figure 14. In the graph produced by the ERG (Figure 20) the fitted curve does not seem to overlap the KM curve as much as in the graph produced by Celgene.

Reproduction of Figure 14. KM plot and fitted log-logistic model for TTF (Figure 26 in Celgene submission)

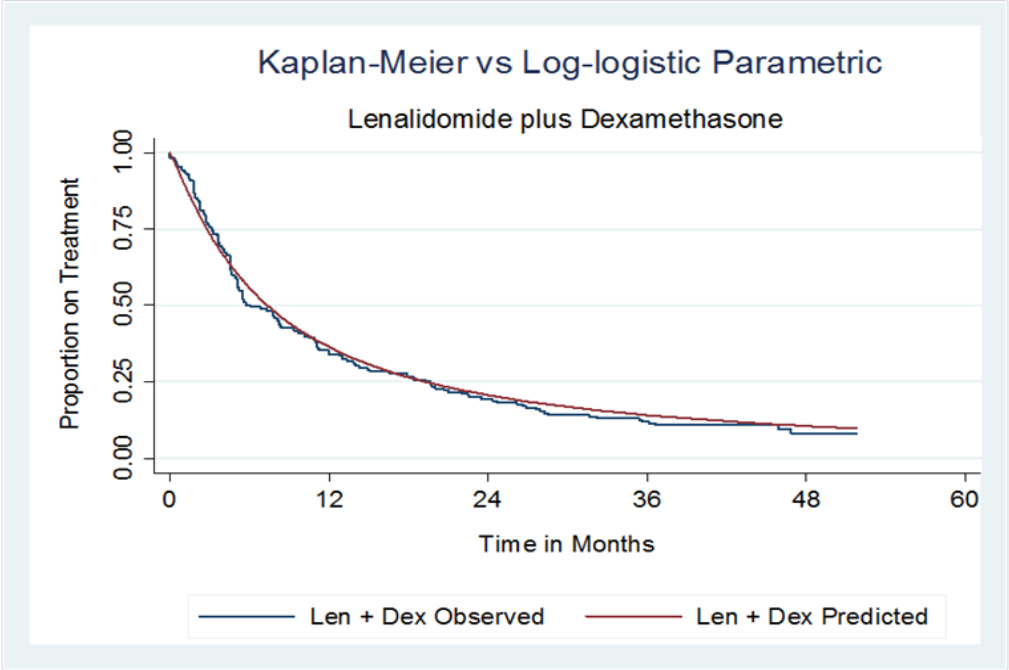
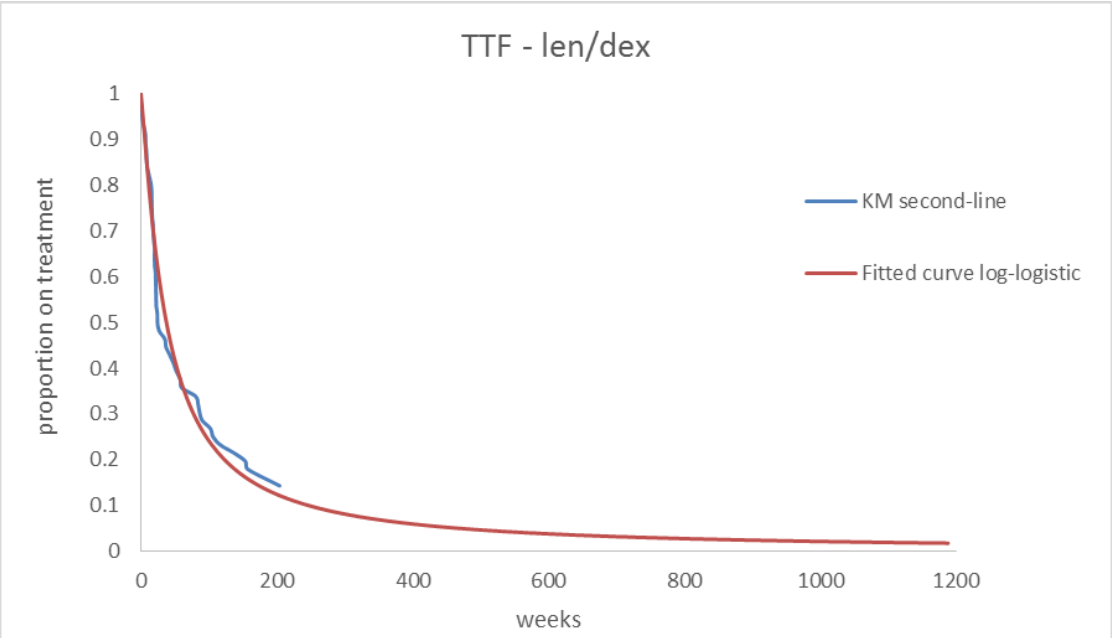


Figure 2. KM plot and fitted log-logistic curve for TTF over 25 years produced by the ERG



Source: produced by the ERG

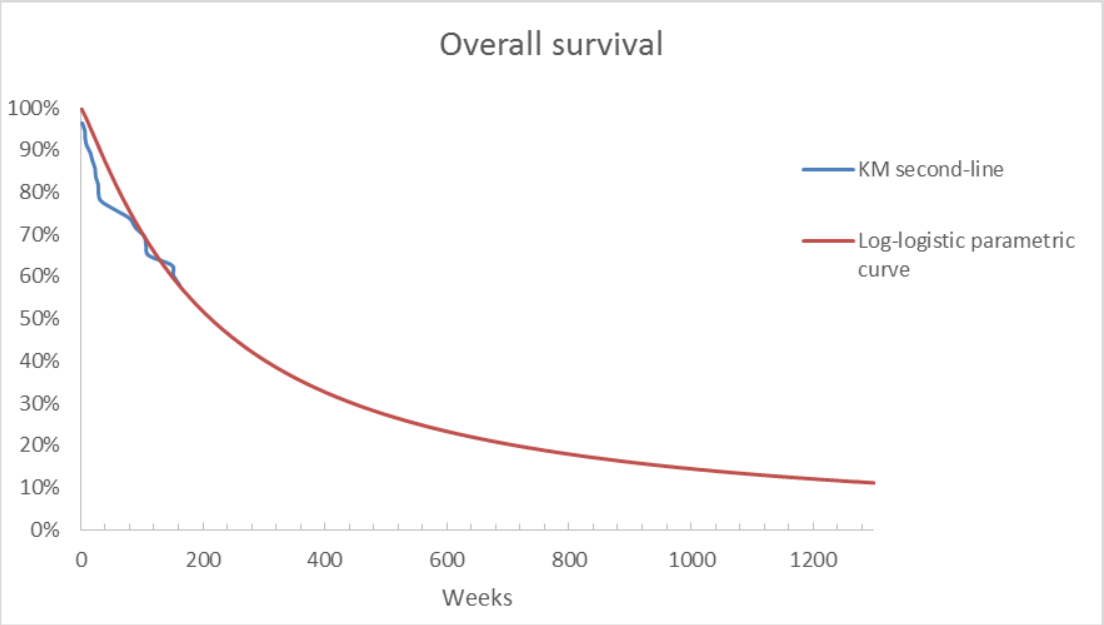
The ERG understand that crossing survival curves are a possible complication arising from fitting data to different distributions. However, when this is observed, a different approach needs to be taken which prevents the curves from crossing for example, using flexible models on the hazard ratios (e.g.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

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

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

Based on visual inspection of the curves, the fitted curve seems to be overestimating OS, especially until week 100.

Figure 3. KM plot and fitted log-logistic curve for OS over 25 years –Len/Dex

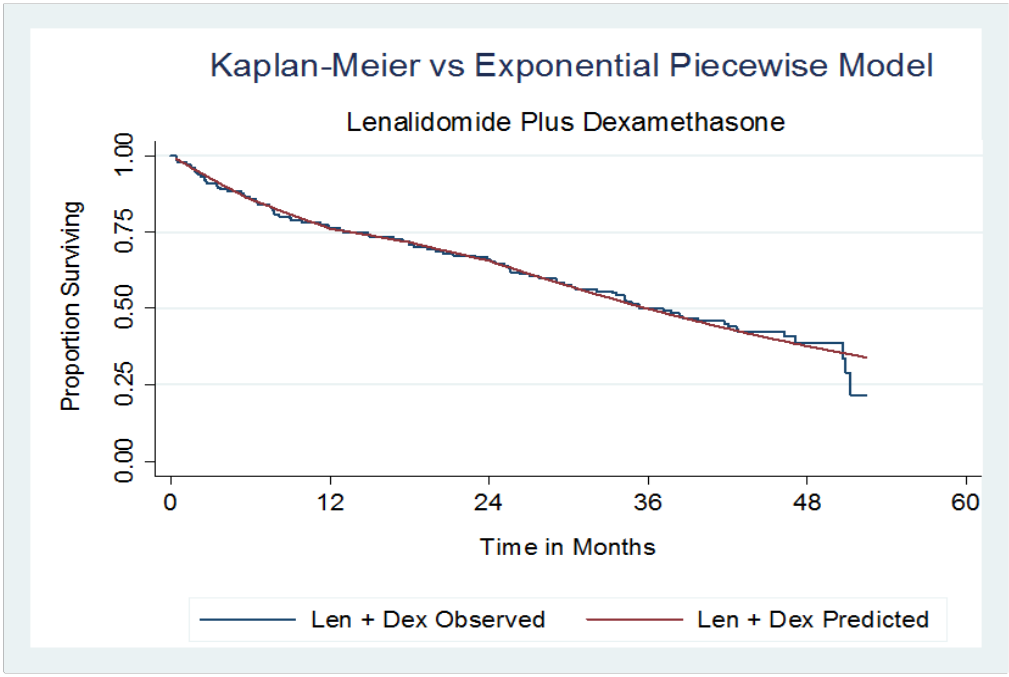


Source: produced by the ERG

Additionally, the ERG tried to replicate Figure 15 (replicated below), which was taken from the original submission (Figure 28) and shows the KM curve for OS as well as the extrapolated curve produced by fitting an exponential piecewise model to OS data. The resulting curves are presented in Figure 27. Unfortunately it was not possible to replicate Figure 15 (the same problem was found for PFS and TTF original graphs) and based on Figure 27 produced by the

ERG, even though the exponential piecewise curve seems a better fit until around week 50 it seems to be a poor fit as time progresses.

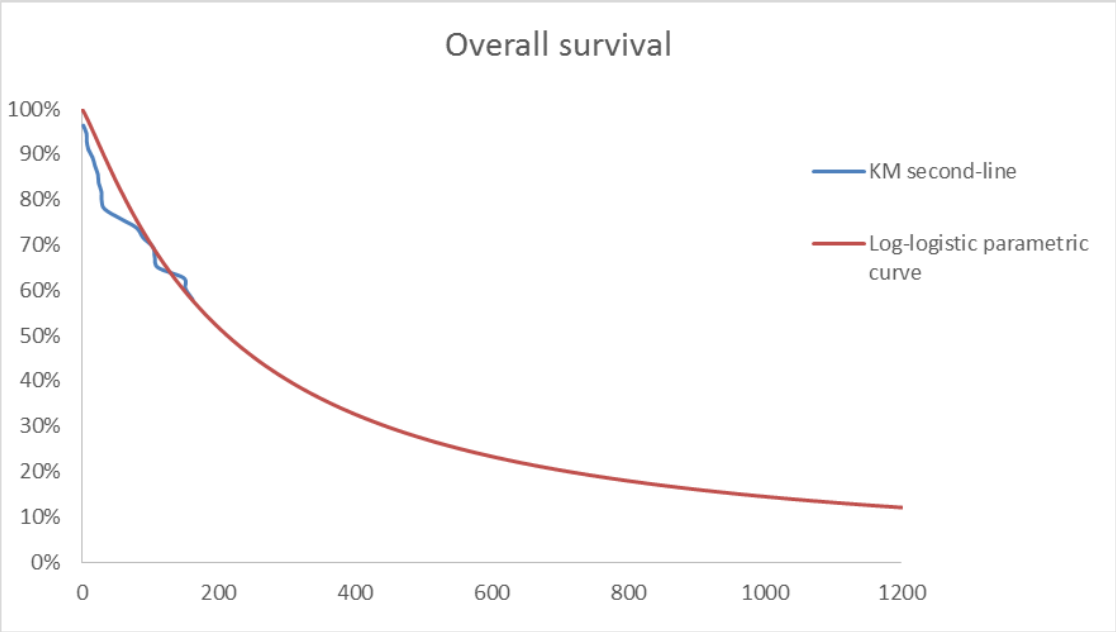
Reproduction of Figure 15. KM plot and fitted exponential piecewise model for OS (Figure 28 in Celgene submission)



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

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

Figure 4. KM plot and fitted piecewise exponential curve for OS over 25 years – Len/Dex



Source: Produced by the ERG

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

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

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

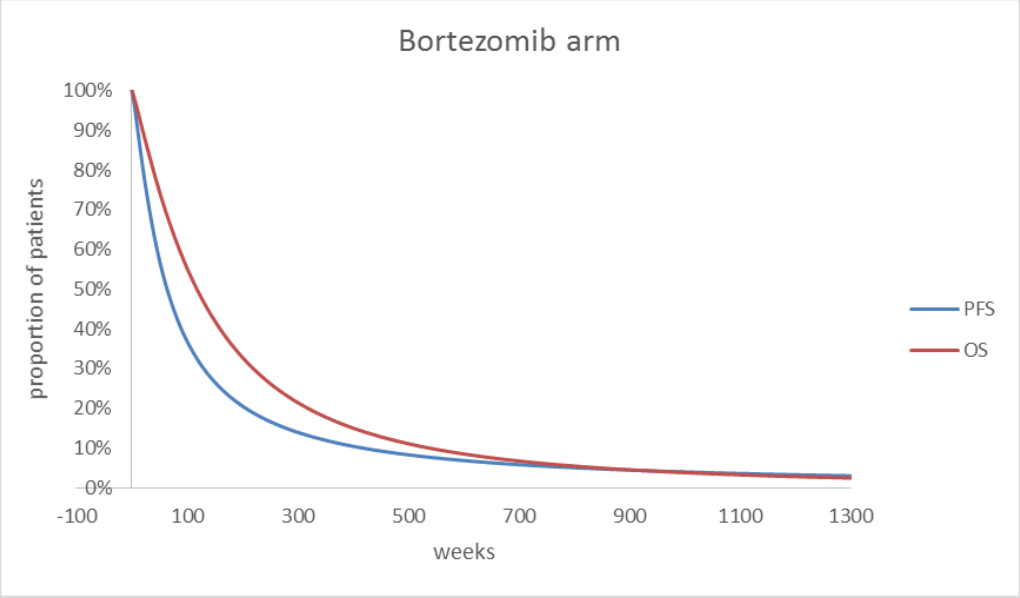
In summary, the ERG do not feel confident that the explanations and approaches followed by the manufacturer truly addressed the initial problems raised.

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

Even though the OS curves do not cross the PFS and the TTF curves in the intervention arm of the model anymore, Figure 28 and Figure 29 show how this is still a problem in the comparator arm of the model.
The curves now cross later in time (in the original submission the curves crosses around week 600) with PFS and OS curves crossing each other around week 900 (19 years) and TTF and OS curves crossing each other around week 1290, which corresponds to approximately 25

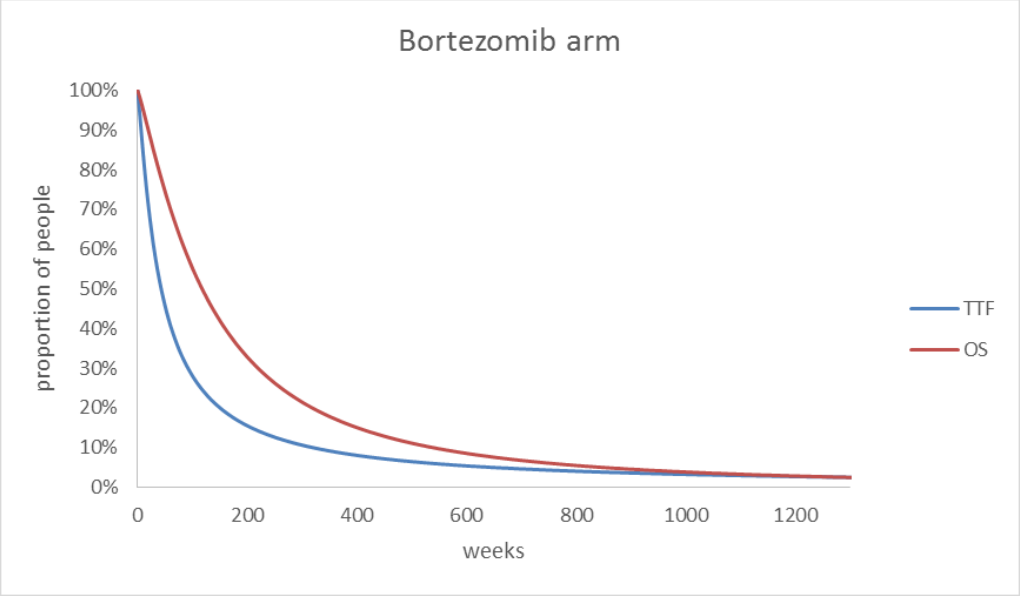
years (note that the economic analysis lasts for 25 years). However this is still an implausible scenario for the reasons explained before.

Figure 5. PFS and OS curves in the Bort arm of the model



Source: produced by the ERG

Figure 6. TTF and OS curves in the Bort arm of the model



Source: produced by the ERG