Additional clarification to NICE arising from ERG initial assessment of the Celgene submission

Introduction:

After an initial assessment of the manufacturer submission, the ERG have communicated to NICE some concerns related with the validity of the economic results presented. As a result, NICE have requested the ERG provide detail of the initial findings which raised such concerns.

The ERG is particularly concerned that in the manufacturer excel model, the Overall Survival (OS) curves cross with the Progression-Free Survival (PFS) and the Time-to-treatment Failure (TTF) curves.

Other technical problems were found within the Markov models. The calculations used to estimate the number of people in each health state per cycle of the Markov models are not consistent between the intervention and the comparator arms and include significant mistakes.

Furthermore, the manufacturer did not provide clear details on a substantial part of the data used in the economic model on their submission. This point was raised to the manufacturer in the questions sent for clarification.

Consequently, based on the initial assessment completed by the ERG to date, below we provide (some) detail on the methods used by the manufacturer and the outcomes of their analysis, and explore the potential concerns associated with these.

Estimation of the OS and PFS curves in the model:

The OS curve used to model mortality in the Len/Dex arm of the model was obtained through a piecewise exponential model which allows the hazard ratio (HR) to vary over time. Additionally, the OS curve seems to have been adjusted using the mean of covariates method, in which average values of covariates (like for example the beta-2 microglobulin count and presence of bone lesions for the baseline MM-10 population) are entered into a proportional hazards regression equation.

The OS curve for bortezomib was then estimated by applying a HR to the Len/Dex OS curve over time. The HR estimation seems to have been adjusted to take into consideration the population characteristics of the Taverna study (which is the study used to assess the effectiveness of Bortezomib).

The PFS curve used to model disease progression in the Len/Dex arm of the model was obtained by fitting a log-logistic distribution to the MM-010 data. This curve also seems to have been adjusted with the mean of covariates method. However, the PFS curve was only adjusted to one covariate (beta-2 microglobulin count at baseline for the MM-010 population).

The PFS curve for bortezomib was then estimated by applying a HR to the Len/Dex PFS curve over time. The HR estimation does not seem to have been adjusted to take into consideration

the population characteristics of the Taverna study (which is the study used to assess the effectiveness of Bortezomib).

TTF curves for Len/Dex and comparator were obtained in a similar way to that of the PFS curves.

The resulting curves are shown in Figure 1 for the Len/Den arm of the model, while Figure 2 presents the OS and PFS curves for the bortezomib arm of the model. Finally, Figure 3 and Figure 4 present the OS and the TTF curves for Len/Dex arm of the model and the comparator arm of the model, respectively.

The OS and the PFS curves cross each other both in the Len/Dex arm of the model as well as in the bortezomib arm. Rationally, such crossing is not possible as the OS curve determines the proportion of people alive at each cycle of the model. It is therefore impossible to have a greater number of people alive and free from disease progression than the total number of people alive at the same point in time.

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

Figure 1

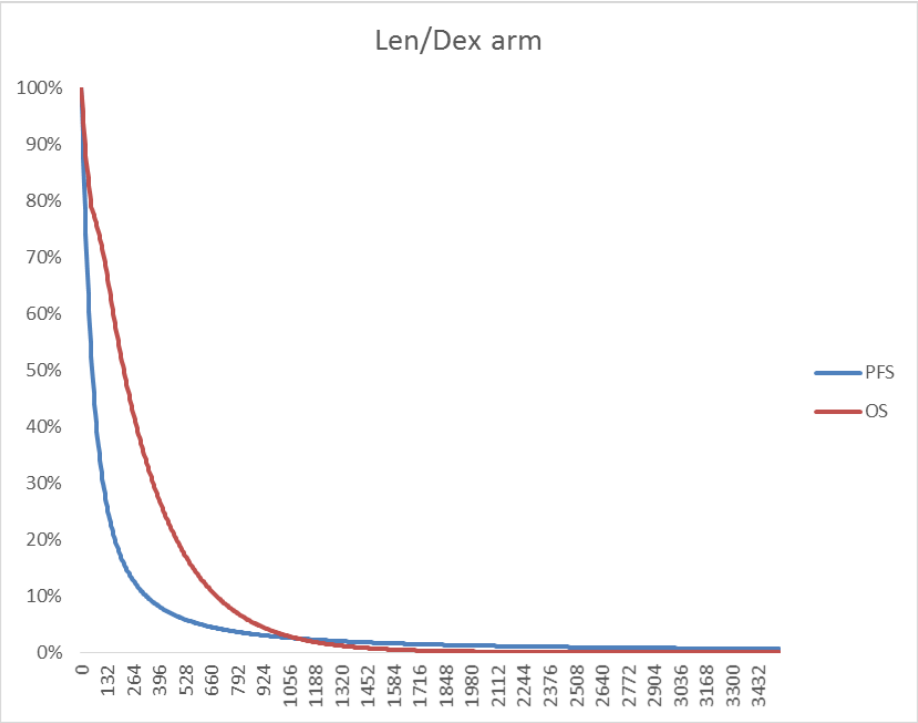


Figure 2

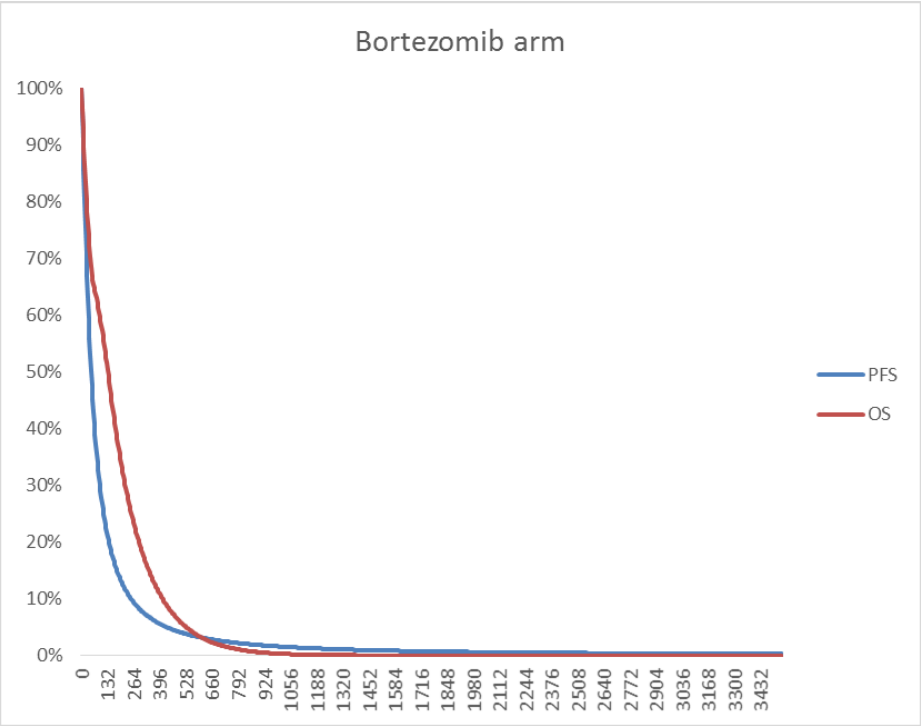


Figure 3

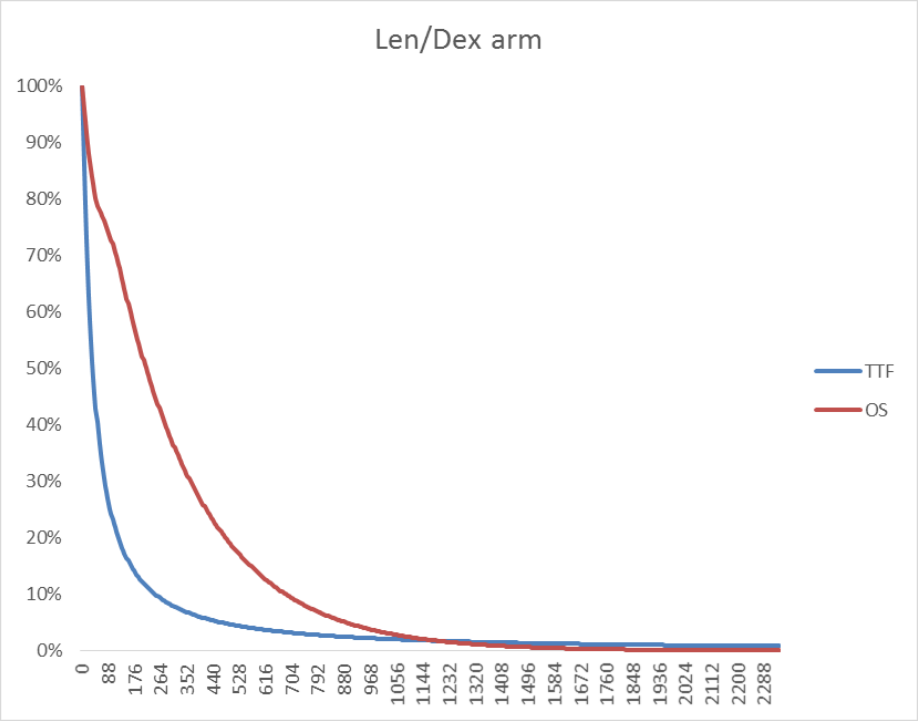
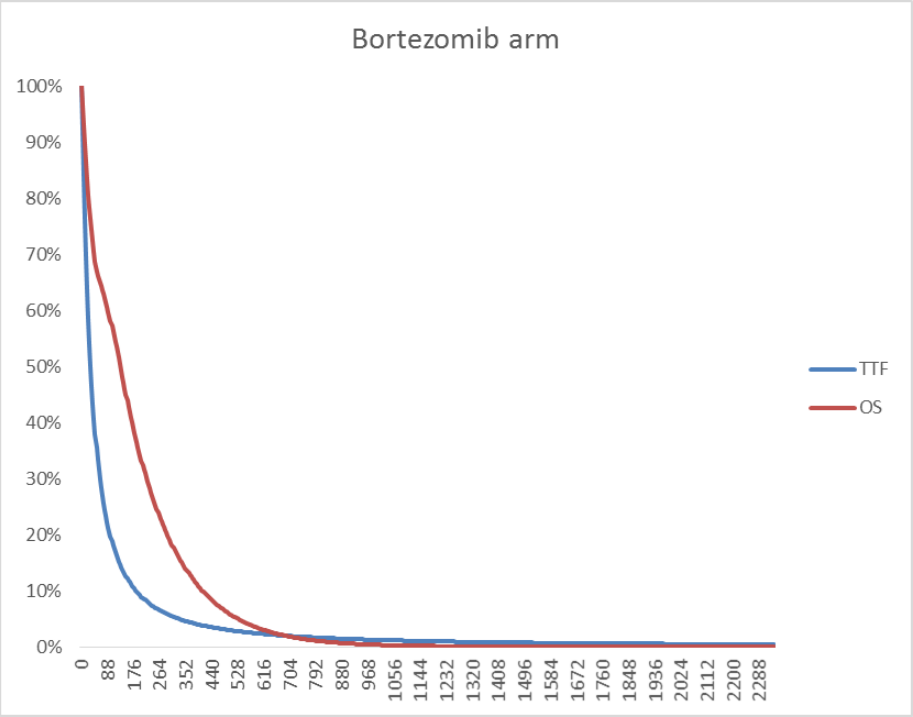


Figure 4



Possible explanations/suggested approaches:

It is not clear to the ERG what is the root of the problem described above. Overall, it seems that the OS curves are overestimated as per visual inspection of the Markov trace for OS provided in the submission (appendix O) with the KM curve in Figure 28 in the main submission.

The use of the mean of covariates method to adjust the OS curve might potentially be overestimating survival. In fact this method has been criticized for the validity of the resulting estimated curves (Ghali et al, 2001). One of the underlying reason is that using baseline mean characteristics to adjust survival curves might skew the curve if the mean values are also skewed. Alternative approaches could be used by the manufacturer to adjust for baseline characteristics (Ghali et al, 2001; Bradburn et al 2003).

Furthermore, there is the possibility that PFS is also overestimated. The manufacturer did not adjust the HR to the Taverna study (note that this was the approach taken for OS) and also did not report the adjusted median PFS. It would be recommend that a consistent approach is taken and that the PFS HR is adjusted, similarly to the OS HR.

Finally there is also the possibility that one of the causes for the OS curve to cross both the PFS and TTF curves is related to censoring. As noted in the initial questions sent to the manufacturer, little detail was provided on censored data and so it would be useful if the manufacturer could provide the number of patients at risk throughout time.

Excel model:

It is not clear to the ERG the rationale behind several calculations in the Markov model (both in the intervention and the comparator arm). These are outlined below by excel tab, together with the ERG suggested approach to mitigate the problem (whenever possible).

Pf.Len tab:

1. The manufacturer often used the “minimum” excel formula to choose either from OS data or PFS/TTF data.
   Suggestion: This is possibly related with the fact that the 3 curves are crossing each other. However this does not seem to be a sound approach. The curves should not cross and therefore the minimum function should not be used under these circumstances.

2. The manufacturer multiplied the PFS by the OS estimates per cycle.
   Suggestion: This is does not seem to be correct. The ERG suggest that the manufacturer take the population in each curve without multiplying these by each other. For example, to calculate the “Pre-progression” population in each cycle, this should be done simply by taking the PFS transition probabilities (estimated in tab “MM10 Progression & Survival”). These estimates should not be multiplied by any other parameters.

3. The calculations to estimate the proportion of people under 3rd line and 4th line treatment do not seem to be correct. Furthermore, when we add the populations on 2nd, 3rd and 4th line treatments together with the number of dead patients, the final estimate does not add to 100% of the cycle population. This is not correct.
   Suggestion: To calculate the number of patients under 3rd and 4th line treatment, the manufacturer should be using the data provided on the column “Patients Alive & Off Therapy” (AH16:AH334) instead of using other data and calculations (as currently done).

   Should the manufacturer wish to split the patients off therapy who have progressed and the patients off therapy who have not progressed in order to (possibly) isolate a group of patients moving to further treatments, it is also possible to do so. By subtracting the “Patients on Therapy” from the “Pre-progression” patients, the manufacturer can obtain the number of patients off treatment who haven’t progressed.

4. The calculations to estimate the duration of 3rd and 4th line treatments and the respective cost (for 4 cycles), calculated on columns CU and columns DE16, do not seem to be right. These seem to be resulting in an overestimation of costs. Since deaths between treatment cycles should be considered, the calculations can get slightly complex. The ERG does not have a specific suggestion for dealing with this at the moment.

Pf.Comparator tab:

1. Overall, there is a lack of consistency between the calculations on the intervention and the comparator tab.
   Suggestion: The population in the “Pre-progression 2nd line” health state should be estimated from the bortezomib PFS curve directly (as for the intervention arm) and the

"Patients on Therapy" and the "Patients Alive & Off Therapy" should also be calculated in similar fashion to the intervention arm of the model (only using different data).

Overall the ERG would suggest that more consistency across health state calculations in the intervention and comparator arm would add clarity to the model and would solve some of the mistakes found on the calculation in the comparator tab.

It should be noted that the ERG did not have enough time to explore the calculations related with 3rd and 4th line treatment options and respective costs in the comparator tab.

2. The ERG noted that the costs of dexamethasone are not being considered in the comparator arm, even though more than 60% of patients in the Taverna study were receiving this drug.
   Suggestion: The ERG would suggest that the manufacturer include this cost for the proportion of patients receiving the drug in the Taverna study.

Conclusion:

As indicated by the ERG in the request for clarification to the manufacturer, a considerable number of parameters used in the economic model are not mentioned in the submission report and therefore it is impossible for the ERG to know the sources of such parameters and what are the underlying calculations (if any) to obtain these.

Therefore, the ERG would suggest that the manufacturer provide all data used in the economic model properly sourced and in detail. Additionally, the ERG would suggest that the whole of the raw data used to produce the KM curves and parametric curves are presented, as these are a key piece to the economic model.

The fact that the OS and PFS/TTF curves are crossing, is *per se* a sign of significant methodology weakness. It should also be mentioned that other issues were already identified by the ERG.

Furthermore, the ERG did not have the time to adequately look into all areas of the submission in detail, and so new issues may arise.

Overall, at the moment the ERG feels a lack of confidence in the validity of the economic analysis and the results presented.

11 December 2013
Matrix & Peninsular Technology Assessment Group (PenTAG)

Ghali WA, Quan H, Brant R, van Melle G, Norris CM, Faris PD, Galbraith PD, Knudtson ML. 2001**.** Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models. JAMA. 2001 Sep 26;286(12):1494-7.

M J Bradburn, T G Clark, S B Love and D G Altman1. 2003. Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. Br J Cancer. 2003 August 18; 89(4): 605–611.