National Institute for Health and Clinical Excellence
Centre for Health Technology Evaluation

Pro-forma Response

ERG report

Abiraterone for the treatment of metastatic castration resistant prostate cancer following previous cytotoxic therapy

Please find enclosed the ERG report prepared for this appraisal.

You are asked to check the ERG report from Warwick Evidence to ensure there are no factual inaccuracies contained within it. If you do identify any factual inaccuracies you must inform NICE by the end of Tuesday 6 December using the below proforma comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the Evaluation report.

The attached proforma document should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Issue 1 The ERG suggests that data for subjective outcomes may have been incomplete and that patients were not blinded in the updated analysis, p 12 1st paragraph, p 13 3rd paragraph, p 43 2nd paragraph.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
The ERG report suggests that the PRO data for these subjective outcomes were apparently incomplete and that during the time between the 'Primary' and 'Updated' analyses patients may not have been blinded to treatment. Janssen would like to confirm that the PRO was not incomplete and that missing data in the PRO analysis was minimal. Overall, compliance with expected PRO assessments in COU-AA-301 was good and both arms had a relatively small amount of missing PRO instruments, which is consistent with published PRO oncology results. The overall percentage of missing PRO assessments totalled across all study visits for both study arms was [redacted] for FACT-P, [redacted] for BPI and [redacted] for BFI. The PRO dataset is highly unlikely to contain any data for patients unblinded between the 'Primary' and the 'Updated' analysis time points. The Independent Data Monitoring Committee (IDMC) did not meet to review the interim analysis data until the 20th August 2010 and therefore the decision to stop the study and unblind was not made until after that point. The clinical cut off for the updated analysis occurred on the 20th Sept 2010 and only 8 subjects (2%) discontinued due to unblinding after the IDMCs decision. Therefore it is highly unlikely that the treatment assignment of any patients would have influenced the PRO outcomes in the 'Updated' analysis.	Janssen requests that the ERG report be amended to reflect the robustness of the PRO analysis which contained minimal missing data values and was highly unlikely to contain data from any unblinded patients.	As the report currently stands, it suggests that the PRO data collected in the study was incomplete and that the updated PRO analysis may have contained unblinded patients. Janssen would like to confirm that there was minimal missing data and that the 'Updated' analysis was highly unlikely to contain data from unblinded patients.	Given this information the ERG accept that missing data is unlikely to have affected these analyses. ERG have changed the ERG report text appropriately (page 12, 13, 43, 53) in the amended report sent to NICE 12 Dec.

Issue 2 The ERG suggests that the regression for on treatment utilities has been incorrectly applied in the model: Page 14, paragraph 3, Page 80.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
The ERG has questions regarding the analyses of FACT-P and its mapping to utilities within our submission. Janssen acknowledges that in the original submission the description of how the utilities were applied in the modelling should have been more thorough. We would welcome the opportunity to discuss the application of these utility values with the ERG to ensure the ERG report can clearly reflect how the data was handled and how the economic results were generated, as is outlined below. **Treatment effect of AAP vs. PP:** The key difference between our approach and the ERG's approach to the treatment effect estimate is that our estimate is based on change from baseline rather than the absolute utility. The values in our submission were obtained from a regression analysis on the individual patient level and will also differ slightly from those calculated by the ERG based on aggregate data for example due the handling of missing data. Below we provide a revised Table 38 from the ERG report (titled *The evolution of FACT-P derived utilities and numbers reporting the FACT-P*) to reflect estimates for the one prior chemotherapy group using individual patient data. Also, the table was expanded to include the change in utility from baseline. See Revised Table 38 below. Note that Cycle 1 in the table corresponds to the baseline, i.e. FACT-P at the day of the randomization; therefore change in utility from baseline at cycle 1 is not applicable. From this revised Table 38, it is clear that the treatment effect estimated from change from baseline ([redacted]) is higher than that estimated from the absolute utility ([redacted]). The reason is that although the mean baseline utility is the same between the two groups, patients with lower baseline utility (or those who were relatively sicker) in the PP group were more likely to drop out, when compared to the AAP arm. In our RMME regression model, the estimated treatment effect of 0.0456 is generated from the random intercept model that is suitable for longitudinal data and takes into account "person-effect" on utility; the estimated effect from this model might slightly differ from the conventional OLS model because	Janssen suggests that the ERG apply the [redacted] value for the on treatment patients and not the proposed value of [redacted] or [redacted] as the methodology used to calculate this value is more robust. Similarly, progression free utility for the PP patients should remain at [redacted] as applied in our	It is misleading to suggest that the value of [redacted] is more accurate than [redacted], as the value of [redacted] used in our submission was calculated using more robust methodology. A similar justification applies for the use of [redacted] as the progression free utility for PP patients.	See BOX below.

this model uses different weights to calculate the average. The treatment effect estimate from the RMME model is also adjusted for baseline utility. Without controlling for baseline utility at the patient level, the ERG group underestimated the treatment effect of AAP.
Therefore, we believe our approach in estimating relative utility gain of AAP vs. PP is appropriate.

model as for PP the mean change in utility did not deviate from the baseline value.

Revised Table 38 (page 79) The evolution of FACT-P derived utilities and numbers reporting the FACT-P ; one prior chemotherapy subgroup (PFS columns not included)

Cycle	AAP N	PP N	AAP: Mean Utility	AAP: Mean Change from Baseline	PP: Mean Utility	PP: Mean Change from Baseline	Diff in Mean Change from baseline	Diff in Mean Utility Values
1	491	246	[redacted]	[redacted]	[redacted]	NA	NA	0
4	378	175	[redacted]	[redacted]	[redacted]	-0.02	0.06	0.04
7	298	87	[redacted]	[redacted]	[redacted]	-0.01	0.04	0.02
10	227	54	[redacted]	[redacted]	[redacted]	0	0.03	0.01
13	151	35	[redacted]	[redacted]	[redacted]	0.01	0.03	0.03
16	133	27	[redacted]	[redacted]	[redacted]	0.05	-0.02	-0.03
19	99	22	[redacted]	[redacted]	[redacted]	0	0.04	0.01
average	1286	400	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

PP arm on treatment utility
Our model assumes that PP patients remain at their baseline utility while on treatment or progression free. We believe this is a reasonable assumption because first, PP is a palliative care with little toxicity, and second from the revised Table 38 above, the average mean change from baseline utility for the PP arm is only [redacted] indeed based on the same RMME model the mean utility of PP from the one prior line analysis over the treatment (progression free period) with the Abiraterone treatment indicator set to zero (the mean PP utility) is [redacted].

BOX

Issue 2a: HRQoL treatment effect of AAP vs. PP:

The manufacturer asserts that their RMME regression model appropriately controls for the differences in utility at baseline among those reporting subsequent to baseline. The ERG accepts this proposition. Note that the reason for the ERG presenting table 38 only for the peak reporting cycles is due to the ERG not having access to the individual patient level data, coupled with the obvious peak reporting effects. Given access to the patient level data this restriction should no longer be applied and it would be appropriate for the manufacturer to have presented the table and estimated the difference in mean change from baseline of [redacted] using all the cycles for which data was reported. Also using IPD it would be possible to explore the validity of the manufacturer's model using alternative approaches (e.g. Tobit, CLAD [censored absolute deviation] analyses) since EQ-5D scores are known to exhibit ceiling effects when a large proportion of subjects rate themselves in full health with a utility score of 1 and have the data bounded or censored at 1.

The ERG have changed to text on page 14,

FROM

However, the ERG consider that the regression results were incorrectly applied; when this is corrected, the absolute difference between those on treatment and those off treatment is reduced by [redacted].

TO

However the ERG has some concerns regarding the manufacturer's approach to the regression; in the absence of access to individual patient data, the ERG's exploration of alternative approaches was necessarily limited.

The ERG report does not suggest that [redacted] or [redacted] is more accurate than [redacted]. Indeed the ERG revised baseline retains the [redacted] increment. The ERG report does suggest that there is uncertainty around the utility increment from abiraterone over placebo for those remaining on treatment and that the summary of the derivation of HRQoL within the submission does not dispel this.

The ERG report uses the patient averages supplied by the manufacturer to show that it appears that there are differences in the baseline utilities between the arms among those reporting HRQoL data subsequent to baseline.

The manufacturer confirms that within the patient level data the average at baseline for those reporting subsequent to baseline was lower in the AAP arm than in the placebo arm. The averages among those reporting at baseline, the change from baseline and the resulting utility levels, coupled with the differences in the mean change from baseline and the differences in the mean utility values based upon the patient level data as per the manufacturer response are as below.

	AAP				PP					
Cycle	N	Baseline	Change	Utility	N	Baseline	Change	Utility	Δ mean change	Δ mean utility
4	378	[redacted]	[redacted]	[redacted]	175	[redacted]	[redacted]	[redacted]	0.06	0.04
7	298	[redacted]	[redacted]	[redacted]	87	[redacted]	[redacted]	[redacted]	0.04	0.02
10	227	[redacted]	[redacted]	[redacted]	54	[redacted]	[redacted]	[redacted]	0.03	0.01
13	151	[redacted]	[redacted]	[redacted]	□5	[redacted]	[redacted]	[redacted]	0.03	0.03
16	133	[redacted]	[redacted]	[redacted]	27	[redacted]	[redacted]	[redacted]	-0.02	-0.03
19	99	[redacted]	[redacted]	[redacted]	22	[redacted]	[redacted]	[redacted]	0.04	0.01
average	1286		[redacted]	[redacted]	400		[redacted]	[redacted]	[redacted]	[redacted]

Note that the baseline figures among those reporting in a given cycle have been calculated by the ERG as the cycle utility minus the cycle change.

It should also be borne in mind that the modelling in effect attempts to model what the experience of the representative patient would be in either arm. The utility values at baseline for those reporting HRQoL values in both arms are higher than the assumed placebo baseline average.

Issue 2b: HRQoL PP arm on treatment

The manufacturer response appears to switch back to the individual patient level data rather than the statistical model of utility changes for this argument. There may be some conflict between the two approaches of the manufacturer and it would be reassuring to know that the two are broadly in line.

The manufacturer response does not appear to contradict the ERG report calculation that the utility regression (table 77 appendix 15 of the manufacturer report) implies an average on treatment utility of [redacted] for the all patient placebo group given a baseline of [redacted]:

[redacted]

But there is a logical error in the ERG application of this to the revised baseline for the 1-prior patient population. This should have applied the 1-prior patient population regression. For a baseline of [redacted] this appears to imply:

[redacted]

But in the light of this it is not entirely clear to the ERG why the regression implies an unchanged HRQoL in the placebo arm for the 1-prior group remaining on treatment subsequent to baseline as suggested in the manufacturer response.

The manufacturer suggestion may be that the appropriate method of estimation is to apply the regression equation to the baseline values of the individual data points within the individual patient data, and doing so results in the average utility under placebo among those reporting of [redacted] as per the manufacturer response. But if so this would seem to beg the question of what the estimate for the average utility among those within the abiraterone arm is when applying a similar methodology.

The manufacturer response also ignores the baseline estimates from the Wu reference of utility estimates among mHRPC patients. This is a pity as these patients were at baseline of a similar average age of 72 years and very similar in terms of the average FACT-P scores for the various dimensions used within the manufacturer regression mapping function.

Average values	AAP	PP	Wu
Fact P Subscale	[redacted]	[redacted]	29.8
Emotional Well Being	[redacted]	[redacted]	17.2
Functional Well Being	[redacted]	[redacted]	16.6
Physical Well Being	[redacted]	[redacted]	20.9
Social Well Being	[redacted]	[redacted]	21.0

While the mapping function of Wu has been misreported within the paper, there is no reason to believe that the baseline utility values calculated using the UK preference weights have been misrepresented. This provides a direct estimate of EQ-5D utility among mHRPC patients using the UK preference weights for the patient group with the above concurrent FACT-P values. As summarised within the ERG report, Wu reports a mean baseline EQ-5D utility among 276 mHRPC patients based upon the UK social tariff of 0.635, though this is skewed with a median of 0.73 and a skewness of -1.53.

Applying the manufacturer mapping function to the Wu average FACT-P values results in a utility estimate of [redacted] which is, as would be anticipated given the similarity of FACT-P values, very similar to the estimated baseline utilities within the submission.

Issue 3 The ERG report suggests there may have been an imbalance in the number of bone metastases in the two arms of the COU-AA-301 study, p 18 bullet 8.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
The ERG suggests that the details of the number bone metastasis were not collected and may have been imbalanced between the arms. In response to clarification question A3, Janssen provided the proportion of patients with evidence of bone metastases in each arm at study baseline ([redacted]% of AAP and [redacted]% of PP). The clinical utility of using the Soloway classification for the quantification of tumour burden is not validated and was not collected in the study, however (and as with any other unobserved confounders) due to the large sample size in the COU-AA-301 study and the stratification factors used (ECOG status and presence of pain), it is highly unlikely that the tumour burden of patients at baseline would differ between the two arms.	Janssen requests that this bullet be removed from the report as the evidence suggests that it is unlikely that a difference in tumour burden would be observed between the two study arms.	It would be misleading to suggest that baseline tumour burden may have been different between the two study arms, as the study arms appear to be well balanced in all other respects.	The ERG accepts that randomisation was likely to eliminate imbalance. In the amended ERG report sent to NICE 12 Dec ERG have removed the text referring to imbalance, but have retained the statement that the number of bone metastases was not recorded.

Issue 4 In a sensitivity analysis, the ERG suggests a low utility value of 0.630 should be applied to the progression free state, p20 and p89.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
In a sensitivity analysis, the ERG explores the impact of assigning a low utility value of 0.630 to the progression free state; however this value is not aligned with the literature. The value used in the original submission ([redacted]) is aligned with several other sources; patients >16months from death in the Sandblom study (0.77), the UK subgroup of the Sullivan study (2007) 0.715, and the progression free utility values used in the NICE submission for cabazitaxel. It should also be noted that the majority of the utility studies were published prior to the introduction of docetaxel and therefore do not include any improvements in quality of life that may be observed due to improving the outcomes of these patients with this treatment. It should be noted that as in the study patients are eligible for treatment based on biochemical or radiological progression, reflected in the fact that 55% of patients in the COU-AA-301 study did not have pain present at study baseline and 90% of patients	Janssen suggests that the ERG amend this sensitivity analysis to reflect these post-progression utility values and also proposes that the ERG explore a sensitivity analysis whereby the utility assigned to patients in the progression free state is 0.85 as reported in the Krahn (2003) and Volk (2004) studies.	Amendment of the sensitivity analysis to reflect published utility values in this patient population permits a more robust estimate of uncertainty. Values below 0.70 for the progression free state are not aligned with the published literature in prostate cancer or indeed other analogous metastatic cancers.	These are not errors, but legitimate sensitivity analyses. The manufacturer cites as alternative sources: 1. >16months from death in the Sandblom study with a utility of 0.77, 2. the UK subgroup of the Sullivan study with a utility of 0.715, and 3. the progression free utility values used in the NICE submission for cabazitaxel. For the first point it should be borne in mind that the manufacturer model simulates an average life expectancy in the AAP arm of 22 months and in the PP arm of 17 months. These are admittedly over the 16 months to death, the period from which Sandblom reports utility as a function of time to death and as summarised within the ERG report. The Sandblom study was undertaken specifically to estimate the HRQoL among men with prostate cancer in the year prior to death, the method being to estimate their HRQoL at a given time point then relate this to the date of death among those who had died over the period of the study: Sep 199 to Jan 2001. All patients within an area of Sweden with any form of prostate cancer (n=1,442) were sent the questionnaire of whom 86% (n=1243) responded. Only 167 had died by the end of study. It seems reasonable to use the Sandblom estimates to characterise the progression of utility in the last 16 months of life. It seems less reasonable to assume that

had an ECOG performance status of 1.			the baseline utility of the 1,046 patients who had not died by study end were representative of mHRPC patients in the PFS phase with a life expectancy of 16 to 22 months. As summarised within the ERG report the UK subgroup of the mHRPC within the Sullivan paper do have a utility estimate of 0.715 (n=29) but this also has to be read in conjunction with an estimate across all patients of 0.635 (n=280), and with the FACT-P subscore in the manufacturer data being more in line with that across all patients in the Sullivan study than the UK subset. The ERG report for the cabazitaxel submission reports decrements for moving from progression free utility to survival with progression of 0.070 for the base case and 0.085 as a sensitivity analysis. This compares with the manufacturer applied decrement of [redacted]-0.500=[redacted] in the PP arm. Admittedly those receiving cabazitaxel or mitoxantrone may be experiencing adverse events, but it seems likely that not all patients within either arm of the cabazitaxel STA will be being modelled as still being on chemotherapy at the point of progression. The manufacturer also notes that 0.85 as reported in the Krahn (2003) and Volk (2004) studies should be applied as sensitivity analyses. Volk (2004) is an interesting, exploratory paper among men who may be eligible for prostate cancer screening but is not relevant to the decision problem. Krahn (2003) is among prostate cancer patients but it appears to include any ambulatory prostate cancer patients and so be of less relevance to the decision problem than the values summarised within the ERG report. The average standard gamble values reported within Krahn are 0.86 for the PORPUS-U_{SG} and 0.80 for the HUI. The sensitivity analyses undertaken by the ERG are:

			1. A utility for prednisolone progression free survival of 0.500 + ([redacted]) = [redacted] 2. A utility for survival with progression of [redacted] 3. A utility for prednisolone on treatment of 0.715 and for survival with progression of 0.630 The first reflects the decrement between the baseline and end of therapy FACT-P average values reported by the manufacturer in the PP arm to adjust the PP PFS value, the second applies this more straightforwardly. The utility decrement going from PP PFS to PP survival with progression is the same at [redacted] for both these sensitivity analyses. The last applies the upper sensitivity analysis utility decrement of 0.085 from the cabazitaxel STA, even though the central estimate within this STA was a decrement of only 0.070.

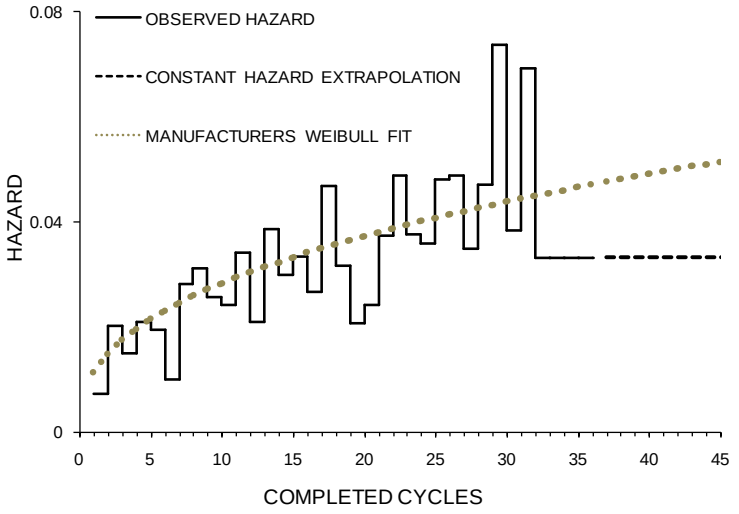
Issue 5 In a sensitivity analysis the ERG suggest assigning a weekly unplanned resource use cost of £106 compared to a fixed cost of [redacted], p20 and p88-89.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
In sensitivity analyses, the ERG applies an unplanned resource use cost weekly (In £106) instead of a fixed cost of [redacted]. This analysis contradicts the results of the empirical medical resource use analysis from the COU-AA-301 study and the advice of the experts consulted by the ERG. Evidence from the COU-AA-301 study and the clinical experts suggests that unplanned resource use is most likely to be focused around the point of progression and therefore it does not seem sensible to apply these costs weekly during whilst patients are in the progression free state. Assigning a weekly cost biases against AAP as patients are in this state longer than PP patients, and contradicts the results of the MRU analysis which demonstrates that patients in the study incurred similar unplanned resource use costs regardless of treatment duration.	Janssen suggests that the ERG remove this sensitivity analysis from the report.	Inclusion of this analysis in the report indicates to the reader that assigning a weekly cost for unplanned resource use is reflective of what would happen in UK clinical practice, whereas both clinical expert opinion and evidence from the MRU analysis in the COU-AA-301 study suggests that these costs are weighted towards the progression event.	The application of the £106 is again a legitimate sensitivity analysis rather than a factual error. The ERG requested resource use data from the manufacturer for the progression free state. This was in order to permit both a cross check and to assess the degree to which the costed unplanned resource use is likely to have in effect been at progression as per the manufacturer assumption. This data was not provided. As a consequence the only sensitivity analysis available relating to this data is to assume a constant average weekly cost over the progression free period. The ERG accepts that this unplanned resource use will be concentrated at the end of the progression free period. The ERG does not accept that it has been demonstrated that there will be no unplanned resource during the progression free period prior to progression. The derivation of the average weekly £106 has not been cross checked by the ERG. But if the value is accepted, applying it equally over the progression free period is accepted by the ERG as a worst case scenario.

Issue 6 Application of Weibull parametric fits to the OS and PFS curves, p 73-77 and p87-88.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
The ERG base case applies Weibull parametric fits to the OS and PFS curves instead of direct use of K-M data from the clinical trial and extrapolating this data using a constant hazard assumption as presented in the original submission. Janssen suggests that the assumptions regarding a constantly increasing hazard implicit in the Weibull are also questionable given the evidence, whilst the log normal and log logistic approaches we agree suggest an untenable assumption of a decreasing hazard in the extrapolation. From a biological plausibility perspective we believe that the increasing hazard is appropriate in the early part of the curve as initial treatment fails and patients progress, however thereafter the force of 'depletion of susceptibles' (high risk patients dying), competes with, or predominates over, the force of increasing age or disease progression – this effect is necessarily evident in the PP arm before the AA arm due to earlier progression and reflected in the curves. Hence, we feel it is not justified to propose the Weibull as the singular preferred analysis. We continue to believe that the constant hazard provides a reasonable intermediate option (between Weibull and log logistic/normal), but feel that at very least the	Janssen proposes that the within this section on pages 73-77 that ERG report acknowledge that the constant hazard extrapolation may also provide a reasonable estimate and that the most plausible ICERs may fall between the estimates generated using the Weibull and those generated using the constant hazard.	This acknowledgement would ensure that the rationale for choosing either approach is not concrete and that there are pros and cons of using either approach.	The rationale for the approach adopted is the AIC and BIC provided by the manufacturer, coupled with the NICE DSU technical support document no. 14 as outlined in the ERG report. The ERG agree that amongst the various models of overall survival offered in the submission the Weibull parametric fit and the constant hazard extrapolation represent the most reasonable options. It follows that, within the context of the models offered, plausible ICERs might fall between estimates made with these models. The convention of identifying a base case necessitates the selection of a "preferred" model. The manufacturer's selection of a constant hazard suffers several weaknesses: [a] arbitrariness of the hazard when 10% patients remain at risk; this is only one of several equally arbitrary values that could be selected, the large affect this has on ICERs is illustrated in Fig 15 of the ERG report (ranging from ~£52k to ~£58k/QALY. The wide fluctuation in hazard around this region of observed data is illustrated in Fig 1 below. [b] The constant hazard is based on only two data (one averaged over a 21 day cycle). In contrast parametric fits make use of all data up to the arbitrary cut off. [c] The calculated constant hazard derives from data toward the end of the observed survival where increasing uncertainty is evident. In contrast the parametric fit additionally uses data from the more robust part of the observed data. The ERG's preferred model used Weibull fits supplied by the manufacturer. Of those made available to ERG these were arguably the most reasonable while observing DSU advice about using similar models for both arms (AAP and PP). It is true that parametric fits to data up to 10% patients at risk uses an arbitrary cut off. In order to explore fits to more or less extensive parts of the observed data required the availability of IPD—unfortunately, despite repeated and early requests

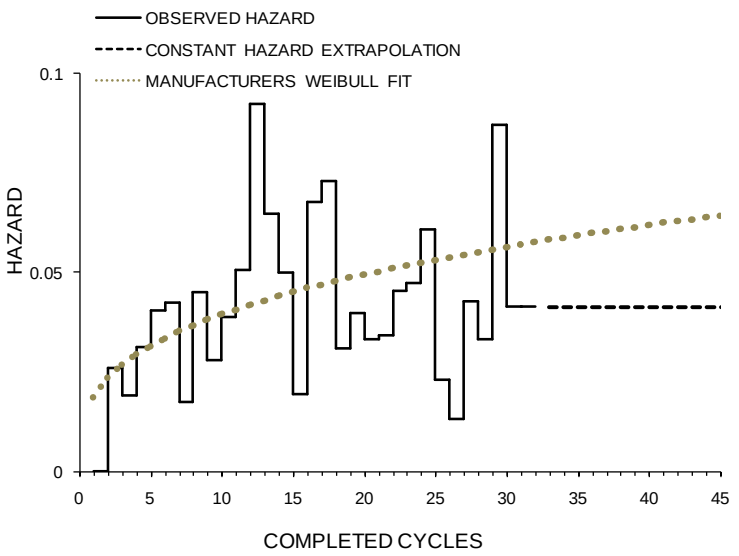
reader should be advised to consider the assumptions implicit in the Weibull and constant hazard and the likelihood that the most plausible ICERs may lie between these two estimates. With regard to the KM curve over-fitting the data, this is a fair criticism if we believe that the underlying mortality mechanisms can be described by the limited set of parameters for a given function. The discussion above suggests this may not be a defensible assumption, in which case the KM plot itself is the best available estimate we have of survival.			for IPD this was not supplied [for information on the exchange of emails see the box below]. The manufacturer's submission did not explore the relative merits of these alternatives beyond stating: "*choice of functional form for the extrapolation could not be unequivocally made based on tests of fit or from visual examination of the curves the exponential appears a reasonable middle ground to optimise the fit"* and, importantly in the context of this issue, "*Data from other mCRPC studies*[11, 12, 29, 56] *support an assumption of a monotonically increasing hazard, such a Weibull function*". ERG do not believe that the ERG report unfairly represents the relative merits of the constant hazard model versus the Weibull parametric; however ERG concede that there are pros and cons for each and that the ERG report did not specifically state that the Weibull model exerted a monotonically increasing hazard. Full analyses are best undertaken using IPD, however as a fall back procedure, given the 21 day cycle K-M data provided in the economic model sheets, it is possible to start to explore the relationship between time and hazard as modelled according to constant hazard and Weibull distribution. The figure below shows the observed hazard, the manufacturer's constant hazard extrapolation, and the manufacturer's Weibull predicted hazard for the AAP one prior chemotherapy population. **FIG 1 Hazard for the AAP one chemo population**



The observed hazard exhibits increasing hazard through time with the rate of increase gradually attenuating. Over the last few observed cycles there is more fluctuation in hazard as the data becomes sparse and uncertainty increases; however the hazard for the last cycle (used in the constant hazard extrapolation) is rather low relative to most of the immediately preceding cycles (e.g. cycle 21 onwards) and as such is probably an underestimate. On the other hand the Weibull fit follows the observed hazard quite well, but a continuing rise in hazard beyond cycle 37, as predicted by the Weibull fit, is only moderately supported. On balance the ERG tend to favour the Weibull fit, however the uncertainty associated with the data means the ERG concede that such a choice cannot be made with great confidence.

Fig 2 provides similar information for the PP one prior chemotherapy group.

FIG 2 Hazard for the PP one chemo population



Because of the smaller number of patients (compared with the AAP arm) there is more fluctuation associated with the observed hazard; this initially increases. It is not easy to choose between a continued rise in hazard or a constant hazard beyond the observed data.

On the basis of hazard alone it is arguable that a Weibull fit should be applied for the AAP arm and constant hazard for the PP arm. This combination would not favour abiraterone relative to the manufacturer's base case. Also this would be contrary to the advice of the NICE DSU regarding survival modelling.

On balance ERG continue to favour the Weibull model, however in view of the uncertainties outlined above, and the manufacturer's comments regarding hazard, in the amended ERG report sent to NICE 12 Dec the ERG have modified the following text on page 15 and page 74 to make

			explicit the fact that the Weibull model imposes monotonically increasing hazard. **Page 15** "For this reason and others described above, the ERG is of the opinion that a more reasonable choice for the base case model would be to use Weibull parametric fits rather than the exponential extrapolation selected by the manufacturer". **Changed to** "For this reason and others described above, the ERG is of the opinion that on balance a more preferred choice for the base case model would be to use Weibull parametric fits rather than the exponential extrapolation selected by the manufacturer. However it is conceded that both models are based on data associated with considerable uncertainties. In contrast to the constant hazard extrapolation the Weibull fit imposes monotonically increasing hazard beyond the observed data so that abiraterone ceases to deliver survival gain relative to placebo beyond 5.47 years (Appendix 7), whereas with the constant hazard model abiraterone continues to deliver survival gain relative to placebo for up to 10 years. The manufacturer has commented on page 97 of the submission that "*Data from other mCRPC studies*[11, 12, 29, 56] *support an assumption of a monotonically increasing hazard, such a Weibull function*". **and page 74** From this the Weibull appears the best fit for modeling OS for the AAP arm, but the log normal appears to be a better fit for modeling OS in the PP arm as can be seen below over the first 37 cycles or 2 years 2 months of the model ([redacted]) . Note that from cycle 29 the number at risk declines to less than 10%, while from cycle 32 the number at risk declines to less than 5%.

			Changed to From this the Weibull appears the best fit for modeling OS for the AAP arm, but the log normal appears to be a better fit for modeling OS in the PP arm as can be seen below over the first 37 cycles or 2 years 2 months of the model ([redacted]). Note that from cycle 29 the number at risk (PP) declines to less than 10%, while from cycle 32 the number at risk declines to less than 5%. Note that in contrast to the constant hazard extrapolation the Weibull fit imposes monotonically increasing hazard beyond the observed data so that abiraterone ceases to deliver survival gain relative to placebo beyond 5.47 years (Appendix 7), whereas with the constant hazard model abiraterone continues to deliver survival gain relative to placebo for up to 10 years.

BOX: **emails referring to requests for IPD.**

Before the manufacturer's submission arrived ERG requested NICE ask for IPD data because ERG anticipated that fitting survival curves would be one critical aspect of the submission. Janssen declined to supply this. The emails to and from NICE about this are shown below

A] From Warwick Evidence to NICE 23 August

I have been given two queries regarding the STA by our team. These are listed below:

1. Since their submission will be based largely on their unpublished longer follow up data can we ask if they would be willing to provide IPD (individual patient data) at the time of the submission rather than us having to wait for this information in response to clarification request? 2. Will their analyses make any adjustment for cross over from placebo to Abiraterone ?

B] Reply from NICE to Warwick Evidence 22 Sept

Dear Amy, Here are the responses we have received from Janssen to your two queries below:

1. Regarding the IPD, Janssen believe that all analyses relating to the Primary Study (COU-AA-301) are incorporated within the submission with sufficient detail and transparency for NICE's decision making requirements; Janssen would be happy to discuss any additional analyses required with the NICE technical team – is this likely to cause any further problems for Warwick evidence?
2. Janssen have indicated that there will not be any adjustment for crossover in this analysis, as crossover to abiraterone was not permitted until after the updated data cut off point.

If you have any further queries please let me know. Many thanks, Matthew

C] From Warwick Evidence to NICE 26 Sept

For cost effectiveness Janssen will need to extrapolate survival data to a 10 year horizon (this is the time horizon stated in their SCOPE document to NICE); ERG will not be able to check any fits submitted without having access to the IPD, furthermore ERG will not be able to explore alternative fits should those provided in the manufacturer's submission be considered inappropriate or insufficiently exploratory. The IPD may well be available within the model Excel sheets; if this is the case then that's fine, ERG will not require additional information, however if these sheets are hidden or are otherwise unavailable ERG would need the trial IPD.

D] Reply from NICE to Warwick Evidence 28 Sept

Please note that if Warwick Evidence still require any IPD that is not available within the economic model and can provide a suitable justification for the request then it can be included as part of your clarification letter to the manufacturer.

Clarification question A2 read

Priority request. Please provide a copy of the full trial report including the updated analysis for trial COU-AA-301. The full trial report has more detailed information than that contained in the submission.

Janssen response read: The CSR for the 'Primary' analysis and report for the 'Updated' analysis are attached. All data within these reports should be considered as AIC.

However IPD for the survival analyses was not included in the 700 page pdf document sent.

Issue 7 Reference to 'Lancet' publication on last paragraph of page 41.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
The primary publication from the COU-AA-301 study was not published in Lancet, but was published in the New England Journal of Medicine.	Please change the 'Lancet' to 'New England Journal of Medicine'	This is incorrect.	ERG have changed the ERG report text appropriately in the amended report sent to NICE 12 Dec.

Issue 8 Commercial in confidence information.

Description of problem	Description of proposed amendment	Justification for amendment	ERG RESPONSE
p72, first paragraph: The statement outlining the ICER without the PAS should be highlighted as CIC.	This paragraph relating to the ICER without the PAS should be highlighted and CIC.	Results generated using the without PAS should be marked as CIC in the document.	ERG have changed the ERG report text appropriately in the amended report sent to NICE 12 Dec.

(please cut and paste further tables as necessary)