janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Janssen's Response to the Appraisal Consultation Document

Prostate cancer (hormone relapsed, metastatic) - enzalutamide (after docetaxel)

Please find below Janssen's response to the Appraisal Consultation Document (ACD) for enzalutamide. We are pleased to have the opportunity to provide our comments in relation to the interpretation of the cost-effectiveness evidence within the ACD and these are detailed below.

Our detailed response to the ACD is split into two main sections. In section 1 we highlight our response with respect to the evidence base used in this appraisal. Specifically, the dataset used by the manufacturer in the indirect comparison between enzalutamide and abiraterone and the assumptions made about hazard ratios and survival projection are all conducted in such a way as to have biased the results against abiraterone. Consequently, Janssen's view is that the summaries of the comparative clinical and cost-effectiveness of enzalutamide and abiraterone within the ACD are not reasonable interpretations of the evidence (Section 2). We propose that the guidance should detail the limitations and uncertainty of the analysis conducted by the manufacturer to ensure that the evidence base is accurately reflected.

Section 1: Has all of the relevant evidence been taken into account?

1.1 The Indirect comparison presented by the manufacturer does not utilise the most comparable datasets (Section 3.11, 3.12 and 3.18 of the ACD)

The robustness of the indirect comparison conducted by the manufacturer is questionable when the comparability of the datasets used and key differences of the two studies (AFFIRM and COU-AA-301) are considered. Clinical evidence used in the indirect comparison of overall survival (OS) should be taken from time points in the two trials (AFFIRM and COU-AA-301) with a similar follow-up duration given that treatment effects have been shown to be time-dependent in this stage of mCRPC (see details below). In addition, any indirect comparison should account for differences in the patient populations and treatment experience between the two studies that might impact treatment effect.

The indirect comparison presented by the manufacturer compared the final analysis from the AFFIRM study with the final analysis from COU-AA-301 (see Table 1). However, the follow-up duration is 5.2 months longer in the COU-AA-301 trial compared to the AFFIRM study.

Several key studies, including the COU-AA-301 trial, have demonstrated that the treatment effect on overall survival (OS) is time-dependent (COU-AA-301, TROPIC and TAX327[1-4]). This is demonstrated by converging overall survival curves, which indicated a diminishing treatment effect in late stage mCRPC. Therefore, hazard ratios (HR) derived over different follow-up durations are not comparable. For the purposes of conducting an indirect comparison, a similar follow-up period for comparison is available from the interim analysis or re-analysis of the final data at similar time points. At the interim analysis time points, the proportion of patients who had died in the placebo arms (55% and 53%) and the average length of follow-up (12.8 vs. 14.4 months) were more comparable.

Table 1 Overall Survival Hazard Ratios from Phase III Clinical Trials ITT populations

	Analysis (average follow-up)	Number (%) of total patients who had died in the placebo arm	Difference in median OS between study arms	HR (95% CI)
Enzalutamide vs. BSC, AFFIRM (n=1199)	Interim (14.4 months)	212 out of 339 (53%)	4.8mth (18.4 vs. 13.6)	0.63 (0.53, 0.75)
	Final (15.0 months)	232 out of 399 (58%)	4.5mth (17.8 vs. 13.3)	0.62 (0.52, 0.73)
Abiraterone vs. BSC, COU-AA-301 (n=1195)	Interim (12.8 months)	219 out of 398 (55%)	3.9mth (14.8 vs. 10.9)	0.65 (0.54, 0.77)
	Final (20.2 months)	274 out of 398 (69%)	4.6mth (15.8 vs. 11.2)	0.74 (0.64, 0.86)

As the follow-up for Final analysis of COU-AA-301 is much longer (and far more complete) than the AFFIRM Final analysis, an indirect analysis of based on the final data is biased against abiraterone. Janssen proposes that using the interim datasets is the appropriate approach as survival for the treatment arm and control arm tend to converge over a longer follow-up.

Janssen also suggests that the manufacturer's indirect comparison does not accurately report the key differences between the trials. The differences relate to patient baseline characteristics, the treatments permitted as best supportive care (BSC) and subsequent treatment. Recent independent analysis[5] has explored adjusting the survival curves from these studies to account for differences in study populations between the studies; an approach which is valid considering:

- The inclusion criteria specified for AFFIRM may have lead to a fitter group of patients than were enrolled in the COU-AA-301 study. Specifically, the AFFIRM study had stricter exclusion criteria for cardiovascular and neurological disease and a minimum life expectancy of 6months was a screening criterion.
- The ERG commented that patients in AFFIRM and COU-AA-301 were generally comparable, but noted some differences in baseline characteristics, with COU-AA-301 having slightly older patients and more patients who had received more than one chemotherapy regimen. In addition, a larger proportion of patients had a Brief

Pain Inventory (BPI) score ≥ 4 and radiographic progression at baseline in the COU-AA-301 study compared with AFFIRM and the median baseline PSA was greater in COU-AA-301.[6] All these factors are suggestive of a patient population with more advanced disease in the COU-AA-301 study.

- In the COU-AA-301 study, BSC did *not* allow the use of radiopharmaceuticals: in contrast radiopharmaceuticals (e.g. strontium-89 or samarium-153) were permitted as part of BSC in the AFFIRM study. Although radiopharmaceuticals are not proven to offer a clinical benefit in survival or delayed disease progression in this patient population, it would be useful to describe how many patients received this treatment to ensure the BSC permitted in each of the studies was comparable.

Utilising the results at comparable time points shows there is no difference between the two treatments, even without controlling for the more severe population in COU-AA-301. In the ITT population, an HR of 0.63 (95% CI) 0.53 -0.75 for enzalutamide and 0.65 (95% CI) 0.54 -0.77 for abiraterone acetate were reported at the interim data cut. An indirect treatment comparison based on this data cut estimated an HR of 1.03; 95% CI 0.80-1.32 (abiraterone acetate vs. enzalutamide) (Janssen Data on File) which shows that the efficacy on OS is not different. These results are in concordance with the opinion of several clinical experts consulted for the manufacturer's submission (p187 of the manufacturer submission); in that both treatments have an equal clinical benefit for PFS and OS. In an alternative analysis, an indirect comparison of the Interim datasets from the two trials has recently been presented (Li et al (2013)).[6] This analysis addressed the difference in BSC comparator arms, by using the data from all COU-AA-301 patients and data from patients receiving corticosteroids concurrently in the AFFIRM trial. The results of the ITC for abiraterone vs. enzalutamide showed no significant difference in OS, using both the Bucher method (HR 0.949; 95% CI 0.712-1.265) or Bayesian statistics (HR 0.948; 95% CI 0.711-1.26)).

Janssen suggests that more details surrounding the indirect comparison (briefly discussed in Section 4.6), specifically around the limitations surrounding this analysis, should be provided. In Section 4.6 of the ACD, the only limitation of the indirect comparison that is discussed is the use of corticosteroids. Janssen suggests that other limitations of the manufacturer's indirect comparison, namely differences in follow-up duration of the datasets used and differences between the trials (patient baseline characteristics, the treatments permitted as best supportive care (BSC) and subsequent treatment) be presented. Specifically, a more robust analysis would be needed, utilising comparable follow-up time points from COU-AA-301 and AFFIRM to make firmer conclusions.

1.2 Assumptions made by the manufacturer about the HR for OS and treatment duration in the economic analysis are subject to significant uncertainty (Sections 4.9 and 4.10 of the ACD)

To estimate OS and progression-free survival, the manufacturer uses BSC as a 'reference' treatment. Alternative parametric functions were then fitted (exponential, Weibull, log-logistic, log-normal and gamma) to patient-level data for the placebo group in AFFIRM (representing best supportive care) and curves extrapolated beyond the end of the trial. The manufacturer stated that, unlike the hazard ratio for abiraterone compared with placebo, the hazard ratio for enzalutamide compared with placebo remained constant over time (meeting the proportional hazards assumption). The time between the interim and final analyses of the AFFIRM trial is too short (0.6 months) to support the manufacturer's argument. The follow-up duration at the final analysis of the AFFIRM trial is only 15.0 months, and unless the manufacturer has additional data with longer-term follow-up to support this argument, conclusions about the AFFIRM HR ratio remaining constant over a longer time are not supported. The HR for COU-AA-301 for the interim OS analysis also remained fairly constant over a similar follow-up period (12.8 months) and it is only when longer follow-up is observed that the proportional hazards requirement is less likely to be met. Assuming a constant HR for enzalutamide based on the short follow-up, while assuming otherwise for abiraterone based on the long-term follow-up is not a robust approach in the absence of further data.

The application of AFFIRM trial OS data is inconsistent in that the interim data is used to derive the OS curve for the placebo arm but the HR (0.618) from the final data cut is applied on this reference curve to derive the OS curve for the enzalutamide arm. In the base case, the manufacturer estimated the OS for abiraterone by applying a time-dependent hazard ratio to the survival function for BSC (the 'reference' treatment). In one scenario, the manufacturer assumed a HR of 0.74 for abiraterone based on a 20 months follow-up but a HR of 0.618 for enzalutamide based on 15 months' follow-up, which as previously discussed unfairly favours enzalutamide. In the other scenario, the manufacturer estimated a HR based on mean OS, which would favour enzalutamide in the analysis. As reported by the two phase III clinical trials, more patients in the AFFIRM study received subsequent treatments with a proven beneficial effect on OS (abiraterone or cabazitaxel) compared to COU-AA-301; 31% vs. 0% of patients in the active treatment arms (Table 2).

Table 2. Subsequent Treatments received in the active and placebo arms of COU-AA-301 and AFFIRM trials

	COU-AA-301 (Interim Analysis 12.8 months)		AFFIRM (Interim Analysis 14.4months)	
Subsequent treatment (% of patients)	**AAP (n=797)**	**BSC (n=398)**	**Enzalutamide (n=800)**	**BSC (n=399)**
Total	26%	36%	42%	61%
Abiraterone	0%	0%	*21%*	*24%*
Cabazitaxel	0%	0%	*10%*	*14%*
Docetaxel	*8%*	*9%*	*9%*	*14%*
Mitoxantrone	*5%*	*7%*	*3%*	*11%*

The longer mean OS for enzalutamide arm of the AFFIRM trial compared to abiraterone arm of the COU-AA-301 trial is likely to be caused at least partially by the use of active subsequent treatments. As explained above, the OS assumptions in the scenario analysis were biased against abiraterone, yet the ICERs in these scenario analyses were worse than in the base case which increased the base-case for enzalutamide compared with abiraterone from £14,795 to £19,972 and £18,034 per QALY gained respectively. Therefore, these scenario analyses indicate that the OS projection method in the base case is likely to *underestimate* the OS benefit for abiraterone.

To estimate the treatment duration hazard ratio for abiraterone, the manufacturer used data from abiraterone's post-chemotherapy NICE technology appraisal (TA259). These data, however, related to the subgroup of patients who had received only one chemotherapy regimen and not the whole population that was used for OS. The manufacturer estimated a hazard ratio of 0.52 (95% CI 0.44 to 0.60) for abiraterone, and used this ratio to derive the progression-free survival function for abiraterone. In an analysis of the COU-AA-301 trial data (following a similar approach to derive the HR of treatment discontinuation for enzalutamide vs. placebo) a Weibull function was fitted to treatment discontinuation data to both treatment arms, including treatment as a covariate. The HR ratio vs. placebo would be [redacted] at the interim data cut and [redacted] at the final data cut. The 0.52 estimated by the manufacturer and 0.49 estimated by the committee overestimates the treatment duration for abiraterone.

Janssen is concerned that a number of assumptions made in the manufacturer's submission about the OS projection and the hazard ratios for OS and treatment duration. This has the consequence of overestimating the cost-effectiveness of enzalutamide relative to abiraterone to a significant degree. Janssen requests that the limitations of the flawed indirect comparisons undertaken by the manufacturer be taken into consideration in the FAD and that the interpretation of the modelling should emphasise much more clearly the limitations of the approach taken and the potential biases in favour of enzalutamide.

Section 2: Are the summaries for clinical and cost-effectiveness reasonable interpretations of the evidence?

2.1 Given the issues outlined in Section 1, it is inappropriate for the committee to draw conclusions about the comparative clinical and cost effectiveness of enzalutamide and abiraterone on the basis of the evidence presented to the committee (Section 4.16 of the ACD).

Given the limitations outlined above, Janssen proposes that the only appropriate conclusion is that both abiraterone and enzalutamide are cost-effective options compared to BSC, but that the comparative cost-effectiveness of these two treatments cannot be established without further evidence or a reanalysis of the data as described above.

Section 3: Are the provisional recommendations sound and a suitable basis for guidance to the NHS?

No further comments.

References

1. de Bono,J.S. et al. Abiraterone and increased survival in metastatic prostate cancer. N. Engl. J. Med. 364, 1995-2005 (2011).

2. de Bono,J.S. et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet 376, 1147-1154 (2010).

3. Fizazi,K. et al. Final Overall Survival (OS) Analysis of COU-AA-301, a Phase 3 Study of Abiraterone Acetate Plus Prednisone in Patients with Metastatic Castration-Resistant Prostate Cancer (mCRPC) Pretreated with Docetaxel. ESMO. 2011. Ref Type: Conference Proceeding

4. Tannock,I.F. et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N. Engl. J. Med. 351, 1502-1512 (2004).

5. Wilson,L. et al. New Therapeutic Options in Metastatic Castration-Resistant Prostate Cancer: Can Cost-Effectiveness Analysis Help in Treatment Decision. 2013. Ref Type: Conference Proceeding

6. Li,T. et al. Abiraterone and Enzalutamide for the Treatment of Metastatic Castration-Resistant Prostate Cancer Post Chemotherapy: An Indirect Comparison and Budget Impact Analysis. ISPOR 16th Annual European Congress. 2013. 2-11-2013. Ref Type: Conference Proceeding