in collaboration with:

Abiraterone for the treatment of chemotherapy naïve metastatic castration-resistant prostate cancer

Produced by Kleijnen Systematic Reviews Ltd in collaboration with Erasmus University Rotterdam and Maastricht University

Authors
Rob Riemsma, Reviews Manager, KSR Ltd
Bram Ramaekers, Health Economist, Maastricht UMC, The Netherlands
Florian Tomini, Health Economist, Maastricht UMC
Robert Wolff, Reviews Manager, KSR Ltd
Thea van Asselt, Health Economist, Maastricht UMC
Manuela Joore, Health Economist, Maastricht UMC
Sohan Deshpande, Systematic Reviewer, KSR Ltd
Gill Worthy, Statistician, KSR Ltd
Steven Duffy, Information Specialist, KSR Ltd
Nigel Armstrong, Health Economist, KSR Ltd
Johan L. Severens, Professor of Evaluation in Health Care, EUR
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to Rob Riemsma, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 18/03/2014

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 11/54/01 STA.

Declared competing interests of the authors
None.

Acknowledgements
None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.
Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Rider on 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:
Riemsma R, Ramaekers BLT, Tomini F, Wolff R, van Asselt ADI, Joore MA, Deshpande S, Worthy G, Duffy S, Armstrong N, Severens JL, Kleijnen J. Abiraterone for the treatment of chemotherapy naïve metastatic castration-resistant prostate cancer: a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd, 2013.

Contributions of authors
Rob Riemsma acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Bram Ramaekers acted as health economic project lead, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Florian Tomini, Thea van Asselt, Manuela Joore and Nigel Armstrong acted as health economists on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Robert Wolff and Sohan Deshpande acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the manufacturer's submission and contributed to the writing of the report. Steven Duffy critiqued the search methods in the submission and contributed to the writing of the report. Jos Kleijnen and Johan Severens provided senior advice and support to the clinical and cost-effectiveness sections, respectively, and were involved in drafting and/or commenting on the report.

ABBREVIATIONS

AA	Abiraterone acetate
AAP	Abiraterone acetate plus prednisone/prednisolone
ADRs	Adverse drug reactions
ADT	Androgen deprivation therapy
AE	Adverse Events
AJCC	American Joint Committee on Cancer
ALT	Alanine aminotransferase
ASCO	American Society of Clinical Oncology
AST	Aspartate aminotransferase
AUA	American Urological Association
BAP	Bone Alkaline Phosphatase
bd/b.i.d	Twice Daily
BSC	Best supportive care
CE	Cost-effectiveness
CEA	Cost-effectiveness Analysis
CEAC	Cost-effectiveness Acceptability Curve
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
CMC	Clinically Meaningful Changes
CR	Complete response
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CRPC	Castration-resistant prostate cancer
CSR	Clinical study report
DES	Discrete event simulation
EAU	European Association of Urology
ECOG PS	Easter Cooperative Oncology Group Performance Status
EMEA	European Medicines Agency
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
EUR	Erasmus University Rotterdam
FACIT-Sp	Functional Assessment of Chronic Illness Therapy Scales – Spirituality subscale
FACT-G	Functional Assessment of Cancer Therapy General
FACT-P	Functional Assessment of Cancer Therapy – Prostate
FKSI-DRS	Functional Assessment of Cancer Therapy Kidney Symptom Index -Disease Related Symptoms
HR	Hazard ratio
HRQL	Health-related Quality of Life
HTA	Health Technology Assessment
IC	Indirect Comparison

ICER	Incremental Cost-effectiveness Ratio
ITT	Intention to Treat
KPS	Karnofsky Performance Status
KSR	Kleijnen Systematic Reviews
LYS	Life Year Saved
MAH	Marketing authorization holder
mCRPC	Metastatic castration-resistant prostate cancer
mg	Milligram
MID	Minimally Important Difference
MRU	Medical resource utilisation
MS	Manufacturer's Submission
MSKCC	Memorial Sloan-Kettering Cancer Centre
MTC	Mixed Treatment Comparison
NCCN	National Comprehensive Cancer Network
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NR	Not Reported
od	Once Daily
OR	Odds Ratio
ORR	Objective response rate
OS	Overall survival
P	Prednisone
PAS	Patient access scheme
PCT	Primary Care Trust
PDGFR	Platelet-derived growth factor receptor
PP	Placebo plus prednisone/prednisolone
PR	Partial response
PRESS	Peer Review of Electronic Search Strategies
PSA	Probabilistic Sensitivity Analyses
PSA	Prostate-specific antigen
PSS	Personal Social Services
QALY(s)	Quality-adjusted Life Year(s)
RCC	Renal Cell Carcinoma
RCT	Randomised Controlled Trial
RECIST	Response Evaluation Criteria in Solid Tumours
RR	Relative Risk
SAE	Serious Adverse Events
SmPC	Summary of Product Characteristics
STA	Single Technology Appraisal
TEAEs	Treatment-Emergent Adverse Events
TNM	Tumour Node Metastases system
TTP	Time to disease progression
UMC	University Medical Centre

VEGFR Vascular endothelial growth factor receptor

Table of Contents

ABBREVIATIONS **3**
TABLE OF TABLES **7**
TABLE OF FIGURES **9**
1. SUMMARY **10**
1.1 Critique of the decision problem in the manufacturer's submission 10
1.2 Summary of clinical effectiveness evidence submitted by the manufacturer 10
1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted 11
1.4 Summary of cost-effectiveness submitted evidence by the manufacturer 12
1.5 Summary of the ERG's critique of cost-effectiveness evidence submitted 14
1.6 ERG commentary on the robustness of evidence submitted by the manufacturer 15
1.6.1 Strengths 15
1.6.2 Weaknesses and areas of uncertainty 15
1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG 16
2. BACKGROUND **17**
2.1 Critique of manufacturer's description of underlying health problem. 17
2.2 Critique of manufacturer's overview of current service provision 19
3. CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM **22**
3.1 Population 24
3.2 Intervention 24
3.3 Comparators 25
3.4 Outcomes 26
3.5 Other relevant factors 28
4. CLINICAL EFFECTIVENESS **29**
4.1 Critique of the methods of review(s) 29
4.1.1 Searches 29
4.1.2 Inclusion criteria 33
4.1.3 Critique of data extraction 35
4.1.4 Quality assessment 36
4.1.5 Evidence synthesis 39
4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these) 39
4.2.1 Study characteristics of the COU-AA-302 trial 39
4.2.2 Patient characteristics in the COU-AA-302 trial 47
4.2.3 Results of the COU-AA-302 trial 48
4.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison 60
4.4 Critique of the indirect comparison and/or multiple treatment comparison 61

4.5 Additional work on clinical effectiveness undertaken by the ERG 61
4.6 Conclusions of the clinical effectiveness section 64
5. COST-EFFECTIVENESS 66
5.1 ERG comment on manufacturer's review of cost-effectiveness evidence 66
5.1.1 Objective of cost effectiveness review 66
5.1.2 Inclusion/exclusion criteria used in the study selection 66
5.1.3 Included/excluded studies in the cost-effectiveness review 67
5.1.4 Conclusions of the cost-effectiveness review 67
5.2 Summary and critique of manufacturer's submitted economic evaluation by the ERG 68
5.2.1 NICE reference case checklist (TABLE ONLY) 71
5.2.2 Model structure 73
5.2.3 Population 76
5.2.4 Interventions and comparators 77
5.2.5 Perspective, time horizon and discounting 77
5.2.6 Treatment effectiveness and extrapolation 77
5.2.7 Health related quality of life 87
5.2.8 Resources and costs 91
5.2.9 Cost effectiveness results 94
5.2.10 Sensitivity and scenario analyses 98
5.2.11 Subgroup analyses 101
5.2.12 Model validation and face validity check 101
5.3 Exploratory and sensitivity analyses undertaken by the ERG 103
5.4 Conclusions of the cost-effectiveness section 106
6. IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG 108
7. END OF LIFE 110
8. OVERALL CONCLUSIONS 113
8.1 Implications for research 113
9. REFERENCES 115
APPENDIX 1: FURTHER CRITIQUE OF MANUFACTURERS SEARCHES 122
APPENDIX 2: PHILLIPS ET AL CHECKLIST 125

TABLE OF TABLES

Table 2.1: Adverse reactions identified in clinical studies and post-marketing 21

Table 3.1: Statement of the decision problem (as presented by the manufacturer) 22

Table 4.1: Inclusion and exclusion criteria used in the systematic review 33

Table 4.2: Data sources for the pivotal RCT, COU-AA-302 36

Table 4.3: Quality assessment of COU-AA-302 36

Table 4.4: Overview of the COU-AA-302 trial 39

Table 4.5: Summary of methodology of the COU-AA-302 trial 39

Table 4.6: COU-AA-302 study planned rPFS and OS analyses 42

Table 4.7: Eligibility criteria of the COU-AA-302 trial 43

Table 4.8: Treatment discontinuations at the third interim OS analysis (22.05.12; Safety population) 47

Table 4.9: Characteristics of participants in the COU-AA-302 trial by randomised group (ITT) 47

Table 4.10: OS of patients treated with either AAP or PP (ITT) 49

Table 4.11: rPFS in patients treated with either AAP or PP (ITT) 50

Table 4.12: Best overall response (confirmed) based on modified RECIST criteria in patients with measurable disease at baseline (investigator review – ITT population – third interim analysis – 22.05.12, 55% data cut-off) 52

Table 4.13: Response to treatment outcomes results ITT population (third interim analysis – 22.05.12, 55% data cut-off) 52

Table 4.14: HRQL outcomes results – ITT population (third interim analysis – 22.05.12, 55% data cut-off) 53

Table 4.15: Pain-related outcomes results – ITT population (third interim analysis – 22.05.12, 55% data cut-off) 54

Table 4.16: Overall safety results – safety population (third interim analysis – 22.05.12, 55% data cut-off) 55

Table 4.17: Incidence of all AEs occurring in ≥5% of patients in either group (COU-AA-302 study third interim analysis, 55% data cut-off) 56

Table 4.18: Incidence of grade 3 or 4 AEs occurring in ≥1% of patients across randomised groups (COU-AA-302 study third interim analysis, 55% data cut-off) 58

Table 4.19: Incidence of AEs of special interest (>5% in either arm) (COU-AA-302 study third interim analysis, 55% data cut-off) 59

Table 4.20: OS of patients treated with either AA or placebo (ITT) (interim analysis 3); subgroup of patients with subsequent docetaxel treatment 61

Table 4.21: rPFS in patients treated with either AA or placebo (ITT) (interim analysis 3); subgroup of patients with subsequent docetaxel treatment 62

Table 5.1: Inclusion criteria for the economic evaluations systematic review 66

Table 5.2: Summary of the manufacturer's economic evaluation (with signposts to MS[1]) 68

Table 5.3: treatment pathways excluded from the model 79

Table 5.4: Summary of model results compared with clinical data (based on the ITT population) 80

Table 5.5: Results of the FACT-P to EQ-5D Mapping Study 87

Table 5.6: Summary of the utility values associated with each model phase 88

Table 5.7: Unit costs associated with the technology in the economic model 91

Table 5.8: Scheduled oncologist MRU costs 92

Table 5.9: Trial-based unplanned MRU costs per month 92

Table 5.10: List of AEs and summary of costs included in the economic model 93

Table 5.11: Summary of proportion of patients and respective duration in each treatment phase 96

Table 5.12: Summary of LY and QALY gain by pre/on/post-docetaxel status 97

Table 5.13: Summary of costs by health state disaggregated by category of cost 97

Table 5.14: Base-case results (with PAS) 97

Table 5.15: ERG revised summary of the PSA (Manufacturer base case)[a] 98

Table 5.16: Comparison of post-docetaxel survival with TA 259 100

Table 5.17: Updated prediction equations[a] 103

Table 5.18: Summary of LY and QALY gain by pre/on/post-docetaxel status (ERG base case) 104

Table 5.19: Summary of the PSA (ERG base case) 105

Table 6.1: Overview of additional analyses undertaken by the ERG 108

Table 6.2: Additional sensitivity analyses (based on ERG base case) 109

Table 7.1: End of life criteria for abiraterone 110

TABLE OF FIGURES

Figure 4.1: COU-AA-302 study patient flow diagram (data from third interim analysis) 46

Figure 4.2: Kaplan–Meier curve of OS – ITT population (COU-AA-302 study third interim analysis) 50

Figure 4.3: Kaplan–Meier curve of rPFS – ITT population (third interim analysis – 22.05.12, 55% data cut-off) 51

Figure 4.4: Overall survival, AA versus placebo, stratified by subsequent docetaxel treatment 62

Figure 4.5: rPFS, AA versus placebo, stratified by subsequent docetaxel treatment 63

Figure 5.1: Model pathway 74

Figure 5.2: Patient flow through model simulation 75

Figure 5.3: Time to AAP/BSC (PP) discontinuation for the ITT population versus the analysable dataset 78

Figure 5.4: Overall survival for the ITT population versus the analysable dataset 78

Figure 5.5: Time from post-docetaxel treatment discontinuation to death 84

Figure 5.6: Time from docetaxel start to docetaxel discontinuation 85

Figure 5.7: OS comparison of model prediction and COU-AA-302 trial results 86

Figure 5.8: ERG revised cost-effectiveness acceptability curve (Manufacturer base case)[a] ... 98

Figure 5.9: Tornado diagram (deterministic 1-way sensitivity analyses) 99

Figure 5.10: Cost-effectiveness acceptability curve (ERG base case) 105

Figure 7.1: Kaplan–Meier curve of OS – ITT population (COU-AA-302 study third interim analysis) (See MS, Figure 6, page 68) 112

1. SUMMARY

1.1 Critique of the decision problem in the manufacturer's submission

The patient population described in the final scope is "Adults with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated".

Abiraterone with prednisone or prednisolone is indicated for the treatment of mCRPC in adult men who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy (ADT) in whom chemotherapy is not yet clinically indicated.

The main deviation from the scope is that docetaxel is not included as a comparator in the manufacturer submission. As the indication is men with mCRPC **in whom chemotherapy is not yet clinically indicated**, it seems reasonable that docetaxel is not considered as a comparator. However, in the final scope, NICE explicitly states that: "Docetaxel is included in the list of comparators because the recommendations in TA101 include patients who are asymptomatic or mildly symptomatic, and clinicians have stated that docetaxel is increasingly used for this patient group, and because of the lack of clear clinical criteria to identify the patient group in the CHMP indication".

Assuming that most patients will end up using docetaxel, which also seems to be implied by the phrase "**not yet** clinically indicated", an important question in this appraisal, according to the ERG, is whether abiraterone followed by docetaxel is more effective than watch-full waiting (BSC) followed by docetaxel. In the COU-AA-302 trial, 239 out of 546 (43.8%) of AAP patients and 304 out of 542 (56.1%) of PP patients received docetaxel as subsequent therapy, following abiraterone or placebo. The results for this specific group of patients are not presented in the MS; therefore, we asked the manufacturer to provide these data in the clarification letter.

According to the manufacturer abiraterone meets the criteria for appraisal of end of life medicines. However, looking at the COU-AA-302 trial data it is unlikely that life expectancy in this patient group will be less than 24 months. According to the manufacturer, patients in the trial are likely to have gone on to receive other clinical trial technologies post-docetaxel and therefore the survival observed for these patients is probably not reflective of the average mCRPC patient in the UK. However, as far as the ERG is aware the "short life expectancy, normally less than 24 months" is based on the normal treatment options available for these patients without the intervention under assessment.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

One RCT (the COU-AA-302 trial) is included for the comparison of abiraterone acetate in combination with prednisolone versus best supportive care.

In the COU-AA-302 trial, a total of 1,088 patients were recruited and randomised to abiraterone acetate plus prednisone/prednisolone (n=546) or placebo plus prednisone

(n=542). One thousand and eighty-two patients received at least one dose of the allocated intervention and constituted the safety population. Patients continued treatment with AAP or PP until disease progression (determined according to radiographic and clinical measures). The median treatment duration was 13.8 months (15 cycles initiated) in the AAP arm and 8.3 months (nine cycles initiated) in the PP arm.

Results presented in the MS[1] are based on the results from the second (data cut-off 20/12/11) and third (data cut-off 22/5/12) interim analyses of the COU-AA-302 study, which were conducted after approximately 40% and 55% of the total OS events had occurred.

Neither the second nor third interim analysis overall survival results met the pre-specified statistical significance levels (HR at third interim analysis: 0.79 (95% CI: 0.66, 0.96). Median overall survival was 35.3 months (95% CI: 31.2, 35.3) in the AAP group and 30.1 months (95% CI: 27.3, 34.1) in the PP group. The manufacturer did not provide mean survival for both groups or mean survival gain, despite explicit questions in the clarification letter.

[redacted].

Treatment with AAP resulted in a 48% relative reduction in the risk of radiographic progression compared with PP (absolute risk reduction 11.5%), and increased PFS by 8.2 months. Significant differences in favour of the AAP group were observed for objective response rate (complete or partial response according to modified RECIST criteria), PSA response and duration of response. HRQL was assessed in the COU-AA-302 study via the FACT-P instrument. However, no results are report by treatment arm for baseline, follow-up or change scores. Time to progression in average pain intensity and worst pain intensity showed no significant differences between treatment arms. All other pain-related outcomes favoured AAP over PP.

Adverse events were significantly more often reported in the AAP arm when compared with the PP arm for treatment-emergent adverse event (TEAEs), Drug-related grade 3–4 TEAEs, treatment-emergent serious adverse event (SAEs) and Grade 3–4 treatment-emergent SAEs. The most frequently reported AEs were fatigue (39.7% AAP vs. 34.6% PP), back pain (33.2% vs. 33.1%), arthralgia (29.3% vs. 24.4%), nausea (24.0% vs. 23.0%), peripheral oedema (26.0% vs. 20.9%), constipation (23.6% vs. 20.4%), diarrhoea (23.4% vs. 18.1%) and hot flush (22.7% vs. 18.3%). AAP resulted in significantly more grade 3 or 4 increased ALT, increased aspartate aminotransferase, and dyspnoea; but less hydronephrosis.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

According to the manufacturer, the Independent Data Monitoring Committee (IDMC) concluded on 27 February 2012, that patients in the abiraterone arm had a 'highly significant

advantage', even though the hazard ratio (HR) for OS had not reached the stringent pre-specified statistical significance level (0.0034). The committee unanimously recommended stopping the study, unblinding, and allowing cross-over. The study was unblinded on 2 April 2012. Cross-over from PP to AAP occurred following unblinding (02.04.12) for three patients by the third interim analysis (22.05.12). Neither the second nor third interim analysis OS results met the pre-specified statistical significance levels (see results below). Because cross-over is now allowed, it is unlikely that the trial will ever show a significant survival benefit.

According to the manufacturer, the population eligible to participate in COU-AA-302 is not mutually exclusive from the population who could receive docetaxel in clinical practice, because the population determined by the license was requested by the regulators on the basis of the study results, rather than as a result of the study being designed to specifically for patients who are not yet suitable for docetaxel. Therefore, it is possible that there are patients included in COU-AA-302 for whom docetaxel may have been considered suitable in routine UK practice. Although specifically asked in the clarification letter, the manufacturer did not provide the number of patients for whom this might be the case.

1.4 Summary of cost-effectiveness submitted evidence by the manufacturer

The literature search for relevant cost-effectiveness studies was appropriate. However, it did not identify any studies on AAP for the treatment of adult men who are asymptomatic or mildly symptomatic after failure of ADT and in whom chemotherapy is not yet clinically indicated. Therefore, a de novo economic analysis was performed.

The manufacturer presented a comparison of AAP versus BSC, by means of a discrete event simulation (DES) model, tracking patients at the individual level. The model follows patients until age 100, which is assumed to reflect a lifetime time horizon. Patients entering the model were assigned to either the AAP or the BSC strategy. Patients who discontinue pre-docetaxel active treatment or progress are monitored in a BSC phase before starting docetaxel. After the docetaxel treatment phase, patients are monitored in a BSC phase for progression again upon which they could receive active treatment (AAP) if deemed appropriate. However, patients who had already received AAP in the 1st line were not eligible for re-treatment with AAP post-docetaxel. After all treatment options had been explored and disease has progressed, patients then enter a palliative stage (before death).

The model consisted of a total of 17 prediction equations for estimating time to treatment discontinuation, time to treatment start, and time to death within the various treatment phases, and also to estimate disease status of the patient. To estimate these prediction equations, study data of 902 patients were used (83% of the intention to treatment population which consisted of 1,088 patients). Various covariates were included in these prediction equations, chosen largely on the basis of statistical significance, although non-significant covariates were inconsistently included in some cases. These prediction equations were combined with the profile/characteristics of individual patients to estimate the exact treatment path, including duration in the various treatment phases, and survival.

Although utility data were obtained (indirectly via mapping FACT-P results) from the COU-AA-302 trial, utility values in the base case model came from a UK mCRPC utility study, performed by means of an online survey among 163 patients. Only the base case on-treatment utility increment of AAP over BSC (pre-docetaxel) was obtained from the COU-AA-302 trial. For all other treatment phases, FACT-P based utilities were included in a scenario analysis. Adverse events were not separately taken into account in the utility score as the safety profile of AAP and BSC is considered similar, and all other effects of treatment (e.g. docetaxel) on HRQoL would have been captured in the treatment-phase specific utility value. For AAP post-docetaxel, unlike in TA259[2], no on-treatment utility increment was applied since post-docetaxel patients in the UK mCRPC utility study were assumed to be receiving AAP (as part of current clinical practice) and therefore the on-treatment benefit was already captured in the utility value.

Costs were subdivided into treatment costs, costs of scheduled medical resource utilisation (MRU), and costs of unplanned MRU (including AEs). Monthly treatment costs for AA are [redacted] (including a PAS discount of [redacted]) which is considerably higher than the cost for BSC, which was represented by 10 mg of prednisolone daily and is therefore negligible. The monthly cost of docetaxel, including administration costs, is £1,550. Scheduled MRU was assessed by means of a survey among 53 UK oncologists, with questions on total outpatient visits, scans, and laboratory tests. For AA patients both pre- and post-docetaxel, a higher MRU is applied until three months after start of treatment because they require additional monitoring. Unplanned events while on treatment were estimated, where possible, based on the COU-AA-301 and COU-AA-302 trial data. However, since these trials did not contain data on pre- and post-docetaxel BSC nor on docetaxel, or BSC before death, unplanned MRU of proxy groups had to be used for these treatment phases. For pre- and post-docetaxel phases, treatment of adverse events was considered to be included in the unplanned MRU. Costs of incremental grade 3 or 4 adverse events for docetaxel as compared to AAP were assigned separately. Resources and medication used for treating these AEs were assessed by means of expert opinion.

The base case deterministic ICER for AAP versus BSC was £46,722, based on incremental costs of £26,404 and incremental QALYs of 0.57. The probability that AAP is cost-effective compared to BSC for thresholds of £30,000, £40,000 and £50,000 is 0%, 10% and 67% respectively. The one-way sensitivity analyses show the most influential parameters to be the post-ADT baseline utility (but the uncertainty of this parameter is limited), and the discount rate for the health benefits. In addition, 11 scenario analyses were performed on various assumptions, such as excluding the PAS discount for AA, or using urologist instead of oncologist inputs for the scheduled MRU costs. When excluding the PAS, and also in the scenario where FACT-P mapping utilities were used instead of EQ-5D from the patient utility study, this resulted in ICERs above £50,000 per QALY. For all other scenarios, ICERs would be lower than £50,000.

1.5 Summary of the ERG's critique of cost-effectiveness evidence submitted

Regarding the model structure, the ERG does not believe that a DES model, simulating individual patients by means of 17 prediction equations, was the most transparent approach possible to address the decision problem defined in the scope. Although the manufacturer provides several reasons for choosing a DES model (e.g. able to reflect multiple courses of therapy, flexibility), these have failed to convince the ERG. A less elaborate Markov model (using more health states than the regular three-state model) would have been as appropriate and would have allowed the ERG much more flexibility in performing additional analyses.

Prediction equations were estimated based on what the manufacturer referred to as the analysable patient sample, which is a subset (n=902) of the ITT population (n=1,088). The manufacturer argued that the ITT population could not be used for estimating prediction equations because for a number of patients baseline data were missing. However, this approach introduced bias in favour of AAP for both TTD and OS (as OS is dependent on TTD). The ERG would have preferred an approach in which the prediction equations are based on the total ITT population and imputing any missing baseline data or to use only treatment as a covariate.

In addition, the process of estimating the prediction equations was not always consistent. For instance, the equation for "Time from AAP/BSC (PP) end to death" was, unlike all other prediction equations, estimated separately by arm, while for all other equations, treatment was used as a covariate. Although requested by the ERG in the clarification phase, the manufacturer could not provide a convincing reason for using this procedure. Furthermore, candidate covariates vary between prediction equations. A rationale for selecting the candidate covariates is absent. Also, interaction terms are sometimes included in an equation despite a non-significant p-value. The ERG would prefer a well-defined and consistently applied procedure on whether or not to stratify, and on including covariates and interaction terms. Without such a procedure, it is difficult to rule out bias caused by these elements. The ERG has performed its own base case analysis to investigate the extent of this bias (see section 1.7).

Although AAP seems to be associated with more grade 3 and 4 AEs, the manufacturer argued that, because AAP and BSC have a similar safety profile, differential utility values for AAP and BSC were not indicated, and the on-treatment utility gain for AAP versus BSC would capture all relevant differences. The only way AEs are explicitly taken into account are in the costs of treating AEs during the docetaxel phase. So AEs are not incorporated separately in HRQoL in any way, nor are they incorporated in the costs in pre- and post-docetaxel phases. In the clarification phase, the ERG requested an additional analysis, removing the on-treatment utility gain and using per-AE utility decrements, as well as pre- and post-docetaxel AE treatment costs. The ICER resulting from this additional analysis was £50,880.

The manufacturer assumes a post-docetaxel utility value of [redacted] for both BSC (post-docetaxel) and AAP (post-docetaxel). Unlike in TA259, there was no post-docetaxel on-treatment utility increment for AAP applied here. The manufacturer argues that to apply this post-docetaxel utility increment of 0.046 (derived from COU-AA-301 trial data) would be

double-counting since the majority of patients in the UK mCRPC Utility Study were assumed to already have been receiving AAP in this setting and so the on-treatment utility gain was captured directly in the utility value. The ERG however could not see any reason why this would not still allow the use of a differential utility value, and requested an analysis incorporating a BSC on-treatment decrement. The manufacturer performed this analysis, alongside with a higher post-docetaxel baseline utility to be more in line with TA259 (also requested by the ERG). This analysis resulted in an ICER of £47,936. The ERG therefore concludes that the results are rather robust with respect to these changes in utility values post-docetaxel.

The post-docetaxel survival in the current model seems very low compared to TA259. Since in the current assessment the post-docetaxel phase solely consists of patients who entered the post-docetaxel active treatment phase, one would expect this population to be comparable with the patients in TA259. Therefore, the ERG performed an additional sensitivity analysis (using the ERG base case as starting point) to assess the impact of incorporating post-docetaxel survival similar to TA 259, which resulted in an ICER of £65,515 per QALY.

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer

1.6.1 Strengths

The manufacturer searched all required databases specified by NICE. The MS[1] provided sufficient detail for the ERG to appraise the searches. Additional searches of conference abstracts were conducted by the manufacturer for the clinical effectiveness and cost-effectiveness sections, and HTA agency websites were searched for cost data. The searches were clearly reported for the most part and the search strategies well translated amongst the different resources searched.

The comparison with BSC was based on a head-to-head comparison in a good quality RCT including more than 500 patients per arm.

The manufacturer's model structure incorporates all clinical pathways in the UK for adult men with mCRPC who were asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy was not yet clinically indicated.

1.6.2 Weaknesses and areas of uncertainty

The main uncertainties regarding clinical effectiveness are that the effectiveness of docetaxel following abiraterone might be reduced and that abiraterone for men with mCRPC in whom chemotherapy is not yet clinically indicated might not meet the end-of-life criteria because the life expectancy in this patient group is likely to be more than 24 months.

As for the economic model, the ERG is not convinced that a DES model, simulating individual patients using 17 prediction equations would have been the most transparent approach to address the decision problem in the scope. The ERG believes that it would have been possible to use a more transparent model with less phases and equations. The ERG also thinks that an individual patient simulation by means of a DES could have been avoided, since acknowledging patient heterogeneity does not necessarily require patient level

simulation.[3] Also, the processes used to estimate the prediction equations in the model were not entirely transparent and consistent within and between the strategies. The latter might have introduced bias in the incremental estimates.

In addition, some of the model parameters were obtained indirectly by means of mapping or assumption (e.g. utility values, scheduled and unplanned MRU). The ERG considers this a source of uncertainty which might have been reduced by collecting these data empirically within the COU-AA-302 trial.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

Due to the above mentioned concerns, the ERG questions the validity of the ICER provided by the manufacturer. The ERG was able to resolve some of the issues highlighted by using an on-treatment utility for post-docetaxel active treatment and non-stratified prediction equations based on the ITT population using treatment as the only covariate. This resulted in an ICER of £56,463 for the ERG base case. However, the ERG acknowledges that there are remaining uncertainties concerning the reliability of the cost-effectiveness evidence which are not handled in the ERG base case, nor could a sensitivity analysis be provided to estimate the impact of these issues on the results. These issues include: censoring patients in the BSC (PP) arm after sequential treatment with AAP and cabazitaxel, not including the possibility of dying during AAP/BSC treatment and post-docetaxel active treatment, not using differential costs and utilities for all AEs for all treatment phases and lack of empirical data to calculate resources and costs for most of the treatment phases.

ICERs calculated in the additional sensitivity analyses ranged between £57,202 and £74,803. Assuming a post-docetaxel survival equal to that in TA 259 (by adjusting the coefficients for "Time from post-docetaxel treatment discontinuation to death") would result in incremental costs, QALYs and life years of £24,159, 0.37 and 0.28 respectively leading to an ICER of £65,515. Finally, replacing the Log-logistic distributions (two prediction equations) with Weibull distributions would result in incremental costs, QALYs and life years of £19,620, 0.26 and 0.21 respectively leading to an ICER of £74,803.

2. BACKGROUND

This report provides a review of the evidence submitted by Janssen in support of abiraterone acetate (trade name Zytiga®) for the treatment of metastatic castration resistant prostate cancer (mCRPC) in adult men who are asymptomatic or mildly symptomatic after failure of Androgen Deprived Therapy (ADT) in whom chemotherapy is not yet clinically indicated.

2.1 Critique of manufacturer's description of underlying health problem.

In section 2.1, the manufacturer explained the underlying health problem, including diagnosis, symptoms, and possible treatment options.

Prostate cancer requires active surveillance; in early stages patients can have symptoms such as pain and problems with urination or can be asymptomatic.[1] Generally, prostate cancer is localised to the prostate gland in early stages and can be cured locally with a surgery or radiotherapy. However, the disease may slowly progress to a chronic stage and over a period of time can rapidly progress to a more advanced and/or metastatic stage.[1] At this stage, patients get on to the treatments such as surgical castration or ADT to reduce the testosterone levels which helps in slowing down the tumour growth and delays progression. However, after 1-2 years the tumour stops responding to the castration therapy and resumes growth.[1] Thus, at this stage it is termed as 'castration-resistant' prostate cancer (CRPC). The patients diagnosed with CRPC are likely to be metastatic (mCRPC) which means the tumour has spread outside the prostate. According to the MS, in the past it was thought that the tumours grow during ADT as they become 'hormone-refractory' or 'androgen-independent'. However, the current knowledge suggests that these tumours still rely on hormones such as testosterone for their growth but, they become dependent on sources outside the prostate such as the adrenal cortex and synthesis within the tumour itself.[1, 4]

Prostate cancer progression can be assessed using prostate-specific antigen (PSA) levels, clinically and radio graphically.[1] Generally, a combination of these techniques is used to determine disease progression. PSA levels can be measured using a simple blood test and it often rises when prostate cancer progresses. Pain is a significant component in the progression of mCRPC, defined by the occurrence of disease progression typically with associated rising serum PSA levels despite surgical or medical castration.[1]

The MS[1] states the most common complaints reported by symptomatic patients including, lower extremity pain, loss of appetite and weight loss, skeletal-related events, renal failure due to obstruction of the urethra and oedema due to obstruction of venous and lymphatic tributaries by nodal metastases.[5, 6]

ERG comment: While the ERG believes the overview presented in section 2.1 of the MS[1] to be accurate, it should be noted that not all statements, e.g. in the first two paragraphs, are fully supported by the presented references. However, this is unlikely to distort the overview in a relevant way.

Section 2.2 of the MS[1] estimated that *"a total of 7,172 patients would be eligible for treatment with AAP in both the post-ADT, pre-chemotherapy and post-chemotherapy settings*

in 2014". The number of men with mCRPC in England and Wales was taken from the costing template for NICE technology appraisal 101 "Docetaxel for the treatment of hormone-refractory metastatic prostate cancer" published in 2006[7] and extrapolated to 2014 population estimates. In a next step, estimates of *"patients eligible for AAP in the post-ADT, pre-chemotherapy setting"* and those *"eligible for AAP in the post-chemotherapy setting"* are calculated.[8-10]

For 'patients eligible for AAP in the post-ADT, pre-chemotherapy setting', the MS[1] stated: *"Clinical opinion estimates that 40% of the mCRPC population will receive treatment with docetaxel,[8] which equates to 4,434 men in 2014. This estimate is aligned with a recent publication, highlighting that the number of men receiving chemotherapy increased from 11% in 2002 to 33% in 2008 within the Thames Valley Cancer Network.[9] The remaining 60% of patients (6,651) were assumed to be chemotherapy-naïve. Of these, it is estimated that 70.5%[10] are asymptomatic or mildly symptomatic. Therefore, in 2014, it is estimated that there will be 4,689 men with mCRPC in whom ADT has failed who are mildly or asymptomatic and who do not yet require chemotherapy"*

For 'patients eligible for AAP in the post-chemotherapy setting', the MS[1] stated: *"Of the 4,434 men estimated to receive docetaxel in 2014, approximately 70% would be eligible for treatment with AAP in the post-chemotherapy setting, which equates to 3,104 men). The 70% accounts for patients who may die on docetaxel treatment, may have rapid deterioration on docetaxel (not suitable for further treatment) or those men in whom AAP may be contraindicated or unsuitable. Four oncologists were consulted to estimate the percentage of patients who would be eligible for AAP following treatment with docetaxel. The responses varied from 55% to 85% (individual responses are collated in appendix 1) and, therefore, we assumed the midpoint of 70% of patients would be eligible for treatment with AAP. This equates to 3,104 patients. As mCRPC patients will be treated earlier in the patient pathway with AAP it is assumed that the number of patients eligible for treatment with AAP in the post-chemotherapy setting will decline as patients who receive AAP in the post-ADT, pre-chemotherapy setting are not eligible for AAP re-treatment post-chemotherapy. It is estimated that there would be approximately 2,483 patients eligible for treatment with AAP in the post-chemotherapy setting in 2014. This is calculated using a phased reduction in the number of eligible patients in ensuing years (please refer to section 8.1 for further details)".*

ERG comment: The 'estimated total cases of hormone-refractory metastatic prostate cancer requiring treatment' were reported as 10,448 in a total male population of 24,220,813 (based on 2003 and 2004 data).[7] This represents 0.0431% of the male population while the MS[1] gave *"0.0195% of the population"*. As a result, all calculations by the manufacturer are approximately 10% lower than the estimates by the ERG. The estimated number of patients in England and Wales eligible for AAP in the MS[1] is 4,689. The ERG calculated this number as 5,158. It should be noted that a number of assumptions have been made based on clinical and expert opinion which makes it possible that the 'real' number of patients for the respective treatment setting might be somewhat different.

Section 2.3 of the MS[1] presented *"natural estimates for the life expectancy in England and Wales for asymptomatic or mildly symptomatic patients with mCRPC for whom chemotherapy is not yet clinically indicated. (...) Among men with mCRPC, estimated survival is around 9–27 months, therefore, the majority of this patient population meet the NICE criteria for short life expectancy (i.e. end-of-life) of 24 months.[11] This is supported by 5-year survival rates of only 26–31%.[12-14] These data are also supported by results of chemotherapy trials in mCRPC patients. In a recent meta-analysis of 12 trials of docetaxel-based regimens, median OS was 18–22 months, depending on whether patients received docetaxel alone or in combination with other treatments.[15]"*

ERG comment: The MS[1] presented a table (Table 6 of the MS[1]) with "estimated natural survival of patients with hormone refractory prostate cancer by clinical criteria". As acknowledged in the MS, some of the survival estimates ("9-27 months") do not meet the NICE criteria for short life expectancy.[11]

2.2 Critique of manufacturer's overview of current service provision

In section 2.4 of the MS, the *"most recent NICE guidance documents on the diagnosis and treatment of prostate cancer (clinical guideline 175)"* was summarised[16] while other relevant guidance was mentioned.[2, 17, 18]

"According to NICE guidelines, chemotherapy should usually be reserved for men with symptomatic progression, and the combination of docetaxel + prednisolone can cause substantial side effects.[16] Although these guidelines add that asymptomatic men with metastatic disease and a rapidly rising PSA level may also benefit from chemotherapy, there is no definition of rapidly rising PSA in the guideline. A study by Armstrong et al.[19] found that a PSA doubling time <55 days in conjunction with a baseline PSA ≥114 ng/mL had a significant negative impact on overall survival (OS). Only a minor proportion (approximately 10%) of the population in the COU-AA-302 study met these criteria, so chemotherapy would not be a suitable treatment for the majority of patients. However, it should be noted that PSA doubling time could only be roughly estimated in COU-AA-302."

ERG comment: As stated in the MS, the "PSA doubling time could only be roughly estimated in COU-AA-302" which means that theoretically a higher proportion of asymptomatic men might be suitable to receive chemotherapy as recommend by NICE guidance.[16]

As shown in Figure 1 (section 2.5 of the MS[1]), abiraterone acetate plus prednisone/ prednisolone (AAP) *"would provide a treatment option for asymptomatic or mildly symptomatic mCRPC patients in whom chemotherapy is not yet clinically indicated"*. In addition, information on AAP from various guidelines (EAU, ESMO, AUA, NCCN) are presented.[20-23]

Section 2.6 of the MS stated that *"while NICE,[16] EAU,[20] ESMO,[21] AUA[22] and NCCN[23] guidelines all agree that first- and second-line treatments for patients with metastatic prostate cancer should be androgen withdrawal and anti-androgens, respectively, once the*

disease has become castration resistant, the guidelines are less clear. Options include corticosteroids,[16, 21, 23] oestrogenic compounds,[20, 21, 23] adrenolytic drugs,[20] novel compounds (e.g. enzalutamide,[20, 23] AA[20, 22, 23] or sipuleucel-T[22, 23]), ketoconazole[20-23] and chemotherapy[16, 20-23]".

ERG comment: The place in which AAP has been added to the existing treatment pathway is in line with the final scope.[24]

According to section 2.7 of the MS, *"the NICE scope identified two possible comparators: best supportive care (BSC; this may include corticosteroids, radiotherapy, radiopharmaceuticals, analgesics, bisphosphonates, further hormonal therapies, and mitoxantrone with or without corticosteroids) or docetaxel. Of these, BSC is the appropriate comparator for AA. As BSC can include corticosteroids, the placebo arm of the COU-AA-302 study can be considered an appropriate comparison population, as patients in the placebo arm all received corticosteroids as part of supportive care"*. The section concluded that *"based on the lack of available comparative evidence in an aligned patient population and on current UK clinical practice, docetaxel was not considered to be a comparator of interest for this submission. Therefore, BSC (corticosteroid) should be considered the most appropriate comparator for AAP in this patient population"*.

ERG comment: As the indication is men with mCRPC in whom chemotherapy is not yet clinically indicated, it seems reasonable that docetaxel was not considered as comparator. However, in the final scope, NICE explicitly stated that: *"Docetaxel is included in the list of comparators because the recommendations in TA101 include patients who are asymptomatic or mildly symptomatic, and clinicians have stated that docetaxel is increasingly used for this patient group, and because of the lack of clear clinical criteria to identify the patient group in the CHMP indication"*.[24]

Taking all of this into account and assuming that most patients will end up using docetaxel, which also seems to be implied by the phrase "not yet clinically indicated", an important question in this appraisal seems to be whether abiraterone followed by docetaxel is more effective than watch-full waiting (BSC) followed by docetaxel.

Section 2.8 of the MS described *"hypertension, hypokalaemia or fluid retention"* as adverse events *"in a small number of cases"* where *"co-administration of prednisolone reduces the frequency and severity of these AEs"*.[25] The following sections (sections 2.9 and 2.10) described the main resource use to the NHS and if additional infrastructure is required. While according to the MS, no additional infrastructure is required and *"as AAP is self-administered orally in the patient's home, there are no anticipated costs due to location of care, staff or administration"*, some measurements might need to be taken:

"With administration of AAP, blood pressure, serum potassium and fluid retention should be monitored before treatment and at least monthly thereafter.[25] Patients with a significant risk of congestive heart failure (exemplified in the SPC for AA[25] as those with a history of cardiac failure, uncontrolled hypertension, or cardiac events such as ischaemic heart disease) should

be monitored every 2 weeks for the first 3 months of treatment by measuring blood pressure and serum potassium levels and checking for signs of oedema. To monitor for hepatotoxicity, serum transaminases should be measured in all patients before treatment with AA and every 2 weeks for the first 3 months, and then monthly. Monitoring would be carried out as an outpatient visit to an oncology clinic.[8] After the initial more frequent monitoring needs to determine hepatotoxicity and potentially congestive heart failure parameters, the frequency of follow-up visits (monthly) would be similar to that of other treatment options in this patient population.[8]"

ERG comment: The list of therapies that may be prescribed to manage adverse reactions associated with abiraterone acetate is incomplete. As shown in Table 2.1, the summary of product characteristics lists further adverse events which might require additional treatment after a potential occurrence:[25]

Table 2.1: Adverse reactions identified in clinical studies and post-marketing

Infections and infestations	very common: urinary tract infection common: sepsis
Endocrine disorders	uncommon: adrenal insufficiency
Metabolism and nutrition disorders	very common: hypokalaemia common: hypertriglyceridaemia
Cardiac disorders	common: cardiac failure*, angina pectoris, arrhythmia, atrial fibrillation, tachycardia
Vascular disorders	very common: hypertension
Respiratory, thoracic and mediastinal disorders	rare: allergic alveolitis[a]
Gastrointestinal disorders	very common: diarrhoea common: dyspepsia
Hepatobiliary disorders	common: alanine aminotransferase increased, aspartate aminotransferase increased
Skin and subcutaneous tissue disorders	common: rash
Musculoskeletal and connective tissue disorders	uncommon: myopathy, rhabdomyolysis
Renal and urinary disorders	common: haematuria
General disorders and administration site conditions	very common: oedema peripheral
Injury, poisoning and procedural complications	common: fractures**

* Cardiac failure also includes congestive heart failure, left ventricular dysfunction and ejection fraction decreased

** Fractures includes all fractures with the exception of pathological fracture

[a] Spontaneous reports from post-marketing experience

3. CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

Table 3.1: Statement of the decision problem (as presented by the manufacturer)

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope
Population	Men with mCRPC who have not received prior cytotoxic chemotherapy or biologic therapy	Men with mCRPC who are asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy is not yet clinically indicated	AA received a marketing authorisation for an indication in this patient population from the EMA in December 2012
Intervention	AA in combination with prednisolone	As per scope	
Comparator(s)	Docetaxel BSC (this may include radiotherapy, radiopharmaceuticals, analgesics, bisphosphonates, further hormonal therapies, and mitoxantrone with or without steroids or steroids alone)	The appropriate comparator for AAP is BSC. In COU-AA-302, AAP was compared with PP; supportive care was permitted in both arms during the treatment phase as per institutional guidelines. The following agents were permitted: LHRH agonists, multivitamins, selenium and soy supplements, 'stress dose' glucocorticoids, transfusions, haematopoietic growth factors	Due to their palliative benefit (pain relief[26]), prednisone or prednisolone can be considered a form of BSC. The co-administration of one of these agents with AA is also necessary to suppress ACTH drive to reduce the incidence and severity of potential AEs such as hypertension, hypokalaemia and fluid retention[25] Docetaxel is not an appropriate comparator for AA in the intended patient population for the reasons discussed in section 2.7; briefly: • AAP is licensed for men with mCRPC who are mildly or asymptomatic while docetaxel is generally reserved for symptomatic patients[16, 21, 25, 27] • UK clinical practice is currently aligned with NICE clinical guideline 175,[16] whereby docetaxel is usually reserved for the more symptomatic patient • AAP demonstrates a clinically relevant benefit and significant advantage to patients who require additional therapeutic options[28] but are not in immediate need of chemotherapy • There is a lack of clinical evidence supporting the use of docetaxel in a truly

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope
			asymptomatic or mildly symptomatic patient population[29] • There is no evidence that starting chemotherapy when the patient is still asymptomatic or mildly symptomatic is more effective than waiting until the patient is more highly symptomatic
Outcomes	• OS • PFS • Response rate • PSA response • AEs of treatment • HRQL	The following outcomes were prespecified in the protocol: • Co-primary: OS, rPFS • Secondary: times to: opiate use, chemotherapy, ECOG PS deterioration, PSA progression • Other: objective response rate, PSA response rate, duration of response, HRQL, time to pain progression, time to analgesic progression • Safety: AEs, SAEs, laboratory tests	Additional endpoints were included in the COU-AA-302 study
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	As per scope	–
Subgroups to	BPI 0□1	BPI 0□1	BPI 0–1 as per scope; ECOG 0

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope
be considered		ECOG 0	was also felt to be clinically relevant
Special considerations, including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation	As per scope	–

AA, abiraterone acetate; AAP, abiraterone acetate plus prednisone/prednisolone; ACTH, adrenocorticotropic hormone; ADT, androgen deprivation therapy; AE, adverse event; BSC, best supportive care; BPI, brief pain inventory; CHMP, Committee for Medicinal Products for Human Use; ECOG, Eastern Cooperative Oncology Group; EMA, European Medicines Agency; HRQL, health-related quality of life; LHRH, luteinising hormone-releasing hormone; mCRPC, metastatic castration-resistant prostate cancer; NHS, National Health Service; NICE, National Institute for Health and Clinical Excellence; OS, overall survival; PFS, progression-free survival; PS, performance status; PSA, prostate-specific antigen; QALY, quality-adjusted life year; rPFS, radiographic progression-free survival; SAE, serious adverse event.

3.1 Population

The patient population described in the final scope is as follows: "Adults with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated".[24] This is in line with the patient population included in the manufacturer submission[1] and in the main trial for this submission, the COU-AA-302 study.[30] The table above seems to be based on the draft scope issued by NICE.

3.2 Intervention

Abiraterone received marketing authorisation in the UK on 18 December 2012. Abiraterone with prednisone or prednisolone is indicated for the treatment of mCRPC in adult men who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy (ADT) in whom chemotherapy is not yet clinically indicated.

The following information is based on the EMA Summary of Product Characteristics (SmPC).[31] The recommended dose is 1000 mg (four 250 mg capsules) as a single daily dose that must not be taken with food. Abiraterone is to be taken with low dose prednisone or prednisolone. The recommended dose of prednisone or prednisolone is 10 mg daily.

Mechanism of action: Abiraterone acetate is converted *in vivo* to abiraterone, an androgen biosynthesis inhibitor. Specifically, abiraterone selectively inhibits the enzyme 17α-hydroxylase/C17,20-lyase (CYP17). This enzyme is expressed in and is required for androgen biosynthesis in testicular, adrenal and prostatic tumour tissues. CYP17 catalyses the conversion of pregnenolone and progesterone into testosterone precursors, DHEA and androstenedione, respectively, by 17α-hydroxylation and cleavage of the C17,20 bond. CYP17 inhibition also results in increased mineralocorticoid production by the adrenals (see section 4.4).

Androgen-sensitive prostatic carcinoma responds to treatment that decreases androgen levels. Androgen deprivation therapies, such as treatment with LHRH analogues or orchiectomy, decrease androgen production in the testes but do not affect androgen production by the adrenals or in the tumour. Treatment with abiraterone acetate decreases serum testosterone to undetectable levels (using commercial assays) when given with LHRH analogues (or orchiectomy).

The most common adverse reactions seen are peripheral oedema, hypokalaemia, hypertension and urinary tract infection. Other important adverse reactions include, cardiac disorders, hepatotoxicity, fractures, and allergic alveolitis.

3.3 Comparators

The main deviation from the scope is that docetaxel is not included as a comparator in the manufacturer submission. The reasons for this deviation are discussed in chapter 2.7 of the MS and summarised in the table above.

As the indication is men with mCRPC **in whom chemotherapy is not yet clinically indicated**, it seems reasonable that docetaxel is not considered as a comparator. However, in the final scope, NICE explicitly states that: "Docetaxel is included in the list of comparators because the recommendations in TA101 include patients who are asymptomatic or mildly symptomatic, and clinicians have stated that docetaxel is increasingly used for this patient group, and because of the lack of clear clinical criteria to identify the patient group in the CHMP indication".[24]

Assuming that most patients will end up using docetaxel, which also seems to be implied by the phrase "**not yet** clinically indicated", an important question in this appraisal is whether abiraterone followed by docetaxel is more effective than watch-full waiting (BSC) followed by docetaxel. In the COU-AA-302 trial, 239 out of 546 (43.8%) of AAP patients and 304 out of 542 (56.1%) of PP patients received docetaxel as subsequent therapy (MS, Table 21, page 68). The results for this specific group of patients are not presented in the MS; therefore, we asked the manufacturer to provide these data in the clarification letter. At first, the manufacturer did not provide these data with the following explanation:

"Janssen is unable to answer this question. The data requested is a *post-hoc* analysis of patients in the COU-AA-302 trial who subsequently receive docetaxel. This group of patients progressed more quickly, and therefore moved onto docetaxel treatment earlier than the other patients in the trial. This *post-hoc* analysis violates the principles of randomisation, and in effect, selects for the patients with the worst prognosis (ie those that progress quickly and move onto chemotherapy), which renders any interpretation of these results meaningless."

Four days before the deadline of this report, the ERG did receive data from the manufacturer. These are discussed in Chapter 4.5.

ERG Comment: There is very little evidence regarding the effectiveness of docetaxel after abiraterone. At least one study seems to suggest that the effectiveness of docetaxel following abiraterone might be seriously reduced.[32] This study reported the following results:

> "Of the 54 patients treated with abiraterone, 35 subsequently received docetaxel. Docetaxel resulted in a prostate-specific antigen (PSA) decline of ≥50% in nine patients [26%, 95% confidence interval (CI) 13% to 43%], with a median time to PSA progression of 4.6 months (95% CI 4.2% to 5.9%). PSA declines ≥30% were achieved by 13 patients (37%, 95% CI 22% to 55%). The median overall survival was 12.5 months (95% CI 10.6–19.4). All patients who failed to achieve a PSA fall on abiraterone and were deemed abiraterone-refractory were also docetaxel-refractory (N = 8). In the 24 patients with radiologically evaluable disease, partial responses were reported in four patients (11%), none of whom were abiraterone-refractory".[32]

Based on these results, the authors concluded that "the activity of docetaxel post-abiraterone appears lower than anticipated and no responses to docetaxel were observed in abiraterone-refractory patients." However, this is only a small single arm study. A *post-hoc* analysis using data from the COU-AA-302 trial might provide a more reliable estimate.

3.4 Outcomes

The COU-AA-302 trial examined two co-primary outcomes: overall survival (OS) and radiographic progression-free survival (rPFS) and several secondary outcomes. rPFS was defined as 'time from randomisation to one of the following: progression by bone scan (adapted PCWG2 criteria) or CT or MRI (modified RECIST criteria) or death'.

The manufacturer states: "Although PFS is not routinely used in clinical practice, it is a common endpoint used in oncology trials. A substantial improvement in rPFS is clinically meaningful to mCRPC patients as it delays the time to increased tumour burden, which may ultimately lead to the decision to initiate opiates or cytotoxic chemotherapy. Progression of metastatic bone disease is of paramount importance because of pain and skeletal morbidity" (MS, Table 15, page 56).

Secondary outcomes assessed in the COU-AA-302 trial are:

- Time to opiate use for cancer pain: The time from randomisation to opiate use for prostate cancer pain
- Time to initiation of cytotoxic chemotherapy: The time from randomisation to initiation of cytotoxic chemotherapy for prostate cancer
- Time to deterioration in ECOG PS by ≥1 point: The time from randomisation to first ≥1 grade worsening in the ECOG PS scale
- Time to PSA progression: The time from randomisation to PSA progression, assessed by observation of trends in serial PSA measurements, according to adapted PCWG2 criteria

Other outcomes assessed in the COU-AA-302 trial are:

- PSA response rate: The proportion of patients achieving a PSA decline ≥50% according to PCWG2 criteria

- Objective response rate in patients with measurable disease: The proportion of patients with measurable disease achieving a complete or partial response according to RECIST criteria (baseline lymph node size ≥2 cm to be considered a target lesion)
- Duration of response: The time from first response (in patients with measurable disease, according to modified RECIST criteria) to progression
- QoL total score and each subscale: HRQL as measured by the FACT-P instrument
- Time to pain progression: The time from randomisation to first increase in pain (with the exact definition varying depending on the pain progression outcome measured)
- Time to analgesic progression: Defined as a ≥30% increase in analgesic usage score from baseline that was observed at two consecutive evaluations ≥4 weeks apart. Analgesic scores were assessed according to the WHO scale (0 for no medication, 1 for non-opiate pain medication, 2 for opiates for moderate pain, 3 for opiates for severe pain)

ERG Comment: rPFS is not an established end point in metastatic prostate cancer. This is the first time that rPFS has been used as a co-primary end point in this setting and it is not clear if rPFS superiority translates into clinical benefit.

Regarding quality of life, the manufacturer states that "HRQL was assessed in the COU-AA-302 study via the FACT-P instrument, and results of the second interim analysis have recently been published[33]."

The cited reference,[33] which has not been submitted by the manufacturer, states that the "FACT-P questionnaire is validated and accepted for metastatic castration-resistant prostate cancer...". This claim is supported by two references.[34, 35]

One of these studies[35] presents reliability data for a newly developed tool, the "NCCN/FACT-P Symptom Index". This "National Comprehensive Cancer Network/Functional Assessment of Cancer Therapy-Prostate Symptom Index" includes 17 items. However, the tool used in the paper submitted by the manufacturer,[35] FACT-P (version 4),[36] consists of 39 items (27 core items supplemented by 12 prostate-specific items). Therefore, it is unclear whether and how the reliability data reported by Victorson et al[35] are applicable to the submission by the manufacturer.

The objective of the second study is to "determine clinically meaningful changes (CMC) for the Functional Assessment of Cancer Therapy–Prostate (FACT–P)".[34] For that purpose, "anchor-based differences using Karnofsky Performance Status (KPS), bone alkaline phosphatase (BAP), hemoglobin, time to disease progression (TTP), adverse events (AE), and survival" were calculated. It should be noted that the derived minimally important difference (MID) is partly based on laboratory parameters (BAP, haemoglobin) assessed at baseline which makes it hard to assess the relevance of this parameter for patients. Furthermore, a range of "6 to 10" was reported as "clinically distinguishable score". It is unclear why the (higher) value of 10 was chosen for the submission. The reported results do not allow an assessment whether use of a different MID, e.g. the lower value of six, could have had an influence on the findings. In addition, Cella et al fail to report which version of

the FACT-P instrument was assessed in the study. Therefore, it is unclear if the results are applicable to the submitted data.

In Table 24 ('HRQL outcomes results'), the manufacturer presents a "summary of the time to a decrease of ≥10 points for all FACT-P subscale results at the time of the third interim OS analysis". The median time to progression, defined as "decrease of ≥ 10 points", is presented for abiraterone acetate plus prednisone/prednisolone (AAP) and placebo plus prednisone (PP). No justification is given on why the outcome was operationalised in this way when a simpler approach could have been used.

3.5 Other relevant factors

According to the MS, "Janssen has previously agreed to supply AA with a PAS involving a confidential discount. Under the terms of the PAS, the cost of AA used in the model is £[redacted] per month ([redacted] discount). This PAS is in place for all current and future indications under consideration by NICE and is therefore used in the base-case analysis." (See MS, Section 7.5.5, page 147).[1]

In addition the manufacturer also claims that abiraterone meets the criteria for appraisal of end of life medicines (see MS, Section 7.7.6, page 167).[1] This is discussed in Chapter 7 of this ERG report.

4. CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

An evidence based checklist for the Peer Review of Electronic Search Strategies (PRESS), developed by McGowan et al was used to inform this critique.[37] The submission was checked against the Single Technology Appraisal (STA) specification for manufacturer/sponsor submission of evidence.[38] The ERG has presented only the major limitations of each search strategy in the main report. Further criticisms of each search strategy can be found in Appendix 1.

4.1.1 Searches

Searches were reported for all databases required by NICE guidance: Medline, Medline In-Process, Embase and the Cochrane Library.[38] The database provider for each database was listed; the date span of the databases searched and the specific date the searches were run on were provided. The manufacturer additionally searched conference proceedings for specific conferences in specific years.

The manufacturer reported that the searches were an update of previous searching undertaken to support an earlier submission relating to NICE TA259: abiraterone in mCRPC after docetaxel.[2] The search strategies used for the earlier submission were not amended, as the study population (men with prostate cancer) and interventions included (standard of care in England and Wales, and investigational interventions) were appropriate for this submission.

The manufacturer translated the research question into appropriate search strategies and the ERG considered the searches to be adequate. Searches were clearly structured and divided into population and intervention facets. Study design limits to identify RCTs and non-RCTs were applied, and the manufacturer stated that the search strategies for clinical effectiveness (6.1) were used for the non-RCT evidence (6.8) and adverse events (6.9) sections of the submission. The study design filters were not referenced, so it was unclear whether the filters used were published objectively derived filters. The filters contained a combination of subject heading terms and free text terms and the ERG deemed them to be adequate. In response to the ERG points of clarification (POC) letter[39], the manufacturer reported that the RCT and non-RCT (observational) search filters used in the current submission[1] were based on those provided by the Scottish Intercollegiate Guidelines Network (SIGN).[40]

The ERG noted that the manufacturer searched Embase and Medline simultaneously using a single database provider (embase.com) and search strategy. This has limitations when using subject heading terms which could affect recall of results. Embase subject heading terms (Emtree) were used in the search strategy, and although simultaneous searching of embase.com should automatically identify and search for equivalent Medline subject heading terms (MeSH), it is not clear if this is the case for all potentially useful MeSH terms. Given the potential limitations of this approach, the ERG considered it preferable to search each database separately, or at least to ensure inclusion of both Emtree and MeSH terms in the search strategy.

The manufacturer reported that there were no restrictions on the search start dates (Table 7 and 10.2.3), however the search strategies presented all had specified start dates. The ERG assumed that the manufacturer was including the searches undertaken for the earlier submission, although this was not reported clearly. The manufacturer presented five update search strategies for Embase and Medline in 10.2.4.1. The search strategy in Table 100 had an error in the date span used indicating that only one day had been searched (search line #101). In response to the ERG POC letter,[39] the manufacturer explained that this was a typographical error, and that the updated literature search was conducted from 27-02-2012 to 27-09-2012.[40]

Indirect and mixed treatment comparisons

Searches were not carried out as no indirect or mixed treatment comparisons were performed.

Non-RCT Evidence

The same search strategies and databases used for the clinical evidence section (6.1/10.2) were used for non-RCT evidence. The search strategies included a study design filter for non-RCTs.

Adverse events

The same search strategies and databases used for the clinical evidence section (6.1/10.2) were used to identify adverse events data. CRD guidance recommends that if searches have been limited by an RCT filter, additional searches should be undertaken to ensure that adverse events that are long-term, rare or unanticipated are not missed.[41] Despite the addition of a non-RCT filter the ERG considered that it was possible that some relevant evidence may not have been identified as a consequence of the study design limits.

Cost-effectiveness

Searches were carried out for all of the databases required by NICE: Medline, Medline In-Process, Embase, NHS EED and EconLit.[38] The database provider for each database was reported; the date span of the databases searched and the specific date the searches were run were provided. The manufacturer additionally searched conference proceedings, and health technology assessment organisation websites.

As with the clinical effectiveness searches, this was an update of previous searches undertaken to support an earlier submission relating to NICE TA259: abiraterone in mCRPC after docetaxel.[2] The search strategies used for the earlier submission were not amended, as the study population (men with prostate cancer) and interventions included (standard of care in England and Wales, and investigational interventions) were appropriate for this submission.

The manufacturer translated the research question into appropriate search strategies and the ERG considered the searches to be adequate. Searches were clearly structured and divided into population and intervention facets. A study design filter to identify cost-effectiveness studies was applied and the manufacturer stated that this was based on standard filters developed by SIGN.

The ERG noted that the manufacturer searched Embase and Medline simultaneously using a single database provider (embase.com) and search strategy. This has limitations when using subject heading terms which could affect recall of results. Embase subject heading terms (Emtree) were used in the search strategy, and although simultaneous searching of embase.com should automatically identify and search for equivalent Medline subject heading terms (MeSH), it is not clear if this is the case for all potentially useful MeSH terms. Given the potential limitations of this approach, the ERG considered it preferable to search each database separately, or at least to ensure inclusion of both Emtree and MeSH terms in the search strategy.

Of more concern to the ERG was the absence of any subject heading terms, Emtree or MeSH, in the interventions facet of search terms for both the Embase/Medline and Cochrane Library (NHS EED) search strategies. The manufacturer responded to the ERG POC letter[39] with the following explanation:

"In the literature search most relevant to this decision problem, an extensive list of interventions was included. All the interventions were searched as text terms using brand, generic and nomenclature used during the research and development phase. This is an extremely comprehensive list that was used to search the title and abstracts. In addition to this, a number of Emtree terms were also used, including corticosteroid, antiandrogen, gonadroreline agonist, LHRH agonist and cancer immunotherapy. Due to the comprehensive nature of the search terms used in combination with the Emtree terms, it is believed that no studies relevant to the decision problem were excluded from the search. It should be noted that in addition to the structured search, a bibliographic search was also conducted. This methodology ensures that all studies pertinent to the decision problem have been included."[40]

Despite this explanation the ERG considered that the absence of subject heading terms might have impaired sensitivity and recall.

The search strategy included some unusual intervention search terms that would be unlikely to work in any search interface due to the inappropriate use of brackets, square brackets, hyphens and numbers, e.g. '5-fluoro-2,4(1h, 3h)-pyrimidinedione'.ab,ti and '1-hydroxy-7beta,10beta-dimethoxy-9-oxo-5beta,20-epoxytax-11-ene-2alpha,4,13alpha-triyl 4-acetate 2-benzoate 13-[(2r,3s)-3-([(tertbutoxy)carbonyl]amino)-2-hydroxy-3-phenylpropanoate]':ab,ti. The ERG could not confirm whether this is the case in embase.com as it did not have access to this database provider.

The manufacturer reported that the date span for the cost-effectiveness searches ran from 2000 (7.1 and 10.10.3), although it was not clear why. The searches undertaken for the previous submission ran until 2010. The manufacturer explained why in response to the ERG POC letter[39] as follows:

"The cost-effectiveness searches were conducted from 2000 as this timeframe would capture those studies most relevant to the decision problem. Abiraterone acetate received marketing

authorisation in 2011. A search spanning 10 years earlier than this date is sufficient to identify any publications pertinent to the economic analysis in this decision problem." [40]

The ERG noted inconsistencies in the date limitations reported, which were raised for clarification.[39] The manufacturer's clarification response[40] stated that all date inconsistencies were typographical errors. The manufacturer reported that additional conference searches were conducted for ISPOR European Congress and ISPOR International Congress (2006-2012) in 7.1.1, but no details were provided in 10.10.5. Details of HTA organisation website searches were provided in the submission: dates of searches, names of organisations, and website addresses. No details of the search strategies used or the results of the searches were reported. Details of the ISPOR congress website addresses and search terms used were subsequently provided in response[40] to the ERG POC letter.[39]

Measurement and valuation of health effects

Searches were carried out for all the databases required by NICE. The database provider for each database was listed; the date span of the databases searched and the specific date the searches were run were provided. The searches were reproducible.

The manufacturer reported that the searches updated a systematic review undertaken for NICE TA101.[17] It was reported that the searches were conducted from 2005 to 2013 to account for this. However, the search strategies for embase.com and the Cochrane Library (Tables 130-131) indicated that the date span used was 2005 to 2012. In response to the ERG POC letter, the manufacturer explained that this was a reporting error, and that the Cochrane Library HRQoL update search strategies were missing from the submission. The two missing Cochrane Library search strategies, one conducted on 30 May 2013 and the second on 2 September 2013, were provided in the response to the ERG POC letter.

The search strategies used for the previous systematic review were not amended as they combined the study population of interest (men with prostate cancer) with search terms for measurement and valuation of health effects.

There appeared to be a mistake in the final search line of the Cochrane Library (NHS EED) search strategy (Table 131, search line #8). It was not clear whether this was a reporting error or if the search lines #1 and #2 were combined for this search. In response to the ERG POC letter the manufacturer explained that this was a typographical error, and that search line #8 should have been '#7 AND [2005-2012]/py [NHS EED only]'.

It was unclear whether EconLit was searched or not. EconLit was listed as one of the databases searched in Table 44, but in 10.12.1 it was reported that searches of EconLit were only conducted during the previous review. There was no EconLit search strategy, search date, or date span reported anywhere in the current submission. The manufacturer confirmed in response to the ERG POC letter that EconLit was not searched for HRQoL data.

Resource identification, measurement and valuation
The same search strategies and databases used for the cost-effectiveness section (7.1/10.10) were used for Resource identification.

Summary of searching
The searches in the MS were, in the main, well documented, clearly presented and reproducible. Search strategies did not report the number of records retrieved by each line or for each database. Inclusion of this information would have aided the ERG in assessment of the searches, making it easier to see where errors might have occurred, what impact amendments made to the strategies, and to ensure that the methods were transparent.

4.1.2 Inclusion criteria
The updated review for this submission utilised a broad set of inclusion criteria (See Table below) and included all studies in mCRPC.

Table 4.1: Inclusion and exclusion criteria used in the systematic review

	Clinical effectiveness	Rationale
Inclusion criteria	**Population** Age: Adults (≥18 years) Gender: Any Race: Any Disease/disease stage: mCRPC	• The patient population included in this review was a broader population of patients with mCRPC. The subset of studies including post-ADT, pre-chemotherapy patients was identified prior to data extraction • In line with marketing approval and the decision problem of this submission studies including children or adolescents were excluded
	Interventions **Standard of care (England and Wales):** docetaxel, mitoxantrone, estramustine, cyclophosphamide, 5-fluorouracil, doxorubicin, carboplatin, etoposide, paclitaxel, vinorelbine, vinblastine, dexamethasone, hydrocortisone, prednisone, strontium, zoledronate **Investigational interventions:** abiraterone, sipuleucel-T, MDV3100, bevacizumab, atrasentan, dasatinib, ZD4054, patupilone, AS1404, ipilimumab, sunitinib, IMC-A12, aflibercept, cabazitaxel (XRP6258), JM216, alpharadin	• The list covers common interventions used for the treatment of mCRPC in the UK
	Comparator Another included intervention, best supportive care (includes radiotherapy, corticosteroids, oxygen, analgesics), or placebo	• A wide range of chemotherapy-based combinations are being investigated besides BSC. These comparators were selected to enable the inclusion of all relevant citations • Exclusion of studies based on the comparator used was applicable only for RCTs. RCTs have high internal validity and are considered to represent the gold standard of

	Clinical effectiveness	Rationale
		clinical evidence • Observational studies and non-RCTs were included regardless of the comparator treatment evaluated, given the external validity of these type of studies
	Study design RCTs Non-RCTs Single-arm studies Dose-finding studies Observational studies, including: • Cohort studies (prospective and retrospective) • Case-control studies • Cross-sectional study/survey • Analysis of database/clinical records	• The review included RCTs, as they are the gold standard of clinical evidence, minimising the risk of confounding and allowing the comparison of the relative efficacy of interventions • Observational studies were also included in the review as they include broad patient populations and reflect real world evidence and thus have external validity
	Language restrictions English only	• The restriction would not limit results substantially due to publication availability in the English language and within the NICE scope of the current submission
	Publication timeframe **Original review:** All publications up to and including the cut-off date of 30.05.11 All conference proceedings from 01.01.06 until the cut-off date of 08.06.11 **Review update:** All publications from 31.05.11 to 02.09.13 All conference proceedings from 08.06.11 to 02.09.13	• Searches of conference proceedings were limited to the previous 7 years as studies presented at conferences are usually published in journals within 6 years (conference data older than 7 years which have not been published in journals after this duration are unlikely to be useful for the purposes of this review)
	Outcomes of interest Studies should report at least one of the following outcomes of interest: • OS • PFS (including both symptomatic and radiographic PFS) • Time to progression (according to PSA and RECIST criteria) • WHO PS improvement • Overall mortality • Response rate (according to prostate-specific antigen and RECIST criteria) • Duration of response • Prostate-specific antigen measurements • EORTC QLQC30 • EQ-5D • FACT-P score and its subscale • BPI score	• Studies that did not report outcomes of interest were excluded • These outcomes were chosen since these are frequently measured and reported in the trials involving advanced prostate cancer patients and include those outcomes specified in the decision problem.

	Clinical effectiveness	Rationale
	• PPI • Bone pain • Pain response • Time to pain progression • Time to opiate use • VAS pain score • Analgesic score • Time to first SRE • SMR • Vertebral fractures • Non-vertebral fractures • AEs • Withdrawals and discontinuations	
Exclusion criteria	**Population** Disease: Prostate cancer other than mCRPC, Secondary prostate cancer **No subgroup analysis** Studies reporting no subgroup data for population of interest (mCRPC) were excluded. However, studies including mixed patient population with the proportion of mCRPC patients being ≥90% were included in the review **Study design** Case studies, case series, case reports, and studies in Phase I	• Case-series and case-reports were not included in the review as they are generally smaller, non-comparative studies that have a higher risk of bias • Phase I studies were excluded as they aim to establish the safety profile rather than clinical effectiveness

Abbreviations: ADT, androgen deprivation therapy; BPI, brief pain inventory; BSC, best supportive care; EORTC QLQ, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; EQ-5D, EuroQoL Five Dimensions; FACT-P: Functional Assessment of Cancer Therapy – Prostate; mCRPC, metastatic castration-resistant prostate cancer; NICE, National Institute for Health and Clinical Excellence; PFS, progression-free survival; PPI, present pain intensity; PS, performance status; PSA, prostate-specific antigen; RCT, randomised controlled trial; RECIST, Response Evaluation Criteria In Solid Tumors; SMR, skeletal morbidity rate; SRE, skeletal-related event; VAS, visual analogue scale

ERG comment: No specific search was done for this appraisal. Instead, the manufacturer performed one broad search for all studies in mCRPC. The full systematic review presented in the MS had previously been conducted to evaluate the clinical efficacy, safety and tolerability of AA in mCRPC patients and was used in a previous NICE STA submission (NICE TA259), focussing on a subset of studies in patients with mCRPC who had disease progression despite treatment with docetaxel.[42] For the current submission, an update to September 2013 was conducted. In this update, the focus was on those studies specific to the indication stated within the decision problem of this submission: patients with mCRPC who are asymptomatic or mildly symptomatic (defined as ECOG 0–1 and brief pain inventory – short form [BPI-SF] score of 0–1 [asymptomatic] or 2–3 [mildly symptomatic]) after failure of ADT in whom chemotherapy is not yet clinically indicated.

4.1.3 Critique of data extraction

One RCT was included, the COU-AA-302 trial. The most recent data from this trial (third interim analysis) were extracted from the updated clinical study report,[43] a review article[44]

and conference abstracts[45-51]. The only full journal publication for the trial was based on the second interim analysis.[30]

Table 4.2: Data sources for the pivotal RCT, COU-AA-302

COU-AA-302 analysis point	Cut-off date for the analysis	Data availability	Data source
First interim analysis	20.12.10	Unpublished	Clinical study report[52]
Second interim analysis	20.12.11	Published	Ryan et al. 2013[30] Review article[44] Patient-reported outcomes[33] Conference abstracts (ASCO, ESMO)[53, 54]
		Unpublished	Clinical study report[52]
Third interim analysis	22.05.12	Unpublished	Updated clinical study report[43]
		Published	Review article[44] Conference abstracts (ASCO, ASCO GU)[45-51]
Final analysis	N/A	N/A	N/A

Abbreviations: ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology.

4.1.4 Quality assessment

The quality assessments of the COU-AA-302 trial can be found in **A**ppendix 3, Section 10.3 of the MS, and in the table below.

The methods used to generate random allocation sequence and for concealment of allocation sequence were reported and were judged as adequate. Blinding status was clear and the study did not show any evidence of selective reporting. Overall, the COU-AA-302 trial was rated as being at a low risk of bias.

ERG Comment: The ERG agrees with the manufacturer's assessment on most items.

Disagreements with the manufacturer assessment of study quality were as follows:

- Imbalances in drop-outs between groups: No imbalances; but large numbers of drop-outs in both groups.
- Missing outcomes: No, all outcomes were reported. However, no data were reported for QoL scores by arm (baseline, follow-up and change scores).
- Did the analysis include an intention-to-treat analysis: No. The ITT population did not include all patients randomised into the study, but those who received at least 1 dose of the allocated intervention.

Table 4.3: Quality assessment of COU-AA-302

Study question	Manufacturer's explanation: How is the question addressed in the study?	Manufacturer's assessment	ERG comment
Was randomisation carried out appropriately?	The randomisation schedule was generated by an independent statistician at Almac Clinical Technologies. Patients were	Yes	Low risk of bias

Superseded - see Erratum

Study question	Manufacturer's explanation: How is the question addressed in the study?	Manufacturer's assessment	ERG comment
	assigned randomly in a 1:1 ratio to receive either abiraterone acetate plus prednisone or placebo plus prednisone. Patient eligibility was verified by the investigators, who then entered the stratification factor (i.e. baseline ECOG PS grade [0 versus 1]) into the Almac IWRS/IVRS system.		
Was the concealment of treatment allocation adequate?	All patients, family members, study personnel (at the study site, the sponsor, or participating Clinical Research Organization), and members of the IDMC were to remain blinded to treatment assignment until completion of the study with the exception of the circumstances described in the text below regarding blinding of treatment allocation. The matched placebo tablets given to patients in the placebo arm were also visually indistinguishable from the abiraterone acetate tablets.	Yes[a]	Low risk of bias
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	With a few exceptions, demographics and disease characteristics were balanced between the two treatment groups. The few differences in demographics and disease characteristics were not considered clinically relevant.	Yes	Low risk of bias
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	All patients, family members, study personnel (at the study site, the sponsor, or participating Clinical Research Organization), and members of the IDMC were to remain blinded to treatment assignment until completion of the study with the following exceptions: The Independent Biostatistician and Independent Statistical Programmer (employed by Novella) responsible for preparing interim tables, listings, and graphs for IDMC review who had no other responsibilities associated with the study. The IDMC, in order to evaluate whether the study should be stopped early for efficacy/futility or safety. Laboratory personnel performing plasma concentration assays for	Yes[a]	Low risk of bias

Accepted - see Erratum

Study question	Manufacturer's explanation: How is the question addressed in the study?	Manufacturer's assessment	ERG comment
	pharmacokinetic analysis. The Independent Biostatistician provided laboratory personnel with patients' randomisation codes without sponsor involvement. This process was undertaken to avoid futile pharmacokinetic analysis of placebo specimens that did not contain abiraterone. Laboratory personnel received no other data associated with the patients, with the exception of deviation listings pertaining to the collection of the pharmacokinetic samples.		
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No imbalances in dropouts between groups were observed	No imbalances in dropouts between groups were observed	No imbalances; but large numbers of drop-outs in both groups:
Is there any evidence to suggest that the authors measured more outcomes than they reported?	The clinical study report and associated journal and conference publications for the COU-AA-302 study were available and were reviewed. There was no indication that the clinical study report did not include all the measured outcomes	No	No, all outcomes were reported. However, no data were reported for QoL scores by arm (baseline, follow-up and change scores)
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	The ITT population included all patients randomised into the study; patients were to be classified according to assigned treatment group, regardless of the actual treatment received. The ITT population was used for all efficacy analyses, and all analyses of disposition, demographic, and baseline disease characteristics.	Yes	No. The ITT population did not include all patients randomised into the study, but those who received at least 1 dose of the allocated intervention.

Source: MS, Table 114, page 236[1]

Abbreviations: ECOG, Eastern Cooperative Oncology Group; IDMC, Independent Data Monitoring Committee; ITT, intent-to-treat; IVRS, Interactive Voice Response System; IWRS, Interactive Web Response System; N/A, not applicable; PS, performance status.

[a] It should be noted that an error resulted in the wrong AA study drug tablets being distributed by the sponsor. From 15.12.11 through 29.0312, these tablets were dispensed to 62 subjects assigned to the AAP group at 24 sites in the US and Canada. The affected tablets contained the proper dosage and formulation of AA, but were debossed with the text "AA250." The correct study tablets were not marked. This error resulted in the possibility that two subjects may have imputed their treatment assignment from the de-bossed tablets 5 days prior to the 20.12.11 cut-off date. The last bottle with de-bossed tablets was dispensed on 29.03.12. The issue was resolved.

Superseded See Erratum

4.1.5 Evidence synthesis

No evidence synthesis is included in the submission. As discussed in chapter 3.3, docetaxel was considered not appropriate as a comparator by the manufacturer. The remaining comparator was BSC (prednisone or prednisolone) was included in the trial.

ERG comment: The ERG agrees that for the comparison of abiraterone acetate in combination with prednisolone versus best supportive care in adults with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated, the COU-AA-302 trial is most likely the best source of clinical effectiveness evidence.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

In this section we will present the results from the COU-AA-302 trial. The study characteristics are presented in Table 4.4, and a summary of the methodology in Table 4.5.

4.2.1 Study characteristics of the COU-AA-302 trial

Table 4.4: Overview of the COU-AA-302 trial

Study	Intervention	Comparator	Population	Study refs.
COU-AA-302	AA (1 g q.d.) + prednisone/ prednisolone (5 mg b.i.d.)	Placebo + prednisone/ prednisolone (5 mg b.i.d.)	Asymptomatic or mildly symptomatic patients with an ECOG PS score of 0 or 1 with mCRPC after failure of ADT in whom chemotherapy was not yet clinically indicated	2nd interim analysis: • Original CSR COU-AA-302[52] • Ryan et al. 2013[30] • ASCO and ESMO abstracts[53, 54] 3rd interim analysis • Updated CSR COU-AA-302[43] • ASCO and ASCO GU abstracts[45-51]

Source: MS, Table 10, page 48[1]

Abbreviations: AA, abiraterone acetate; ASCO, American Society of Clinical Oncology; ASCO GU, American Society of Clinical Oncology Genitourinary Cancers; CSR, clinical study report; ECOG, Eastern Cooperative Oncology Group; ESMO, European Society for Medical Oncology; mCRPC, metastatic castration-resistant prostate cancer; PS, performance status

Table 4.5: Summary of methodology of the COU-AA-302 trial

Trial title	COU-AA-302
Location	151 sites worldwide in Australia, Belgium, Canada, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, the UK, and the US (12 study sites and 9.0% of patients were from the UK)
Design	Phase III, randomised, double-blind, placebo-controlled study of AAP versus PP (1:1)
Patient population	Asymptomatic or mildly symptomatic patients with mCRPC after failure of ADT in whom chemotherapy was not yet clinically indicated

Duration of study	Patients treated until disease progression, as defined by radiographic progression or unequivocal clinical progression (e.g. need for alternative anti-cancer therapy) The first patient was enrolled on 28.04.09 and the last patient was enrolled on 23.06.10. Follow-up is ongoing
Method of randomisation	Eligible patients were randomised (1:1) using a centralised IWRS/IVRS and were stratified by baseline ECOG PS grade (0 versus 1)
Method of blinding (care provider, patient and outcome assessor)	Double blind: patients and investigators were blinded to the study drug. Placebo matched the AA tablets in size, colour and shape. All patients, family members, study personnel, and members of the IDMC were to remain blinded to treatment assignment until completion of the study
Intervention(s) (n =) and comparator(s) (n =)	• AAP: AA (1 g q.d.) + prednisone/prednisolone (5 mg b.i.d.) until disease progression (N=546) • PP: placebo + prednisone/prednisolone (5 mg b.i.d.) until disease progression (N=542) Each treatment cycle was 28 days
Primary outcomes	The co-primary efficacy endpoints of this study were rPFS and OS. The outcome rPFS was the time from randomisation to the occurrence of one of the following, whichever occurred first: progression by bone scan (according to adapted PCWG2 criteria), progression by CT or MRI (according to modified RECIST criteria), or death (but not rising PSA). The rPFS distribution, median rPFS, and the 95% CIs were estimated using the Kaplan–Meier method. OS was measured from the date of randomisation to the date of all-cause death. OS data were collected throughout the study treatment phase and during follow-up. Survival time of living patients was censored at the last date a patient was known to be alive or lost to follow up. The OS distribution curve, median OS, and 95% CIs were estimated using the Kaplan–Meier method. Statistical inference was evaluated according to the group sequential testing design.
Secondary outcomes	• Time from randomisation to first opiate use for cancer pain • Time from randomisation to initiation of cytotoxic chemotherapy for prostate cancer • Time from randomisation to first established clinical deterioration in terms of ECOG PS by ≥1 grade • Time from randomisation to first established PSA progression
Other endpoints	• PSA response rate • Objective response rate • Duration of response • Time from randomisation to first established analgesic progression • Functional status • Time from randomisation to first established functional status deterioration • Time from randomisation to first established progression in average pain intensity (BPI-SF) • Time from randomisation to first established progression in worst pain intensity (BPI-SF) • Time from randomisation to first established progression in pain interference
Planned analyses	• Single rPFS analysis at 100% of total expected rPFS events (378 events) • First interim OS analysis at 15% of total OS events (116 events) • Second interim OS analysis at 40% of total OS events (311 events) • Third interim OS analysis at 55% of total events (425 events) • Final OS analysis at 100% of total events (773 events)
Duration of follow-up	Patients were to be treated until disease progression. After discontinuing study treatment, patients were to be contacted every 3 months. Follow-up was to continue for up to 60 months or until the patient died, was lost to follow-up, or withdrew informed consent

Source: MS, Table 12, pages 51-52[1]
Abbreviations: AA, abiraterone acetate; AAP, abiraterone acetate plus prednisone/prednisolone; BPI-SF, brief pain inventory short form; CI, confidence interval; CT, computed tomography; ECOG, Eastern Cooperative Oncology Group; IDMC, Independent Data Monitoring Committee; IVRS, Interactive Voice Response System; IWRS, Interactive Web Response System; mCRPC, metastatic castration-resistant prostate cancer; MRI, magnetic resonance imaging; OS, overall survival; PCWG, Prostate Cancer Working Group; PP, placebo plus prednisone; PS, performance status; PSA, prostate-specific antigen; RECIST, Response Evaluation Criteria In Solid Tumors; rPFS, radiographic progression-free survival.

Statistical analysis

The primary hypothesis was that patients receiving AAP would have improved OS and/or rPFS compared with patients receiving PP. Therefore, OS and rPFS were co-primary efficacy endpoints. The intent-to-treat (ITT) population was used for all efficacy analyses, and analyses of disposition, demographics, and baseline disease characteristics. Safety analyses were summarised using the Safety Population (all patients in the randomised population who received any study medication).

Time-to-event analyses (OS, PFS, time to progression) were compared between the two treatment groups using the log-rank test procedure in the ITT population according to the stratification factors specified at the time of randomisation.

Estimates of hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) were provided using a Cox proportional hazard model stratified by the same stratification factors specified at randomisation. Kaplan–Meier survival curves were generated. The chi square or Fischer's exact test methods were used to compare proportions.

The original protocol and statistical analysis plan for OS included two interim analyses (after 15% and 50% of 773 required OS events) and a final analysis (after 100% of the required OS events). However, this was amended, before the first interim OS analysis had been done, to three interim analyses (after 15%, 40%, and 55% of the required OS events). The additional interim analysis after 40% of OS events was added because a survival advantage was observed with AAP versus PP in study COU-AA-301 (**post**-chemotherapy).[55, 56] Only one statistical analysis (by Independent Review) was planned for rPFS, when 378 rPFS events had occurred. This coincided with the first OS interim analysis and therefore, the first interim analysis included both OS and rPFS analyses. Updated analyses of rPFS to include additional events based on Investigator Review were also conducted at each of the subsequent OS interim analyses.

As the COU-AA-302 trial had co-primary endpoints of rPFS and OS, the p-value of 0.05 (i.e. the probability that an observed difference could have occurred by chance, 5% in this case) had to be shared amongst rPFS (which received 0.01, or 1%) and OS (which received 0.04, or 4%). In addition, the overall p-value of 0.04 for OS needed to be spread over multiple time points (after 15%, 40% and 55% of deaths). Therefore, the p-value at any of these time points had to be much lower than 0.04. Using the O'Brien-Fleming statistical stopping boundaries, as implemented by the Lan-DeMets alpha spending method, p values of <0.0001, 0.0005, and

0.0034 were required at the first, second, and third interim analyses, respectively, to show statistical significance for OS (Table 4.6).

Table 4.6: COU-AA-302 study planned rPFS and OS analyses

Variable	rPFS analysis	OS analyses			
	Single planned analysis	Interim 1[a]	Interim 2	Interim 3	Final
Percentage of total events	~100	~15	40	55	100
Projected observed events	378	116	311	425	773[b]
Efficacy boundary (HR)	0.667	0.336	0.672	0.751	0.861
Projected cumulative stop probability under (H_0)	–	<0.0001	0.0005	0.0034	0.04
Actual data cut-off date for the analysis	20.12.10	20.12.10	20.12.11	22.05.12	Projected to be 31.03.14

Source: MS, Table 16, page 59[1]
[a]At the time of rPFS analysis.
[b]Required to detect a difference between a median OS of 22 months in the placebo group and a median OS of 27.5 months in the abiraterone acetate group (HR=0.80) at the 2-tailed significance level of 0.04 with a power of 85%.
Abbreviations: HR, hazard ratio; OS, overall survival; rPFS, radiographic progression-free survival.

ERG Comment: According to the manufacturer, the Independent Data Monitoring Committee (IDMC) concluded on 27 February 2012, that patients in the abiraterone arm had a 'highly significant advantage', even though the hazard ratio (HR) for OS had not reached the stringent pre-specified statistical significance level (0.0034). The committee unanimously recommended stopping the study, unblinding, and allowing cross-over. The study was unblinded on 2 April 2012. Cross-over from PP to AAP occurred following unblinding (02.04.12) for three patients by the third interim analysis (22.05.12). Neither the second nor third interim analysis OS results met the pre-specified statistical significance levels (see results below). Because cross-over is now allowed, it is unlikely that the trial will ever show a significant survival benefit.

Sample size, power calculation

The planned sample size of approximately 1,000 patients (randomised 1:1 to AAP or PP) provided 85% power to detect a difference between a median OS of 27.5 months in the AAP group and a median OS of 22 months in the PP group (HR=0.80) under the assumption of a two-tailed significance level of 0.04 and required 773 deaths to have occurred. This planned sample size also provided 91% power to detect a difference between median rPFS of six months for the AAP group and a median rPFS of four months in the PP group (HR=0.667) under the assumption of an exponential model with proportional hazards and a two-tailed level of significance of 0.01 and required 378 rPFS events to have occurred.

Discontinuation and censoring

Reasons for patient discontinuation included: unequivocal clinical progression; sustained side effects; initiation of new anticancer treatment; administration of prohibited medications; and patient withdrawal of consent.

Survival time of living patients was censored at the last date they were known to be alive or lost to follow-up as of the cut-off date for the interim analysis database lock. In the analysis of rPFS, the following censoring rules applied:

- If the patient did not have a baseline scan or on-study scans, the patient was to be censored on the date of randomisation
- If the patient did not show progression according to modified RECIST or bone scan, the patient was to be censored on the date of the last scheduled scan
- Patients were to be censored on the date of the last scan that showed no disease progression if the patient received another therapy (i.e. cytotoxic chemotherapy) known or intended for the treatment of mCRPC during the study; or the patient missed ≥2 planned radiographic scans or had ≥2 consecutive unreadable scans.

Patients with no opiate use at the time of analysis were censored at the last known date of no opiate use; patients with no opiate use assessment were censored at the date of randomisation. Similarly for: no cytotoxic chemotherapy, no ECOG PS deterioration, no PSA progression, no pain progression, no worst pain intensity progression, and no progression in analgesic use.

Eligibility criteria

The full inclusion and exclusion criteria for the COU-AA-301 trial are presented in the table below.

Table 4.7: Eligibility criteria of the COU-AA-302 trial

Inclusion criteria	Exclusion criteria
• Male, ≥18 years of age • Histologically or cytologically confirmed adenocarcinoma of the prostate • Metastatic disease documented by positive bone scan or metastatic lesions, other than liver or visceral metastasis, on CT or MRI. If lymph node metastasis was the only evidence of metastasis, it must have been ≥2 cm in diameter • Surgical or medical castration, as demonstrated by serum testosterone levels <50 ng/dL (<2.0 nM). If the patient was treated with LHRH agonists, the therapy must have been initiated ≥4 weeks prior to Cycle 1 Day 1 and must have continued throughout the study • Documented prostate cancer progression by PSA, according to adapted PCWG2, or radiographic	• Use of opiate analgesics for cancer-related pain, including codeine and dextropropoxyphene, within 4 weeks of Cycle 1 Day 1 • Prior cytotoxic chemotherapy or biological therapy for the treatment of CRPC • Radiation therapy for treatment of the primary tumour within 6 weeks of Cycle 1 Day 1 • Radiation or radionuclide therapy for treatment of mCRPC • Prior therapy with ketoconazole for prostate cancer lasting >7 days • Prior systemic therapy with an azole drug (e.g. fluconazole, itraconazole) within 4 weeks of Cycle 1 Day 1 • Prior flutamide treatment within 4 weeks of Cycle 1 Day 1 (patients whose PSA did not decline for ≥3 months in response to antiandrogen given as a second-line or later intervention required only a 2-week washout prior to Cycle 1 Day 1) • Prior bicalutamide or nilutamide within 6

progression according to modified RECIST criteria • Asymptomatic or mildly symptomatic from prostate cancer, as defined by a score of 0–1 (asymptomatic) or 2–3 (mildly symptomatic) for BPI-SF Question #3 • Previous antiandrogen therapy followed by documented PSA progression after discontinuing the antiandrogen (≥4 weeks since last flutamide, ≥6 weeks since last bicalutamide or nilutamide) prior to enrolment • ECOG PS grade 0 or 1 • Haemoglobin ≥10.0 g/dL, independent of transfusion • Platelet count ≥100,000/μL • Serum albumin ≥3.5 g/dL • Serum creatinine <1.5 x ULN or a calculated creatinine clearance ≥60 mL/min • Serum potassium ≥3.5 mmol/L • Adequate liver function as defined by: Serum bilirubin <1.5 x ULN (except for patients with documented Gilbert's disease) AST or ALT <2.5 x ULN • Ability to swallow the study medication whole as a tablet • Life expectancy ≥6 months • Patients who had partners of childbearing potential must have been willing to use a method of birth control with adequate barrier protection as determined to be acceptable by the principal investigator and sponsor during the study and for 13 weeks after the last study medication administration • Able to provide written informed consent • Able to provide written Authorisation for Use and Release of Health and Research Study Information (US sites only) or Data Protection Consent (European sites only)	weeks of Cycle 1 Day 1 (patients whose PSA did not decline for ≥3 months in response to antiandrogen given as a second-line or later intervention required only a 2-week washout prior to Cycle 1 Day 1) • Active infection or other medical condition that would have made prednisone (corticosteroid) use a contraindication • Any chronic medical condition that required a higher dose of corticosteroid than 5 mg b.i.d. prednisone • Pathological finding of small cell carcinoma of the prostate • Known liver, brain, or visceral organ metastasis • Uncontrolled hypertension (systolic BP ≥160 mmHg or diastolic BP ≥95 mmHg). Patients with a history of hypertension were allowed, provided BP was controlled by antihypertensive therapy • Active or symptomatic viral hepatitis or chronic liver disease • History of pituitary or adrenal dysfunction • Clinically significant heart disease as evidenced by myocardial infarction, or arterial thrombotic events in the 6 months prior to screening, severe or unstable angina, or NYHA Class II IV heart disease or cardiac ejection fraction measurement of <50% at baseline • Atrial fibrillation, or other cardiac arrhythmia requiring medical therapy • Other malignancy, except non-melanoma skin cancer, with a ≥30% probability of recurrence within 24 months • Current enrolment in an investigational drug or device study or participation in such a study within 30 days of Cycle 1 Day 1 • Condition or situation which, in the investigator's opinion, might have put the patient at significant risk, confounded the study results, or interfered significantly with the patient's participation in the study

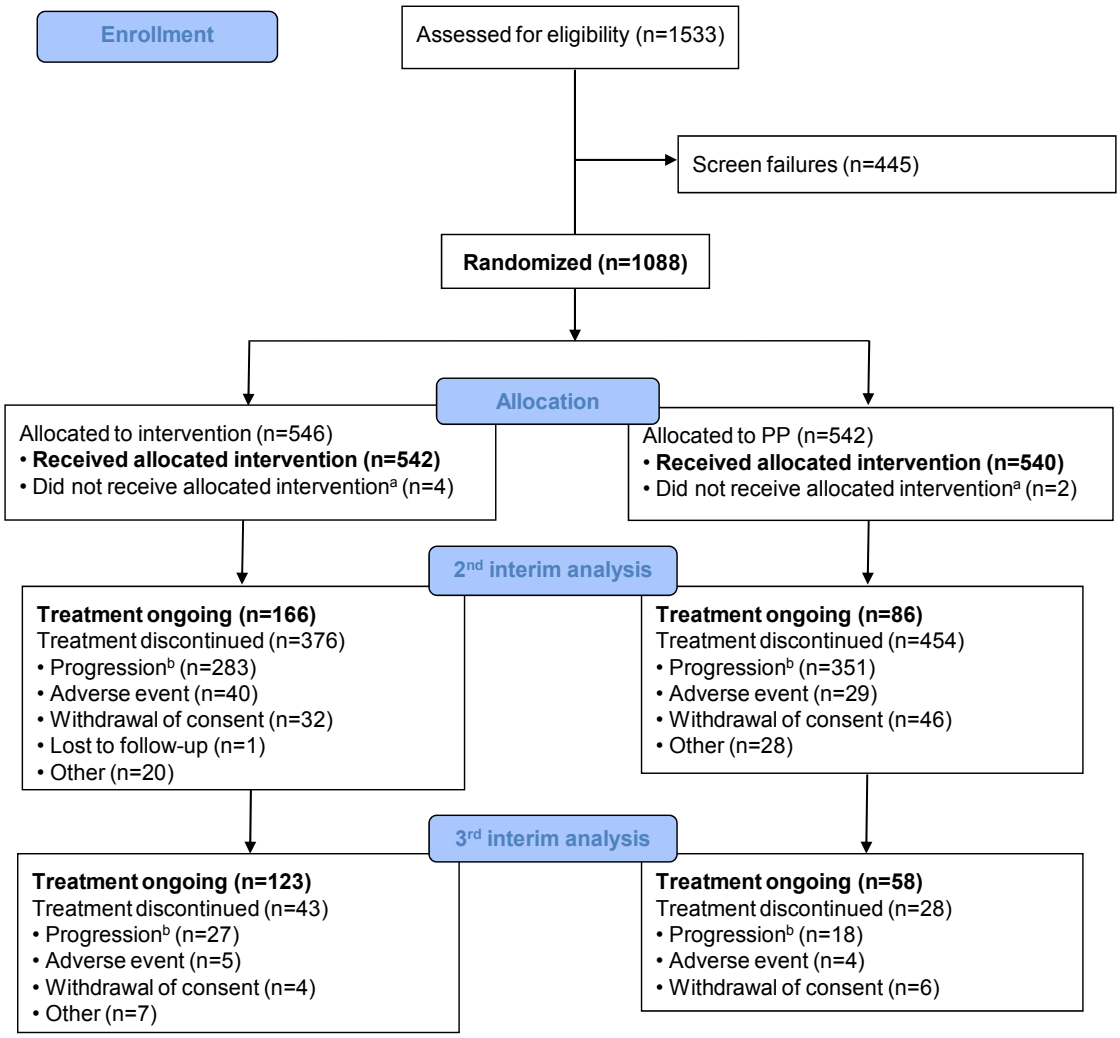
Source: MS, Table 13, pages 53-54[1]
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BP, blood pressure; BPI-SF, brief pain inventory short form; CRPC, castration-resistant prostate cancer; CT, computed tomography; ECOG, Eastern Cooperative Oncology Group; LHRH, luteinising hormone-releasing hormone; mCRPC, metastatic castration-resistant prostate cancer; MRI, magnetic resonance imaging; NYHA, New York Heart Association; PCWG, Prostate Cancer Working Group; PS, performance status; PSA, prostate-specific antigen; RCT, randomised controlled trial; RECIST, Response Evaluation Criteria In Solid Tumors; ULN, upper limit of normal.

In the clarification letter we asked the manufacturer whether the population who were eligible to participate in COU-AA-302 were mutually exclusive from the population who could receive docetaxel in clinical practice. According to the manufacturer, the population eligible to participate in COU-AA-302 is not mutually exclusive from the population who could receive docetaxel in clinical practice, because the population determined by the license was requested by the regulators on the basis of the study results, rather than as a result of the study being designed to specifically for patients who are not yet suitable for docetaxel. Therefore, it is possible that there are patients included in COU-AA-302 for whom docetaxel may have been considered suitable in routine UK practice. Although specifically asked in the clarification letter, the manufacturer did not provide the number of patients for whom this might be the case.[40]

In the COU-AA-302 trial, a total of 1,088 patients were recruited and randomised to abiraterone acetate plus prednisone/prednisolone (n=546) or placebo plus prednisone (n=542). 1,082 patients received at least one dose of the allocated intervention and constituted the safety population.

The flow of patients through the study is presented in Figure 4.1.

Figure 4.1: COU-AA-302 study patient flow diagram (data from third interim analysis)



Source: MS, Figure 3, page 61[1]

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; ITT, intent-to-treat; PP, placebo plus prednisone.

[a]Five patients were mistakenly randomised instead of indicating them as screening failures; one patient (AAP arm) withdrew consent after randomisation but before starting treatment.

[b]Radiographic and/or unequivocal clinical progression.

Treatment exposure

Patients continued treatment with AAP or PP until disease progression (determined according to radiographic and clinical measures). The median treatment duration was 13.8 months (15 cycles initiated) in the AAP arm and 8.3 months (nine cycles initiated) in the PP arm.

Treatment discontinuation

The most common reason for discontinuation by the third interim analysis was disease progression, which was observed in 57% and 68% of patients in the AAP and PP groups, respectively (Table 4.8). AEs that led to discontinuation of AAP or PP (not including discontinuations of prednisone/prednisolone only) were observed in 8% and 6%, respectively.

Table 4.8: Treatment discontinuations at the third interim OS analysis (22.05.12; Safety population)

	AAP (N=542)	**PP (N=540)**
Patients treated, n (%)	542 (100.0)	540 (100.0)
Treatment discontinued	419 (77.3)	482 (89.3)
Treatment ongoing	123 (22.7)	58 (10.7)
Reasons for discontinuation, n (%)		
Radiographic and unequivocal clinical progression	66 (12.2)	56 (10.4)
Radiographic progression only	126 (23.2)	172 (31.9)
Unequivocal clinical progression only	118 (21.8)	141 (26.1)
AE	45 (8.3)	33 (6.1)
Withdrawal of consent to treatment	36 (6.6)	52 (9.6)
Other	27 (5.0)	28 (5.2)
Lost to follow-up	1 (0.2)	0

Source: MS, Table 17, page 63[1]

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; AE, adverse event; OS, overall survival; PP, placebo plus prednisone/prednisolone.

4.2.2 Patient characteristics in the COU-AA-302 trial

The demographics, baseline disease characteristics and medical history of patients in both treatment arms are presented in Table 4.9. Overall, both treatment arms were well balanced.

Table 4.9: Characteristics of participants in the COU-AA-302 trial by randomised group (ITT)

COU-AA-302 Baseline characteristic	**AAP (N=546)**	**PP (N=542)**
Age, years, mean (SD)	70.5 (8.8)	70.1 (8.7)
Male, %	100	100
White race N n (%)	 545 520 (95.4)	 540 510 (94.4)
Weight, kg N Mean (SD)	 527 88.5 (15.1)	 520 89.6 (17.0)
Height, cm N Mean (SD)	 539 174.7 (7.5)	 536 175.2 (7.9)
Time from initial diagnosis to first dose, years N Mean (SD)	 542 6.7 (4.9)	 540 6.5 (4.8)
PSA at initial diagnosis, ng/mL N Mean (SD)	 470 174.0 (540.4)	 454 219.7 (888.8)
Baseline PSA, ng/mL N Mean (SD)	 546 133.4 (323.6)	 539 127.6 (387.9)

COU-AA-302 **Baseline characteristic**	**AAP** **(N=546)**	**PP** **(N=542)**
TNM stage at initial diagnosis, n (%)		
Stage I	1 (0.2)	2 (0.4)
Stage II	86 (15.8)	70 (12.9)
Stage III	52 (9.5)	63 (11.6)
Stage IV	201 (36.8)	191 (35.2)
Incomplete reporting	206 (37.7)	216 (39.9)
Gleason score at initial diagnosis, n (%)		
N	488	508
<7	65 (13.3)	64 (12.6)
7	160 (32.8)	190 (37.4)
≥8	263 (53.9)	254 (50.0)
Extent of disease, n (%)		
N	544	542
Bone	452 (83.1)	432 (79.7)
Bone only	274 (50.4)	267 (49.3)
Soft tissue or node	267 (49.1)	271 (50.0)
Bone, soft tissue, or node	544 (100.0)	542 (100.0)
Other	4 (0.7)	7 (1.3)
Baseline BPI-SF #3 pain score (worst pain over last 24 hours), n (%)		
N	539	534
0–1	370 (68.6)	346 (64.8)
2–3	129 (23.9)	147 (27.5)
≥4	40 (7.4)	41 (7.7)
Mean (SD)	1.2 (1.7)	1.2 (1.6)
ECOG PS, n (%)		
0	413 (75.6)	409 (75.5)
1	133 (24.4)	133 (24.5)
Previous prostate cancer therapy, n (%)		
N	544	542
Surgery	256 (47.1)	244 (45.0)
Radiotherapy	283 (52.0)	303 (55.9)
Hormonal	544 (100.0)	542 (100.0)
Orchiectomy	20 (3.7)	24 (4.4)
Other	82 (15.1)	63 (11.6)

Source: MS, Table 14, pages 54-55[1]

Abbreviations: BPI-SF, brief pain inventory short form; ECOG, Eastern Cooperative Oncology Group; ITT, intent-to-treat; PS, performance status; PSA, prostate-specific antigen; RCT, randomised controlled trial; SD, standard deviation; TNM, tumour□node□metastasis.

4.2.3 Results of the COU-AA-302 trial

The final scope lists the following outcome measures: overall survival, progression-free survival including radiographic progression-free survival, response rate, prostate specific antigen (PSA) response, adverse effects of treatment and health-related quality of life. These results will be discussed below.

Results presented in the MS are based on the results from the second (data cut-off 20/12/11; published[30]) and third (data cut-off 22/5/12; unpublished CSR[43] and published abstracts[45-51]) interim analyses of the COU-AA-302 study, which were conducted after approximately 40% and 55% of the total OS events had occurred. Efficacy analyses were performed using the ITT population, which included all randomised patients who received at least 1 dose of the allocated intervention.

Overall Survival

At the third interim analysis cut-off date (22 May 2012), 434 death events had been observed (200 [36.6%] in the AAP group and 234 [43.2%] in the PP group). Treatment with AAP was associated with an increase in OS compared with PP treatment, representing a 21% decrease in the relative risk of death (absolute risk reduction 6.5%), and an increase in OS of 5.2 months (Table 4.10). The Kaplan–Meier plot for OS (third interim analysis) is presented in Figure 4.2.

Neither the second nor third interim analysis OS results met the pre-specified statistical significance levels. The study was halted by the IDMC between these two analyses to allow cross-over from PP to AAP. Three patients had crossed over from PP to AAP by the third interim analysis.

Table 4.10: OS of patients treated with either AAP or PP (ITT)

	Interim analysis 2		**Interim analysis 3**	
	AAP (N=546)	**PP (N=542)**	**AAP (N=546)**	**PP (N=542)**
Number of deaths, n (%)	147 (26.9)	186 (34.3)	200 (36.6)	234 (43.2)
OS[a] (months), median (95% CI)	NR (NR, NR)	27.2 (26.0, NR)	35.3 (31.2, 35.3)	30.1 (27.3, 34.1)
HR (95% CI)[b]	0.75 (0.61, 0.93)		0.79 (0.66, 0.96)	
p value[c]	0.0097		0.0151	
p value required for significance	0.0005		0.0034	

Source: MS, Table 20, page 67[1]

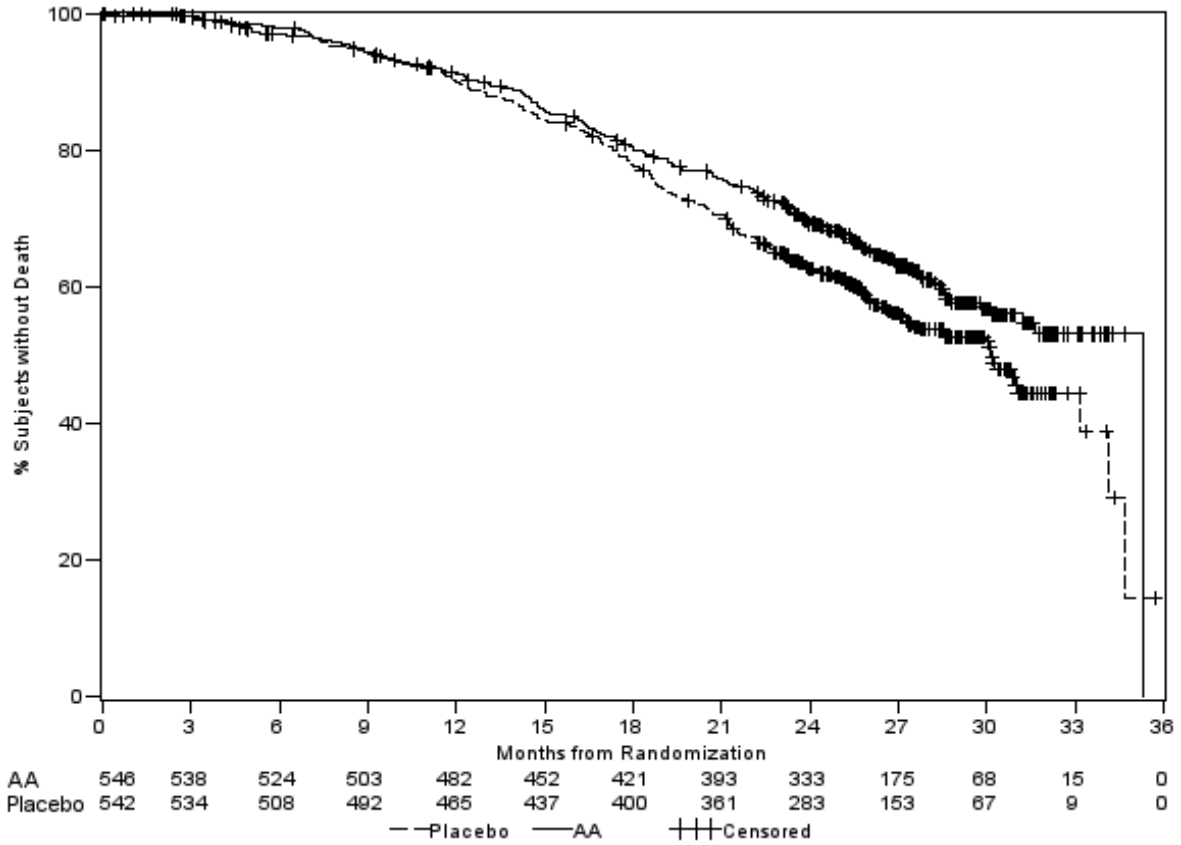
[a]Survival time of living patients was censored at the last date a patients was known to be alive or lost to follow-up as of the cut-off date for the interim analysis.

[b]HRs from a stratified proportional hazards Cox model. HRs <1 favour AAP.

[c]p value from a log-rank test stratified by ECOG PS score (0 or 1).

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; NR, not reached; OS, overall survival; PP, placebo plus prednisone/prednisolone.

Figure 4.2: Kaplan–Meier curve of OS – ITT population (COU-AA-302 study third interim analysis)



Source: MS, Figure 6, page 68[1]

Abbreviations: AA, abiraterone acetate plus prednisone/prednisolone; ITT, intent-to-treat; OS, overall survival; Placebo, placebo plus prednisone/prednisolone.

Progression-free survival (including rPFS)

Treatment with AAP resulted in a 48% relative reduction in the risk of radiographic progression compared with PP (absolute risk reduction 11.5%), and increased PFS by 8.2 months (Table 4.11). The Kaplan–Meier plot for rPFS (third interim analysis) is presented in Figure 4.3.

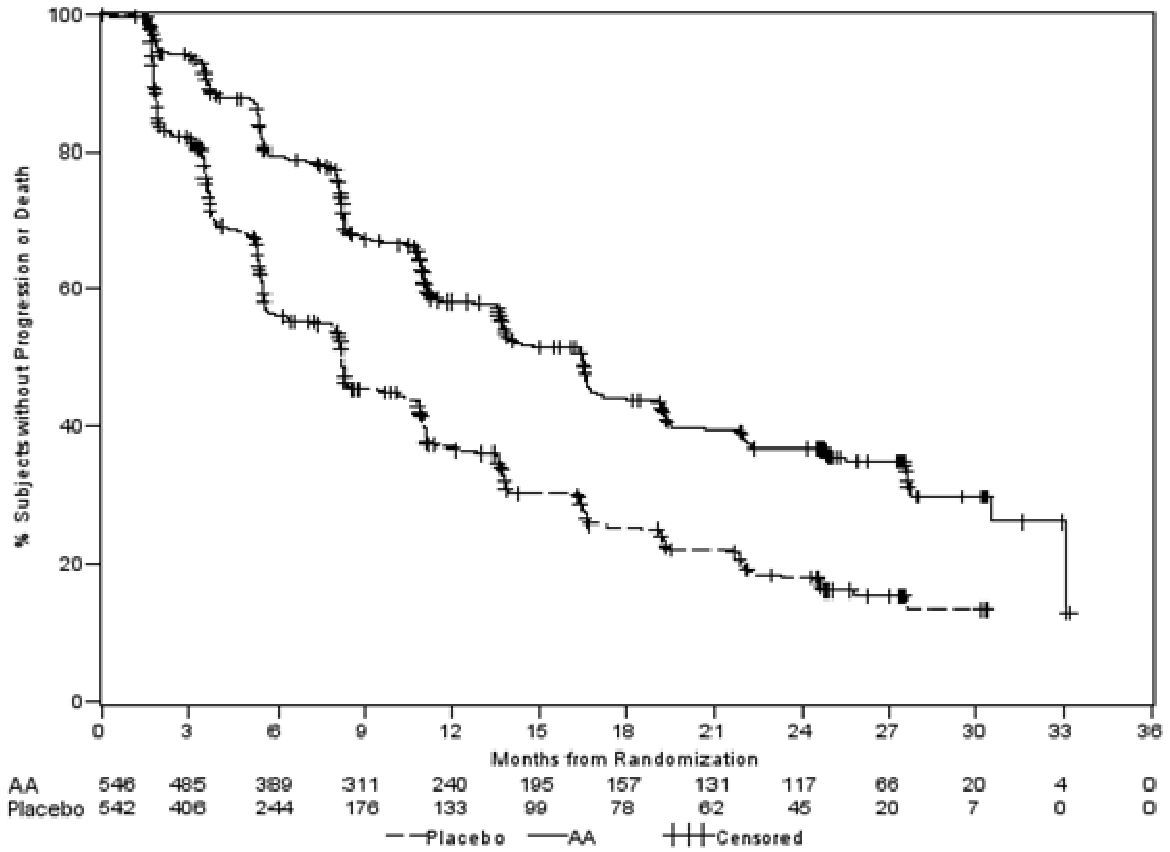
Table 4.11: rPFS in patients treated with either AAP or PP (ITT)

	Interim analysis 2		**Interim analysis 3**	
	AAP (N=546)	**PP (N=542)**	**AAP (N=546)**	**PP (N=542)**
Number of patients with PFS event, n (%)	271 (49.6)	336 (62.0)	292 (53.5)	352 (64.9)
Time-to-event[a] (months), median (95% CI)	16.5 (13.8, 16.8)	8.3 (8.1, 9.4)	16.5 (13.8, 16.8)	8.2 (8.0, 9.4)
HR (95% CI)[a]	0.53 (0.45, 0.62)		0.52 (0.45, 0.62)	
p value[b]	<0.0001		<0.0001	

Source: MS, Table 19, page 66[1]

[a]HR is from a stratified proportional hazards Cox model. HRs <1 favour AAP.
[b]p value is from a log-rank test stratified by ECOG PS score (0 or 1).
Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; PFS, progression-free survival; PP, placebo plus prednisone/prednisolone.

Figure 4.3: Kaplan–Meier curve of rPFS – ITT population (third interim analysis – 22.05.12, 55% data cut-off)



Source: MS, Figure 5, page 66[1]
Abbreviations: AA, abiraterone acetate plus prednisone; ITT, intent-to-treat; Placebo, placebo plus prednisone; rPFS, radiographic progression-free survival.

Response rate (including prostate specific antigen (PSA) response)

Response rates are reported in Appendix 14 of the MS (MS, Tables 135 and 136, pages 275-276). Significant differences in favour of the AAP group were observed for objective response rate (complete or partial response according to modified RECIST criteria), PSA response and duration of response.

Table 4.12: Best overall response (confirmed) based on modified RECIST criteria in patients with measurable disease at baseline (investigator review – ITT population – third interim analysis – 22.05.12, 55% data cut-off)

Best overall response	**AAP (N=546)**	**PP (N=542)**
Patients with measurable disease at baseline	[redacted]	[redacted]
Patients with best overall response, n (%)		
Complete response	[redacted]	[redacted]
Partial response	[redacted]	[redacted]
Stable disease	[redacted]	[redacted]
Progressive disease	[redacted]	[redacted]
Not evaluable	[redacted]	[redacted]

Source: MS, Table 136, page 276[1]
Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; PP, placebo plus prednisone/ prednisolone.

Table 4.13: Response to treatment outcomes results ITT population (third interim analysis – 22.05.12, 55% data cut-off)

Outcome	**Median (95% CI) months to outcome**		**HR (95% CI)**	**p value**
	AAP (N=546)	**PP group (N=542)**		
PSA response[a]	[redacted]	[redacted]	[redacted]	[redacted]
Confirmed PSA response	[redacted]	[redacted]	[redacted]	[redacted]
Duration of response[b] in patients with measurable disease at baseline	[redacted]	[redacted]	[redacted]	[redacted]
Objective response rate in patients with measurable disease at baseline[c]	[redacted]	[redacted]	[redacted]	[redacted]

Source: MS, Table 135, page 275[1]
Data are median (95% CI) or n/N (%).
[a]Confirmed and not confirmed.
[b]By modified RECIST and bone progression criteria.
[c]Proportion of patients with measurable disease achieving a complete or partial response according to modified RECIST criteria (baseline lymph node size was required to be ≥2 cm to be considered a targeted lesion).
Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; CI, confidence interval; HR, hazard ratio; NR, not reached; PP, placebo plus prednisone; PSA, prostate-specific antigen.

Health-related quality of life

HRQL was assessed in the COU-AA-302 study via the FACT-P instrument. However, no results are report by treatment arm for baseline, follow-up or change scores. Instead, the manufacturer presents a summary of the time to a decrease of ≥10 points for the FACT-P total score and all FACT-P subscale results at the time of the third interim OS analysis and Kaplan–Meier plots of the HRQL data. The summary of the time to a decrease of ≥10 points

for all FACT-P subscale results at the time of the third interim OS analysis is copied in the table below.

Table 4.14: HRQL outcomes results – ITT population (third interim analysis – 22.05.12, 55% data cut-off)

FACT-P Subscale	Median (95% CI) time to progression[a] (months)		HR of AAP/PP (95% CI)	p value
	AAP	PP		
FACT-P (Total Score)	12.7 (11.1, 14.0)	8.3 (7.4, 10.6)	0.79 (0.67, 0.93)	0.0046
PCS	11.1 (8.6, 13.8)	5.8 (5.5, 8.3)	0.72 (0.61, 0.84)	<0.0001
TOI	13.9 (12.0, 16.5)	9.3 (8.3, 11.1)	0.77 (0.65, 0.91)	0.0018
FACT-G	16.6 (13.8, 19.4)	11.1 (8.5, 14.0)	0.76 (0.64, 0.91)	0.0023
PWB	14.8 (13.6, 16.8)	11.1 (9.1, 13.8)	0.76 (0.64, 0.91)	0.0019
SFWB	18.4 (13.8, 24.8)	16.6 (11.1, NE)	0.95 (0.78, 1.15)	0.5774
EWB	22.5 (17.4, 27.9)	14.2 (13.3, 19.5)	0.73 (0.61, 0.89)	0.0017
FWB	13.3 (11.0, 15.7)	8.4 (6.5, 10.1)	0.77 (0.65, 0.91)	0.0016

Source: MS, Table 135, page 275[1]

[a]Decrease of ≥10 points.

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; CI, confidence interval; EWB, emotional well being; FACT-G, Functional Assessment of Cancer Therapy General; FACT-P: Functional Assessment of Cancer Therapy – Prostate; FWB, functional well being; HR, hazard ratio; ITT, intent-to-treat; PCS, Prostate Cancer Scale; PP, placebo plus prednisone; PWB, physical well being; SFWB, social/family well being; TOI, Total Outcome Index.

ERG comment: As reported in Chapter 3.4, the ERG has concerns about the validation of the FACT-P and the way the FACT-P instrument was used in this submission. In addition the results are only presented as time to event data and not as change scores by treatment arm. It is not clear whether the data presented here constitute significant differences in experienced quality of life by patients between treatment arms.

As stated by the manufacturer, "the main drivers of reduced health-related quality of life (HRQL) reported by patients with mCRPC are bone pain, fatigue, sexual disturbances and interrupted social relationships".[57, 58] The only component reported in the MS is pain, which was measured using the Brief Pain Inventory-Short Form (BPI-SF). However, BPI scores were only reported as time to event data (see Table 4.15 below).

Time to progression in average pain intensity and worst pain intensity showed no significant differences between treatment arms. All other pain-related outcomes favoured AAP over PP.

Table 4.15: Pain-related outcomes results – ITT population (third interim analysis – 22.05.12, 55% data cut-off)

Outcome	**Median time to progression in months (95% CI)**		**HR (95% CI)[a]**	**p value[b]**
	AAP	**PP**		
Time to progression in				
Average pain intensity[c]	26.7 (19.3, NE)	18.4 (14.8, 24.9)	0.83 (0.68, 1.01)	0.0612
Average pain intensity using 2-point increase threshold (post-hoc analysis)[d]	[redacted]	[redacted]	[redacted]	[redacted]
Worst pain intensity[e]	25.8 (NE, NE)	20.3 (NE, NE)	0.85 (0.69, 1.04)	0.1134
Worst pain intensity using 2-point increase threshold (post-hoc analysis)[f]	[redacted]	[redacted]	[redacted]	[redacted]
Pain interference[g]	10.3 (NE, NE)	7.4 (NE, NE)	0.80 (0.68, 0.93)	0.0049
Analgesic use	[redacted]	[redacted]	[redacted]	[redacted]

Source: MS, Table 137, page 276[1]

[a]HR is from stratified proportional hazards model. HRs <1 favour AAP.

[b]p value from a log-rank test stratified by ECOG PS grade (0 or 1).

[c]Time from randomisation to the first date the patient experienced a BPI-SF increase by ≥30% from baseline in the average of the BPI-SF pain intensity item scores (#3, #4, #5, #6) that was observed at two consecutive evaluations ≥4 weeks apart without a decrease in analgesic usage score.

[d]Time from randomisation to the first date the patient experienced an increase by 2 points from baseline in the BPI-SF average pain intensity (average of BPI-SF items #3, #4, #5, #6) observed at two consecutive evaluations ≥4 weeks apart without decrease in analgesic usage score.

[e]Time from randomisation to the first date the patient experienced an increase by ≥30% from baseline in the BPI-SF worst pain intensity item (#3) observed at two consecutive evaluations ≥4 weeks apart without a decrease in analgesic usage score.

[f]Time from randomisation to the first date the patient experienced an increase by 2 points from baseline in the BPI-SF worst pain intensity item (#3) observed at two consecutive evaluations ≥4 weeks apart without decrease in analgesic usage score.

[g]Time from randomisation to the first date the patient experienced an increase at any visit in baseline BPI-SF pain interference score of one half the baseline SD of BPI-SF.

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; BPI-SF, brief pain inventory short form; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; ITT, intent-to-treat; NE, not estimable; PP, placebo plus prednisone/prednisolone; PS, performance status; SD, standard deviation.

Adverse events of treatment

All adverse events data presented in the MS are from the COU-AA-302 trial. In the COU-AA-302 trial, the median observation duration (including follow-up) up to the third data cut-off was 27.1 months. The median treatment time (minimum; maximum) was 13.8 (0.3; 34.9) months in the abiraterone arm, and 8.3 (0.1; 32.4) months in the placebo arm. Overall safety and tolerability data are shown in Table 4.16 below.

Table 4.16: Overall safety results – safety population (third interim analysis – 22.05.12, 55% data cut-off)

Safety outcome, n (%)	**AAP (N=542)**	**PP (N=540)**	**RR (95% CI)[a]**
Number of patients with TEAEs[b]	538 (99.3)	524 (97.0)	**1.02 (1.01, 1.04)**
Drug-related[c]	426 (78.6)	414 (76.7)	1.03 (0.96, 1.09)
Number of patients with grade 3–4 TEAEs	267 (49.3)	235 (43.5)	1.13 (1.00, 1.29)
Drug-related[c]	127 (23.4)	97 (18.0)	**1.30 (1.03, 1.65)**
Number of patients with treatment-emergent SAEs[b]	188 (34.7)	146 (27.0)	**1.28 (1.07, 1.54)**
Drug-related[c]	63 (11.6)	55 (10.2)	1.14 (0.81, 1.61)
Grade 3–4	156 (28.8)	123 (22.8)	**1.26 (1.03, 1.55)**
Number of patients with TEAEs leading to treatment discontinuation[d]	58 (10.7)	53 (9.8)	1.09 (0.77, 1.55)
Drug-related[c]	32 (5.9)	24 (4.4)	1.33 (0.79, 2.22)
Number of patients with TEAE leading to death	21 (3.9)	16 (3.0)	1.31 (0.69, 2.48)
Drug-related[c]	6 (1.1)	6 (1.1)	1.00 (0.32, 3.07)
All deaths within 30 days of last dose	18 (3.3)	11 (2.0)	1.63 (0.78, 3.42)
Other	11 (2.0)	6 (1.1)	1.83 (0.68, 4.90)
Death due to prostate cancer	6 (1.1)	4 (0.7)	1.49 (0.42, 5.27)
Unknown	1 (0.2)	1 (0.2)	1.00 (0.06, 15.89)

Source: MS, Table 25, pages 75-76[1]

[a]Calculated using http://www.hutchon.net/confidrr.htm. Values in bold indicate statistically significant results.

[b]Does not include grade 5 events.

[c]AEs reported as unlikely, possibly, or related to AA, prednisone/prednisolone, or placebo are classified as drug-related AEs.

[d]Discontinuation of study medication includes discontinuation of AA, prednisone/prednisolone, or placebo.

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; PP, placebo plus prednisone/prednisolone; RR, relative risk; SAE, serious adverse event; TEAE, treatment-emergent adverse event.

The incidences of individual AEs that occurred in ≥5% of patients in either group are summarised in Table 4.17. The most frequently reported AEs were fatigue (39.7% AAP vs. 34.6% PP), back pain (33.2% vs. 33.1%), arthralgia (29.3% vs. 24.4%), nausea (24.0% vs. 23.0%), peripheral oedema (26.0% vs. 20.9%), constipation (23.6% vs. 20.4%), diarrhoea (23.4% vs. 18.1%) and hot flush (22.7% vs. 18.3%). The majority of these events were grade 1–2.

Table 4.17: Incidence of all AEs occurring in ≥5% of patients in either group (COU-AA-302 study third interim analysis, 55% data cut-off)

System organ/class/AEs, n (%)	**AAP (N=542)**	**PP (N=540)**	**RR (95% CI)[a]**
Total	538 (99.3)	524 (97.0)	**1.02 (1.01, 1.04)**
Musculoskeletal and connective tissue disorders	406 (74.9)	409 (75.7)	0.99 (0.92, 1.06)
Back pain	180 (33.2)	179 (33.1)	1.00 (0.85, 1.19)
Arthralgia	159 (29.3)	132 (24.4)	1.20 (0.98, 1.46)
Bone pain	113 (20.8)	103 (19.1)	1.09 (0.86, 1.39)
Pain in extremity	93 (17.2)	87 (16.1)	1.07 (0.82, 1.39)
Musculoskeletal pain	88 (16.2)	81 (15.0)	1.08 (0.82, 1.43)
Muscle spasms	77 (14.2)	111 (20.6)	**0.69 (0.53, 0.90)**
Groin pain	38 (7.0)	22 (4.1)	**1.72 (1.03, 2.87)**
Myalgia	35 (6.5)	32 (5.9)	1.09 (0.68, 1.73)
Muscular weakness	32 (5.9)	42 (7.8)	0.76 (0.49, 1.18)
Flank pain	27 (5.0)	17 (3.1)	1.58 (0.87, 2.87)
General disorders and administration site conditions	361 (66.6)	314 (58.1)	**1.15 (1.04, 1.26)**
Fatigue	215 (39.7)	187 (34.6)	1.15 (0.98, 1.34)
Oedema peripheral	141 (26.0)	113 (20.9)	**1.24 (1.00, 1.54)**
Pyrexia	52 (9.6)	34 (6.3)	**1.52 (1.01, 2.31)**
Asthenia	47 (8.7)	47 (8.7)	1.00 (0.68, 1.47)
Gastrointestinal disorders	356 (65.7)	329 (60.9)	1.08 (0.98, 1.18)
Nausea	130 (24.0)	124 (23.0)	1.04 (0.84, 1.30)
Constipation	128 (23.6)	110 (20.4)	1.16 (0.93, 1.45)
Diarrhoea	127 (23.4)	98 (18.1)	**1.29 (1.02, 1.63)**
Vomiting	77 (14.2)	61 (11.3)	1.26 (0.92, 1.72)
Dyspepsia	60 (11.1)	29 (5.4)	**2.06 (1.34, 3.16)**
Abdominal pain	42 (7.7)	51 (9.4)	0.82 (0.56, 1.21)
Infections and infestations	305 (56.3)	212 (39.3)	**1.43 (1.26, 1.63)**
Upper respiratory tract infection	72 (13.3)	43 (8.0)	**1.67 (1.17, 2.39)**
Nasopharyngitis	60 (11.1)	45 (8.3)	1.33 (0.92, 1.92)
Urinary tract infection	51 (9.4)	41 (7.6)	1.24 (0.84, 1.84)
Bronchitis	30 (5.5)	16 (3.0)	**1.87 (1.03, 3.39)**
Sinusitis	28 (5.2)	6 (1.1)	**4.65 (1.94, 11.14)**
Influenza	27 (5.0)	18 (3.3)	1.49 (0.83, 2.68)
Vascular disorders	253 (46.7)	183 (33.9)	**1.38 (1.19, 1.60)**
Hot flush	123 (22.7)	99 (18.3)	1.24 (0.98, 1.57)
Hypertension	118 (21.8)	73 (13.5)	**1.61 (1.23, 2.10)**
Nervous system disorders	240 (44.3)	210 (38.9)	1.14 (0.99, 1.31)
Headache	74 (13.7)	66 (12.2)	1.12 (0.82, 1.52)
Dizziness	72 (13.3)	74 (13.7)	0.97 (0.72, 1.31)
Metabolism and nutrition disorders	235 (43.4)	222 (41.1)	1.05 (0.92, 1.21)
Hypokalaemia	93 (17.2)	69 (12.8)	**1.34 (1.01, 1.79)**
Hyperglycaemia	47 (8.7)	43 (8.0)	1.09 (0.73, 1.62)
Anorexia	40 (7.4)	38 (7.0)	1.05 (0.68, 1.61)
Decreased appetite	33 (6.1)	32 (5.9)	1.03 (0.64, 1.65)

System organ/class/AEs, n (%)	AAP (N=542)	PP (N=540)	RR (95% CI)[a]
Respiratory, thoracic and mediastinal disorders	213 (39.3)	181 (33.5)	**1.17 (1.00, 1.37)**
Cough	98 (18.1)	74 (13.7)	**1.32 (1.00, 1.74)**
Dyspnoea	68 (12.5)	55 (10.2)	1.23 (0.88, 1.72)
Injury, poisoning and procedural complications	195 (36.0)	151 (28.0)	**1.29 (1.08, 1.53)**
Contusion	74 (13.7)	50 (9.3)	**1.47 (1.05, 2.07)**
Fall	35 (6.5)	21 (3.9)	1.66 (0.98, 2.81)
Renal and urinary disorders	194 (35.8)	159 (29.4)	**1.22 (1.02, 1.44)**
Haematuria	60 (11.1)	31 (5.7)	**1.93 (1.27, 2.93)**
Pollakiuria	55 (10.1)	55 (10.2)	1.00 (0.70, 1.42)
Nocturia	34 (6.3)	28 (5.2)	1.21 (0.74, 1.97)
Urinary incontinence	34 (6.3)	25 (4.6)	1.35 (0.82, 2.24)
Investigations	190 (35.1)	145 (26.9)	**1.31 (1.09, 1.56)**
Alanine aminotransferase increased	65 (12.0)	27 (5.0)	**2.40 (1.56, 3.70)**
Aspartate aminotransferase increased	60 (11.1)	26 (4.8)	**2.30 (1.47, 3.59)**
Weight decreased	38 (7.0)	26 (4.8)	1.46 (0.90, 2.36)
Weight increased	28 (5.2)	39 (7.2)	0.72 (0.45, 1.15)
Skin and subcutaneous tissue disorders	180 (33.2)	146 (27.0)	**1.23 (1.02, 1.47)**
Rash	46 (8.5)	21 (3.9)	**2.18 (1.32, 3.61)**
Psychiatric disorders	144 (26.6)	123 (22.8)	1.17 (0.95, 1.44)
Insomnia	79 (14.6)	62 (11.5)	1.27 (0.93, 1.73)
Anxiety	28 (5.2)	23 (4.3)	1.21 (0.71, 2.08)
Depression	27 (5.0)	19 (3.5)	1.42 (0.80, 2.52)
Cardiac disorders	94 (17.3)	80 (14.8)	1.17 (0.89, 1.54)
Atrial fibrillation	26 (4.8)	27 (5.0)	0.96 (0.57, 1.62)
Blood and lymphatic system disorders	80 (14.8)	68 (12.6)	1.17 (0.87, 1.58)
Anaemia	61 (11.3)	52 (9.6)	1.17 (0.82, 1.66)

Source: MS, Table 26, pages 76-77[1]

[a]Calculated using http://www.hutchon.net/confidrr.htm. Values in bold indicate statistically significant results.

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; AE, adverse event; CI, confidence interval; PP, placebo plus prednisone; RR, relative risk.

Grade 3 or 4 AEs reported in ≥1% of patients in either group are summarised in Table 4.18. The most frequently reported grade 3 or 4 AEs were hypertension, back pain, and increased alanine aminotransferase (ALT). AAP resulted in significantly more grade 3 or 4 increased ALT, increased aspartate aminotransferase, and dyspnoea; but less hydronephrosis.

Table 4.18: Incidence of grade 3 or 4 AEs occurring in ≥1% of patients across randomised groups (COU-AA-302 study third interim analysis, 55% data cut-off)

System organ/class/AEs, n (%)	**AAP (N=542)**	**PP (N=540)**	**RR[a] (95% CI)**
Total	267 (49.3)	235 (43.5)	**1.13 (1.00, 1.29)**
Metabolism and nutrition disorders	61 (11.3)	42 (7.8)	**1.45 (1.00, 2.10)**
Hyperglycaemia	14 (2.6)	11 (2.0)	1.27 (0.58, 2.77)
Hypokalaemia	14 (2.6)	10 (1.9)	1.39 (0.63, 3.11)
Hyponatraemia	9 (1.7)	8 (1.5)	1.12 (0.44, 2.88)
Dehydration	7 (1.3)	3 (0.6)	2.32 (0.60, 8.94)
Anorexia	6 (1.1)	1 (0.2)	5.98 (0.72, 49.49)
Hypophosphataemia	5 (0.9)	7 (1.3)	0.71 (0.23, 2.23)
Investigations	51 (9.4)	27 (5.0)	**1.88 (1.20, 2.95)**
Alanine aminotransferase increased	30 (5.5)	4 (0.7)	**7.47 (2.65, 21.07)**
Aspartate aminotransferase increased	17 (3.1)	5 (0.9)	**3.39 (1.26, 9.12)**
Blood alkaline phosphatase increased	6 (1.1)	5 (0.9)	1.20 (0.37, 3.89)
Musculoskeletal and connective tissue disorders	48 (8.9)	60 (11.1)	0.80 (0.56, 1.14)
Back pain	15 (2.8)	21 (3.9)	0.71 (0.37, 1.37)
Arthralgia	10 (1.8)	10 (1.9)	1.00 (0.42, 2.37)
Bone pain	7 (1.3)	11 (2.0)	0.63 (0.25, 1.62)
Musculoskeletal pain	7 (1.3)	6 (1.1)	1.16 (0.39, 3.44)
Muscular weakness	3 (0.6)	6 (1.1)	0.50 (0.13, 1.98)
Infections and infestations	43 (7.9)	35 (6.5)	1.22 (0.80, 1.88)
Urinary tract infection	9 (1.7)	3 (0.6)	2.99 (0.81, 10.98)
Pneumonia	7 (1.3)	4 (0.7)	1.74 (0.51, 5.92)
Nervous system disorders	36 (6.6)	23 (4.3)	1.56 (0.94, 2.60)
Syncope	9 (1.7)	6 (1.1)	1.49 (0.54, 4.17)
Vascular disorders	36 (6.6)	31 (5.7)	1.16 (0.73, 1.84)
Hypertension	23 (4.2)	17 (3.1)	1.35 (0.73, 2.49)
Deep vein thrombosis	8 (1.5)	6 (1.1)	1.33 (0.46, 3.80)
Renal and urinary disorders	31 (5.7)	28 (5.2)	1.10 (0.67, 1.81)
Haematuria	7 (1.3)	4 (0.7)	1.74 (0.51, 5.92)
Urinary retention	7 (1.3)	3 (0.6)	2.32 (0.60, 8.94)
Hydronephrosis	1 (0.2)	8 (1.5)	**0.12 (0.02, 0.99)**
General disorders and administration site conditions	30 (5.5)	33 (6.1)	0.91 (0.56, 1.46)
Fatigue	13 (2.4)	10 (1.9)	1.30 (0.57, 2.93)
General physical health deterioration	6 (1.1)	2 (0.4)	2.99 (0.61, 14.74)
Asthenia	1 (0.2)	7 (1.3)	0.14 (0.02, 1.15)
Cardiac disorders	28 (5.2)	13 (2.4)	**2.15 (1.12, 4.10)**
Atrial fibrillation	9 (1.7)	5 (0.9)	1.79 (0.60, 5.32)
Gastrointestinal disorders	28 (5.2)	25 (4.6)	1.12 (0.66, 1.89)
Diarrhoea	6 (1.1)	5 (0.9)	1.20 (0.37, 3.89)
Abdominal pain	3 (0.6)	9 (1.7)	0.33 (0.09, 1.22)
Respiratory, thoracic and mediastinal disorders	28 (5.2)	23 (4.3)	1.21 (0.71, 2.08)
Dyspnoea	14 (2.6)	5 (0.9)	**2.79 (1.01, 7.69)**

System organ/class/AEs, n (%)	AAP (N=542)	PP (N=540)	RR[a] (95% CI)
Pulmonary embolism	11 (2.0)	15 (2.8)	0.73 (0.34, 1.58)
Blood and lymphatic system disorders	21 (3.9)	19 (3.5)	1.10 (0.60, 2.02)
Anaemia	13 (2.4)	10 (1.9)	1.30 (0.57, 2.93)
Neoplasms benign, malignant and unspecified (including cysts and polyps)	20 (3.7)	22 (4.1)	0.91 (0.50, 1.64)
Cancer pain	5 (0.9)	9 (1.7)	0.55 (0.19, 1.64)

Source: MS, Table 27, page 78[1]

[a]Calculated using http://www.hutchon.net/confidrr.htm. Values in bold indicate statistically significant results.

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; AE, adverse event; CI, confidence interval; PP, placebo plus prednisone; RR, relative risk.

AEs of special interest (as described in the EPAR for AA) include events related to mineralocorticoid excess (hypertension, hypokalaemia, and fluid retention), cardiac disorders, and hepatotoxicity. These AEs were reported in a higher proportion of patients in the AAP group than in the PP group (68.6% vs. 51.3%; see Table 4.19 below).

The most commonly reported subcategories of AEs of special interest were fluid retention (29% AAP vs. 24% PP), hypertension (22% vs. 14%), cardiac disorders (21% vs. 18%), hepatotoxicity (19% vs. 11%), and hypokalaemia (17% vs. 13%). The incidence of grade 3 and above events was <10% for all individual special events (see Table below).

Table 4.19: Incidence of AEs of special interest (>5% in either arm) (COU-AA-302 study third interim analysis, 55% data cut-off)

n (%)	AAP (N=542)					PP (N=540)				
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
AE of special interest	151 (27.9)	109 (20.1)	97 (17.9)	11 (2.0)	3 (0.6)	130 (24.1)	81 (15.0)	57 (10.6)	6 (1.1)	3 (0.6)
Fluid retention/oedema	116 (21.4)	37 (6.8)	5 (0.9)	0	0	85 (15.7)	36 (6.7)	8 (1.5)	1 (0.2)	0
Peripheral oedema	107 (19.7)	31 (5.7)	2 (0.4)	0	0	75 (13.9)	33 (6.1)	5 (0.9)	0	0
Hypertension	46 (8.5)	49 (9.0)	23 (4.2)	0	0	27 (5.0)	29 (5.4)	17 (3.1)	0	0
Cardiac disorders	46 (8.5)	28 (5.2)	32 (5.9)	4 (0.7)	3 (0.6)	38 (7.0)	35 (6.5)	16 (3.0)	3 (0.6)	3 (0.6)
Arrythmias	38 (7.0)	19 (3.5)	18 (3.3)	3 (0.6)	1 (0.2)	33 (6.1)	24 (4.4)	11 (2.0)	0	2 (0.4)
Hepatotoxicity	25 (4.6)	34 (6.3)	38 (7.0)	5 (0.9)	0	30 (5.6)	16 (3.0)	13 (2.4)	2 (0.4)	0
ALT increased	14 (2.6)	21 (3.9)	27 (5.0)	3 (0.6)	0	18 (3.3)	5 (0.9)	3 (0.6)	1 (0.2)	0
AST increased	24 (4.4)	19 (3.5)	17 (3.1)	0	0	13 (2.4)	8 (1.5)	5 (0.9)	0	0
Hypokalaemia	74 (13.7)	5 (0.9)	12 (2.2)	2 (0.4)	0	59 (10.9)	0	10 (1.9)	0	0

Source: MS, Table 27, page 78[1]

Abbreviations: AAP, abiraterone acetate plus prednisone/prednisolone; AE, adverse event; ALT, alanine aminotransferase; AST, aspartate aminotransferase; PP, placebo plus prednisone/prednisolone.

ERG comment: The Committee for Medicinal Products for Human Use (CHMP) assessment report for abiraterone, which was based on data from the second interim analysis, included the following comments regarding adverse events:

- Overall the adverse event profile is generally consistent with previous observations from study COU-AA-301. A review performed by the marketing authorization holder (MAH) to identify any clinically meaningful new imbalances between studies identified dyspepsia, increased aspartate aminotransferase (AST), rash and haematuria as new adverse drug reactions (ADRs).
- In study 302, a higher rate of deaths within 30 days of last dose was observed in the abiraterone group and was considered of concern. Of them, death causes categorised as 'other' were higher in the abiraterone group [10 (1.8%) vs 4 (0.7%)]. Focusing on AEs leading to death, infections seem to be the most relevant AE; 5 patients in abiraterone arm (0.9%) vs none in the control.
- Although hepatotoxicity was associated with abiraterone treatment during the assessment of the marketing authorisation application, it appeared to be more notable in the population of study 302. Considering that Treatment-Emergent Adverse Events (TEAEs) AST/ALT increases as well as Grade 3 and 4 AST/alanine aminotransferase (ALT) increases were higher in study 302 and that liver metastasis was an exclusion criterion in study 302 (not in 301), differences were considered of importance. However, the higher hepatotoxicity could be related to the fact that patients were treated with abiraterone for a longer time in study 302. The actual mechanism of abiraterone hepatotoxicity is unknown and further information has been included in the SmPC to manage this risk.
- A higher rate of cardiac disorder events was noted and a warning regarding use of abiraterone in patients with a history of cardiovascular disease is included in section 4.4 of the SmPC.
- Events of renal toxicity were reported for higher proportions of abiraterone treated patients versus placebo treated patients.
- The small proportions of non-white subjects enrolled preclude any meaningful comparisons of AE profiles analyzed by race.
- Rates of AEs were generally higher in the subgroups with more advanced age, higher baseline ECOG performance status grade, baseline hemoglobin concentration <12.5 g/dL and baseline LDH >1xULN.
- Conclusion: Treatment with abiraterone was tolerable for the majority of subjects and the safety profile was consistent with previous experience (except for four new ADR identified - dyspepsia, AST increased, rash and haematuria). Adverse events were generally manageable and no major safety concerns have been raised by this application.

The ERG found no evidence to contradict this conclusion from the Committee for Medicinal Products for Human Use (CHMP).

4.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

No indirect comparison or mixed treatment comparison evidence synthesis was included in the submission.

4.4 Critique of the indirect comparison and/or multiple treatment comparison

No indirect comparison or mixed treatment comparison evidence synthesis was included in the submission.

4.5 Additional work on clinical effectiveness undertaken by the ERG

As discussed in Chapter 3.3, there is some literature suggesting that docetaxel might be less effective following abiraterone. Assuming that most patients will end up using docetaxel, which also seems to be implied by the phrase "**not yet** clinically indicated", an important question in this appraisal is whether abiraterone followed by docataxel is more effective than watch-full waiting (BSC) followed by docetaxel. In the COU-AA-302 trial, 239 out of 546 (43.8%) of AAP patients and 304 out of 542 (56.1%) of PP patients received docetaxel as subsequent therapy (MS, Table 21, page 68). The results for this specific group of patients are not presented in the MS; therefore, it is important that these data are presented to the Appraisal Committee.

The manufacturer provided these data four days before the deadline of this ERG report, and the results are summarised below.

The manufacturer stressed that "the data requested is a *post-hoc* analysis of patients in the COU-AA-302 trial who subsequently receive docetaxel. This group of patients progressed more quickly, and therefore moved onto docetaxel treatment earlier than the other patients in the trial. This *post-hoc* analysis violates the principles of randomisation, and in effect, selects for the patients with the worst prognosis (ie those that progress quickly and move onto chemotherapy), which renders any interpretation of these results meaningless." The manufacturer also requested that these remain commercial-in-confidence.

The data for OS in the subgroup of patients treated with subsequent docetaxel are presented in Table 4.20 and Figure 4.4. The data for rPFS in the subgroup of patients treated with subsequent docetaxel are presented in Table 4.21 and Figure 4.5.

Median time on docetaxel treatment was [redacted]_months in the AA group, and [redacted] in the placebo group.

Table 4.20: OS of patients treated with either AA or placebo (ITT) (interim analysis 3); subgroup of patients with subsequent docetaxel treatment

	AA (n=239)	Placebo (n=304)
Number of patients with OS[a] event, n (%)	[redacted]	[redacted]
Time to event (months), median (95% CI)	[redacted]	[redacted]
HR (95% CI)	[redacted]	
p-value	[redacted]	

Source: Additional information provided by Janssen on 13 March 2014

[a]) The ERG assumed this should be 'OS event', instead of 'PFS event'.

[b]) The ERG assumed [redacted] was the correct HR as reported in the graph.

Figure 4.4: Overall survival, AA versus placebo, stratified by subsequent docetaxel treatment

Table 4.21: rPFS in patients treated with either AA or placebo (ITT) (interim analysis 3); subgroup of patients with subsequent docetaxel treatment

	AA (n=239)	Placebo (n=304)
Number of patients with PFS event, n (%)	[redacted]	[redacted]
Time to event (months), median (95% CI)	[redacted]	[redacted]
HR (95% CI)	[redacted]	
p-value	[redacted]	

Source: Additional information provided by Janssen on 13 March 2014

Figure 4.5: rPFS, AA versus placebo, stratified by subsequent docetaxel treatment

ERG Comment: Although the OS results at the second and third interim analysis did not meet the pre-specified statistical significance levels, they showed a considerable advantage for abiraterone. [redacted]

Median time on docetaxel treatment was only [redacted] months in the AA group, and [redacted] in the placebo group. It could be argued that this period is too short for a proper analyses for this question; it could also be argued that, if the effectiveness of docetaxel is reduced after treatment with abiraterone, a longer period on docetaxel would be even less favourable for abiraterone followed by docetaxel compared to watch-full waiting followed by docetaxel.

Finally, regarding the cost-effectiveness consequences for the subgroup of patients with subsequent docetaxel treatment, the incremental effectiveness most likely decreases (for AAP versus BSC) compared to the base case analysis. However, the incremental costs might also decrease (e.g. due to a potential smaller difference in time on first line treatment and smaller

difference in life years for AAP versus BSC). Therefore, the cost-effectiveness consequences for these data are unclear.

4.6 Conclusions of the clinical effectiveness section

One RCT (the COU-AA-302 trial) is included for the comparison of abiraterone acetate in combination with prednisolone versus best supportive care.

In the COU-AA-302 trial, a total of 1,088 patients were recruited and randomised to abiraterone acetate plus prednisone/prednisolone (n=546) or placebo plus prednisone (n=542). 1,082 patients received at least one dose of the allocated intervention and constituted the safety population. Patients continued treatment with AAP or PP until disease progression (determined according to radiographic and clinical measures). The median treatment duration was 13.8 months (15 cycles initiated) in the AAP arm and 8.3 months (nine cycles initiated) in the PP arm.

Results presented in the MS are based on the results from the second (data cut-off 20/12/11) and third (data cut-off 22/5/12) interim analyses of the COU-AA-302 study, which were conducted after approximately 40% and 55% of the total OS events had occurred.

Neither the second nor third interim analysis overall survival results met the pre-specified statistical significance levels (HR at third interim analysis: 0.79 (95% CI: 0.66, 0.96). Median overall survival was 35.3 months (95% CI: 31.2, 35.3) in the AAP group and 30.1 months (95% CI: 27.3, 34.1) in the PP group. Janssen was unable to provide mean survival for both groups or mean survival gain, despite explicit questions in the clarification letter.

Treatment with AAP resulted in a 48% relative reduction in the risk of radiographic progression compared with PP (absolute risk reduction 11.5%), and increased PFS by 8.2 months. Significant differences in favour of the AAP group were observed for objective response rate (complete or partial response according to modified RECIST criteria), PSA response and duration of response. HRQL was assessed in the COU-AA-302 study via the FACT-P instrument. However, no results are report by treatment arm for baseline, follow-up or change scores. Time to progression in average pain intensity and worst pain intensity showed no significant differences between treatment arms. All other pain-related outcomes favoured AAP over PP.

Adverse events were significantly more often reported in the AAP arm when compared with the PP arm for treatment-emergent adverse event (TEAEs), Drug-related grade 3–4 TEAEs, treatment-emergent serious adverse event (SAEs) and Grade 3–4 treatment-emergent SAEs. The most frequently reported AEs were fatigue (39.7% AAP vs. 34.6% PP), back pain (33.2% vs. 33.1%), arthralgia (29.3% vs. 24.4%), nausea (24.0% vs. 23.0%), peripheral oedema (26.0% vs. 20.9%), constipation (23.6% vs. 20.4%), diarrhoea (23.4% vs. 18.1%) and hot flush (22.7% vs. 18.3%). AAP resulted in significantly more grade 3 or 4 increased ALT, increased aspartate aminotransferase, and dyspnoea; but less hydronephrosis.

[redacted]

5. COST-EFFECTIVENESS

5.1 ERG comment on manufacturer's review of cost-effectiveness evidence

5.1.1 Objective of cost effectiveness review

The main objective of the cost-effectiveness review was to identify all primary studies in mCRPC that involved an economic evaluation, a burden of illness study, or an evaluation linked to a health technology assessment (HTA) submission. The search strategy for relevant economic studies was detailed in appendix 10, section 10.10 of the MS. The databases searched were PubMed/MEDLINE, EMBASE, Medline (R) In-Process, Cochrane Library, EconLit, NHS EED. All searches were conducted on 2 September 2013 and covered the period between 1 January 2000 and 2 September 2013. A description of the search strategies is given in appendix 10 (section 10.10.4) of the MS. In addition, conference proceedings from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) International and ISPOR European conferences (2006–2012) as well as national HTA websites (UK, US, Australia, Canada, Global) were searched.

ERG Comment: The ERG believes that the objective of the cost effectiveness review was appropriate. The quality of the search strategy is discussed in Chapter 4.1.1 of this report.

5.1.2 Inclusion/exclusion criteria used in the study selection

The inclusion criteria were reported in Table 30 of the MS (Section 7.1.1). Those that did not meet the eligibility criteria were excluded.

Table 5.1: Inclusion criteria for the economic evaluations systematic review

Inclusion criteria for the economic evaluations systematic review			
Study designs	Budget impact analyses, Resource use studies, Cost/economic burden of illness studies, Cost analyses, Cost-minimisation analyses, Cost-effective analyses, Cost-utility analyses, Cost-benefit analyses, Clinical trial-based analyses		These types of economic studies were seen as potential sources to input into the development of the economic model relevant to the decision problem
Population	Adults (≥18 years), Males, Any race, Confirmed diagnosis of advanced or metastatic prostate cancer. All lines of therapy		Only studies relating to advanced or metastatic prostate cancer were relevant to the decision problem
	5-FU, abarelix, abiraterone acetate aflibercept, aminoglutethimide, AS1404, atrasentan bevacizumab, bicalutamide, CAB, cabazitaxel, carboplatin, cyproterone acetate, dasatinib, degarelix	IMC-A12, ipilimumab, JM216, ketoconazole, leuprorelin, mdv3100, megestrol, mitoxantrone, nilutamide, paclitaxel, pamidronate, patupilone, prednisone, samarium, sipuleucel T,	All commonly used interventions in mCRPC were included in the search. All of these interventions in combination or as monotherapy

Inclusion criteria for the economic evaluations systematic review		
	dexamethasone, diethystilbesterol, docetaxel, doxorubicin, dutasteride, e7389/eribulin mesylate, estramustine, etoposide, finasteride, flutamide, goserelin, hydrocortisone, strontium, sunitinib, TAK-700, triptorelin, vinblastine, vinorelbine, zd4054, zoledronic acid	
Language	English language only	The restriction would not limit results substantially due to data availability in English language

5-FU, 5-fluorouracil; CAB, combined androgen blockade; mCRPC, metastatic castration-resistant prostate cancer.
Source: MS, Table 30 pg. 82.[1]

ERG comment: The ERG considers that the inclusion and exclusion criteria used in the study selection are appropriate.

5.1.3 Included/excluded studies in the cost-effectiveness review

The systematic literature review identified 45 economic evaluations (displayed in Table 140, Appendix 15 of the MS[1]) and 12 additional economic evaluations associated with HTA appraisals (displayed in Table 141, Appendix 15 of the MS[1]).

ERG comment: The ERG considers the studies displayed in MS Table 140 (appendix 15) and the evaluations included as part of HTA appraisals described in MS Table 141 (appendix 15) are the more relevant ones from the search performed. However, none of these studies investigated AAP for the treatment of adult men who are asymptomatic or mildly symptomatic after failure of ADT and in whom chemotherapy is not yet clinically indicated and therefore their findings are less relevant to the current submission.

5.1.4 Conclusions of the cost-effectiveness review

No specific conclusions from the economic review were provided in the submission.

ERG comment: None of the studies in the economic review investigated AAP for the treatment of adult men who are asymptomatic or mildly symptomatic after failure of ADT and in whom chemotherapy is not yet clinically indicated. For this reason the manufacturer has provided a de novo analysis. The ERG agrees with this approach.

5.2 Summary and critique of manufacturer's submitted economic evaluation by the ERG

Table 5.2: Summary of the manufacturer's economic evaluation (with signposts to MS[1])

	Approach	Source / Justification	Signpost (location in MS[1])
Model	A discrete event simulation (DES) approach tracking patients and their experiences at the individual level. The model adopted a treatment pathway structure, simulating multiple courses of therapy from the start of AAP/BSC (PP) until death.		7.2.2 (pg. 86)
States and events	The pathway structure allowed for the tracking of AAP subsequent treatments after AAP - if AAP was received as first line treatment - as well as previous and subsequent treatments before and after AAP – if AAP was received after docetaxel. Treatment phases included: • Time before receiving an active treatment where patient received treatments that palliate symptoms of disease (consisting in BSC pre-/post-docetaxel) • Time on active treatment (consisting in AAP, BSC [PP] and docetaxel) • End-of-life phase where patients were near death and did not receive additional active treatments that may impact survival, but are managed for their pain or other symptoms (consisting of BSC before death).		7.2.2 (pg. 86)
Comparators	The comparator was best supportive care (BSC; this may include corticosteroids, radiotherapy, radiopharmaceuticals, analgesics, bisphosphonates, further hormonal therapies, and mitoxantrone with or without corticosteroids).	Docetaxel was not considered as a comparator by the manufacturer based on the reasoning that: *"Whilst both AAP and docetaxel are indicated for the treatment of mCRPC for adult men following ADT failure, AAP is indicated for the treatment of those in whom chemotherapy is not yet clinically indicated, the asymptomatic or mildly symptomatic patient … a patient*	2.7 (pg. 32)

	Approach	Source / Justification	Signpost (location in MS[1])
		population for whom chemotherapy is not yet clinically indicated"	
Treatment pathways	Asymptomatic or mildly symptomatic mCRPC post-ADT patients enter the model and are assigned to one of the two treatment pathways (i.e. monitoring and use of BSC or AAP in post-ADT patients). Patients for whom pre-docetaxel treatment is discontinued or in whom disease is progressed are monitored in a BSC (pre-docetaxel) phase prior to commencing docetaxel treatment. They start docetaxel only if Karnofsky PS ≥60%, (ECOG PS <2). After docetaxel treatment is completed, patients are again monitored for disease progression (active treatment is given if benefits outweigh the risks). Post-docetaxel treatment may include enzalutamide, cabazitaxel and BSC. In this submission, the predicted use of post-docetaxel treatment is restricted to BSC (post-docetaxel) and based on the observations from the COU-AA- 302 trial. Furthermore it was assumed that if patients received AAP prior to docetaxel they would not be eligible for AAP retreatment post-docetaxel (whereas BSC patients were allowed to receive AAP post-docetaxel).		7.2.2 (pg. 86 -91)
Treatment effectiveness	Overall survival and progression free survival state.	COU-AA-302 trial	
Adverse events	The most frequently reported AEs were fatigue (39.7% AAP vs. 34.6% PP), back pain (33.2% vs. 33.1%), arthralgia (29.3% vs. 24.4%), nausea (24.0% vs. 23.0%), peripheral oedema (26.0% vs. 20.9%), constipation (23.6% vs. 20.4%), diarrhoea (23.4% vs. 18.1%) and hot flush (22.7% vs. 18.3%). The majority of these events were grade 1–2. The most frequently reported grade 3 or 4 AEs were hypertension, back pain, and increased alanine aminotransferase (ALT). AAP resulted in significantly more grade 3 or 4 increased ALT, increased aspartate aminotransferase, and dyspnoea; but less hydronephrosis.	COU-AA-302 trial	6.9.2 (pg. 73 – 78)

	Approach	Source / Justification	Signpost (location in MS[1])
	The most commonly reported subcategories of AEs of special interest were fluid retention (29% AAP vs. 24% PP), hypertension (22% vs. 14%), cardiac disorders (21% vs. 18%), hepatotoxicity (19% vs. 11%), and hypokalaemia (17% vs. 13%).		
Health related QoL	Despite the mapping algorithms to convert FACT-P values collected in COU-AA-302 study into EQ-5D, the approach taken for the model was to gather UK-specific EQ-5D data over a broader range of the treatment phases than captured in the COU-AA-302 study. EQ-5D-5L utility values for the four mCRPC stratified treatment phases are shown in Table 40 of the MS. A utility increment of [redacted] (derived from the FACT-P mapped COU-AA-302 trial data) was applied to those patients who received AAP post-ADT. A utility value of 0.50 (s.e. 0.08) was assumed for patients in the BSC group before death following Sandblom et al [59]. **Error! Reference source not found.**	COU-AA-302 study	7.4.3. (pg. 108 -113; 126-127)
Resource utilisation and costs	The costs of scheduled, disease related patient follow-up consisted of clinical visits, imaging diagnostic tests and clinical laboratory tests to monitor the status of disease in patients with mCRPC. They were estimated through a cost study involved 53 oncologists and 50 urologists from Medeconnect. COU-AA-301 and COU-AA-302 trials recorded resources consumed as a result of unplanned events (such as AEs) while on treatment. Patients on BSC (pre and ppost docetaxel) were assumed to have the same unplanned event-related MRU as BSC (PP) treated patients from the COU-AA-301 study. For docetaxel, it is assumed that baseline resource utilisation was equivalent to the placebo arm of the COU-AA-301. Under the terms of the PAS, the cost of AA used in the model is £[redacted] per month ([redacted] discount). Terminal cost was £3,598 per patient	COU-AA-301 and COU-AA-302 studies. Unit costs for the regularly scheduled follow-up procedures were determined using the NHS Reference Costs, 2012–2013.[60]	7.5.4 (pg. 133 – 142)

	Approach	Source / Justification	Signpost (location in MS[1])
	and was applied as a one-off cost in the economic model.		
Discount rates	3.5 % for utilities and costs	According to NICE reference case	7.2.6 (pg.95 – 96)
Sub groups	Two subgroup analyses were undertaken, of patients with baseline BPI question #3 of 0 or 1 (accounting for 67% of the COU-AA-302 trial population) and of patients with ECOG PS = 0 (76% of COU-AA-302).		7.9 (pg. 163 – 166)
Sensitivity analysis	Numerous scenario analyses were run investigating the effect of changing the base case assumptions. Deterministic analyses were undertaken using 10 trial replications of the DES model to ensure stability. Probabilistic sensitivity analysis was also undertaken (200 iterations).		7.7.7 – 7.7.10` (pg. 153 - 159

5.2.1 NICE reference case checklist (TABLE ONLY)

Attribute	Reference case and TA Methods guidance	Does the *de novo* economic evaluation match the reference case
Comparator(s)	Therapies routinely used in the NHS, including technologies regarded as current best practice	Yes (partially). The main deviation from the scope is that docetaxel is not included as a comparator in the MS. As the indication is men with mCRPC ***in whom chemotherapy is not yet clinically indicated,*** it seems reasonable that docetaxel is not considered as a comparator.
Patient group	As per NICE scope	Yes. The patient population described in the final scope is: "*Adults with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated*". [24] This is in line with the patient population included in the MS and in the main trial for this submission, the COU-AA-302 study
Perspective costs	NHS and Personal Social Services (PSS)	Yes.
Perspective benefits	All health effects on individuals	Yes.
Form of economic evaluation	Cost-effectiveness analysis	Yes
Time horizon	Sufficient to capture differences in costs and outcomes	Yes. All patients enter the model when treatment begins and exit the model at death or once the maximum age of 100

Attribute	Reference case and TA Methods guidance	Does the *de novo* economic evaluation match the reference case
		years is reached, whichever comes first.
Synthesis of evidence on outcomes	Systematic review	The main comparison with BSC relies upon the pivotal head to head COU-AA-302. Section 2.7 of the MS argued that "*… as BSC can include corticosteroids, the placebo arm of the COU-AA-302 study can be considered an appropriate comparison population, as patients in the placebo arm all received corticosteroids as part of supportive care"*
Outcome measure	Quality adjusted life years (QALYs)	Yes
Health states for QALY	Described using a standardised and validated instrument	Yes (partially). Health-related quality of life (HRQoL) data were collected during the pivotal trial (COU-AA-302 study) through FACT-P questionnaires and mapped to EQ-5D values. However, the MS base-case uses UK-specific EQ-5D data gathered through a UK utility study. The UK utility study did not allow for direct comparisons of the utility impact of AAP with that of BSC but collected utility values for different treatment phases. Mapped values (derived from the COU-AA-302 trial data) were used in the MS base case for the utility increment of [redacted] applied to those patients who received AAP post-ADT. The average utility of 0.50 (s.e. 0.08) was assumed for patients in the BSC group before death.
Benefit valuation	Time-trade off or standard gamble	Yes. The EQ-5D utility from the UK utility study scores were calculated using the UK time trade-off (TTO) value set. FACT-P values from COU-AA-302 study were mapped to EQ-5D values.
Source of preference data for valuation of changes in HRQL	Representative sample of the public	Yes.
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes.
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Probabilistic modelling	Probabilistic modelling	Yes. However the probabilistic sensitivity analysis was undertaken using only 200 Monte Carlo simulations.
Sensitivity analysis		Yes. A range of sensitivity analyses were performed.

5.2.2 Model structure

The model consisted of a discrete event simulation (DES) evaluating the cost-effectiveness of AAP in comparison to BSC (PP) in adult men with mCRPC who were asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy was not yet clinically indicated. The approach allowed for tracking patients and their experiences at the individual level. The comparator, BSC (PP), was a proxy for an active monitoring strategy based on the PP group in the COU-AA-302 trial.

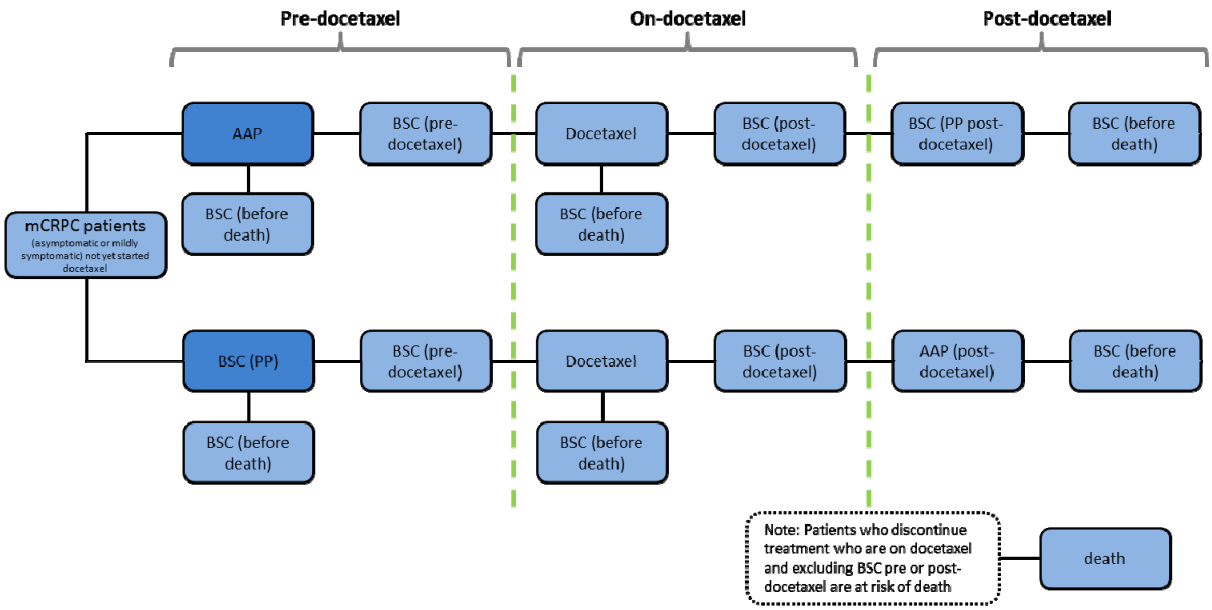
Patients diagnosed with asymptomatic or mildly symptomatic mCRPC post-ADT entered the model and were assigned to one of the two treatment pathways (BSC and AAP) as in Figure 5.1 below. Patients for whom pre-docetaxel treatment was discontinued or in whom disease was progressed were monitored in a BSC (pre-docetaxel) phase prior to commencing docetaxel treatment. They started docetaxel only if ECOG PS <2, (corresponding to Karnofsky PS ≥60%).

After docetaxel treatment was completed, patients were again monitored for disease progression and other active treatment (AAP) was given if benefits outweigh the risks. In this submission, the predicted use of post-docetaxel treatment was restricted to BSC and based on the observations from the COU-AA- 302 trial. Furthermore it was assumed that if patients received AAP prior to docetaxel they would not be eligible for AAP retreatment post-docetaxel, whereas BSC patients were allowed to receive AAP post-docetaxel. Throughout the model, patients may receive additional treatments, but these are not expected to impact survival (i.e. no evidence exists demonstrating a statistically significant impact on survival) and are not explicitly considered in the model.

In the model structure different types of BSC can be distinguished:

- **BSC (PP)**, active monitoring comparator treatment arm where patients are not receiving active treatments such as AAP before docetaxel that impact survival
- **BSC (pre-docetaxel / post-docetaxel)**, time before receiving an active treatment that has shown to impact overall survival where patients are still receiving treatments that palliate symptoms (e.g., corticosteroids) of disease. This phase aimed to capture the slow progression of the disease during which time patients received treatments to alleviate worsening symptoms
- BSC before death involves palliative care, until death. This consists of the “end of life” phase where patients are near death and will not receive additional active treatments that may impact survival, but instead are managed for their pain or other symptoms

Figure 5.1: Model pathway

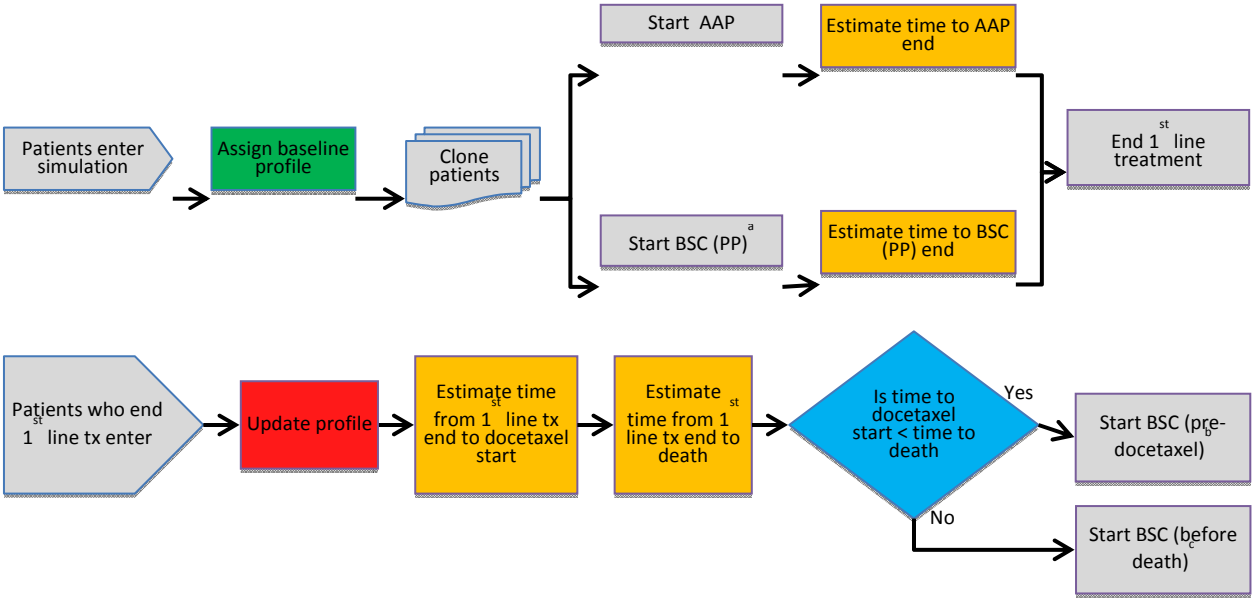


Note: AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone.
The size of the boxes does not reflect active treatment/BSC duration.
Patients only receive licensed products or those with positive reimbursement appraisal.

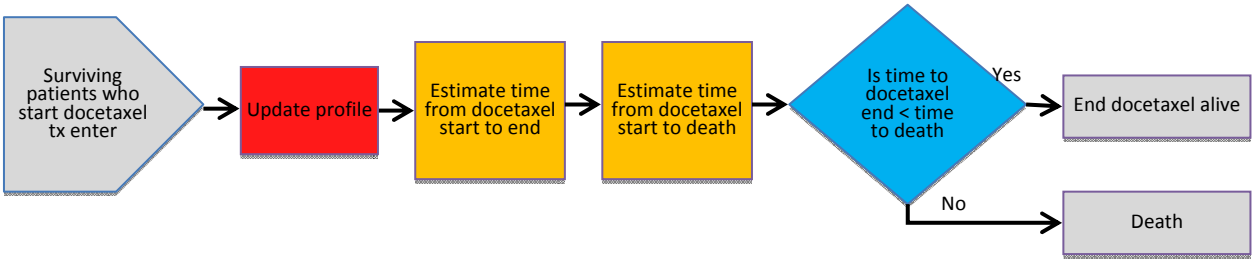
After the ERG request for clarification, the manufacturer has provided a modified schematic overview of the schematic patient flow through the simulations (Figure 11 in the MS[1]) in order to include a description of the various phases of BSC. The updated scheme of the patient flow is given in Figure 5.2 (below). First, in the pre-docetaxel phase, patients with mCRPC (asymptomatic or mildly symptomatic) were assigned a baseline profile, and the profiles were duplicated to allow patients with the same characteristics to progress through treatment arms (AAP and BSC [PP]). When patients finished AAP or BSC (PP) treatment, their profiles (e.g. age, ECOG score) were updated, and times to docetaxel start and to death were estimated. If the estimated time to docetaxel start was less than time to death, a patient received docetaxel and their estimated time to docetaxel determined the duration of BSC (before docetaxel). Otherwise, patients moved to BSC (before death) until death.

Figure 5.2: Patient flow through model simulation

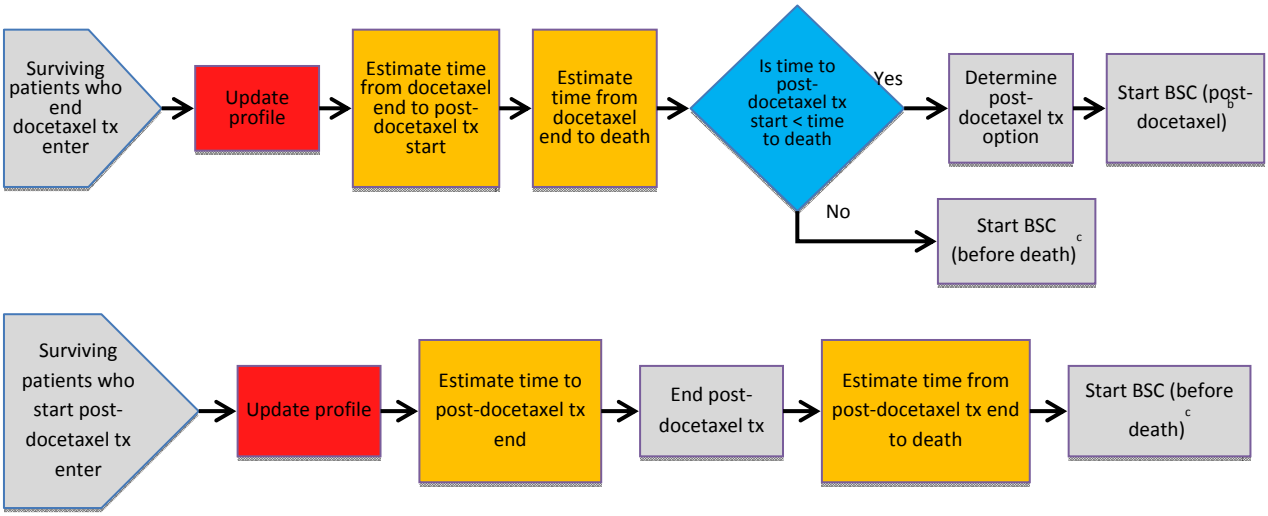
A



B



C



AAP, abiraterone acetate + prednisolone; BSC, best supportive care; PP, placebo + prednisolone; tx, treatment.
[a]BSC (PP), active monitoring comparator treatment arm where patients are not receiving active treatments such as AAP before docetaxel that impact survival
[b]BSC (pre-docetaxel / post-docetaxel), time before receiving an active treatment that has shown to impact overall survival where patients are still receiving treatments that palliate symptoms (e.g., corticosteroids) of disease
[c]BSC (before death), best supportive care "end of life" phase where patients are near death and will not receive additional active treatments that may impact survival, but instead are managed for their pain or other symptoms
Source: Response to request for clarification from the ERG; pg. 34 [40]

Patient profiles were updated at the start of docetaxel treatment (Figure 5.2 B). Times from docetaxel start to treatment end and to death were estimated. Patients could die during treatment with docetaxel (i.e. time to docetaxel end was longer than time to death) or fail docetaxel treatment alive. The latter group's profiles were updated at docetaxel discontinuation, and so was their time to death (Figure 5.2 C). Time to post-docetaxel treatment was estimated and compared with time to death. Patients whose time to death was less than time to next treatment moved on to BSC (before death) until death, while the others received post-docetaxel treatment after a period of time receiving BSC (post-docetaxel). These patients had their profiles updated upon starting the next treatment, and their times to treatment discontinuation and from treatment discontinuation to death were estimated. The simulation terminated when the patient died or the analysis time horizon was reached (up to age 100).

ERG comment: While the manufacturer considers the model presented as "a simple discrete event simulation (DES) model" (pg. 18, MS[1]) the ERG does not believe that a DES model, simulating individual patients using 17 prediction equations would have been the simplest and most transparent approach. The ERG believes instead that it would have been possible to use a more transparent model, for instance a Markov model consisting of health states according to the treatment phases included in the current model and a sufficiently short cycle time. This model would also allow reflection of the clinical pathways in the UK and to produce results for subgroups with varying baseline characteristics.[3] Also, the ERG is not convinced by the manufacturer's arguments that a patient level simulation would be necessary for the decision problem defined during the scope. It should be noted that acknowledging patient heterogeneity does not necessarily require patient level simulation.[3] Transparency is a key aspect of modelling and in this specific case a more transparent model would be more convenient for an external reviewer to assess face validity and internal validity of the model.

In addition, in its request for clarification, the ERG has raised the concern that the structure of the model does not fully reflect the situation of adult men with mCRPC who were asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy was not yet clinically indicated. This was simply because the DES model does not include the possibility of dying during AAP and BSC (PP). As some patients (N=5) died before AAP or BSC (PP) treatment end, it is appropriate that probabilities to die are included during all phases in the model. The same applies for death during post-docetaxel treatment.

5.2.3 Population

The patient population described in the NICE final scope[24] is: "Adults with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated".[24] The target population for treatment with AAP is in adult men who are asymptomatic or mildly symptomatic mCRPC after failure of ADT in whom chemotherapy is not yet clinically indicated. The manufacturer argues that: "*The population included in the model, as an approximation of the UK target population, is closely based on the COU-AA-302 trial*" (pg.

85, MS[1]). The characteristics of the patients in the COU-AA-302 trial are given in Table 31 of the MS (pg 86, MS[1]). All patients had previously been treated with hormonal therapy, 54.0% had received prior radiotherapy, and 46.0% prior surgical therapy. The use of anti-androgen therapies, including bicalutamide, nilutamide, flutamide, and azole drugs, was prohibited during the trial period.

ERG comment: The ERG notes that the population in Table 31 of the MS (pg 86, MS[1]) is not the same as the population used in the model. The population analysable in the model consisted of a total of 902 patients (459 for AAP and 443 for PP) from the COU-AA-302 trial. A total of 186 patients were excluded due to missing baseline data that were used as predictors. The manufacturer did not provide the characteristics of this subpopulation.

5.2.4 Interventions and comparators

The main intervention is abiraterone acetate plus prednisone/prednisolone (AAP). The NICE scope identified two possible comparators: best supportive care (BSC; this may include corticosteroids, radiotherapy, radiopharmaceuticals, analgesics, bisphosphonates, further hormonal therapies, and mitoxantrone with or without corticosteroids) or docetaxel. The manufacturer argues that docetaxel is not a valid comparator as: *"Whilst both AAP and docetaxel are indicated for the treatment of mCRPC for adult men following ADT failure, AAP is indicated for the treatment of those in whom chemotherapy is not yet clinically indicated, the asymptomatic or mildly symptomatic patient."* (pg. 32, MS[1]).

ERG comment: The main deviation from the NICE scope is that docetaxel is not included as a comparator in the MS[1]. However, as the indication is men with mCRPC in whom chemotherapy is not yet clinically indicated, it seems reasonable that docetaxel is not considered as a comparator.

5.2.5 Perspective, time horizon and discounting

The model takes the perspective of the NHS in England and Wales. Costs were considered from the NHS and Personal Social Services (pg. 95, MS[1]). The manufacturer argues that a societal perspective was not relevant and indirect costs were not included in the evaluation since productivity loss is negligible due to the disease severity and older nature of the patient population (pg. 96, MS[1]). A lifetime horizon is adopted with a discount rate of 3.5%. In the model the time horizon can be varied to 5 and 10 years.

ERG comment: The perspective and discount rates are in line with the NICE reference case. The lifetime horizon is considered appropriate.

5.2.6 Treatment effectiveness and extrapolation

The section on treatment effectiveness and extrapolation is subdivided in a section considering prediction equations for estimating baseline characteristics, time to treatment discontinuation, time to treatment start and time to death and a section considering adverse events and skeletal related events.

Analysable dataset for prediction equations

The discrete event simulation (DES) model consists of a total of 17 prediction equations (listed in below). To estimate these prediction equations, 902 patients were used (83% of the ITT population which consisted of 1088 patients) as 186 patients were excluded because baseline data were missing. The manufacturer stated, based on Figures 5.3 and 5.4 below, *"we believe that missing baseline information is missing completely at random and therefore does not bias the results"*.

Figure 5.3: Time to AAP/BSC (PP) discontinuation for the ITT population versus the analysable dataset

Source: MS[1] Figure 37 (Appendix 16)
Abbreviations: AA, abiraterone acetate plus prednisolone; Placebo, placebo plus prednisolone; str, strata; TRTP, treatment arm.

Figure 5.4: Overall survival for the ITT population versus the analysable dataset

Source: MS Figure 38 (Appendix 16)

Abbreviations: AA, abiraterone acetate plus prednisolone; Placebo, placebo plus prednisolone; str, strata; TRTP, treatment arm.

In the analysable dataset, 85% of the patients followed the treatments pathway specified in the DES mode. The remaining 15% of the patients, who did not follow the specified treatment pathways (e.g. placebo patients started AAP before docetaxel), were censored at the time they deviated from the pathway. For the ITT population, 91% (993/1088) of the patients followed the specified treatment pathways (MS Figures 39 and 40). After AAP pathways were excluded because of AA retreatment, cabazitaxel treatment before receiving docetaxel and docetaxel retreatment. In addition, after BSC (PP) pathways were excluded because of AA treatment or cabazitaxel before receiving docetaxel, sequential AA and cabazitaxel treatment (post-docetaxel) and docetaxel retreatment (Table 5.3).

Table 5.3: treatment pathways excluded from the model

Patients deviating from specified treatment pathway (ITT population; N=1,088)		
Excluded pathways for AAP ITT population (N=546)	**N**	**% (of AAP arm)**
AAP → AA retreatment	10	1.8%
AAP → cabazitaxel	4	0.7%
AAP → docetaxel → AA retreatment	19	3.5%
AAP → docetaxel → cabazitaxel → Docetaxel retreatment	1	0.2%
AAP → docetaxel → cabazitaxel → AA retreatment	7	1.3%
AAP → docetaxel → docetaxel retreatment	8	1.5%
Excluded pathways for BSC (PP) ITT population (N=542)	**N**	**% (of BSC arm)**
BSC (PP) → AAP/AA	16	3.0%
BSC (PP) → cabazitaxel	3	0.6%
BSC (PP) → docetaxel → cabazitaxel → AA	10	1.8%
BSC (PP) → docetaxel → AA → cabazitaxel	7	1.3%
BSC (PP) → docetaxel → AA → docetaxel retreatment	5	0.9%
BSC (PP) → docetaxel → docetaxel retreatment	5	0.9%

Source: MS[1] Figures 39 and 40 (Appendix 16)

AA, abiraterone acetate, AAP abiraterone acetate plus prednisone/prednisolone; BSC, best supportive care.

ERG comment: It was concluded by the manufacturer that time to AAP/BSC (PP) discontinuation (TTD) in the selected group (analysable dataset of 902 patients with information on baseline characteristics) is similar to the ITT group. However, based on Figure 5.3 is seems that TTD for BSC is similar for both groups, but overestimation of TTD for AAP in the selected group. In the clarification letter (question B4a) the manufacturer states that this difference is not statistically significant (Logrank p=0.3117; note that the titles for Figures 5-8 in the manufacturer's clarification letter are incorrect). However, regardless of its significance, the effectiveness of AAP, in terms of TTD, is overestimated as stated by the manufacturer: *"analysed patients who were treated with AAP had slightly longer TTD"* (clarification letter question B4a). This suggests that by excluding patients with missing baseline information, bias in favour of AAP has been introduced in the model for both TTD and OS (as OS is dependent on TTD in the discrete event simulation (DES) model). This is also illustrated by Table 5.4, comparing model and clinical trial results, indicating an increased median TTD and an increased median OS by 1.05 and 0.27 months respectively favouring AAP. Despite potential higher costs due to a longer treatment duration, using the

selected population instead of the ITT cannot be regarded as conservative (as mentioned by the manufacturer in response to clarification question B4a). Therefore, the ERG believes it would be more plausible to use the ITT population (despite missing baseline information) instead of the selected subset of patients.

Table 5.4: Summary of model results compared with clinical data (based on the ITT population)

Outcome	Model result			Clinical trial result			Difference
	AAP	BSC	Δ	AAP	BSC	Δ	Δ - Δ[a]
Treatment duration							
≥ 6 months	83.6%	65.5%	18.1%	81.0%	59.6%	21.4%	-3.3%
≥ 12 months	60.3%	34.8%	25.5%	55.7%	34.1%	21.6%	3.9%
≥ 18 months	42.3%	20.4%	21.9%	38.4%	21.7%	16.7%	5.2%
≥ 24 months	30.6%	12.6%	18.0%	28.4%	14.1%	14.3%	3.7%
Median (months)	15.10	8.53	6.57	13.80	8.28	5.52	1.05
OS							
Median (months)	35.11	29.68	5.43	35.29	30.13	5.16	0.27

Source: MS[1] Table 67

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; OS, overall survival.

[a] A positive result indicates an overestimation of the model (compared to the clinical trial result) and a negative result indicates an underestimation.

Excluding treatment pathways that are not relevant to the UK setting seems a reasonable approach to reflect the UK context. However, it is unclear why patients in the BSC (PP) arm were censored after sequential treatment with AAP and cabazitaxel (while treatment with either AAP and cabazitaxel was allowed as well as cabazitaxel for the patients receiving AAP in the 1st line; MS[1] Figures 39 and 40). Despite requested (clarification question 8), no additional analyses were provided by manufacturer allowing for sequential use of AAP and cabazitaxel. Although, in the response to the clarification letter, the manufacturer states: *"To correct our description in the submission report, patients who received another active treatment after post-docetaxel active treatment are not censored. Their time spent with these treatments was included in the time from post-docetaxel treatment end to death"*, the ERG is not convinced that these patients are not censored as this is not consistent with MS[1] Figure 40 (Appendix 16) which indicates these patients were censored.

<u>*Estimation and selection of prediction equations*</u>

To incorporate the "pre-docetaxel", "on-docetaxel" and "post-docetaxel" treatment phases in the DES model, multiple prediction equations were included for, time to treatment discontinuation, time to treatment start and time to death. In on instances stratified prediction equations for BSC and AAP were used. Additionally, prediction equations were estimated to update patient disease status throughout the DES model. The following prediction equations were included:

Time to treatment discontinuation:

1. Time to AAP/BSC (PP) treatment discontinuation

2. Time from docetaxel start to docetaxel discontinuation
3. Time from third-line treatment start to post-docetaxel treatment discontinuation.

Time to treatment start:

4. Time from AAP/BSC (PP) treatment discontinuation to docetaxel start
5. Time from docetaxel discontinuation to post-docetaxel treatment start

Time to death:

6. Time from AAP discontinuation to death before docetaxel start
7. Time from BSC (PP) discontinuation to death before docetaxel start
8. Time from docetaxel start to death before docetaxel discontinuation
9. Time from docetaxel discontinuation to death before third-line treatment starts
10. Time from post-docetaxel treatment (third-line treatment) discontinuation to death

Patient disease status:

11. ECOG status at AAP/BSC (PP) discontinuation
12. ECOG status at start of docetaxel
13. ECOG status at docetaxel discontinuation
14. ECOG status at start of post-docetaxel treatment (third-line treatment)
15. PSA progression at discontinuation of AAP/BSC (PP)
16. Radiographic progression at discontinuation of AAP/BSC (PP)
17. Opiate use at discontinuation of AAP/BSC (PP).

The following steps were undertaken to estimate and select these prediction equations:

1. Decide whether treatment (AAP versus BSC (PP)) should be included as a predictor in the model or stratified models should be estimate for each predictor. This was done based on visual inspection of the Kaplan-Meier curves and cumulative hazard functions.
2. Decide on the parametric distribution of the model (Exponential, Weibull, log-normal and log-logistic) based on the ITT population (N=1088). This was done based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as presented in Table 142 of the MS. To validate the choice of parametric distribution, the observed data was plotted against the predicted distributions by treatment group.
3. Decide on the covariates to be included based on a subset of patients without missing baseline data (N=902). This was done based on a stepwise approach using varying candidate covariates per prediction equation.
    a. Firstly, all candidate predictors were individually tested in an univariate analysis.
    b. Secondly, all significant predictors (p-value ≤ 0.10) were combined in a multivariate model, which was then manually trimmed to exclude predictors that become non-significant (p-value > 0.10). The manufacturer justifies the use of a p-value threshold of 0.10 in Appendix 16 of the MS: *"The p value of 0.10 is often used when the purpose of the analyses is for prediction and not for causal inference.*[61, 62] *... Given that these prediction equations were used in*

the discrete event simulation model (i.e. for each patient we estimated a series of time-to-event predictions), it was very important not to miss any potential predictors that could influence the time-to-event predictions, so it was preferable to be more inclusive".

4. Decide on the interactions to be included. The following interactions were tested for the time to AAP/BSC (PP) discontinuation and included if significant (p-value ≤ 0.20): treatment and age, treatment and baseline BPI, treatment and baseline ECOG, and treatment and bone metastasis present at entry.
5. Finally the prediction equations were compared to the observed data.

It should be noted that, for two prediction equation covariates were included despite these were not statistically significant (α = 0.10). This concerned the inclusion of Baseline BPI (and its interaction term with AAP treatment) for "Time to AAP/BSC (PP) End" and the inclusion of the treatment arm covariate (AAP versus BSC (PP)) for "Time from post-docetaxel tx end to death". The inclusion of Baseline BPI regarded the manufacturer as justified since it *"is important to evaluate AAP impact on this subgroup"* and for the inclusion of treatment arm this was since *"We believe that the insignificance of the treatment arm as predictor is due to small sample size and the treatment should be included as a predictor in this equation to avoid potential bias".*

ERG comment: Based on the initial manufacturer submission it was difficult to validate the results of the above mentioned steps. However, based on responses by the manufacturer to clarification questions B4 and B5, we were able to validate most of the steps for the estimation and selection of the prediction equations. Nevertheless, we were not able to validate the prediction equations for "patient disease status", i.e. compare observed and predicted ECOG status, PSA, radiographic progression at discontinuation and opiate use.

It is noticeable that for the time on AAP/BSC treatment, no prediction equation for time to death was estimated and no probability of dying was included in the DES model. Despite this concerning a low number of patients (N=5), it would be more plausible to incorporate the possibility of dying during all phases in the DES model (e.g. by calculating the rate if a prediction equation is not feasible given the low number of events). These deaths were incorporated post AAP/BSC treatment as clarified by the manufacturer (response to question B4c of the clarification letter): *"instead these deaths were considered in the derivation of survival post AAP/BSC (PP) treatment discontinuation (with time of death post AAP/BSC (PP) discontinuation set to 1 day for these 5 patients)"* Consistently, no probability of dying is assumed during time on post-docetaxel treatment (third-line treatment). However, these deaths were incorporated post 3rd line active treatment as clarified by the manufacturer (response to question B4c of the clarification letter): *"The deaths observed while on post-docetaxel treatment* [N=13] *were considered by setting the time of death to day 1 after discontinuation of post-docetaxel active treatment."* Neglecting these probabilities of dying implicitly extends survival and assumes that patients during AAP/BSC and post-docetaxel treatment have a lower probability of dying than the general population. This cannot be regarded as a conservative assumption. Despite requested (clarification question B4c), no

additional analyses were provided by the manufacturer to incorporate the probabilities of dying. The following is a list of suggestions as to how the modelling could have been done better:

1. Decide whether treatment should be included as a predictor in the model or stratified models should be estimate for each predictor

The prediction equation for "Time from AAP/BSC (PP) end to death" was, unlike all other prediction equations, estimated separately by arm (see MS[1] Table 148). After receiving the observed KM and cumulative hazard functions (response to clarification question B5) which were used by the manufacturer to decide whether treatment should be included as a predictor in the model or stratified models should be estimated for each predictor, it was unclear to the ERG why the "Time from AAP/BSC (PP) end to death" prediction equation stands out and a stratified model was used for this prediction equation. Therefore, to be consistent with other prediction equations, a non-stratified equation would be preferred by the ERG.

2. Decide on the parametric distribution of the model

AIC and BIC were not consistent in two cases for "time from docetaxel end to post-docetaxel treatment start" and "time from post-docetaxel treatment end to death". In both cases a Weibull distribution was selected based on the following justification (response to clarification question B5c): *"For time from docetaxel end to post-docetaxel treatment start, Weibull has the best AIC while exponential has the best BIC. As exponential is a special form of Weibull, we selected Weilbull. For time from post-docetaxel treatment end to death, exponential has the best BIC and lognormal has the best AIC. Weibull is used in this case given the small difference in AIC and BIC across the parametric functions and is clinically more plausible since the long tail of a lognormal distribution is often criticised for offering unrealistic survival benefit."* The justification provided by the manufacturer seems reasonable for selecting a Weibull distribution in case of inconsistencies between AIC and BIC.

It could be argued that the order of the steps to estimate and select the prediction equations should be changed; i.e. selecting the parametric distribution of the model (step 2) after the covariates and interaction terms were selected (steps 3 and 4). Thus the AIC and BIC statistics could be calculated based on prediction models, which include covariates (and interaction terms). However, the manufacturer does *"not expect that including predictors to parametric functions will change the decision on the parametric function selections"* (response to clarification question B5c). It is however difficult for the ERG to validate/speculate whether this is true.

3. Decide on the covariates to be included

Based on the manufacturer's response to clarification question B5b, it becomes clear that candidate covariates vary between prediction equations without providing a rationale for selecting the candidate covariates. The ERG would prefer a consistent approach or a clear rationale for using different candidate covariates. The impact of varying candidate covariates on the cost-effectiveness estimates is unclear.

Adding covariates or interaction terms even when they were not statistically significant for "Time to AAP/BSC (PP) End" and "Time from post-docetaxel tx end to death" could not be regarded as conservative as this increased the effectiveness of AAP versus BSC in both instances (see MS[1] Tables 143 and 151). In particular, adding the covariate for treatment (AAP versus BSC (PP)) to the prediction equation for "Time from post-docetaxel tx end to death" could be questioned, although there seems to be a (non-significant) difference between AAP and BSC (PP) as illustrated in Figure 5.5. Nevertheless, the subjective justification provided by the manufacturer might also apply to other prediction equations; particularly "Time from docetaxel start to docetaxel discontinuation" as illustrated in Figure 5.6. In contrast with the non-significant covariates added by the Manufacture, adding a treatment covariate in the latter case (for "Time from docetaxel start to docetaxel discontinuation") would be in favour of BSC. For these reasons, the ERG believes that the addition of non-significant covariates biased the results in favour of AAP and hence, a consistent approach should preferably be adopted without exceptions. Despite being requested (clarification question B4b), no additional analyses were provided by the manufacturer, which excluded all non-significant covariates.

Figure 5.5: Time from post-docetaxel treatment discontinuation to death

Source: response to clarification question B7

Figure 5.6: Time from docetaxel start to docetaxel discontinuation

Source: response to clarification question B7

4. Decide on the interactions to be included
Based on the manufacturer's response to clarification question B5b, it seems that interaction terms were only tested for "Time to AAP/BSC (PP) treatment discontinuation" without providing a rationale for selecting the candidate interaction terms. The ERG would prefer a consistent approach or a clear rationale for this.

5. Comparing the prediction equations to the observed data
Based on visual inspection, the estimated prediction equations seem to be consistent with the observed data (KM-curves) that were used to estimate these equations, i.e. the analysable dataset (response to clarification question B7). The main issues raised above (using analysable dataset instead of the ITT population and inconsistencies in the use of stratified models, covariates and interaction terms) are however still maintained.

In addition, to correct for post-docetaxel cabazitaxel use in the COU-AA-302 study (which is not recommended in UK clinical practice), a negative treatment effect was applied by the manufacturer (MS Table 35). It is however unclear how this negative treatment effect was calculated.

In conclusion, except for the above mentioned inconsistencies the estimated prediction equations seem plausible based on the analysable dataset, but most likely overestimate the effectiveness of AAP when considering the ITT dataset (as discussed above). This is also illustrated in Figure 5.7, where the difference between AAP and BSC (PP) predicted based on the model (dotted lines) seems an overestimation of the difference observed in approximately the first 1.5 year in the COU-AA-302 trial (ITT population). Therefore, prediction equations

for "Time to treatment discontinuation", "Time to treatment start" and "Time to death" were based on the ITT population and consistently including treatment as the only covariate for the ERG base case (thus all prediction equations were non-stratified). Moreover, prediction equations for patient disease status were used to incorporate the ECOG restriction for docetaxel treatment (i.e. only patients with ECOG < 2 are allowed to switch to docetaxel after first line treatment discontinuation).

Figure 5.7: OS comparison of model prediction and COU-AA-302 trial results

Source: response to clarification question B5C

Adverse events

The selection of AEs to be included was driven by the docetaxel AE profile, since the pre-docetaxel treatments were said to have a relatively tolerable safety profile. Furthermore, in the pre- and post docetaxel phases, no incremental AE effects were considered, because the COU-AA-302 trial indicated that the AE rates were similar pre-docetaxel (grade 3 or 4 AEs occurred in 49.3% of AAP patients and in 43.5% of PP patients), and cost impacts were assumed to be implicitly captured by the event-related MRU. For docetaxel treatment, the incremental grade 3 or 4 AE rates for docetaxel vs AAP (post-docetaxel) were used, based on published literature and considered for cost implications only.

ERG comment: The post-docetaxel AEs indeed seemed to be very similar between AAP and BSC, as apparent from COU-AA-301, with an HR of [redacted] for the total number of AEs and also for only grade 3 or 4 AEs (TA259), although it is difficult to say what the impact of the different setting (e.g. patients already treated with AA in 1st line) in this submission

would be. However, the ERG thinks that at least for the pre-docetaxel phase a differential AE rate would have been indicated, since the HR for grade 3 or 4 AEs was 1.13 (1.00, 1.29) (Table 25 of the MS[1]) and even higher (and statistically significant) for only drug related AEs. Moreover, using AE rates for docetaxel versus post-docetaxel AAP as a proxy for the incremental AEs in the docetaxel phase may not be justified, since it is not comparable to the setting in which docetaxel is administered in the present submission. The ERG was not able to investigate the impact of these assumptions on the ICER.

Skeletal related events

In addition to treatment related AEs, TA259 also took into account skeletal related events (SREs), which are a consequence of metastatic bone disease, and were defined as pathological fracture, spinal cord compression, palliative radiation to bone, or surgery to bone. Skeletal related events were not mentioned in this submission.

ERG comment: Since SREs are typically a consequence of metastatic bone disease, they will probably not be relevant in the pre-docetaxel phase. In the post-docetaxel phase they will however be present, and it would have been useful to see the impact of including SREs in the model. Moreover, TA259 demonstrated that, [redacted]). Therefore, it can be questioned whether not including SREs in the present submission can considered a conservative approach.

5.2.7 Health related quality of life

The HRQL of patients in COU-AA-302 was measured using the FACT-P (FACT – general and prostate cancer subscale) questionnaire on the first day of cycles 1, 3, 5, and 7, and then on the first day of every third cycle and at treatment discontinuation. The MS[1] reports that algorithms were developed to map FACT-P values onto EQ-5D utilities using an ordinary least squares (OLS) regression algorithm. The results of this mapping exercise are displayed in Table 5.5

Table 5.5: Results of the FACT-P to EQ-5D Mapping Study

State	Base case (all patients)	
	Utility value	SE
Post-ADT baseline (asymptomatic and mildly symptomatic)	[redacted]	[redacted]
AAP on-treatment utility increment	[redacted]	[redacted]
BSC (pre-docetaxel)	[redacted]	[redacted]
Docetaxel	[redacted]	[redacted]
Docetaxel utility decrement	[redacted]	[redacted]
BSC (post-docetaxel)	[redacted]	[redacted]
BSC (PP post-docetaxel)	[redacted]	[redacted]
AAP (post-docetaxel)	[redacted]	[redacted]
BSC (before death)	[redacted]	[redacted]

AAP, abiraterone acetate plus prednisolone; AE, adverse event; BPI, brief pain inventory; BSC, best supportive care; SE, standard error.

However, the MS[1] base-case did not use the EQ-5D mapped data, except for the utility increment of AAP over BSC (PP) in the pre-docetaxel phase. EQ-5D utility values mapped from the COU-AA-302 study (section 7.4.4) were also used in a scenario analysis (scenario 4; sections 7.6.1 and 7.7.9). For all other base-case utility values, UK-specific EQ-5D data gathered through an UK mCRPC Utility Study were used. In the UK mCRPC Utility Study the utility and HRQL data, along with background and medical history data were collected via an online survey among participants recruited through a specialist patient recruitment agency and patient advocacy groups. The patient sample consisted of 163 UK men with mCRPC who had previously taken anti-androgen tablets for >1 month but had since stopped (unless they had undergone orchiectomy/orchidectomy). The study allowed distinguishing between the four treatment phases as below:

1. Mildly or asymptomatic after failure of ADT; chemotherapy not yet clinically indicated
2. Symptomatic after failure of ADT; chemotherapy clinically indicated but not started
3. After failure of ADT; receiving chemotherapy
4. After failure of ADT; post-chemotherapy.

Table 5.6 gives a summary of the utility values for each treatment phase which were used in the manufacturer base case.

Table 5.6: Summary of the utility values associated with each model phase

Utilities	Value	SE (distribution)	Source
Post-ADT baseline	[redacted]	[redacted]	UK mCRPC patient utility study
AAP on-treatment utility increment	[redacted]	[redacted]	COU-AA-302 mapping study
BSC (pre-docetaxel)	[redacted]	[redacted]	UK mCRPC patient utility study
Docetaxel	[redacted]	[redacted]	UK mCRPC patient utility study
Post-docetaxel	[redacted]	[redacted]	UK mCRPC patient utility study
BSC (before death)	0.500	0.08	Sandblom et al.[59]

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone.

Post-ADT baseline AAP or BSC - In section 7.4.9 of the MS[1] the manufacturer argues that the mCRPC utility study captured the baseline utility value for those patients who are asymptomatic/mildly symptomatic having failed ADT. The post ADT baseline utility [redacted] (s.e. [redacted]) was assumed to be the same for both post-ADT AAP and post-ADT BSC and was varied in the scenario analysis. Differences between BPI subgroups were accounted for in the DES model (baseline utility was [redacted] for the BPI #3 0–1 subgroup; and [redacted] for the ECOG PS subgroup).

AAP on-treatment utility increment- The manufacturer has applied an on-treatment utility increment while on AAP of [redacted] (s.e. [redacted]) (which was derived from the mapping study of the COU-AA-302 trial data), since the UK mCRPC Utility Study was non-comparative and consequently did not allow for direct comparison of the utility impact of AAP with that of BSC.

BSC (pre-docetaxel) – A utility of [redacted] (s.e. [redacted]) was applied for symptomatic post-ADT patients in the BSC (pre-docetaxel) arm. The manufacturer argues that this value accurately

reflected the impact of treatment and AEs. Furthermore, since AAP has a safety profile similar to PP in the COU-AA-302 trial, application of additional AE utility decrement over BSC (PP) was not seen necessary by the manufacturer.

On docetaxel - A utility of [redacted] (s.e. [redacted]) was applied for patients currently receiving chemotherapy. The manufacturer argues that this value accurately reflected the impact of treatment and AEs.

BSC (post- docetaxel) or AAP (post-docetaxel)- The manufacturer assumes a post-docetaxel utility value of [redacted] (s.e. [redacted]) for both BSC (post- docetaxel) and AAP (post-docetaxel). Unlike in TA259, there was no post-docetaxel on-treatment utility increment for AAP applied here. The manufacturer argues that to apply this post-docetaxel utility increment of 0.046 (derived from COU-AA-301 trial data) would be double-counting since the majority of patients in the UK mCRPC Utility Study were assumed to already have been receiving AAP in this setting and so the on-treatment utility gain was already captured directly. Again, because AEs were considered similar between AAP and BSC, the application of a differential AE utility was not considered necessary by the manufacturer.

BSC (before death) - A utility value of 0.50 (s.e. 0.08) was assumed by the manufacturer for patients in the BSC group before death. Another study by Sandblom et al [59] estimates such utility ranging from 0.58 (patients with 8–12 months of remaining survival) to 0.46 (patients with <4 months survival remaining). The manufacturer claims that the chosen level of utility was assumed given that mCRPC patients were likely to spend their last 6–8 months of life in the progressed health state.

ERG comment: The ERG agrees that using the EQ-5D utility value from the UK mCRPC Utility Study is the preferred approach given the uncertainty of the mapped utility values based on FACT-P responses. However, the manufacturer did not report on the treatment received by patients enrolled in the mCRPC utility study. In its request for clarification the ERG required the manufacturer to provide a description of the treatments received by the patients in the mCRPC utility study post ADT.

In the Response to the Request for Clarification from the ERG[40] the manufacturer has argued that *"... no medication use was recorded in the UK mCRPC Utility Study, apart from chemotherapy treatment as part of the screening/eligibility criteria"*.

AAP on-treatment utility increment – In the request for clarification letter the ERG has raised the concern that the concept of the on-treatment utility increment in favour of abiraterone may seem questionable as it is clear that AAP leads to significantly more adverse events (both overall and grade 3/4) than BSC. The MS[1] states that the utility increment assigned to patients in the AAP was to capture "patient benefits experienced on AAP compared with BSC (PP) with respect to pain and fatigue" (MS pg. 94[1]) The ERG believes that the on-treatment utility increment for AAP it is questionable and that instead separate utility decrements for each separate AEs should be incorporated in the model.

In the Response to the Request for Clarification from the ERG[40] the manufacturer has argued that "*Sample sizes for individual clinically significant AEs i.e. grade 3/4 are too small for meaningful analysis and that … regression analysis of the COU-AA-302 mapped utility data exploring the baseline utility over time, (have provided) the AE decrement effect of –0.06487*" (Response to the Request for Clarification from the ERG pg.13[40]). However the manufacturer has provided the ERG with separate scenario analyses for: (i) removing the on treatment utility increment (ICER £50,120), (ii) incorporating utility decrements for each separate AE (ICER £47,415), and (iii) incorporating per-event costs for AEs for the pre- and post-docetaxel treatment phases (ICER £46,686). The ICER for the scenario where all the above changes were incorporated resulted in £50,880. So, of these three factors, the on-treatment utility increment is the largest driver of change in the ICER.

<u>*BSC (post- docetaxel) or AAP (post-docetaxel)*</u> - in the request for clarification letter the ERG has raised the concern that in a previous STA (TA259) the utility value at baseline at which time patients had received docetaxel was based on FACT-P data from COU-AA-301 and was 0.78 i.e. higher than that the utility value of [redacted] in the current submission for the same treatment phase (post-docetaxel) and which was derived from the UK mCRPC Utility Study.

In the Response to the Request for Clarification from the ERG[40] the manufacturer argues that the differences could be because of: (i) "*TA259 utility value was FACT-P mapped to EQ-5D from patients in an RCT whereas the UK mCRPC Utility Study used EQ-5D directly*"; (ii) "*COU-AA-301 study population may have differed from the UK mCRPC Utility Study population*" due to different settings for which they were designed (post-chemotherapy and pre-chemotherapy); and (iii) the FACT-P scores in COU-AA-301 were all elicited at the start of the study (i.e.., patients were chosen to meet the inclusion criteria prior to the commencement of study medication) whereas in the utility study, "*the inclusion criteria were not controlled in this way and thus patients could have been further along their period of progression*".

The ERG is aware that the above mentioned factors may have influenced the differences in-between the studies and considers the utility value of [redacted] from the UK mCRPC Utility Study for post-docetaxel patients on active treatment as appropriate. Also, one-way sensitivity analysis shows that varying the post-docetaxel utility value only has a minor impact on the ICER (increases to £48,316).

The ERG has further asked the manufacturer to clarify why the fact that the majority of patients in the UK mCRPC Utility Study were assumed to have already been receiving AAP prohibited the use of differential utility scores for AAP vs BSC in the post-docetaxel phase. As a result the ERG has requested the manufacturer to provide results of the scenario using the value resulting from the UK mCRPC Utility Study as the utility for patients receiving AAP and subtract the 0.046 for patients receiving BSC in the model

In the Response to the Request for Clarification from the ERG[40] the manufacturer agreed this as *"a valid scenario to explore*" and has presented two different scenario results where the AAP post-docetaxel on treatment utility increment of 0.046 is applied by either adding it to

the TA259 baseline utility value of 0.78 (for AAP post-docetaxel) or subtracting it from the 0.78 utility value (for BSC post-docetaxel). These additional analyses resulted in ICERs of £48,316 and £47,936 respectively. The ERG believes this last scenario to be the most realistic one, assuming that indeed the majority of patients in the UK mCRPC Utility Study sample were on AAP treatment.

5.2.8 Resources and costs

Direct costs of the technologies

Table 5.7 reflects the costs per month (including, for docetaxel, administration costs) for the intervention and comparator technologies in the economic model. AA was dosed at 1,000 mg daily, BSC at 10 mg daily. Docetaxel was dosed at 75 mg/m2, applying a body surface area of 2.08 m^2, based on observed weight and height in the COU-AA-302 trial. The base case considers wastage of docetaxel and also a PAS discount [redacted] for the monthly cost of AA. A compliance rate of 98% was considered for AA patients, based on the COU-AA-302 trial. The cost of BSC was represented by a 10mg daily dose of prednisolone. Both AA and docetaxel were assumed to be administered with concurrent prednisolone, 10 mg daily.

Table 5.7: Unit costs associated with the technology in the economic model

Items	AA cost, £	BSC (PP) cost, £	Docetaxel cost, £	Ref
Cost of technology per month	[redacted]	2.63	1,550.14	BNF online (accessed 12.13)
Administration costs per 3 weeks	0	0	214	National Schedule of Reference Costs - NHS Trust Administration 2012-2013 (Code SB12Z)
Total cost per month	[redacted]	2.63	1,550.14	

Source: MS, Table 52, page 147[1]
AA, abiraterone acetate; BSC, best supportive care; PP, placebo plus prednisolone.

Scheduled costs

The costs of scheduled, disease related patient follow-up consisted of clinical visits, imaging diagnostic tests and clinical laboratory tests to monitor the status of disease in patients with mCRPC. To obtain information on resource use for an average mCRPC patient in the UK, an online survey was performed among 53 oncologists (used in base-case) and 50 urologists (used in scenario analysis). Questions covered total outpatient visits, scans (CT, MRI, bone, ultrasound, ECG) and laboratory tests (full blood count, liver function, renal function, PSA and 'other'). Based on the SPC, which states that patients receiving AA require additional monitoring for the first 3 months of treatment, higher MRU is applied in AA patients both pre- and post docetaxel until 3 months after start of treatment. NHS reference costs were used, which together with the results from the survey resulted in total costs per treatment, per phase (see Table 5.8)

Table 5.8: Scheduled oncologist MRU costs

	Median, £	25th percentile, £	75th percentile, £
AAP	217.22*	129.24	266.77
BSC (PP)	82.40	43.56	158.26
BSC (pre-docetaxel)	82.40	43.56	158.26
Docetaxel	203.46	107.22	248.79
BSC (post-docetaxel)	116.01	47.69	159.14
BSC (PP post-docetaxel)	116.01	47.69	159.14
AAP (post-docetaxel)	198.76*	96.76	223.37
BSC (before death)	39.96	21.54	124.59

Source: MS, Table 55, page 151[1]
*Cost applied for the first 3 months, and then the equivalent cost of BSC (PP) thereafter (section 7.5.4 of MS)
AAP, abiraterone acetate plus placebo; AE, adverse event; BSC, best supportive care; MRU, medical resource utilisation; PP, placebo plus prednisolone.

Unplanned costs

Unplanned events while on treatment were estimated based on the COU-AA-301 and COU-AA-302 trial data. However, since these trials did not contain data on pre- and post-docetaxel BSC nor on docetaxel, assumptions had to be made. Patients on BSC (pre-docetaxel) were assumed to have the same unplanned event-related MRU as BSC (PP) treated patients in COU-AA-302. For docetaxel, baseline resource utilisation was equivalent to the placebo arm of the COU-AA-301, with incremental costs from grade 3 or 4 AEs. BSC (post-docetaxel) was also assumed to be similar to placebo in COU-AA-301.

In addition, no MRU data were collected for BSC (before death) strategies. Since patients in the COU-AA-301 trial had advanced disease and fairly short life expectancy, the MRU costs from the COU-AA-301 trial was used as a proxy for BSC (before death) in the model.

Unplanned MRU while on-treatment in the COU-AA-302 population seemed to be mostly associated with treatment duration but not with type of treatment. In the COU-AA-301 trial, MRU was independent of both treatment arm and treatment duration. Table 5.9 summarises the unplanned MRU cost inputs used in the model and the impact the unplanned MRU cost has on the application of AE cost.

Table 5.9: Trial-based unplanned MRU costs per month

	Unplanned MRU cost, £	Source	Impact on application of AE cost
AAP	93.79	302 trial unplanned MRU	Already reflected in the trial unplanned MRU data; no need to consider additional AE cost
BSC (PP)	93.79	Assumed to be the same as PP arm of 302 trial	AE costs are included
BSC (pre-docetaxel)	93.79	Assumed to be the same as PP arm of 302 trial	AE costs are included
Docetaxel	380.29	Assumed to be the same as PP arm of 301 trial	The baseline cost of AEs was similar to that of the 301 unplanned MRU. The cost of incremental AEs were also considered in the model.

	Unplanned MRU cost, £	Source	Impact on application of AE cost
BSC (post-docetaxel)	380.29	Assumed to be the same as PP arm of the 301 trial	Treatment is similar to BSC (PP), therefore like BSC (PP) unplanned MRU, the AE costs are included
BSC (PP post-docetaxel)	380.29	301 trial unplanned MRU	Already reflected in the trial unplanned MRU data, no need to consider additional AE cost
AAP (post-docetaxel)	380.29	301 trial unplanned MRU	Already reflected in the trial unplanned MRU data, no need to consider additional AE cost
BSC (before death)	380.29	Assumed to be the same as the 301 model post-progression cost	Treatment is similar to BSC (PP post-docetaxel), therefore like BSC (PP post-docetaxel) unplanned MRU, the AE costs are included

Source: MS, Table 56, page 151[1]
AAP, abiraterone acetate plus placebo; AE, adverse event; BSC, best supportive care; MRU, medical resource utilisation; PP, placebo plus prednisolone.

Adverse-event costs

As mentioned in the previous section, the baseline MRU for docetaxel in the model was assumed to be equal to post-docetaxel AAP in the COU-AA-301 trial. However, to address the impact of incremental grade 3 or 4 AE rates associated with docetaxel, costs of these AEs were assigned separately, using the raw incidence rate. Resources utilised in treating grade 3/4 AEs were identified based on an advisory board composed of 5 UK oncologists familiar with treating such events. The advisory board estimated the percentage of patients treated in each setting (inpatient, day case, GP, or nurse), as well as the likely medications provided (type, dosage, duration). Table 5.10 presents the total costs related to the treatment of each grade 3/4 AE.

The cost of granulocyte colony-stimulating factor (G-CSF) for prevention and treatment of febrile neutropaenia was considered in the economic model as follows: prophylactic use starts in cycle 1 for 15% of the patients and therapeutic use starts in cycle 2 for 3% of the patients.

Table 5.10: List of AEs and summary of costs included in the economic model

AE	Treatment cost per event, £	Medication cost[a] per event, £	Total cost per event, £
Neuropathy	553.50	7.88	561.38
Neutropaenia	808.50	0.00	808.50
Febrile neutropaenia	5,147.50	0.00	5,147.50
Thrombocytopaenia	703.80	0.00	703.80
Anaemia	945.00	233.80	1,178.80
Oedema	891.50	8.43	899.93
Hypokalaemia	1,210.50	3.75	1,214.25
Hypertension	467.50	6.83	474.33
Arthralgia	198.28	13.98	212.26
Asthenia	13.18	16.49	29.67
Diarrhoea	1,356.00	7.51	1,363.51
Dyspnoea	0.00	0.00	0.00
Nausea	693.00	2.13	695.13

AE	Treatment cost per event, £	Medication cost[a] per event, £	Total cost per event, £
Vomiting	2,033.00	9.78	2,042.78

Source: MS, Table 58, page 155[1]
[a]Medication cost is only applied patients who were not treated as inpatient cases.

ERG comment: In general, there are no major issues with the methods for measuring resources and defining unit prices. However, there are a few things worth mentioning

- There is no valid justification for the assumptions that needed to be made with respect to the unplanned MRU in the treatment phases for which empirical data were lacking (i.e. BSC pre- and post-docetaxel phases, docetaxel, and BSC before death). The manufacturer acknowledges this is a limitation of the model in section 7.10.3 of the MS. For instance, using the complete COU-AA-301 population as a proxy for BSC (before death) since this trial included patients with an already fairly short life expectancy seems rather arbitrary and might underestimate true unplanned MRU in BSC patients in the before death stage.
- the way AEs are dealt with is slightly confusing. In the clarification phase, the ERG requested an additional analysis in which a differential AE rate (between AA and BAS) was applied with respect to the costs. The manufacturer has performed this analysis and the results showed a very minimal decrease in the ICER (from £46,722 in the base case to £46,686), implying that differential costs of AEs are a minor factor in the cost-effectiveness analysis.
- It is not clear whether the incorporation of G-CSF costs in the model is extra to the treatment costs for neutropaenia already taken into account with the adverse event rate (if so, G-CSF costs may have been double-counted), or that the costs of G-CSF are included in these AE costs. It is somewhat confusing that G-CSF costs are mentioned separately, and are also used for prophylaxis.

5.2.9 Cost effectiveness results

In the base case analyses AAP and BSC (PP) resulted in a total of [redacted] and [redacted] undiscounted life years respectively (Table 5.11). Disaggregated by treatment phase, AAP accumulated more life years during the pre-docetaxel phase ([redacted] versus [redacted] undiscounted life years), this was mainly due to a longer time before 1st line active treatment discontinuation ([redacted] versus [redacted] undiscounted life years; MS[1] Table 68). Mean survival during the on-docetaxel and post-docetaxel phases was short after treatment with AAP compared with BSC (PP), [redacted] versus [redacted] and [redacted] versus [redacted] undiscounted life years respectively. However, during the post-docetaxel phase, mean survival after active treatment was longer after 1st line treatment with AAP compared with BSC (PP); [redacted] versus [redacted] undiscounted life years (MS Table 68).

Consistent with mean survival, AAP accumulated more QALYs than BSC (PP) resulting in an incremental QALY of 0.57 ([redacted] versus [redacted]; Table 5.12). However AAP was also more expensive than BSC (PP) because of higher costs during the pre-docetaxel phase ([redacted] versus [redacted]); leading to incremental costs of [redacted] (of which [redacted] are attributed to AAP medication and administration). This cost difference between AAP and BSC (PP) was

to some extent compensated during the on- and post-docetaxel phases where AAP was less expensive than BSC (PP) ([redacted] versus [redacted]; Table 5.13). This resulted in a total cost difference of £26,404 between AAP and BSC (PP). Thus AAP was both more effective and more expensive than BSC (PP) leading an ICER of £46,722 per QALY (Table 5.14).

Note that PSA results reported by the manufacturer (MS section 7.7.8) are incorrect. The probability that AAP is cost-effective compared to BSC for thresholds of £30,000, £40,000 and £50,000 is 0%, 10% and 67% respectively (Figure 5.8 and Table 5.15).

Table 5.11: Summary of proportion of patients and respective duration in each treatment phase

Survival (years)	Mean survival of those who entered the phase		% entered this phase	Mean survival	
	Undiscounted	Discounted		Undiscounted	Discounted
AAP					
Pre-docetaxel survival[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
On-docetaxel survival[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-docetaxel survival[c]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Overall survival (total LYs)				[redacted]	[redacted]
BSC (PP)					
Pre-docetaxel survival[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
On-docetaxel survival[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-docetaxel survival[c]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Overall survival (total LYs)				[redacted]	[redacted]
Incremental (AAP - BSC (PP))					
Pre-docetaxel survival[a]	0.99	0.83	0%	0.99	0.83
On-docetaxel survival[b]	-0.18	-0.18	-6%	-0.17	-0.16
Post-docetaxel survival[c]	-0.04	-0.05	-4%	-0.05	-0.05
Overall survival (total LYs)				**0.77**	**0.62**

Source: MS[1] Tables 69 and DES model

AAP, abiraterone acetate plus prednisolone; LY, life year; BSC, best supportive care; PP, placebo plus prednisolone.

[a] Includes pre-docetaxel treatment and BSC before docetaxel treatment start

[b] Includes time between docetaxel treatment end and post-docetaxel active treatment start or death before post-docetaxel active treatment start

[c] Includes time after post-docetaxel active treatment start

Table 5.12: Summary of LY and QALY gain by pre/on/post-docetaxel status

	AAP		BSC (PP)		Incremental	
	LY	QALY	LY	QALY	LY	QALY
Pre-docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	0.83	0.71
On docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	-0.16	-0.10
Post-docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	-0.05	-0.04
Total	[redacted]	[redacted]	[redacted]	[redacted]	0.62	0.57

Source: MS[1] Tables 69 and 70

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; LY, life year; QALY, quality-adjusted life year.

Table 5.13: Summary of costs by health state disaggregated by category of cost

	AAP, £	BSC (PP), £	Incremental, £	Absolute increment, %
Pre-docetaxel	[redacted]	[redacted]	[redacted]	100
Drug, premedication, and administration	[redacted]	[redacted]	[redacted]	[redacted]
Incremental grade 3/4 AEs	[redacted]	[redacted]	0	0.0
Unplanned, event-related MRU	[redacted]	[redacted]	1,159	3.7
Scheduled, follow-up MRU	[redacted]	[redacted]	1,185	3.8
Terminal	[redacted]	[redacted]	142	0.5
On and post-docetaxel	[redacted]	[redacted]	**–4,960**	100
Drug, premedication, and administration	[redacted]	[redacted]	–3,378	68.1
Incremental grade 3/4 AEs	[redacted]	[redacted]	–21	0.43
Unplanned, event-related MRU	[redacted]	[redacted]	–960	19.4
Scheduled, follow-up MRU	[redacted]	[redacted]	–381	7.7
Terminal	[redacted]	[redacted]	–220	4.4

Source: MS[1] Table 71

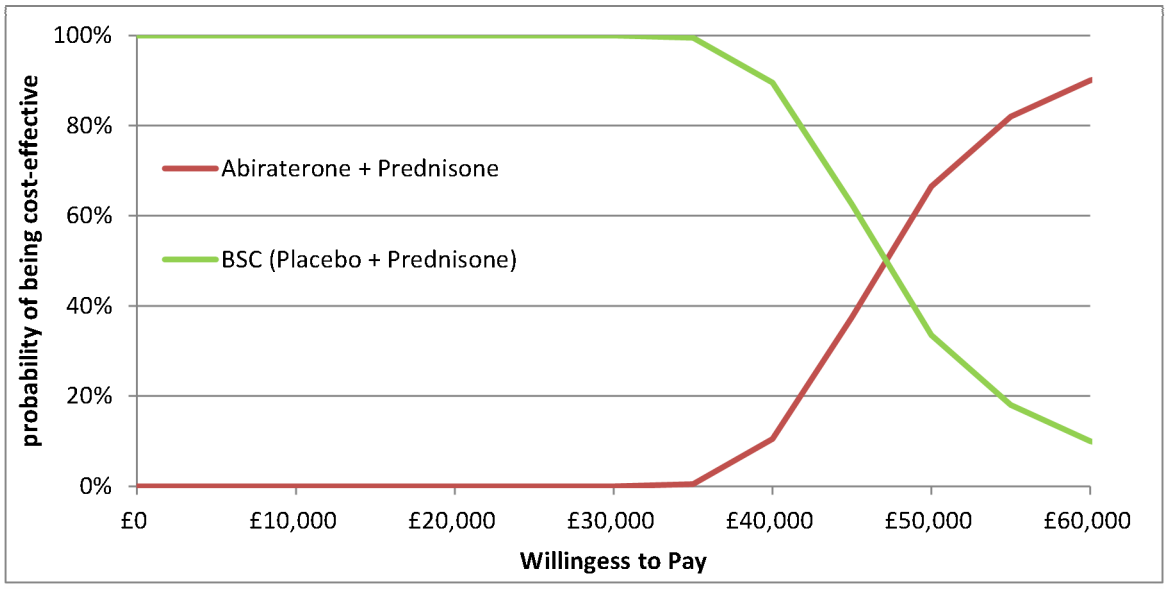
Table 5.14: Base-case results (with PAS)

Technology	Total costs, £	Total LYG	Total QALYs	Incremental costs, £	Incremental LYG	Incremental QALYs	ICER, £/QALY
BSC (PP)	[redacted]	[redacted]	[redacted]	–	–	–	–
AAP	[redacted]	[redacted]	[redacted]	26,404	0.62	0.57	46,722

Source: MS[1] Table 72

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; ICER, incremental cost-effectiveness ratio; LYG, life years gained; PAS, patient access scheme; QALY, quality-adjusted life year.

Figure 5.8: ERG revised cost-effectiveness acceptability curve (Manufacturer base case)[a]



[a]Note that the Cost-effectiveness acceptability curve reported in MS[1] Figure 17 is incorrect

Table 5.15: ERG revised summary of the PSA (Manufacturer base case)[a]

WTP threshold	AAP, %	BSC (PP), %
£30,000/QALY	0	100
£35,000/QALY	0	100
£40,000/QALY	10	89
£45,000/QALY	38	63
£50,000/QALY	67	34

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; QALY, quality-adjusted life year.
[a]Note that the values reported in MS[1] Table 47 are incorrect

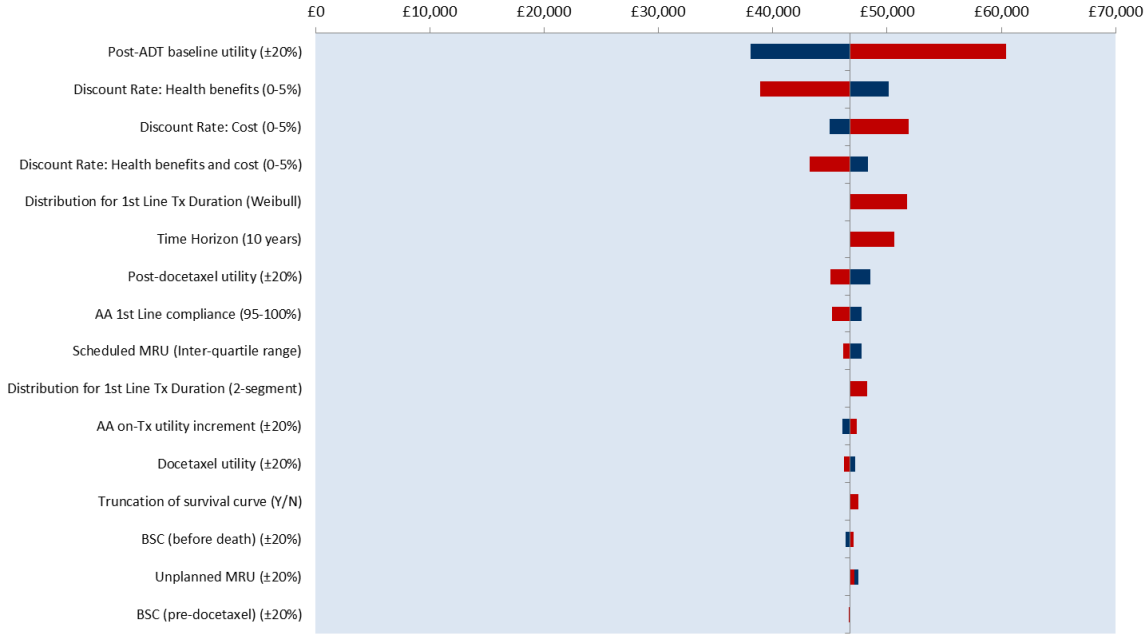
ERG comment: For a willingness to pay threshold of £46,722 or higher AAP can be considered cost-effective based on the manufacturer's base case. However if the ERG recalculates this deterministic ICER based on the economic model provided by the manufacturer, a slightly higher ICER of £46,756 was calculated (consistent among multiple re-runs). Differences between the ICER provided by the manufacturer and the recalculated ICER were due to lower on- and post-docetaxel Grade 3/4 AE costs. As a result, the incremental Grade 3/4 AE costs went from [redacted] to [redacted] ([redacted]). Moreover, it should be noted that the PSA performed by the manufacturer was based on only 200 iterations.

5.2.10 Sensitivity and scenario analyses

Five one-way sensitivity analyses showed an ICER higher than £50,000 (Figure 5.9). For one-way sensitivity analyses based on discount rates (for effects or costs), a shorter time horizon of 10 years and using a Weibull (instead of Log-logistic) distribution for the "Time to AAP/BSC (PP) treatment discontinuation" prediction equation yielded ICERs between

£50,000 and £52,000. The highest ICER of £60,418 was observed when using a post-ADT baseline utility of [redacted] (instead of [redacted]).

Figure 5.9: Tornado diagram (deterministic 1-way sensitivity analyses)



Source: MS[1] Figure 15

AA, abiraterone acetate; ADT, androgen deprivation therapy; BSC, best supportive care; MRU, medical resource utilisation; N, no; PAS, patient access scheme; Tx, treatment; Y, yes.

The following scenario analyses were performed:

1. Base-case analysis without PAS
2. Using urologist scheduled MRU costs (instead of oncologist scheduled MRU costs)
3. Using combined oncologist and urologist scheduled MRU costs (instead of oncologist scheduled MRU costs)
4. Utilities from the FACT-P to EQ-5D mapping study
5. Utilities from the FACT-P to EQ-5D mapping study applied to post-docetaxel setting
6. Using a utility of [redacted] prior to death (consistent with the enzalutamide STA)
7. Substituting prednisolone costs with dexamethasone costs for BSC
8. Modifying prediction coefficients to generate comparable post-docetaxel survival for AAP and BSC (PP)
9. Patients in the BSC (PP) arm do not receive an efficacious active treatment post-docetaxel
10. Enzalutamide included as a post-docetaxel active treatment option for both arms
11. No restriction on patients ECOG status when switching to docetaxel

Excluding the scenario in which the PAS was not applied, the only scenario to increase the ICER to above £50,000/QALY the scenario using mapped FACT-P to EQ-5D utility scores (scenario 4: £50,163/QALY). All other scenarios resulted in ICERs of £45,393–48,833/QALY.

ERG comment: Although several sensitivity and scenario analyses were performed by the manufacturer, the ERG would have preferred additional analyses for:

1. Using a post-docetaxel on treatment utility gain of 0.046 for patients treated with abiraterone compared to patients not on active treatment (consistent with TA 259; clarification question B1c)).
2. Using a post-docetaxel baseline utility of 0.78 instead of [redacted] (consistent with TA 259; clarification question B2)
3. Removing the on treatment utility gain for AAP and including costs and utility decrements for each separate adverse event (clarification question B3d)
4. Using the ITT population (clarification question B4a)
5. Consistently using a p-value threshold of 0.10 without exceptions (clarification question B4b)
6. Including the possibility of dying during post-docetaxel AAP and BSC (PP) treatment and during post-docetaxel treatment (clarification question B4c)
7. When patients for BSC are not censored after sequential post-docetaxel abiraterone and cabazitaxel treatment. This would be consistent with AAP as these patients were not censored after post-docetaxel cabazitaxel treatment (clarification question B8)

The manufacturer provided additional analyses for three (scenarios 1-3) of the above mentioned requested additional analyses. Additional analyses were provided for scenarios using a post-docetaxel on treatment utility of 0.046 and/or using a post-docetaxel baseline utility of 0.78 which resulted in ICERs ranging between £46,917 and £48,316. When the on treatment utility for AAP was removed and separate costs and utility decrements were included for each separate adverse event, this resulted in an ICER of £50,880. This increase of the ICER was mainly due to removing the on treatment utility gain which separately resulted in an ICER of £50,120, whereas including separate costs and utility decrements for each separate adverse event resulted in ICERs of respectively £47,415 and £46,686.

In addition, the ERG noticed (Table 5.16) that post-docetaxel survival in the current model seems very low compared to reported in TA259 (considering post-docetaxel treatment with abiraterone). In the current assessment the post-docetaxel phase solely consists of patients who entered the post-docetaxel active treatment phase (MS Table 68), so this phase seems comparable to the population in TA259. Moreover, the difference with TA259 is in favour of AAP in the current assessment. Therefore, the ERG performed an additional sensitivity analysis (presented in Section 5.3), using the ERG base case as starting point, to assess the impact of using post-docetaxel survival similar to TA259 (done by modifying prediction coefficients for "Time from post-docetaxel treatment discontinuation to death").

Table 5.16: Comparison of post-docetaxel survival with TA 259

Undiscounted LYs	AA[a]	BSC (PP)[a]	Incremental
TA 259 (post-docetaxel N=832)	[redacted]	[redacted]	[redacted]
Current TA(post-docetaxel N=104)[b,c]	[redacted]	[redacted]	[redacted]

Abbreviations: LY, life year; AA, abiraterone acetate; BSC, best supportive care; PP prednisone/prednisolone.

[a] Post-docetaxel treatment; in the current assessment patients treated initially treated with BSC (PP) in the pre-docetaxel phase are eligible for Post-docetaxel treatment with AAP whilst patients treated initially treated with AA in the pre-docetaxel phase are not eligible for retreatment with AA.
[b] Number of patients estimated from the number of patients at risk retrieved from the Kaplan-Meier curve for "Time from post-docetaxel treatment discontinuation to death" which is based on the ITT (N=1088). As the prediction equations were calculated based on the analysable dataset (N=902), this number is multiplied by 83% (902/1088).
[c] Calculated from the economic model (undiscounted LYs for the post-docetaxel phase if 100% of the patients would enter this phase)

5.2.11 Subgroup analyses

Two pre-specified subgroups were considered:

1. patients with baseline BPI question #3 of 0 or 1 (67% of the COU-AA-302 trial population)
2. patients with ECOG PS = 0 (76% of the COU-AA-302 trial population)

These were subgroups where AAP demonstrated better efficacy in terms of OS HR compared to the ITT population. The HR's applied for these subgroups were estimated using non-stratified Cox proportional hazard models.

As expected, both subgroup analyses decreased the base case ICER of £46,722 to approximately £42,000.

ERG comment: The manufacturer states that the HRs estimated for the prespecified subgroups were consistent with MS[1] Figures 7 and 8. However, the HRs and their 95% confidence intervals were not specified. Moreover it was unclear how the estimated HRs were exactly incorporated in the DES model. Thus the validity of the subgroup analyses could not be established.

5.2.12 Model validation and face validity check

Face validity

The manufacturer reports that the model approach and structure were reviewed by UK clinicians. They indicated that the UK scenario treatment pathway reflected the current practice in UK. UK clinicians also provided constructive feedback on the scheduled MRU for the average patient with mCRPC. Furthermore, the basic concept of the model design was presented to an internal peer review board comprised of senior scientists who are familiar with prostate cancer. This process helped to ensure that the model design was grounded with an appropriate clinical basis.

Internal validity

The manufacturer reported that the completed Microsoft Excel model was verified by a modelling expert not involved in this study, by examining logical structures in the simulation, mathematical expressions, and sequence of computations. Patients were followed from start to end of the simulation run with their event list, and profiles evaluated after each event occurred to ensure that the correct event sequences were initialised. Expressions were

verified by comparing the simulated values to hand calculations. A variety of stress tests were performed, such as extreme-value analysis.

In addition to the internal model validation by the manufacturer, the ERG performed internal validity checks. Given the computational burden and extensive number of formulas used in the DES model, it was not possible to check all formulas and macros within the available time. Therefore, a sample of formulas (prioritised based on anticipated importance) and macros were checked by the ERG as well as the global methodology and structure of the DES model.

Cross validation

Cross-validation includes examining different models that address the same problem and comparing their results. The manufacturer did not report that they had cross-validated the model. The fact that post-docetaxel survival in the model for this submission was not comparable with the survival in TA259 was addressed by an additional sensitivity analysis performed by the ERG, setting post-docetaxel survival equal to TA259. This increased the ICER to £65,515 per QALY.

External validity

The manufacturer stated to have examined predictive validity of the model by comparing simulated survival curves (e.g. time on treatment, OS) with observed clinical trial data (i.e. Kaplan–Meier survival curves). The ERG considers this type of analysis to be addressing external validity rather than predictive validity.[63]

To generate model predictions, patient profiles from the COU-AA-302 trial (the analysed 902 patients) were used in the simulation to inform baseline patient characteristics. Model predictions including time to AAP/BSC (PP) discontinuation, time from AAP/BSC (PP) to docetaxel start, time from docetaxel start to docetaxel end, and OS were plotted against the Kaplan–Meier survival curves. The manufacturer concludes that model predictions were consistent with the COU-AA-302 trial results.

ERG comment: In general, the ERG feels that the model structure used is overly complicated and could have been more straightforward. This is partly due to the fact that many stages and equations included in the model may not have a relevant contribution to the outcome (e.g. waiting/monitoring time before starting treatment). Also, the opinion of the ERG is that the decision problem did not require an individual patient simulation. For instance, a Markov cohort model (using only relevant health states, which may be more than the regular three state model) would have been more transparent and would have allowed the ERG much more flexibility in performing additional analyses. However, given that the ERG had to work with the DES model provided, it might have been useful to have a technical document alongside it, listing all steps, formulas and macros used.

The ERG did not entirely agree with the manufacturer's statement that model predictions were consistent with the trial results, as it is not the ITT population that is used (but a subset), and therefore model results are probably not fully comparable to the ITT results of the COU-AA-302 trial.

A few minor points from the internal validation check:

- When recalculated, the deterministic ICER deviated slightly from the ICER reported in the MS
- Sometimes *years*365* is used and sometimes *years*365.25* to calculate the number of days (however no big impact expected)
- There was a small error in the presented CEAC (see revised CEAC curves)

5.3 Exploratory and sensitivity analyses undertaken by the ERG

The Base case analyses provided by the manufacturer resulted in an ICER of £46,722. However when recalculating the deterministic analyses, this resulted in an ICER of £46,756. As mentioned above, the ERG would have preferred a base case wherein 1) a disutility (of 0.046) is applied in the post-docetaxel phase for patients not on active treatment (see also clarification question B1c) and 2) prediction equations for time to treatment discontinuation, treatment start and death that are based on the ITT population and consistently include treatment as the only covariate. Incorporating these changes separately would result in ICERs of £46,952 and £57,337 respectively (see Table 5.17 for the updated prediction equations). When these two changes are combined to form the ERG base case this would result in incremental costs, QALYs and life years of £24,757, 0.43 and 0.40 respectively leading to an ICER of £57,668 (Tables 5.18 and 6.1).

Table 5.17: Updated prediction equations[a]

	Intercept	SE	Scale	SE	Treatment	SE
1st line (AAP/BSC (PP)) - on treatment						
Time to AAP/BSC (PP) treatment discontinuation						
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1st line (AAP/BSC (PP)) - post treatment						
Time from AAP/BSC (PP) treatment discontinuation to docetaxel start						
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time from AAP/BSC (PP) discontinuation to death before docetaxel start						
Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2nd line (docetaxel) - on treatment						
Time from docetaxel start to docetaxel discontinuation						
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time from docetaxel start to death before docetaxel discontinuation						
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2nd line (docetaxel) - post treatment						

Time from docetaxel discontinuation to post-docetaxel treatment start						
Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time from docetaxel discontinuation to death before third-line treatment starts						
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3rd line (post-docetaxel active treatment) - on treatment						
Time from third-line treatment start to post-docetaxel treatment discontinuation.						
Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3rd line (post-docetaxel active treatment) - post treatment						
Time from post-docetaxel treatment (third-line treatment) discontinuation to death						
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull[b]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[a] Model distribution was selected based on AIC and BIC. If these two were inconsistent on the 'best fit' the distribution with the lowest average of the AIC and BIC was used.
[b] Only used in sensitivity analyses

Table 5.18: Summary of LY and QALY gain by pre/on/post-docetaxel status (ERG base case)

	AAP		**BSC (PP)**		**Incremental**	
	LY	**QALY**	**LY**	**QALY**	**LY**	**QALY**
Pre-docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	0.61	0.57
On docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	-0.17	-0.10
Post-docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	-0.04	-0.04
Total	[redacted]	[redacted]	[redacted]	[redacted]	0.40	0.43

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; LY, life year; QALY, quality-adjusted life year.

Probabilistic sensitivity analyses were performed for the ERG base case (using 2000 iterations) and the probability that AAP is cost-effective compared to BSC for thresholds of £30,000, £40,000 and £50,000 is 0%, 0% and 6% respectively (Figure 5.10 and Table 5.19).

Figure 5.10: Cost-effectiveness acceptability curve (ERG base case)

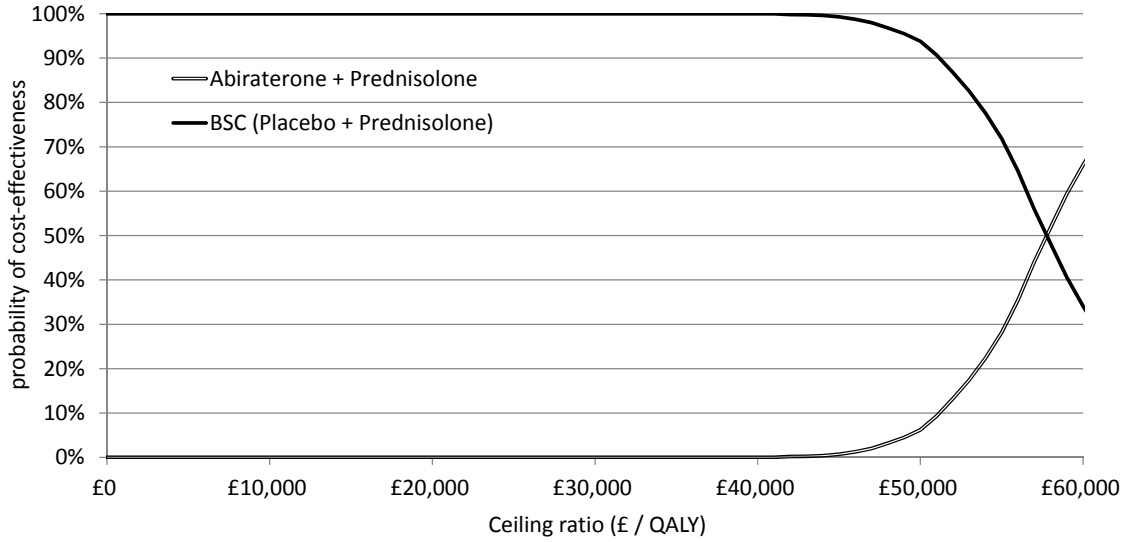


Table 5.19: Summary of the PSA (ERG base case)

WTP threshold	AAP, %	BSC (PP), %
£30,000/QALY	0	100
£35,000/QALY	0	100
£40,000/QALY	0	100
£45,000/QALY	1	99
£50,000/QALY	6	94

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; QALY, quality-adjusted life year.

Starting with the ERG base case, four additional sensitivity analyses were provided. Firstly, it was unclear to the ERG how the negative treatment effect to account for cabazitaxel use in the COU-AA-302 trial (MS Table 35) was calculated. Therefore, the impact of removing this negative treatment was explored. Secondly, as mentioned above it was noticed by the ERG that post-docetaxel survival was substantially lower than reported in TA259, (Table 5.16). Therefore the impact of assuming an equal post-docetaxel survival as TA259 was explored. Thirdly, the log-logistic distributions for estimating time to AAP/BSC (PP) treatment discontinuation and time to death while on docetaxel treatment were replaced by Weibull distributions. This can be regarded as a more conservative approach as the log-logistic distribution is often criticised for its long tail used to extrapolate data (offering for instance an unrealistic survival benefit). Note that although the AIC and BIC criteria, used to select the model distributions, provide data on the goodness of fit compared to the observed data; these information criteria are less helpful when considering the validity of the extrapolation. Fourthly, the above mentioned extrapolation issue is also applicable to the log-normal distribution. Therefore, the log-normal distributions for estimating time to docetaxel treatment discontinuation and time to death after post-docetaxel active treatment were replaced by Weibull distributions.

ICERs calculated in the additional sensitivity analyses ranged between £57,202 and £74,803. The second and third sensitivity analyses resulted in the highest ICERs. Assuming an equal post-docetaxel survival as TA 259 (by adjusting the coefficients for "Time from post-docetaxel treatment discontinuation to death"), would result in incremental costs, QALYs and life years of £24,159, 0.37 and 0.28 respectively, leading to an ICER of £65,515 (Table 6.2). Post-docetaxel survival in this analysis is [redacted] and [redacted] life years (undiscounted) for AAP and BSC (PP) respectively. This corresponds to mean survival of [redacted] and [redacted] life years (undiscounted) for AAP and BSC (PP) respectively for patients that started post-docetaxel active treatment. Finally, replacing the Log-logistic distributions with Weibull distributions would result in incremental costs, QALYs and life years of £19,620, 0.26 and 0.21 respectively leading to an ICER of £74,803 (Table 6.2).

5.4 Conclusions of the cost-effectiveness section

The main deviation from the decision problem defined in the scope is that docetaxel is not included as a comparator. However, as the indication is men with mCRPC in whom chemotherapy is not yet clinically indicated, it seems reasonable that docetaxel is not considered as a comparator.

Due to the above mentioned concerns, the ERG questions the validity of the ICER provided by the manufacturer. The ERG was able to resolve some of the issues highlighted by using an on-treatment utility for post-docetaxel active treatment and non-stratified prediction equations based on the ITT population using treatment as only covariate. This resulted in an ICER of £56,463 for the ERG base case. However, the ERG acknowledges that there are remaining uncertainties about the reliability of the cost-effectiveness evidence, which are not handled in the ERG base case and sensitivity analyses could not be provided to estimate the impact of these issues on the results. These issues include: censoring patients in the BSC (PP) arm after sequential treatment with AAP and cabazitaxel, not including the possibility of dying during AAP/BSC treatment and post-docetaxel active treatment, not using differential costs and utilities for all AEs for all treatment phases and no empirical data to calculate resources and costs for most of the treatment phases.

The ERG is not convinced that a DES model, simulating individual patients using 17 prediction equations would have been the most transparent approach, i.e. the model used is overly complicated and could have been more straightforward. This is partly due to the fact that many stages and equations included in the model may not have a relevant contribution to the outcome (e.g. waiting/monitoring time before starting treatment). The ERG believes that it would have been possible to use a more transparent model, for instance a Markov cohort model using only relevant health states, (which may be more than the regular three state model) and a sufficiently short cycle time. This model would also allow to reflect the clinical pathways in the UK and to produce results for subgroups with varying baseline characteristics.[3] Also, the ERG is not convinced by the manufacturer's arguments that a patient level simulation would be necessary for the decision problem defined during the scope. It should be noted that acknowledging patient heterogeneity does not necessarily require patient level simulation.[3] Moreover, the subgroup analyses provided by the

manufacturer could have been produced using a cohort model. Transparency is a key aspect of modelling and in this specific case a more transparent model would be more convenient for an external reviewer to assess face validity and internal validity of the model.

6. IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

Table 6.1: Overview of additional analyses undertaken by the ERG

Technology	Total costs, £	Total QALYs	Incremental costs, £	Incremental QALYs	ICER, £/QALY
MS Base case					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	26,404	0.57	46,722
Recalculated MS[1] Base case					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	26,423	0.57	46,756
Post-docetaxel on treatment utility[a]					
BSC (PP)	[redacted]	[redacted]			
AAP	[redacted]	[redacted]	26,423	0.56	46,952
Updated prediction equations[b]					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,757	0.43	57,337
ERG Base case[c]					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,757	0.43	57,688

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio.

[a] A disutility of 0.046 was applied in the post-docetaxel phase for patients not on active treatment (i.e. receiving BSC instead of abiraterone).

[b] Prediction equations based on the ITT population and including treatment as only covariate were used (based on the "302 mode Parametric Functions Parameters" file provided by the manufacturer in response to clarification question B4a)

[c] Combination of the two scenarios mentioned above

Table 6.2: Additional sensitivity analyses (based on ERG base case)

Technology	Total costs, £	Total QALYs	Incremental costs, £	Incremental QALYs	ICER, £/QALY
ERG Base case[c]					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,757	0.43	57,688
Remove cabazitaxel negative treatment effect					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,821	0.44	56,671
Equal post-docetaxel survival compared to TA 259					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,159	0.37	65,515
Weibull instead of Log-logistic					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	19,620	0.26	74,803
Weibull instead of Log-normal					
BSC (PP)	[redacted]	[redacted]	–	–	–
AAP	[redacted]	[redacted]	24,565	0.43	57,202

AAP, abiraterone acetate plus prednisolone; BSC, best supportive care; PP, placebo plus prednisolone; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio.

7. END OF LIFE

NICE has issued supplementary advice to the Appraisal Committees for appraising life-extending, end of life treatments. These are treatments which may be life-extending for patients with short life expectancy, and which are licensed for indications affecting small numbers of patients with incurable illnesses.[11]

NICE end of life supplementary advice should be applied in the following circumstances and when all the criteria referred to below are satisfied:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment, and;
- The treatment is licensed or otherwise indicated, for small patient populations.

The end of life criteria used for abiraterone in the manufacturer's submission and the manufacturer's justification for applying the end of life criteria to abiraterone are outlined in section 7.7.6 of the manufacturer's submission (MS, page 167; see Table 7.1 below).[1]

Table 7.1: End of life criteria for abiraterone

Criteria	Justification
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	The prognosis of mCRPC patients is poor. Five-year survival rates of only 26–31% have been reported.[12-14] The control arm of the COU-AA-302 study shows that asymptomatic and mildly symptomatic patients have a short life expectancy of approximately 30 months; however patients in the trial are likely to have gone on to receive other clinical trial technologies post-docetaxel and therefore the survival observed for these patients is probably not reflective of the average mCRPC patient in the UK. The EAU[20] has estimated that the mean survival of patients with asymptomatic disease, dependent on the extent of metastases, is 9–27 months which is indicative that average survival is within the range considered end-of-life criteria (See MS, Table 6). As the disease becomes symptomatic, survival time decreases accordingly. For example, patients in docetaxel trials only have a median OS of 18–22 months.[15]
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment	AAP offers the asymptomatic and mildly symptomatic patient population a 5.2-month increase in median OS compared to BSC with PP[43] which exceeds the survival benefit observed for patients given AAP in the post-docetaxel setting (4.6 months).[56] Patients now have the opportunity to receive AAP earlier in the course of their disease (i.e. post-ADT, pre-chemotherapy rather than post-chemotherapy). If patients receive AAP earlier in their disease, they not only gain additional survival but are also able to preserve their HRQL earlier in the disease process, rather than extending their life at a time when they are likely to have a poorer HRQL.
The treatment is licensed or otherwise indicated, for small patient populations	It is estimated that in 2014, 4,689 patients in England and Wales will have asymptomatic or mildly symptomatic mCRPC post-ADT, pre-chemotherapy (see sections 2.2 and 8). A further 2,483 patients are estimated to be eligible for AAP in the post-chemotherapy setting (section

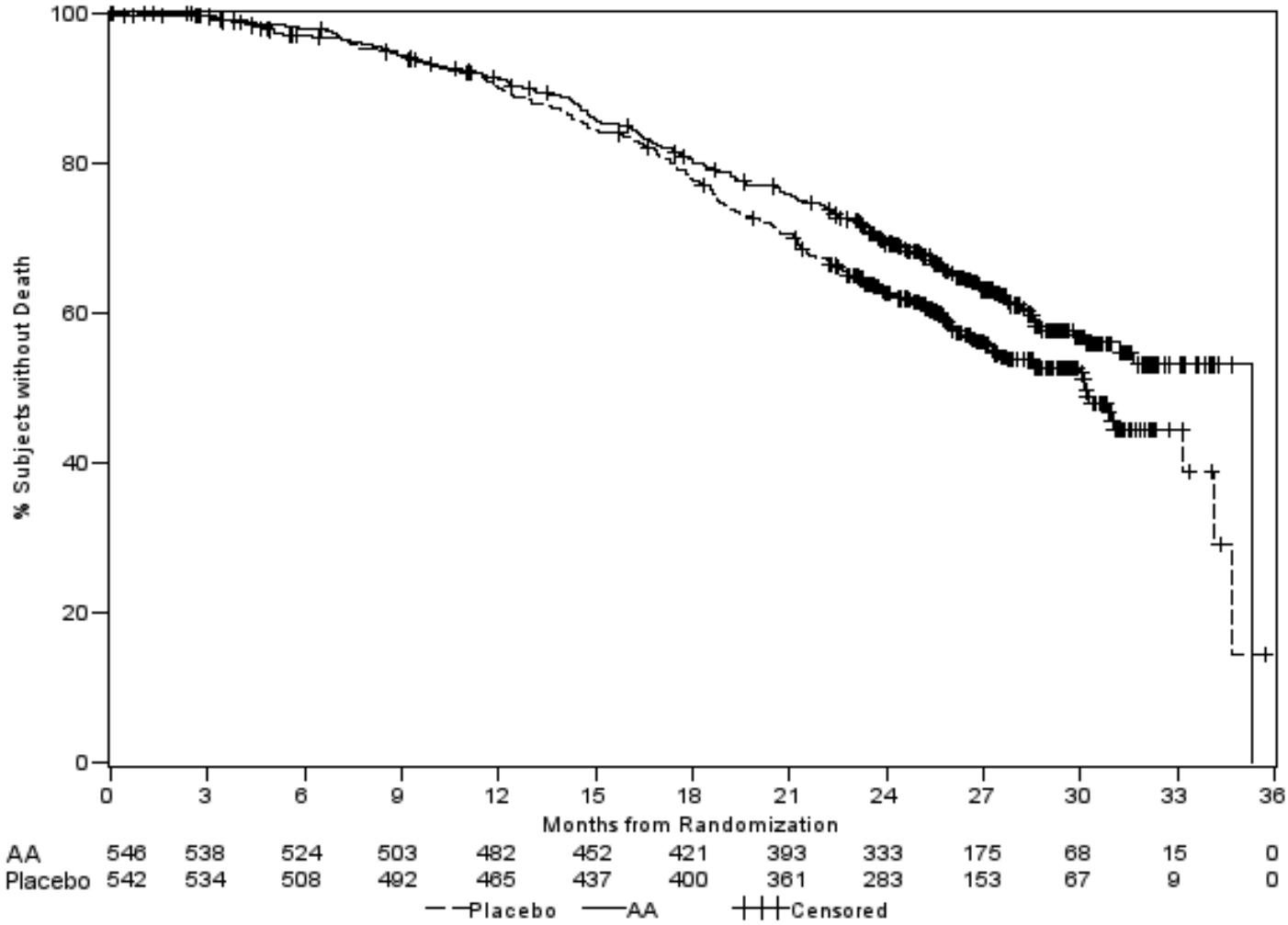
	2.2). Now that AAP is available to patients in the post-ADT, pre-chemotherapy setting it is anticipated that the number of patients receiving treatment in the post-chemotherapy setting will drop in future years, as those who have received AAP in the post-ADT, pre-chemotherapy setting will not be eligible for treatment post-chemotherapy. Once this is accounted for, the total number of patients eligible for treatment with AAP across both indications will remain small; estimated at 7,172 patients in 2014.

ERG comment: The manufacturer's justification for the third criterion seems valid. It is likely that the treatment indicated for a small patient population.

Regarding the second criterion, the manufacturer provides median survival estimates, but not mean survival in the MS. In the clarification letter, the ERG asked the manufacturer to provide the mean survival in the BSC group in COU-AA-302 for the overall population and for the subgroup of patients from UK centres, and the mean survival gain of abiraterone (AAP) compared with BSC (PP) in COU-AA-302 for the overall population and for the subgroup of patients from UK centres. The manufacturer responded that they were unable to answer these questions (See Janssen's response to clarification letter[40])

Regarding the first criterion, Figure 6 in the MS[1] (see Figure 7.1 below) shows that after 24 months, approximately 63% in the control group are still alive; and that the median survival is 30.1 (95% CI: 27.3 to 34.1) months. Therefore, it is unlikely that life expectancy in this patient group will be less than 24 months.

Figure 7.1: Kaplan–Meier curve of OS – ITT population (COU-AA-302 study third interim analysis) (See MS, Figure 6, page 68)



SOURCE: COU-AA-302 study clinical study report (unpublished)[43] and conference poster[51].
AA, abiraterone acetate plus prednisone/prednisolone; ITT, intent-to-treat; OS, overall survival; Placebo, placebo plus prednisone/prednisolone.

According to the manufacturer, patients in the trial are likely to have gone on to receive other clinical trial technologies post-docetaxel and therefore the survival observed for these patients is probably not reflective of the average mCRPC patient in the UK. However, as far as the ERG is aware the "short life expectancy, normally less than 24 months" is based on the normal treatment options available for these patients without the intervention under assessment.

8. OVERALL CONCLUSIONS

The main deviation from the scope is that docetaxel is not included as a comparator in the manufacturer submission. As the indication is men with mCRPC **in whom chemotherapy is not yet clinically indicated**, it seems reasonable that docetaxel is not considered as a comparator. However, in the final scope, NICE explicitly states that: "Docetaxel is included in the list of comparators because the recommendations in TA101 include patients who are asymptomatic or mildly symptomatic, and clinicians have stated that docetaxel is increasingly used for this patient group, and because of the lack of clear clinical criteria to identify the patient group in the CHMP indication".[24]

Assuming that most patients will end up using docetaxel, which also seems to be implied by the phrase "**not yet** clinically indicated", an important question in this appraisal is, according to the ERG, whether abiraterone followed by docetaxel is more effective than watch-full waiting (BSC) followed by docetaxel. In the COU-AA-302 trial, 239 out of 546 (43.8%) of AAP patients and 304 out of 542 (56.1%) of PP patients received docetaxel as subsequent therapy, following abiraterone or placebo (MS, Table 21, page 68). The results for this specific group of patients are not presented in the MS; therefore, we asked the manufacturer to provide these data in the clarification letter. [redacted]

Regarding the end-of-life criteria, Figure 6 in the MS[1] (see Figure 7.1 in this report) shows that after 24 months, approximately 63% in the control group are still alive; and that the median survival is 30.1 (95% CI: 27.3 to 34.1) months. Therefore, it is unlikely that life expectancy in this patient group will be less than 24 months.

The ERG questioned the validity of the ICER of £46,722 provided by the manufacturer, mainly because the manufacturer did not use an on-treatment utility for post-docetaxel active treatment and the ITT population was not used to estimate effectiveness in the model (both non-conservative assumptions). Moreover, there was a potential for bias due to inconsistency in the use of stratified models, covariates and interaction terms when estimating the prediction equations. The ERG was able to resolve some of the issues highlighted in the report and calculated an ICER of £56,463 for the ERG base case. However, the ERG acknowledges that there are remaining uncertainties concerning the reliability of the cost-effectiveness evidence, which are not handled in the ERG base case nor in additional sensitivity analyses.

8.1 Implications for research

An important question for this appraisal is the relative effectiveness of abiraterone followed by docetaxel in comparison with watch-full waiting (BSC) followed by docetaxel in adults

with metastatic castration-resistant prostate cancer who are asymptomatic or mildly symptomatic after failure of androgen deprivation therapy in whom chemotherapy is not yet clinically indicated. Therefore, a trial exploring that question is warranted. Moreover, this type of research would also facilitate economic analyses of abiraterone for adult men with mCRPC who were asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy was not yet clinically indicated.

9. REFERENCES

[1] Janssen. Abiraterone acetate (Zytiga®) for the treatment of metastatic castration-resistant prostate cancer in men not previously treated with chemotherapy: submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA). Janssen, 2014: 312.

[2] National Institute for Health and Clinical Excellence. *Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen. NICE technology appraisal guidance 259 [Internet]*. London: NICE, 2012 [accessed 17.01.14]. 51p. Available from: http://www.nice.org.uk/nicemedia/live/13775/59723/59723.pdf

[3] Grutters JP, Sculpher M, Briggs AH, Severens JL, Candel MJ, Stahl JE, et al. Acknowledging patient heterogeneity in economic evaluation : a systematic literature review. *Pharmacoeconomics* 2013;31(2):111-23.

[4] Vasaitis TS, Bruno RD, Njar VCO. CYP17 inhibitors for prostate cancer therapy. *J Steroid Biochem Mol Biol* 2011;125(1-2):23-31.

[5] eMedicine from WebMD. Prostate cancer - biology, diagnosis, pathology, staging and natural history [Internet]. *Medscape: eMedicine* 2011 [accessed 10.04.14].

[6] Rentsch CA, Cecchini MG, Thalmann GN. Loss of inhibition over master pathways of bone mass regulation results in osteosclerotic bone metastases in prostate cancer. *Swiss Med Wkly* 2009;139(15-16):220-5.

[7] National Institute for Health and Clinical Excellence. *Costing template for NICE technology appraisal 101-Docetaxel for the treatment of hormone-refractory metastatic prostate cancer [Internet]*. London: NICE, 2006 [accessed 04.03.14]

[8] GfK Healthcare. Healthcare resource utilisation and the treatment-related care-pathways within the current and future setting of metastatic castration resistant prostate cancer (mCRPC). 2011.

[9] Harris V, Lloyd K, Forsey S, Rogers P, Roche M, Parker C. A population-based study of prostate cancer chemotherapy. *Clin Oncol (R Coll Radiol)* 2011;23(10):706-8.

[10] Janssen Cilag. Data on File. Medeconnect Market Research: mCRPC patient split. Abiraterone_DoD_22November2012_OR_IL_02 2013.

[11] National Institute for Health and Clinical Excellence. *Appraising life-extending end of life treatments [Internet]*. London: NICE, 2009 [accessed 14.03.14]. 3p. Available from: http://www.nice.org.uk/media/E4A/79/SupplementaryAdviceTACEoL.pdf

[12] Hoschke B, Kindt B, Tillack A, Schicke B. *Update epidemiologie prostatakarzinom in Deutschland [Internet]*: Kooperationsverbund Qualitatssicherung durch Klinische Krebsregister (KoQK), Arbeitsgemeinschaft Deutscher Tumorzentren e.V., 2014 [accessed 04.03.14]. 17p. Available from: http://www.tumorzentren.de/tl_files/dokumente/Qualitaetskonferenzen/Fachvertraege%20DKK%202012/Prostata_Update%20Epi_DKK2012.pdf

[13] National Cancer Institute. *SEER Stat Fact Sheets: Prostate Cancer [Internet]*. Bethesda, MD: National Cancer Institute, 2013 [accessed 04.03.14] Available from: http://seer.cancer.gov/statfacts/html/prost.html

[14] Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010;60(5):277-300.

[15] Serpa NA, Tobias-Machado M, Kaliks R, Wroclawski ML, Pompeo AC, Del GA. Ten years of docetaxel-based therapies in prostate adenocarcinoma: a systematic review and meta-analysis of 2244 patients in 12 randomized clinical trials. *Clinical Genitourinary Cancer* 2011;9(2):115-23.

[16] National Institute for Health and Care Excellence. *Prostate cancer: diagnosis and treatment. NICE clinical guideline 175 [Internet]*. London: NICE, 2014 [accessed 04.03.14]. 45p. Available from: http://guidance.nice.org.uk/cg175

[17] National Institute for Health and Clinical Excellence. *Docetaxel for the treatment of hormone-refractory metastatic prostate cancer. NICE technology appraisal guidance 101 [Internet]*. London: NICE, 2006 [accessed 17.01.14]. 35p. Available from: http://www.nice.org.uk/nicemedia/live/11578/33348/33348.pdf

[18] National Institute for Health and Clinical Excellence. *Cabazitaxel for hormone-refractory metastatic prostate cancer previously treated with a docetaxel-containing regimen. NICE technology appraisal guideline 255*. London: NICE, 2012 [accessed 04.03.14] Available from: http://www.nice.org.uk/nicemedia/live/13731/59186/59186.pdf

[19] Armstrong AJ, Garrett-Mayer ES, Yang YC, de Wit R, Tannock IF, Eisenberger M. A contemporary prognostic nomogram for men with hormone-refractory metastatic prostate cancer: a TAX327 study analysis. *Clin Cancer Res* 2007;13(21):6396-403.

[20] Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Mason MD, et al. *Guidelines on prostate cancer [Internet]*. Arnhem, The Netherlands: European Association of Urology, 2013 [accessed 04.03.14] Available from: http://www.uroweb.org/gls/pdf/09_Prostate_Cancer_LR.pdf

[21] Horwich A, Parker C, Bangma C, Kataja V, ESMO Guidelines Working Group. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2010;21 Suppl 5:v129-v133.

[22] Cookson MS, Roth BJ, Dahm P, Engstrom C, Freedland SJ, Hussain M, et al. Castration-Resistant Prostate Cancer: AUA Guideline. *J Urol* 2013;190(2):429-38.

[23] National Comprehensive Cancer Network (NCCN). *NCCN Clinical Practice Guidelines in Oncology. NCCN Guidelines. Prostate cancer*. Fort Washington, PA: National Comprehensive Cancer Network, 2013

[24] National Institute for Health and Clinical Excellence. *Abiraterone acetate for the treatment of metastatic hormone relapsed prostate cancer not previously treated with chemotherapy [ID503]. Final scope [Internet]*. London: National Institute for Health and Clinical Excellence, 2012 [accessed 10.03.14] Available from: http://www.nice.org.uk/nicemedia/live/13641/62262/62262.pdf

[25] Janssen-Cilag Ltd. SPC Zytiga 250 mg tablets [Internet]. 2014 [accessed 04.03.14]. Available from: http://www.medicines.org.uk/emc/medicine/24976/SPC/Zytiga+250+mg+tablets

[26] Tannock I, Gospodarowicz M, Meakin W, Panzarella T, Stewart L, Rider W. Treatment of metastatic prostatic cancer with low-dose prednisone: evaluation of pain and quality of life as pragmatic indices of response. *J Clin Oncol* 1989;7(5):590-7.

[27] sanofi-aventis. *Summary of product characteristics: Taxotere [Internet]*, 2011 [accessed 04.03.14] Available from: http://www.sanofi-aventis.co.uk/products/Taxotere%2020mg_SPC.pdf

[28] European Medicines Agency. *Assessment report: Zytiga (abiraterone). Procedure no. EMEA/H/C/002321/II/0004/G [Internet]*. London: European Medicines Agency, 2012 [accessed 04.03.14]. 52p. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/002321/WC500137814.pdf

[29] Berthold DR, Pond GR, Soban F, de Wit R, Eisenberger M, Tannock IF. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer: updated survival in the TAX 327 study. *J Clin Oncol* 2008;26(2):242-5.

[30] Ryan CJ, Smith MR, de Bono JS, Molina A, Logothetis CJ, de Souza P, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med* 2013;368(2):138-48.

[31] European Medicines Agency. *Zytiga. Annex 1: summary of product characteristics [Internet]*. London, 2014 [accessed 10.03.14] Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002321/WC500112858.pdf

[32] Mezynski J, Pezaro C, Bianchini D, Zivi A, Sandhu S, Thompson E, et al. Antitumour activity of docetaxel following treatment with the CYP17A1 inhibitor abiraterone: clinical evidence for cross-resistance? *Ann Oncol* 2012;23(11):2943-7.

[33] Basch E, Autio K, Ryan CJ, Mulders P, Shore N, Kheoh T, et al. Abiraterone acetate plus prednisone versus prednisone alone in chemotherapy-naive men with metastatic castration-resistant prostate cancer: patient-reported outcome results of a randomised phase 3 trial. *Lancet Oncol* 2013;14(12):1193-9.

[34] Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy--Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. *Value Health* 2009;12(1):124-9.

[35] Victorson DE, Beaumont JL, Rosenbloom SK, Shevrin D, Cella D. Efficient assessment of the most important symptoms in advanced prostate cancer: the NCCN/FACT-P Symptom Index. *Psychooncology* 2011;20(9):977-83.

[36] FACIT. FACT-P (Version 4) [Internet]. FACIT, 2007 [accessed 10.3.14]. Available from: http://www.facit.org/FACITOrg/Questionnaires

[37] McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evid Based Libr Inf Pract* 2010;5(1):1-6.

[38] National Institute for Health and Care Excellence. *Single Technology Appraisal: specification for manufacturer/sponsor submission of evidence [Internet]*. London: NICE, 2012 [accessed 17.01.14]. 76p. Available from: http://www.nice.org.uk/media/D54/6E/SpecificationForManufacturerSponsorSubmissionOfEvidenceJune2012.doc

[39] National Institute for Health and Care Excellence. *Abiraterone acetate for the treatment of metastatic hormone relapsed prostate cancer not previously treated with chemotherapy [ID503]. Clarification letter*. Manchester: National Institute for Health and Care Excellence, 2014

[40] Janssen. Abiraterone acetate (Zytiga®) for the treatment of metastatic castration-resistant prostate cancer in men not previously treated with chemotherapy: response to request for clarification from the ERG. Janssen, 2014.

[41] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 13.02.14] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[42] Janssen. *Abiraterone acetate (Zytiga®) for the treatment of metastatic castration resistant prostate cancer following previous cytotoxic therapy: submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA) [Internet]*: Janssen, 2012 [accessed 04.03.14]. 215p. Available from: http://www.nice.org.uk/nicemedia/live/13484/58020/58020.pdf

[43] Janssen Research & Development. A phase 3, randomized, double-blind, placebo-controlled study of abiraterone acetate plus prednisone in asymptomatic or mildly symptomatic subjects with metastatic castration-resistant prostate cancer (22 May 2012 data cutoff). 2012.

[44] Hoy SM. Abiraterone acetate: a review of its use in patients with metastatic castration-resistant prostate cancer. *Drugs Aging* 2013;73(18):2077-91.

[45] Rathkopf DE, Smith MR, de Bono JS, Logothetis C, Shore ND, de Souza PL, et al. Updated interim analysis (IA) of COU-AA-302, a randomized phase III study of abiraterone acetate (AA) in patients (pts) with metastatic castration-resistant prostate cancer (mCRPC) without prior chemotherapy. *J Clin Oncol* 2013;31(Suppl 6):Ab5.

[46] Attard G, de Bono JS, Li W, Molina A, Griffin TW, Kheoh T, et al. ERG rearrangements and association with clinical outcome in patients (pts) receiving abiraterone acetate (AA): Results from the COU-AA-302 study in chemotherapy-naïve metastatic castration-resistant prostate cancer (mCRPC). *J Clin Oncol* 2013;31(Suppl):Ab5004.

[47] Saad F, Fizazi K, Smith MR, Griffin TW, Londhe A, Rathkopf D, et al. Impact of concomitant bone-targeted therapies (BTT) on outcomes in metastatic castration-resistant prostate cancer (mCRPC) patients (pts) without prior chemotherapy (ctx) treated with abiraterone acetate (AA) or prednisone (P). *J Clin Oncol* 2013;31(Suppl):Ab5037.

[48] Ryan CJ, Londhe A, Molina A, Smith MR, de Bono JS, Mulders P, et al. Relationship of baseline PSA and degree of PSA decline to radiographic progression-free survival (rPFS) in patients with chemotherapy-naive metastatic castration-resistant prostate cancer (mCRPC): results from COU-AA-302. *J Clin Oncol* 2013;31(Suppl):Ab5010.

[49] Rathkopf D, Smith MR, de Bono JS, Logothetis C, Shore N, de Souza PL, et al. Long-term safety and efficacy analysis of abiraterone acetate (AA) plus prednisone (P) in metastatic castration-resistant prostate cancer (mCRPC) without prior chemotherapy (COU-AA-302). *J Clin Oncol* 2013;31(Suppl):Ab5009.

[50] Shore ND, Basch E, Ryan CJ, Mulders P, Kheoh T, Fizazi K, et al. The impact of abiraterone acetate therapy on patient-reported pain and functional status in chemotherapy-naïve patients with progressive, metastatic castration-resistant prostate cancer: results from an updated analysis. *American Urological Association (AUA) Annual Meeting*. San Diego, CA, 2013.

[51] Mulders PFA. Updated interim analysis (IA) of COU-AA-302, a randomized phase III study of abiraterone acetate (AA) in patients (pts) with metastatic castration-resistant prostate cancer (mCRPC) without prior chemotherapy. *EAU Annual Congress*, 2013.

[52] Janssen Research & Development. A Phase 3, randomized, double-blind, placebo-controlled study of abiraterone acetate plus prednisone in asymptomatic or mildly symptomatic patients with metastatic castration-resistant prostate cancer (20 December 2010 and 20 December 2011 cutoffs). 2012.

[53] Basch E, Ryan CJ, Khech T, Fizazi K, Logothetis CJ, Rathkopf DE, et al. The impact of abiraterone acetate (AA) therapy on patient-reported pain and functional status in chemotherapy-naive patients with progressive, metastatic castration-resistant prostate cancer (mCRPC). *Ann Oncol* 2012;23(Suppl 9):295.

[54] Ryan CJ, Smith MR, de Bono JS, Molina A, Logothetis C, de Souza PL, et al. Interim analysis (IA) results of COU-AA-302, a randomized, phase III study of abiraterone acetate (AA) in chemotherapy-naive patients (pts) with metastatic castration-resistant prostate cancer (mCRPC). *J Clin Oncol* 2012;30(Suppl):LBA4518.

[55] de Bono JS, Logothetis CJ, Molina A, Fizazi K, North S, Chu L, et al. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med* 2011;364(21):1995-2005.

[56] Fizazi K, Scher HI, Molina A, Logothetis CJ, Chi KN, Jones RJ, et al. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol* 2012;13(10):983-92.

[57] Fosså SD, Aaronson N, Calais da Silva F, Denis L, Newling D, Hosbach G, et al. Quality of life in patients with muscle-infiltrating bladder cancer and hormone-resistant prostatic cancer. *Eur Urol* 1989;16(5):335-9.

[58] Albertsen PC, Aaronson NK, Muller MJ, Keller SD, Ware JE, Jr. Health-related quality of life among patients with metastatic prostate cancer. *Urology* 1997;49(2):207-16.

[59] Sandblom G, Carlsson P, Sennfalt K, Varenhorst E. A population-based study of pain and quality of life during the year before death in men with prostate cancer. *Br J Cancer* 2004;90(6):1163-8.

[60] Department of Health. *NHS reference costs 2012-2013 [Internet]*. London: Department of Health, 2013 [accessed 06.02.14] Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2012-to-2013

[61] Tilling K, Sterne JA, Rudd AG, Glass TA, Wityk RJ, Wolfe CD. A new method for predicting recovery after stroke. *Stroke* 2001;32(12):2867-73.

[62] Craig LE, Wu O, Gilmour H, Barber M, Langhorne P. Developing and validating a predictive model for stroke progression. *Cerebrovascular Diseases Extra* 2011;1(1):105-14.

[63] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Health* 2012;15(6):843-50.

APPENDIX 1: FURTHER CRITIQUE OF MANUFACTURERS SEARCHES

Clinical effectiveness

- The non-RCT (observational) filter duplicated search terms already present in the RCT filter: prospective study/exp, clinical trial/exp and (clinical NEXT/1 trial*).
- Field tags were inconsistently used in the Embase and Medline search strategy. It was not clear what database record fields were searched when field tags were not used.
- CAS Registry numbers for the interventions were not included in the search strategies.
- There was no animal/human limit.
- The Cochrane Library CENTRAL issue number was not reported.
- The MeSH term 'Prostatic Neoplasms' was unexploded in the Cochrane Library CENTRAL strategies, which meant that the narrower term in this MeSH tree was not searched: 'Prostatic Neoplasms, Castration-Resistant'.
- Field tags were not used in the Cochrane Library CENTRAL strategies, so it was not clear which database record fields were searched.
- The wildcard operator '?' was used in the Medline In-process search strategy, e.g. 'mito?antrone' and 'flur?blastin'. This wildcard operator does not work in PubMed, and would not retrieve any records searching for terms using the wildcard. The first update search strategy (Table 108) rectified this problem with an 'Error rectification facet' which replaced search terms that had used the wildcard, e.g. 'mitoxantrone OR mitozantrone'. This facet also corrected for the misspelling of 'onkotrone' and the missing term 'dihydroxyanthracenedione'. However, the Error rectification facet did not correct for two further terms, 'de?amethason*' and 'de?acort'. Further, the corrections were not included in the four subsequent update search strategies (apart from the correct spelling of 'onkotrone').
- Field tags were not used in the Medline In-process strategies.
- Single quotation marks were used for phrase searching in PubMed (e.g. 'prostate'), when double quotation marks should have been used (e.g. "prostate"). Using single quotes does not work, and would not retrieve records with the phrase searched for.
- The manufacturer did not supply website addresses or details of the search strategy or search terms used for the conference searches. However, further details were subsequently provided in the response to the ERG POC letter.
- There were no searches for unpublished and ongoing trials via trials registers, e.g. ClinicalTrials.gov and ICTRP.

Indirect and mixed treatment comparisons

Searches were not carried out as no indirect or mixed treatment comparisons were performed.

Non-RCT Evidence

The same searches for section 6.1 were used for this section; therefore the same comments applied as for clinical effectiveness searches (6.1).

Adverse events

The same searches for section 6.1 were used for this section; therefore no additional strategies were included for this section. Inclusion of a non-RCT (observational) search filter ensured that studies with adverse event data would be retrieved. However, it was possible that a search strategy without any study design filters, or using an adverse event specific filter, might have identified more adverse event data.

Cost-effectiveness

- Field tags were not used in the cost-effectiveness study design filter or the population facet of the Embase and Medline search strategies, so it was not clear which database record fields were searched.
- Numerous synonyms could have been included for 'castration resistant prostate cancer', e.g. 'androgen independent prostate cancer', 'hormone relapsed prostate cancer', 'hormone-refractory prostate cancer', etc.
- There were no subject heading terms (Emtree or MeSH) included for the named drugs in the interventions facet of the Embase and Medline search strategies. This could have resulted in missing relevant references.
- However, subject heading terms (Emtree) were included in the interventions facet for more generic interventions: 'corticosteroids', 'antiandrogens', 'gonadorelin agonist' and 'cancer immunotherapy'. These Emtree terms did not closely match any MeSH terms, and so would probably not have searched for the equivalent subject headings: 'Adrenal Cortex Hormones', 'Androgen Antagonists', 'Gonadotropin-Releasing Hormone'. Whilst there is no MeSH equivalent for 'cancer immunotherapy', an alternative would have been to have searched for 'Neoplasms/exp AND Immunotherapy/exp'. Further, no free text search terms were included for this set of generic intervention search terms.
- CAS Registry numbers for the interventions were not included in the search strategies.
- The trade name for abiraterone (Zytiga) was not included in the Embase/Medline or Cochrane Library search strategies.
- The search term 'megastrol' in line #70 should have used the more popular spelling 'megestrol'. Without using the correct spelling or any subject headings this search line only retrieved 9 records. It was unclear what impact this might have had on the overall recall of results.
- A number of drug names were not included in the Embase/Medline and Cochrane Library strategies, though they were included in the Medline In-process strategy: cyclophosphamide, satraplin and epothilone.
- Enzalutamide appeared in the strategies, but only as 'MDV3100' rather than 'enzulatimde' or 'Xtandi'.
- The Cochrane Library (NHS EED) issue number was not reported.
- Field tags were not used in the population facet of the Cochrane Library strategies, so it was not clear which database record fields were searched.

- The phrase 'prostate tumor' was searched for in the Cochrane Library strategy (search line #2) without using double quotation marks or an adjacency operator. The database would have searched for 'prostate AND tumor' instead. It appeared as though this phrase matched an Emtree subject heading, and might have been copied directly from an Embase strategy.
- The wildcard operator '?' was used in the Medline In-process search strategy, e.g. 'mito?antrone' and 'flur?blastin'. This wildcard operator does not work in PubMed, and would not retrieve any records searching for terms using the wildcard.
- Field tags were not used in the Medline In-process strategies.
- Single quotation marks were used for phrase searching in PubMed, when double quotation marks should have been used. Using single quotes does not work, and would not retrieve records with the phrase searched for.
- It was not clear why there were redundant search terms in the EconLit search strategy. 'Prostate' was used in search line #1, negating the need to include the search terms 'Prostate Cancer' and 'Prostate tumor' in search lines #2 and #6. This may be because the search terms used were subject heading terms, but this was unclear as no field tags or descriptors were reported.

Measurement and valuation of health effects

- The Emtree term 'Prostate cancer/exp' was missing from the prostate cancer facet of search terms in the embase.com search strategy (search line #1).
- There were six redundant search terms in the prostate cancer facet of search terms in the embase.com search strategy, e.g. 'neoplasm' was included as well as 'neoplasm$'.
- An incorrect truncation operator was used in the prostate cancer facet of search terms in the embase.com search strategy: $ was used instead of *.
- It was not clear whether the search facet for utilities/HRQoL was based on a published study design filter.
- The utilities/HRQoL facet included the misspelling 'multiattriattribute*'.
- There were no Emtree (or MeSH) subject headings in the utilities/HRQoL facet of search terms in the embase.com search strategy (search line #2).
- The Cochrane Library (NHS EED) issue number was not reported.
- There were no MeSH subject headings in the utilities/HRQoL facet of search terms in the Cochrane Library search strategy (search lines #2 to #5).

Resource identification, measurement and valuation

- The same searches for section 7.1 were used for this section; therefore the same comments applied as for cost effectiveness searches (7.1).

APPENDIX 2: PHILLIPS ET AL CHECKLIST

Question(s)	***Response (Y, N or NA)***	***Comments***
Is there a clear statement of the decision problem?	Y	
Is the objective of the evaluation and model specified and consistent with the stated decision problem?	Y (partially)	Partially. Final scope issued by NICE includes docetaxel as a comparator was not considered as a comparator by the manufacturer based on the reasoning that: "*Whilst both AAP and docetaxel are indicated for the treatment of mCRPC for adult men following ADT failure, AAP is indicated for the treatment of those in whom chemotherapy is not yet clinically indicated, the asymptomatic or mildly symptomatic patient … a patient population for whom chemotherapy is not yet clinically indicated*" (MS[1] pg.32)
Is the primary decision-maker specified?	Y	NICE
Is the perspective of the model stated clearly?	Y	NHS and Personal Social Services (PSS)
Are the model inputs consistent with the stated perspective?	Y	
Has the scope of the model been stated and justified?	Y	
Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	Y	
Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	Y	
Are the sources of data used to develop the structure of the model specified?	Y	
Are the causal relationships described by the model structure justified appropriately?	Y	
Are the structural assumptions transparent and justified?	Y	
Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	Y	
Is there a clear definition of the options under evaluation?	Y	
Have all feasible and practical options been evaluated?	Y	Although docetaxel is not considered as a comparator in the MS[1], the ERG feels this is sufficiently justified by the manufacturer (see section 3.3 of this report).

Question(s)	***Response (Y, N or NA)***	***Comments***
Is there justification for the exclusion of feasible options?	Y	It is not indicated for the treatment of those in whom chemotherapy is not yet clinically indicated.
Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	Y	The model type was a discrete event simulation (DES) evaluating the cost-effectiveness of AAP in comparison to BSC (PP) in adult men with mCRPC who were asymptomatic or mildly symptomatic after failure of ADT in whom chemotherapy was not yet clinically indicated. Although this may not be the most transparent approach possible, it is still considered adequate for addressing the decision problem at hand.
Is the time horizon of the model sufficient to reflect all important differences between options?	Y	
Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?	Y	
Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	Y	
Is the cycle length defined and justified in terms of the natural history of disease?	NA	No model cycles: not a Markov model
Are the data identification methods transparent and appropriate given the objectives of the model?	Y	
Where choices have been made between data sources, are these justified appropriately?	Y	
Has particular attention been paid to identifying data for the important parameters in the model?	Y	
Has the quality of the data been assessed appropriately?	Y	
Where expert opinion has been used, are the methods described and justified?	Y	
Is the data modelling methodology based on justifiable statistical and epidemiological techniques?	Y	Yes, in general, although the procedure for estimating prediction equations was not followed consistently

Question(s)	***Response (Y, N or NA)***	***Comments***
Is the choice of baseline data described and justified?	Y	
Are transition probabilities calculated appropriately?	N	The time to event was not estimated based the ITT population, rather a subset of patients was used"
Has a half-cycle correction been applied to both cost and outcome?	NA	Half-cycle correction are not relevant to discrete event simulation models.
If not, has this omission been justified?	NA	
If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?	N	Prediction equations were not based on ITT population, but on a subset
Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	Y	
Have alternative extrapolation assumptions been explored through sensitivity analysis?	Y	
Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	N	
Have alternative assumptions regarding the continuing effect of treatment been explored through sensitivity analysis?	N	
Are the costs incorporated into the model justified?	Y	
Has the source for all costs been described?	Y	
Have discount rates been described and justified given the target decision-maker?	Y	
Are the utilities incorporated into the model appropriate?	N (partially)	Since the majority of patients in the UK mCRPC Utility Study sample were on AAP treatment, an on-treatment utility decrement should be incorporated for BSC (post-docetaxel).
Is the source for the utility weights referenced?	Y	

Question(s)	***Response (Y, N or NA)***	***Comments***
Are the methods of derivation for the utility weights justified?	Y	
Have all data incorporated into the model been described and referenced in sufficient detail?	Y	
Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	N (partially)	Prediction equations for "Time to treatment discontinuation", "Time to treatment start" and "Time to death" should be consistently estimated and based on the ITT population
Is the process of data incorporation transparent?	Y	
If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	Y	
If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	Y	
Have the four principal types of uncertainty been addressed?	Y	
If not, has the omission of particular forms of uncertainty been justified?	NA	
Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	Y	
Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	Y	
Has heterogeneity been dealt with by running the model separately for different subgroups?	Y	
Are the methods of assessment of parameter uncertainty appropriate?	Y	
If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	Y	
Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	Y	
Are any counterintuitive results from the model explained and justified?	NA	

Question(s)	***Response (Y, N or NA)***	***Comments***
If the model has been calibrated against independent data, have any differences been explained and justified?	NA	
Have the results of the model been compared with those of previous models and any differences in results explained?	N	

NA=Not Applicable