in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Axitinib for the treatment of advanced renal cell carcinoma after failure of prior systemic treatment

ERRATUM

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

The table below lists the page to be replaced in the original document and the nature of the change:

Page nr:	**Change:**
1	Title: 'systematic' replaced by 'systemic'
2	Same in 'This report should be referenced as follows:'
11	Table 1.1 amended and CiC underlined.
13	CiC underlined, reference to error in MS deleted and 'STA' replaced by 'STC'.
16	CiC underlined, and reference to error in MS amended.
34	Table 4.4: CiC underlined
35	Same
36	Same
37	Same
44	Reference to error in MS amended.
45	Same
56	Paragraph removed
79	Reference to error in MS deleted and text about combining utilities from axitinib and sorafenib arm amended
102	Remark about face validity of the BSC OS clarified.
102-114	Various references to error in MS deleted (including the deletion of two tables) plus new tables and graphs included. (New pages are now 102-112)
109	Redundant remark about distribution of uncertainty was removed
118	Reference to error in MS deleted.
138	'Apixaban' replaced by 'Axitinib'

in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Axitinib for the treatment of advanced renal cell carcinoma after failure of prior systemic treatment

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

Authors
Rob Riemsma, Reviews Manager, KSR Ltd
Maiwenn Al, Health Economics Researcher, EUR
Isaac Corro Ramos, Health Economics Researcher, EUR
Sohan Deshpande, Systematic Reviewer, KSR Ltd
Gill Worthy, Statistician, KSR Ltd
Caro Noake, 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 Ltd
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 16/10/2012

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 11/87/01 STA.

Declared competing interests of the authors
None.

Acknowledgements
The authors are grateful to Saskia de Groot (EUR) for her input in the clarification phase.

Commercial in confidence data are redacted 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, Al M, Corro Ramos I, Deshpande S, Worthy G, Noake C, Armstrong N, Severens JL, Kleijnen J. Axitinib for the treatment of advanced renal cell carcinoma after failure of prior systemic treatment: a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd, 2012.

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. Maiwenn Al acted as health economic project lead, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Isaac Corro Ramos and Nigel Armstrong acted as health economists on this assessment, critiqued the manufacturer's economic evaluation and contributed to the writing of the report. Sohan Deshpande acted as systematic reviewer, 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. Caro Noake critiqued the search methods in the submission and contributed to the writing of the report. Johan L Severens critiqued the manufacturer's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the manufacturer's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Table 1.1: Main outcomes from the AXIS trial (axitinib versus sorafenib)

	All patients (N=361/362)	Cytokine refr. (N=126/125)	Sunitinib refr. (N=194/195)
Overall Survival	HR=0.969 (0.800, 1.174)[1]	HR=0.813 (0.555, 1.191)	HR=0.997 (0.782, 1.270)
PFS	HR=0.665 (0.544, 0.812)[1]	HR=0.464 (0.318, 0.676)[2]	HR=0.741 (0.573, 0.958)[2]
Response	RR=2.056 (1.408, 3.003)[3]	RR=2.392 (1.434, 3.992)[3]	RR=1.477 (0.792, 2.754)[3]
HRQoL (EQ-5D)	NR	Post-Treatment difference: [redacted]	Post-Treatment difference: [redacted]
AEs (grade 3-4)	(N=359/355)[4,5]	(N=126/123)[4,6]	(N=190/190)[4,6]
- Death (all)	31.5% vs 30.7%	NR	NR
- Death due to AE	9.5% vs 6.8%	NR	NR
- Any AE	48.2% vs 52.4%	[redacted]	[redacted]
- Diarrhoea	10.0% vs 7.1%		
- Hypertension	15.6% vs 11.0%		
- Fatigue	9.8% vs 3.7%		
- PPE	5.0% vs 16.1%		

AE=Adverse events, NR=Not reported, PPE=Palmar-plantar erythrodysaesthesia syndrome
[1]) Adjusted for ECOG PS and prior treatment regimen; [2]) Adjusted for ECOG PS; [3]) Risk Ratio for overall confirmed ORR (CR and PR); [4]) Safety population: All patients who received at least one dose of study medication. [5]) Results as reported in the MS. [6]) Results as reported in the response to the clarification letter.

Axitinib versus BSC in cytokine refractory patients:

- For overall survival the Hazard Ratio (HR) for axitinib versus placebo was 0.63 (95% CrI 0.41-0.99). This is based on the ITT population, censored for cross-over in the TARGET trial
- For PFS, the HR for axitinib versus placebo was 0.251 (95% CrI 0.165-0.379)
- Other outcomes, such as response, quality of life and adverse events were not reported.

Axitinib versus BSC in sunitinib refractory patients:

- For overall survival using the ITT RECORD-1 placebo cohort adjusted for cross-over using the RPSFT method, the estimated median OS was 36 weeks (8.3 months) for axitinib-like patients assuming that they received placebo.
- The estimated OS HR for axitinib versus BSC was 0.619 (95% CI 0.384-0.997) using data from a Swedish patient registry.
- For PFS using the ITT RECORD-1 placebo cohort, the estimated median PFS was 6.9 weeks (1.6 months) for axitinib-like patients if they had received placebo.
- Other outcomes, such as response, quality of life and adverse events were not reported.

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

There is no direct evidence for the comparison axitinib versus BSC. Instead the submission relies on an indirect comparison for cytokine refractory population, using evidence from two RCTS and a simulated treatment comparison for the sunitinib refractory population, using evidence from single treatment arms from two trials.

For the indirect comparison in the cytokine refractory population comparison of trial populations used in the analyses is not possible, because patient characteristics are not reported for cytokine refractory patients separately in both trials. Another problem for this comparison is concern about the results for overall survival in the TARGET trial due to treatment switching from placebo to sorafenib at the point of disease progression.

For the sunitinib refractory population, the evidence relies on a simulated treatment comparison, this comparison is not based on randomised treatment allocation, but on a comparison of two single

axitinib PFS, and a Gompertz model to extrapolate axitinib OS. None of the scenarios explored by the manufacturer provided an ICER below £50,000.

For the sunitinib refractory subgroup, the manufacturer concluded that the model is robust to the majority of structural assumptions made. Most of the scenarios examined, including extrapolation assumptions, the STC approach, and source of utility estimate, produced ICERs lower than £50,000/QALY (with PAS). The exception was the method of OS comparison used that produced quite different results ranging from values very close to the base case ICER to a maximum difference larger than £20,000. In particular, the use of the RENCOMP model analysis resulted in ICERs higher than the base case (ranging from between £43,384 and £56,113 with PAS). According to the manufacturer, this indicates that the incremental survival benefit assumed over BSC is a key driver of the model result.

The PSA results with PAS showed a 31% probability that the ICER is below £50,000 per QALY for the cytokine refractory population whereas this was 67% for the sunitinib refractory population. Without PAS, these percentages were [redacted] for the cytokine refractory population and [redacted] for the sunitinib refractory population.

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

The economic model described in the MS is considered by the ERG to meet the NICE reference case[5] and is in-line with the decision problem specified in the scope.

The ERG assessment indicated that the model was generally well presented and reported. Some minor issues were identified regarding the use of standard deviations in the PSA instead of standard errors; whilst these errors have a clear impact on the outcomes of the univariate and probabilistic sensitivity analyses, they do not impact the estimate of the central ICER.

After rerunning the PSA with the correct standard errors, we found that with PAS there is a 19% probability that the ICER is below £50,000 per QALY for the cytokine refractory population whereas this was 83% for the sunitinib refractory population. Without PAS, these percentages were [redacted] for the cytokine refractory population and [redacted] for the sunitinib refractory population.

Various assumptions around the utility estimates were discussed, and some extra scenario analyses showed that changing these assumptions had very minimal impact of the ICERs.

The manufacturer performed the univariate sensitivity analysis by varying all parameters between plus and minus 20%. This is often not very informative, since this 20% may be either a under- or over-estimate of the true uncertainty. Thus, the ERG performed a univariate sensitivity analysis in which parameters were varied between the limits of their 95% confidence interval (as defined for the PSA). This revealed that for the cytokine refractory subgroup, the ICER is extremely sensitive to changes in the HR for the overall survival. At the upper limit of the 95% confidence interval, the ICER would amount to almost £400,000 (with PAS). This is in sharp contrast with the results for the sunitinib refractory subgroup, where changes to input parameters lead to modest changes in the ICER. This is related to the fact that no measures of uncertainty were provided for the STC adjustment factor for the BSC arm, which means that this uncertainty was not considered in the univariate (or probabilistic) sensitivity analysis.

- *"In contrast to AXIS, where all patients included in the study were required to have progressed on first-line therapy by RECIST-defined criteria, in the overall RECORD-1 population, 14% of patients (n=58) discontinued previous TKI therapy because of unacceptable toxicity. Among the subgroup of 58 patients who were intolerant to previous TKI therapy, 45 patients and 13 patients were randomly assigned to everolimus and placebo, respectively. Thus, patients in the RECORD-1 study could have discontinued prior treatment due to intolerance and therefore results would be more reflective of a first-line study." (Source: Pfizer submission, Section 6.7.11, page 108)*
- *"In contrast to the AXIS study, where patients were required to have received only one prior therapy (sunitinib or a cytokine, or bevacuzimab + interferon-α or temsirolimus), patients in the RECORD-1 study were allowed to have received more than one previous therapy and could have been treated with sunitinib or sorafenib, as well as a cytokine in some cases. (Source: Pfizer submission, Section 6.7.2, page 82)*

Regarding this last issue the manufacturer states:

"However, the median OS and patient characteristics have never been reported for those patients in the BSC arm that had progressed on sunitinib after receiving only one line of therapy. The closest available patient populations reporting overall survival data to allow the STC comparison were the ITT BSC population (corrected for crossover using the RPSFT method) and patients receiving everolimus treatment with only prior sunitinib therapy." (Source: Pfizer submission, Section 6.7.11, page 108)

All these issues regarding the clinical effectiveness carry over into the cost effectiveness evaluation.

1.7 Key issues

Overall, the manufacturer's submission is clearly presented and the analyses and underlying assumptions are sound and reasonable. The main issue with this submission is whether a simulated treatment comparison (STC) presents a valid and reliable estimate of the clinical effectiveness of axitinib versus BSC in a sunitinib refractory population. As there is no direct trial evidence it is not possible to compare the results of the STC to any existing evidence so the accuracy and reliability of the results cannot be ascertained. In addition, the uncertainty around the STC results is missing.

The ERG found one error in the description of the data used in the indirect comparison (cytokine refractory population) but the correct data were used in the analyses. Therefore, this did not affect the conclusions. The methods for the STC (sunitinib refractory population) follow the recommended method and it appears to be well-conducted and clearly explained, including considerations of the limitations of the method and available trial evidence. However, the STC is basically a comparison of individual treatment arms and is open to considerable bias. There is no way to assess whether or not the final result is biased.

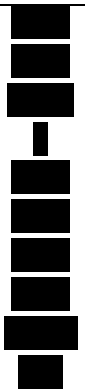
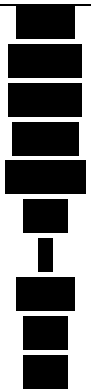
In addition, the reliability of the results of the included studies and the comparability of the trials included are potential issues. These have been adequately reported in the manufacturer's submissions, and have been summarised in this ERG report.

In conclusion, the ICER for axitinib in a cytokine refractory population is based on a well performed indirect comparison. The results seem reliable, and present an ICER (£65,326 with PAS and [redacted] without PAS) with large confidence intervals due to considerable uncertainty. The ICER for axitinib in a sunitinib refractory population (£40,933 with PAS and [redacted] without PAS) is based on an STC which did not included the uncertainty surrounding the estimates used in the analyses, and is basically a comparison of individual treatment arms and is therefore open to considerable bias.

Table 4.2: Results of the three included trials: AXIS, TARGET and RECORD-1

	AXIS		**TARGET**		**RECORD-1**	
	Axitinib	**Sorafenib**	**Sorafenib**	**Placebo**	**Everolimus**	**Placebo**
All Patients	**N=361**	**N=362**	**N=451**	**N=452**	**N=272**	**N=139**
Overall survival	(95% CI)					
- Death	211 (58%)	214 (59%)	171 (38%)	196 (43%)[j]	NR	NR
- Median time to	NR	NR	19.3 m	15.9 m[j]	64.1 wks (14.8 m)	53.4 wks (14.4 m)[f]
- Hazard ratio	axitinib vs sorafenib:	0.969 (0.800, 1.174)[a]	sorafenib vs placebo:	0.77 (0.63, 0.95)[j]	everolimus vs placebo	0.87 (0.65, 1.15)[f]
Progression free	survival (95% CI)					
- PFS event	192 (53%)	210 (58%)	NR	NR	193 (71%)	109 (79%)[g]
- Median time to	6.7m (6.3, 8.6)	4.7m (4.6, 5.6)	5.5 m	2.8 m[j]	4.9 m (4.0, 5.5)	1.9 m (1.8, 1.9)
- Hazard ratio	axitinib vs sorafenib:	0.665 (0.544, 0.812)[a]	sorafenib vs placebo:	0.51 (0.43, 0.60)[j,k]	everolimus vs placebo	0.33 (0.25, 0.43)
Response rates						
- OR event	19.4% (15.4, 23.9)	9.4% (6.6, 12.9)	10% (7, 13)	2% (1, 4)[l]	1.8%	0%
- Duration	11 m (7.4, ne)	10.6 m (8.8, 11.5)	182 days (36, 378)	NR	NR	NR
- OR risk ratio	axitinib vs sorafenib:	2.056 (1.408, 3.003)	NR	NR	NR	NR
Health-related	quality of life					
- TTD[b]	Hazard ratio (95%	0.829 (0.701, 0.981)	NR	NR	NR	NR
- FKSI-DRS	CI):	0.12 (-0.45, 0.69)	NR	NR	Post-Tx difference:	0.82 (0.57, 1.18)
- EQ-5D	Post-Tx difference: NR	NR	NR	NR	NR	NR
Adverse events	All grades[m] Grade 3/4	All grades Grade 3/4	All grades Grade 3/4	All grades Grade 3/4	All grades Grade 3/4	All grades Grade 3/4
- stomatitis	20% 1%	32% 4%	NR	NR	44% 4%	8% 0%
- rash	13% <1%	32% 5%	40% 1%[h]	16% <1%[h]	29% 1%	7% 0%
- fatigue	39% 11%	14% 3%	37% 5%	28% 4%	31% 5%	27% 3%
- asthenia	21% 5%	53% 7%	NR	NR	33% 3%	23% 4%
- diarrhoea	55% 11%	29% 11%	43% 2%	13% 1%	30% 1%	7% 0%
- hypertension						

- decr. appetite	40% 16%	29% 4%	17% 4%	2% <1%	NR	NR
- nausea	34% 5%	22% 1%	NR	NR	NR	NR
- dysphonia	32% 3%	14% 0%	23% <1%	19% 1%	26% 1%	19% 0%
- hypothyroidism	31% 0%	8% 0%	NR	NR	NR	NR
- PPE	19% <1%	51% 16%	NR	NR	NR	NR
- alopecia	27% 5%	32% 0%	30% 6%[i]	7% 0%[i]	NR	NR
- infections	4% 0%	NR	27% <1%	3% 0%	NR	NR
- non-inf pneum.	NR	NR	NR	NR	37% 10%	18% 1%
	NR		NR	NR	14% 4%	0% 0%
First-line cytokine	126 (35%)	125 (35%)	374 (83%)	368 (81%)	--	--
Overall survival	(95% CI)					
- Death	51 (41%)	57 (46%)	NR	NR		
- Median time to	NR	NR	NR	NR		
- Hazard ratio	axitinib vs sorafenib:	0.813 (0.555, 1.191)	sorafenib vs placebo:	0.78 (0.62, 0.97)[d]		
Progression free	survival (95% CI)					
- PFS event	50 (40%)	69 (55%)	NR	NR		
- Median time to	12.1m (10.1, 13.9)	6.5m (6.3, 8.3)	NR	NR		
- Hazard ratio	axitinib vs sorafenib:	0.464 (0.318, 0.676)	sorafenib vs placebo:	0.44 (0.35, 0.55)[c]		
Response (95%	CI)					
- OR event	32.5% (24.5, 41.5)	13.6% (8.1, 20.9)	NR	NR		
- Duration	11m (7.4, ne)	10.6 (5.9, 11.5)	NR	NR		
- OR risk ratio	axitinib vs sorafenib:	2.392 (1.434, 3.992)	NR	NR		
Health-related	quality of life					
- TTD[b]	NR	NR	NR	NR		
- FKSI-DRS[n]	Post-Tx difference:	[redacted]	NR	NR		
- EQ-5D[n]	Post-Tx difference:	[redacted]	NR	NR		
Adverse events	TEAEs[o]	TEAEs[o]	NR	NR		
- stomatitis	[redacted]	[redacted]				
- rash	[redacted]	[redacted]				

- fatigue - asthenia - diarrhoea - hypertension - decr. appetite - nausea - dysphonia - hypothyroidism - PPE - alopecia - infections - non-inf pneum.	 NR NR	 NR NR				
First-line sunitinib	194 (54%)	195 (54%)	--	--	127 (47%)	139*
Overall survival - Death - Median time to - Hazard ratio	(95% CI) 131 (68%) 65.9 wks (15.2 m) axitinib vs sorafenib:	 131 (67%) NR 0.997 (0.782, 1.270)			 NR 54.4 wks (12.6 m) NR	 NR 43.4 wks (10.0 m) NR
Progression free - PFS event - Median time to - Hazard ratio	survival (95% CI) 117 (60%) 4.8m (4.5, 6.4) axitinib vs sorafenib:	 120 (62%) 3.4m (2.8, 4.7) 0.741 (0.573, 0.958)			 NR 16.9 wks (3.9 m) everolimus vs placebo	 NR 7.8 wks (1.8 m) 0.34 (0.23, 0.51)[e]
Response rates - OR event - Duration - OR risk ratio	 11.3% (7.2, 16.7) 11m (5.2, ne) axitinib vs sorafenib:	 7.7% (4.4, 12.4) 11.1 (ne, ne) 1.477 (0.792, 2.754)			 NR NR NR	 NR NR NR
Health-related - TTD[b] - FKSI-DRS[n,o]	quality of life NR Post-Tx difference:	 NR 			 NR NR	 NR NR

- EQ-5D[n,o]	Post-Tx difference:	[redacted]			NR	NR
Adverse events	TEAEs[o]	TEAEs[o]			NR	NR
- stomatitis	[redacted]	[redacted]				
- rash	[redacted]	[redacted]				
- fatigue	[redacted]	[redacted]				
- asthenia	[redacted]	[redacted]				
- diarrhoea	[redacted]	[redacted]				
- hypertension	[redacted]	[redacted]				
- decr. appetite	[redacted]	[redacted]				
- nausea	[redacted]	[redacted]				
- dysphonia	[redacted]	[redacted]				
- hypothyroidism	[redacted]	[redacted]				
- PPE	[redacted]	[redacted]				
- alopecia	[redacted]	[redacted]				
- infections	NR	NR				
- non-inf pneum.	NR	NR				

The indirect comparison was performed using Bayesian Markov Chain Monte-Carlo sampling to determine the relative efficacy of the treatments. Sampling was performed using WinBUGS. A fixed effects model was used due to the limited availability of relevant data for use in the model. In this case because hazard ratios entered to the model and not individual treatment effects, the approach assumes that the relative treatment effect (i.e. HR) for one treatment pair is the same across all trials. Since there was only one trial per pairwise HR, this assumption was appropriate in this analysis. Non-informative prior distributions were used. A non-informative prior assumes that all possible The WinBUGS code for the fixed-effects model is provided in Section 10.14 (Appendix 14 of the Manufacturer's Submission).

Point estimates of the HR for each pair of treatments along with 95% credible intervals (CrI) were calculated from 5,000 simulated draws from the posterior distribution after a burn-in of 20,000 iterations. (MS, Section 6.7.5, page 94)

The data used in the indirect comparison are presented in table 4.6.

Table 4.3: Input data

	AXIS (axitinib vs sorafenib)	**TARGET (sorafenib vs placebo)**
	HR (95% CI)	
PFS (IRC)	***Cytokine refractory population:*** 0.464 (0.318-0.676)	***ITT population:*** 0.44 (0.35-0.55)
OS	***Cytokine refractory population:*** 0.813 (0.555-1.191)	***ITT population censored for cross-over:*** 0.78 (0.62-0.97)

Abbreviations: CI, confidence interval; HR, hazard ratio; IRC, independent review committee; ITT, intent-to-treat; OS, overall survival; PFS, progression-free survival.

The results are presented in table 4.7.

Table 4.4: Results of the indirect comparison in the cytokine refractory population

Treatment comparison	**Median HR**	**95% CrI**
PFS Axitinib vs placebo	0.251	0.165-0.379
PFS Axitinib vs placebo*	0.203*	0.132-0.318*
OS Axitinib vs placebo	0.63	0.41-0.99

* Corrected analysis by ERG.

Comment

The code and data used in the mixed treatment comparisons were checked by the ERG group. An error was found in the description of trial data used in one of the analyses (progression free survival in the prior cytokine group) but the other two analyses were correct (overall survival in the prior cytokine group, and overall survival in the prior sunitinib group). Given the small number of trials included in the analyses the use of fixed effect models was appropriate.

The data used in the indirect comparison for progression free survival is reported in table 17 on page 93 of the manufacturer's submission. This gives a HR of 0.44 (0.35 to 0.55) for PFS from the TARGET trial. This gives a ln HR and SE of -0.82098055 (0.1153). However, the data used in the manufacturer's model for this trial was -0.616186139 (0.090) which is ln 0.54. In the Factual Error Check the manufacturer explained that the relevant HR from the TARGET study for the sorafenib vs. placebo comparison in the cytokine-refractory population is 0.54 (0.45–0.64). This HR was taken from a paper by Negrier et al. (Negrier S, Jager E, Porta C, McDermott D, Moore M, Bellmunt J, et

al. Efficacy and safety of sorafenib in patients with advanced renal cell carcinoma with and without prior cytokine therapy, a subanalysis of TARGET. Med Oncol. 2010 Sep;27(3):899-906). This paper was not mentioned in the manufacturer's submission.

Simulated treatment comparison

As there was no direct or indirect trial evidence the manufacturer performed a simulated treatment comparison (STC). This compared PFS and OS for axitinib vs. everolimus and best standard care using data from the AXIS and RECORD-1 trials.

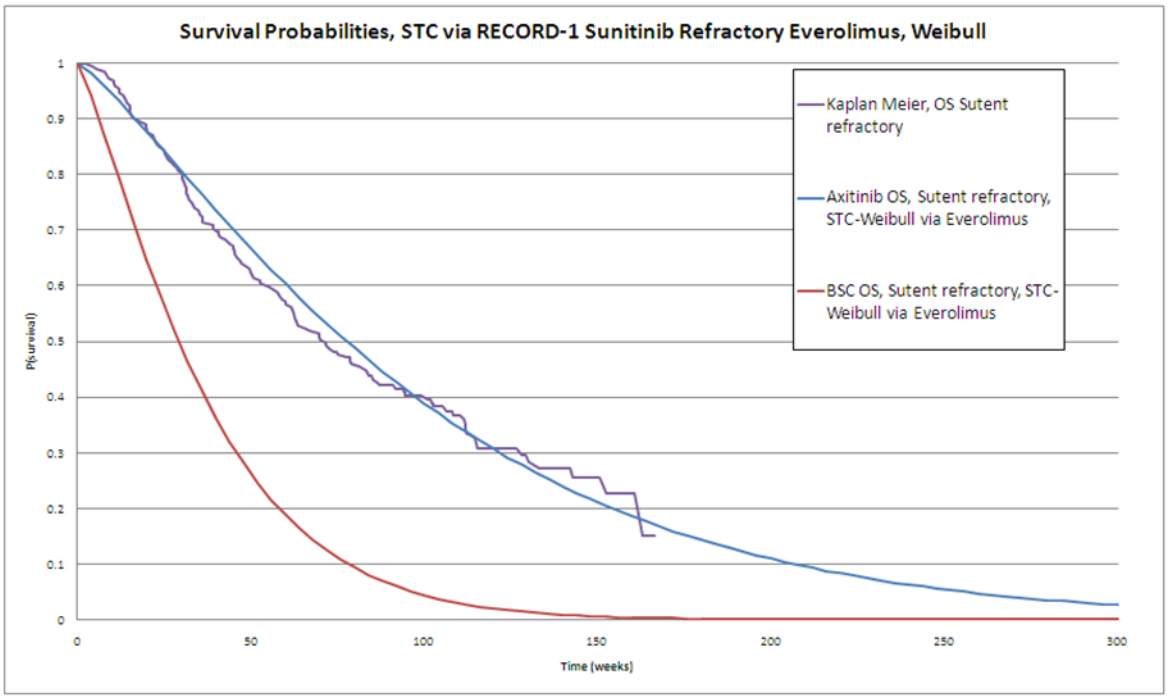
> ***Simulated treatment comparison***
>
> *Simulated treatment comparisons (STCs[27] is a novel technique to derive indirect comparisons between competing treatments (say A and B). Unlike mixed treatment comparisons (MTCs) which provide an average measure of the difference between A and B across all studies, STCs aim to answer a more specific question: what difference could we expect if A and B had been compared in the same trial.*
>
> *STCs rely on individual patient data (IPD) for the treatment from an index trial (e.g., one used as the basis of a submission), and summary data (usually published reports) for the competitor from one or more studies. The studies for the treatment being compared must be generally compatible in terms of the type of population included, measurement methods, timeframe of observation, reporting of information, etc. The studies are not required to be exactly identical in these dimensions, but there must be sufficient overlap so that findings from one study can be assumed to be applicable in the setting of the other trial.*
>
> *Even with close compatibility between the studies, it is unlikely that the characteristics of the patients will be identical, so that comparisons of outcomes between the trials may be confounded by these differences. STCs are specifically designed to adjust for these differences. This is done by using the index trial data to build a predictive equation for each endpoint for which a comparison is desired. We can denote this equation in a general way as having the following form:*
>
> $$\mu = X\beta$$
>
> *where μ represents some parameterization of the outcome variable. For instance, if the outcome of interest is a time-to-event variable like PFS or OS, μ would be the scale parameter in a parametric survival model; X represents a vector of predictors of the outcome and β represents the corresponding coefficients. We note that X may include an indicator for study group, and correspondingly, β would include a treatment effect coefficient. In some applications, the equation may be built from a single (e.g., experimental) treatment; in oncology trials, this may be done when outcomes in the reference arm is biased due to crossover, for example. For the explanations that follow, it is assumed that the equation is built from the primary treatment arm (i.e., A in the current notation).*
>
> *The STC then proceeds with following steps:*
>
> 1. *If the comparator treatment (B) had been included in the index trial the equation would have included a term for a comparison of A vs. B, as follows:*
>
> $$\mu = X\beta + \delta_{B\ vs.\ A} Z_{B\ vs.\ A}$$
>
> *where δ is a coefficient representing the effect of B compared to A (e.g., expressed as a log hazard ratio), and Z is an indicator of treatment group.*

Calibrated OS for axitinib-like patients - RECORD-1 prior sunitinib everolimus patients

For the RECORD-1 prior sunitinib everolimus patients, the derived adjustment factor from the lognormal distribution was -0.37, corresponding to a median OS of 46 weeks (10.6 months). The derived adjustment factor from the Weibull distribution was -0.46, corresponding to a predicted median OS of 45.4 weeks (10.5 months).

Figure 4.7 displays the survival probabilities from the Weibull distribution.

Figure 4.1: Weibull OS distribution via RECORD-1 prior sunitinib everolimus + RPSFT HR if both treatments had been included in AXIS RCT *(Source: Pfizer Submission, page 107)*



Abbreviations: BSC, best supportive care; OS, overall survival; STC, simulated treatment comparison.

To create a modelled placebo arm for the everolimus prior sunitinib population, the RPSFT-adjusted OS hazard ratio from the RECORD-1 study (0.53) was applied to the AXIS-like everolimus curve to generate a modelled AXIS-like, sunitinib refractory placebo curveAs the lognormal model does not support the application of a hazard ratio, the Weibull was the only option explored in the model. Figure 8 displays the survival probabilities calculated using the Weibull distribution.

Table 4.14 presents a summary of the STC results for OS for the sunitinib-refractory population.

treatment withdrawal), but that does not allow for assessing the impact of a smaller or larger percentage of treatment discontinuation. In the current model, such scenario analysis would only impact the costs, as more treatment withdrawal leads to lower treatment costs while still achieving the same effects, thus leading to lower ICERs.

Thus, the ERG feels that ideally, a model would make the impact of treatment withdrawal explicit, not only for the costs, but also for the time-to-event. However, we recognize that due to lack of data, this will often not be possible.

5.2.7 Health related quality of life

Base case – AXIS study

The manufacturer collected utility data in the AXIS trial using the EuroQoL-5D (EQ-5D) instrument, completed by the study patients at Day one, every four weeks afterwards, at the end of study treatment or withdrawal and at follow up Day 28 (28 days after the last dose of active treatment). The quality of life analysis was based on the ITT population (the full analysis set).

The baseline mean (SD) EQ-5D score (Day one of Cycle one) for the axitinib arm was 0.732 (0.01). The mean on-treatment utility for axitinib was calculated by averaging the EQ-5D index value at each time point, weighted by the number of patients still on treatment at that time point, giving a mean (SD) utility of 0.692 (0.275). Finally the progressive disease utility was based on the average end-of-treatment utility, giving a mean (SD) utility of 0.610 (0.316). Note that end-of-treatment utility estimates were based on the axitinib and sorafenib groups together.

The MS highlights the absence of relevant literature on utility values for treatment with BSC. Therefore, the manufacturer made the assumption that BSC patients would experience the same utility as patients receiving active treatment with axitinib. This assumption, tested and confirmed by the main clinical advisor for the economic model, was argued as follows: while patients with axitinib may expect to experience some reduction in health-related quality of life related to the treatment, they will also receive HRQoL benefit in terms of symptomatic control and disease stabilization.

Scenario analysis – previous NICE utility estimates

The manufacturer also carried out a scenario analysis with the utility figures used in several previous NICE appraisals to model second-line mRCC. As these utility estimates are based on consensus between UK experts, the NICE appraisal committee and ERG groups from several appraisals, they allow for "like versus like" comparability between axitinib and other previous advanced/mRCC appraisal.

Adverse events

The MS states that the HRQL estimates included in the AXIS trial reflect the adverse event profile associated with axitinib. Therefore, the utility estimates included in the economic model are expected to reflect the adverse event profile of the treatment. Thus, no specific utilities were included to model adverse events.

Quality-of-life data used in cost-effectiveness analysis

A summary of the QoL values used in the economic analysis is presented in Table 5.11.

and patient opinion as an overestimate of the true BSC overall survival. This was further supported by the systematic review carried out to examine BSC survival post-sunitinib failure (see MS Section 6.7.10) where the majority of the estimates found were in the four to six month range, substantially lower than the RECORD-1 estimate (10 months). Furthermore, this result was consistent with the 5.8 months median OS observed in the BSC arm of the RENCOMP study. These results can be seen as a gauge of the face validity of the model.

Comment

The spreadsheet attached to the clarification letter is not of the form of a checklist as the ERG had expected, i.e. a predefined list of tests to be performed with expected outcomes, to thoroughly test the technical integrity of the model. Instead of that, it shows a post-testing list of bugs found in the model and the corresponding action to fix them. Thus, the ERG cannot assess the quality of the validation process from the model validation spread sheet, i.e. it is not clear how extensively the model was reviewed.

The ERG is of the opinion that the fact that the median estimates predicted by the model are within the 95% confidence intervals of the AXIS trial estimates does not ensure that the model results are comparable to those observed in the clinical trial results. Results could have been more comparable if model results were also accompanied by their corresponding 95% confidence interval. Medians could be judged to be similar in magnitude but this would be a subjective assessment since no formal test for the difference in median was provided. However, using this subjective assessment, the model seems to reasonably reproduce the trial results.

Regarding the face validity of the BSC results, the comparisons made by the manufacturer seem to indicate that for most studies, BSC OS was smaller than that found in this study. This would indeed imply that the overall survival for BSC is overestimated and therefore the OS benefit for axitinib is conservative. But given the higher median OS found in the everolimus appraisal,[47] it cannot be stated with certainty that the current estimate is conservative.

5.3 Additional work undertaken by the ERG

5.3.1 ERG base case

As mentioned in Section 5.2.7 and in Section 5.2.8.1, the SD for the base case utilities used in Table 5.11 (0.275 and 0.316) and the SD for the base case relative dosing intensity used in Table 5.13 (35.2%) do not reflect the true parameter uncertainty since the standard errors (0.0035 and 0.0175 for the utilities and 1.86% for the relative dosing intensity) have to be considered. Moreover, as explained in Section 5.2.8, the cost of death should be considered uncertain with a mean value equal to £3,923 and SE equal to 104.43. With these new values the ERG defined a new ERG base case sensitivity analysis.

[Table 5.23 removed]

The tornado plots presented in figure 5.24 to 5.27 are markedly different from those presented by the manufacturer (see figure 5.8 to 5.15). This is due to the fact that here all parameters have been varied between the lower and upper limit of their 95% confidence interval, rather than between plus and minus 20%. This is especially clear in the cytokine refractory subgroup, where varying the HR for the OS of axitinib versus BSC leads to an ICER up to £423,083 when the upper limit is 0.99.

Figure 5.24: ERG univariate sensitivity analysis – Sunitinib refractory population (with PAS)

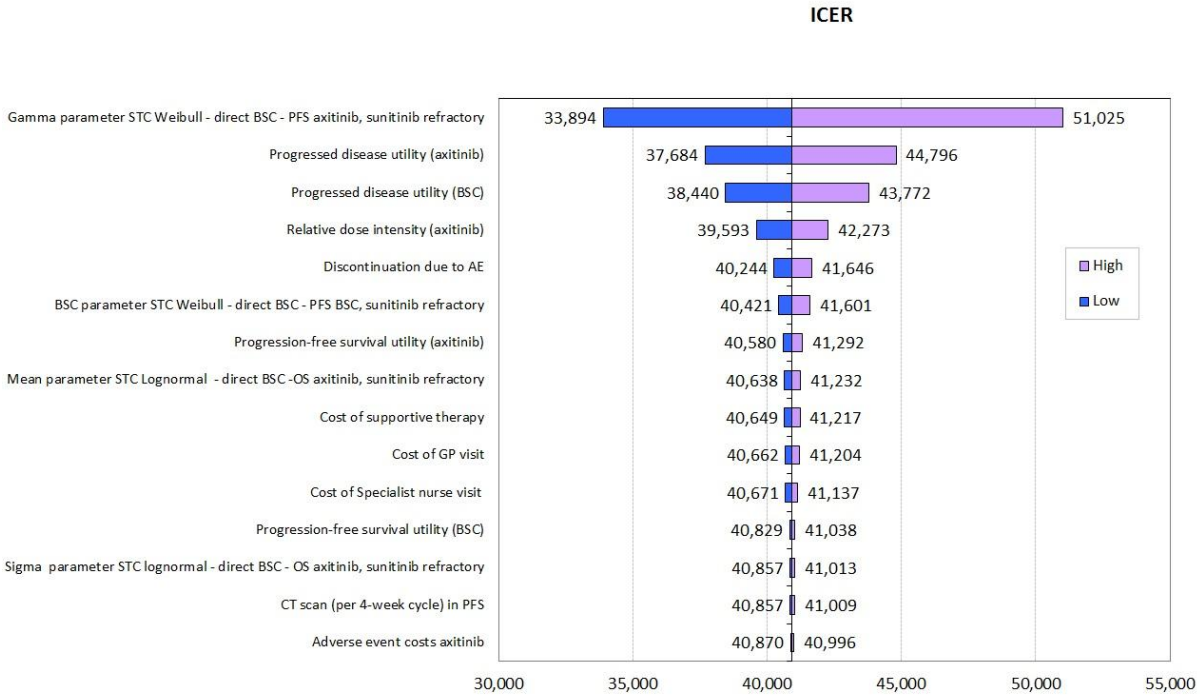


Figure 5.25: ERG univariate sensitivity analysis – Sunitinib refractory population (without PAS)

[redacted]

Figure 5.26: ERG univariate sensitivity analysis – Cytokine refractory population (with PAS)

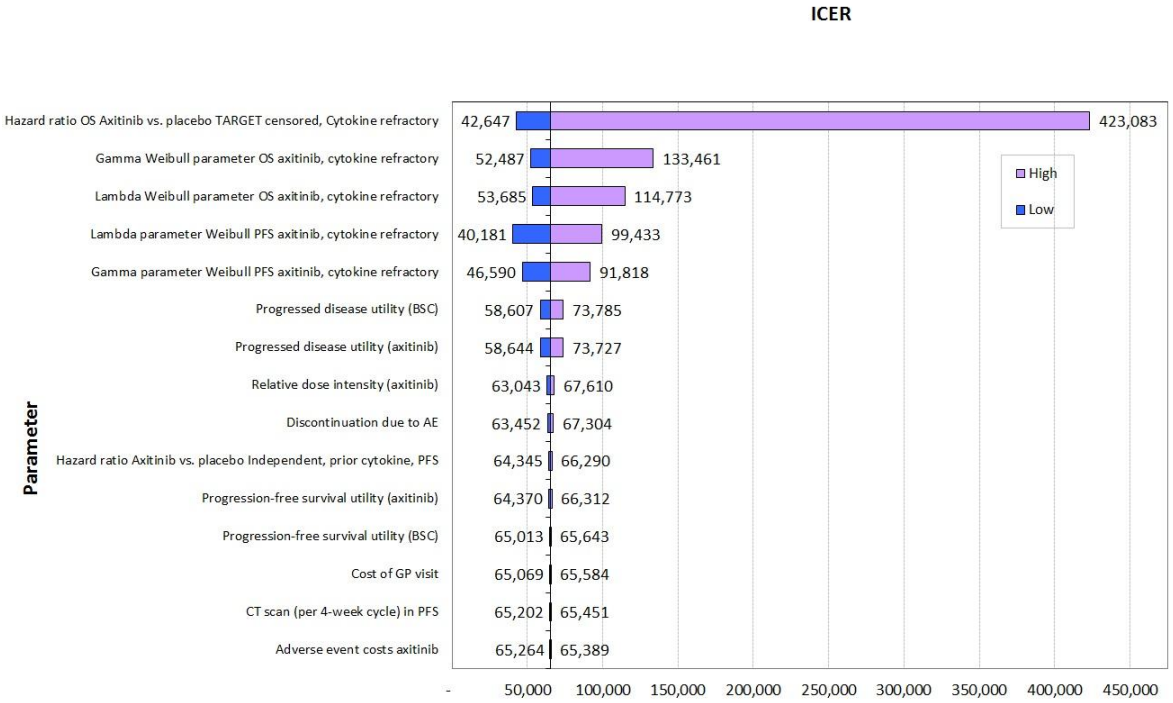


Figure 5.27: ERG univariate sensitivity analysis – Cytokine refractory population (without PAS)

[redacted]

[Table 5.24 removed]

Probabilistic Sensitivity Analysis (cytokine refractory and sunitinib refractory subgroups)
PSA results are first presented with PAS. In this case, for a cost-effectiveness threshold of £50,000 per QALY, axitinib was cost-effective for 20% of the observations (out of 1,000 generated in the PSA) for the cytokine refractory population and 83% for the sunitinib refractory population, respectively (see figures 5.28 and 5.30 below). Note that the manufacturer reported that axitinib was cost-effective for 31% of the observations for the cytokine refractory subgroup and for 67% of the observations for the sunitinib refractory population.

As mentioned in the beginning of Section 5.3, the base case utilities will experience a reduction in their uncertainty since their SD (0.275 and 0.316) has been replaced by the SE (0.0035 and 0.0175). This is expected to reduce the PSA uncertainty along the x-axis. On the costs side, the uncertainty in the relative dosing intensity (RDI) will be also reduced since the SD (35.2%) has also been replaced by the SE (1.86%). However, the inclusion of the cost of death as an uncertain parameter will increase the uncertainty associated to the costs. The overall effect (increase or decrease) on the uncertainty along the y-axis will depend on the magnitude of the uncertainty associated to the RDI and the cost of death.

The PSA outcomes plotted in the CE-plane (see figures 5.29 and 5.31) illustrate this effect for the cytokine and sunitinib refractory subgroups, respectively.

Compared to the manufacturer's base case with PAS in Figure 5.17, it is clear that now the uncertainty on the costs side has been reduced since most of the PSA outcomes are comprised between £30,000 and £40,000 on the y-axis, whereas in the manufacturer's base case with PAS, the PSA outcomes ranged from £10,000 to £80,000 on the y-axis. Therefore, it seems clear that the uncertainty reduction associated to the RDI outweighs the increase in uncertainty associated to the cost of death since the overall uncertainty on the costs side is reduced with respect to the manufacturer's base case. Regarding the effects (x-axis), the PSA outcomes are still spread over the NE quadrant along the x-axis direction. However, only less than 1% of the observations (in contrast to approximately 30% reported by the manufacturer) are now in the NW quadrant which clearly indicates a reduction of the uncertainty on the effects side as well.

For the sunitinib refractory population the uncertainty has been dramatically reduced compared to the manufacturer's base case shown in Figure 5.19 and all the PSA outcomes seem to converge towards the ICER. However, as mentioned in Section 4.6.2 and at the end of Section 5.2.10, the assumptions made by the manufacturer for the STC underestimate in a way that cannot be quantified the overall uncertainty of the model. Nevertheless, given the result of the ERG PSA, we may think that the underestimation is large since the uncertainty estimated for the sunitinib refractory subgroup is much smaller than for the cytokine refractory population. This seems to be counterintuitive because for the cytokine refractory population there is at least a (limited) source of evidence which does not exist for the sunitinib refractory group. For that reason, the ERG base case scenario for the sunitinib refractory population must be interpreted with caution; in fact, it can be seen as a best case scenario or lower limit for the overall uncertainty associated to the model for the sunitinib refractory subgroup.

Figure 5.28: ERG cost-effectiveness acceptability curve – Cytokine refractory population (with PAS)

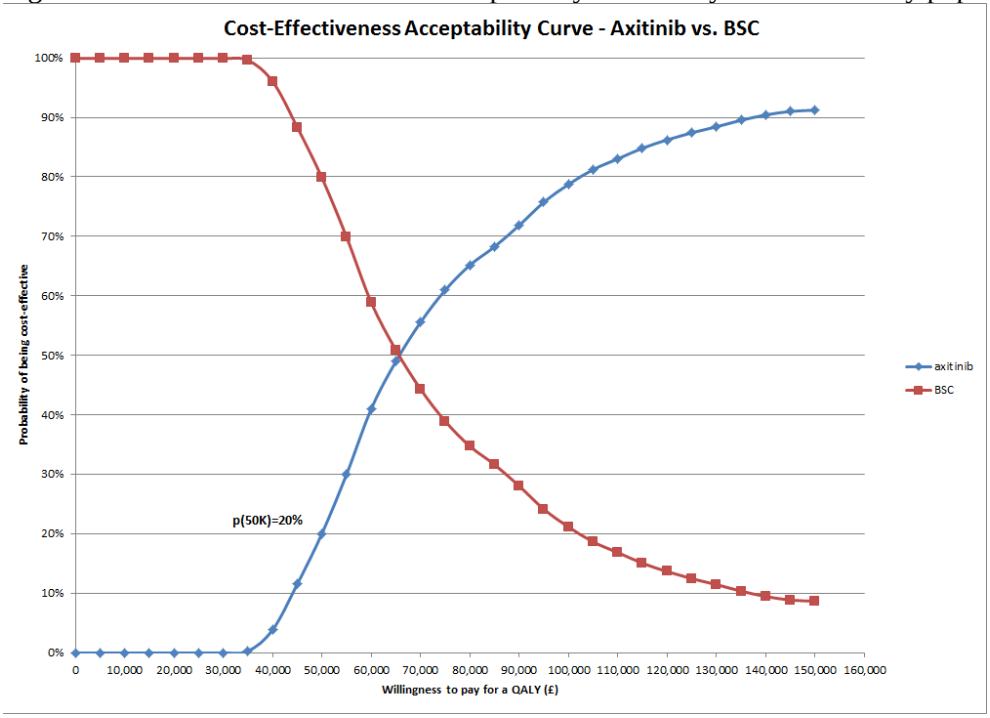


Figure 5.29: ERG PSA scatter plot – Cytokine refractory population (with PAS)

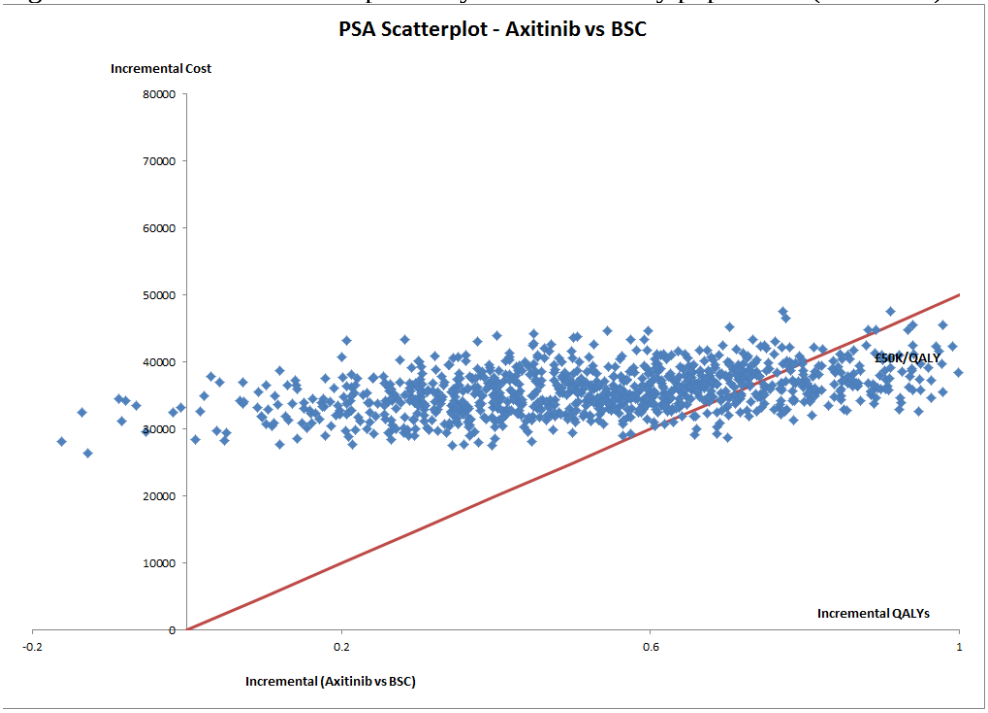


Figure 5.30: ERG Cost-effectiveness acceptability curve – Sunitinib refractory population (with PAS)

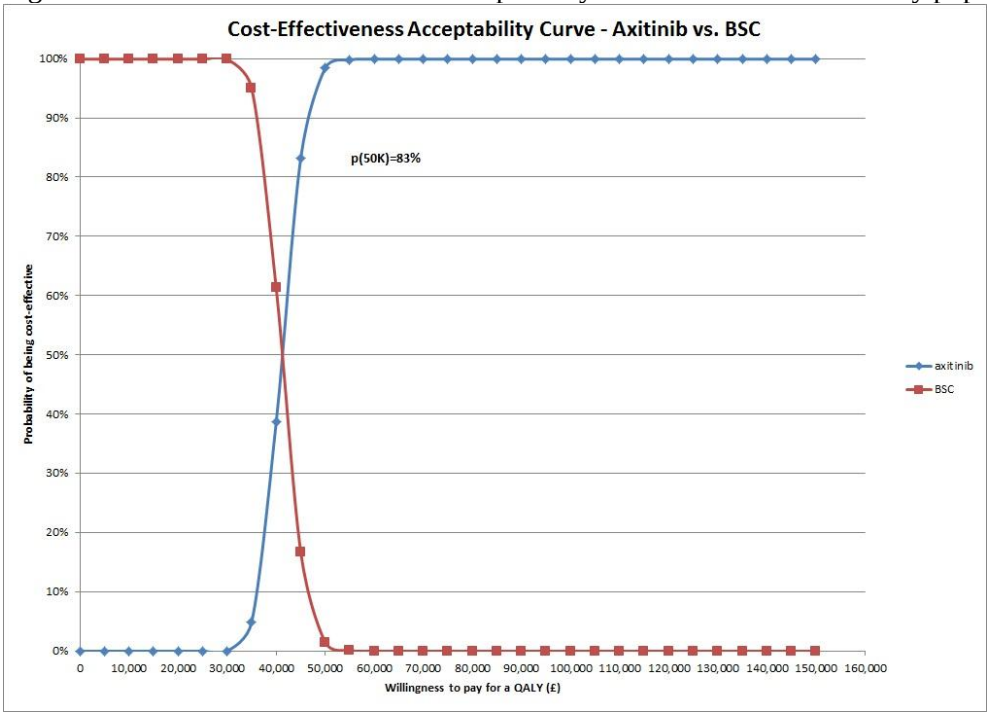
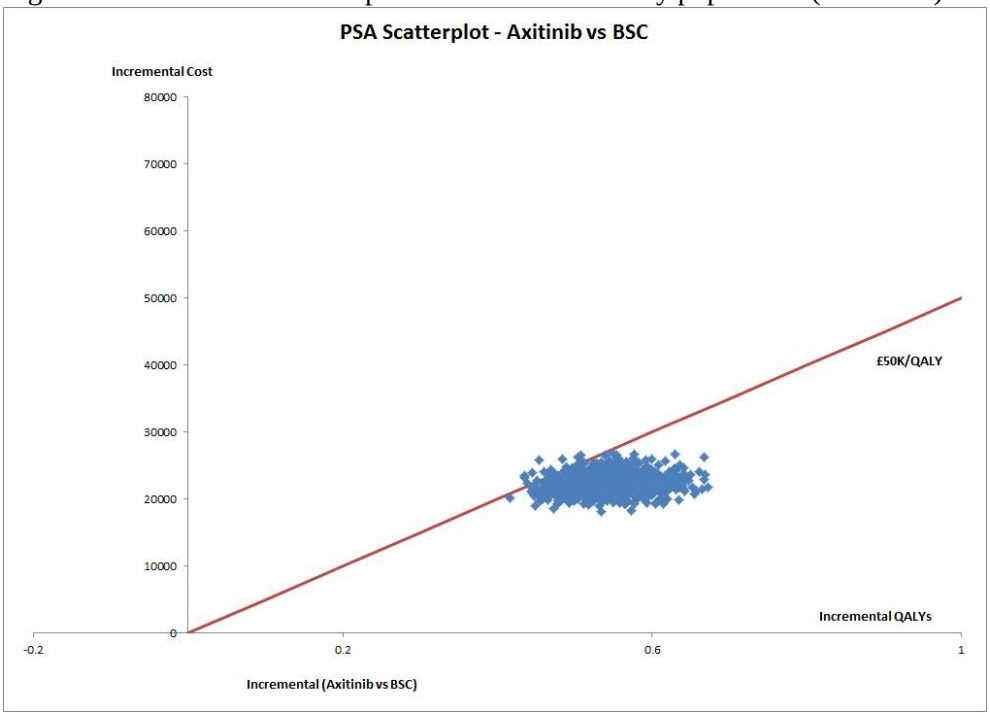


Figure 5.31: ERG PSA scatter plot – Sunitinib refractory population (with PAS)



PSA results without PAS can be seen below. The cost-effectiveness probability of axitinib for a threshold of £50,000 per QALY was [redacted] for the cytokine refractory population (see figures 5.32 and 5.33) and [redacted] for the sunitinib refractory population, respectively (sees figures 5.34 and 5.35). Note that the manufacturer reported that axitinib was cost-effective for [redacted] of the observations for the cytokine refractory subgroup and for [redacted] of the observations for the sunitinib refractory population. Thus, in both cases axitinib is significantly less cost-effective which clearly illustrates the effect of increasing the price of axitinib, i.e. when axitinib costs are higher the PSA outcomes tend move up

into the CE-plane. Therefore, more PSA outcomes are found above the threshold of £50,000 per QALY.

Figure 5.32: ERG Cost-effectiveness acceptability curve – Cytokine refractory population (without PAS)

Figure 5.33: ERG PSA scatter plot – Cytokine refractory population (without PAS)

Figure 5.34: ERG Cost-effectiveness acceptability curve – Sunitinib refractory population (without PAS)

Figure 5.35: ERG PSA scatter plot – Sunitinib refractory population (without PAS)

5.3.2 Additional utility scenarios

Subgroup specific utilities scenario

In response to the clarification letter (Section B – Question 15), the manufacturer provided the utility estimates for the progression free and progressive disease health state for the prior sunitinib and the prior cytokine subgroups separately. These are reported in Table 5.12 in Section 5.2.7.

The ICERs obtained using subgroup specific utilities can be seen in Table 5.25. Note that they are similar to those obtained in the base case (using the overall utility value) but lower ICERs are obtained for the cytokine refractory population whereas the opposite is observed for the sunitinib refractory group. Both scenarios further assumed, as in the base case, that there is no difference in utilities between the axitinib and the BSC arms.

Table 5. 25: Incremental results with and without PAS using subgroup specific utilities.

ICER (£) incremental (QALYs)	Base-case (no PAS)	Subgroup specific utilities (no PAS)	Base case (PAS)	Subgroup specific utilities (PAS)
Prior Cytokine	[redacted]	[redacted]	£65,326	£62,885
Prior Sunitinib	[redacted]	[redacted]	£40,933	£42,095

Lower utilities scenario

In section 5.2.7, it was discussed that the EQ-5D health states as measured in patients during the AXIS trial were valued using a US tariff developed by Shaw et al.[38] Since studies have shown that the US valuation is consistently higher than the UK valuation[39,40] the ERG has assessed the impact of lowering the current utilities (PF 0.692, PD 0.61). Based on table 3 from the paper by Johnson et al. [40] we lower the 0.69 utility value to 0.66, and the 0.61 utility value to 0.54.

The ICERs obtained in this case can be seen in Table 5.26. Note that these are slightly higher than those obtained in the base case (around [redacted] higher without PAS and £3,000 with PAS) for both populations. Both scenarios further assumed, as in the base case, that there is no difference in utilities between the axitinib and the BSC arms.

Table 5.26: Incremental results with and without PAS using lower utilities.

ICER (£) incremental (QALYs)	Base-case (no PAS)	Lower utilities (no PAS)	Base case (PAS)	Lower utilities (PAS)
Prior Cytokine	[redacted]	[redacted]	£65,326	£68,433
Prior Sunitinib	[redacted]	[redacted]	£40,933	£44,125

Higher BSC progression-free utility scenario

As the utility data was collected in the AXIS trial, no data was available in the BSC group. In section 5.2.7 it was discussed that it would be assumed that the utility estimate for axitinib and BSC is equal. While valid arguments were given, it is of course possible that due to for example side effects, the quality of life in the intervention group is a bit higher than in the BSC group. To test to impact of

treatment specific utility estimates before progression, we assumed that the utility in BSC would be a bit higher, i.e. 0.72 instead of the 0.69 used for axitinib.

Using these values we find that this higher utility for BSC has a very minimal impact on the ICERs (see Table 5.27).

Table 5.27: Incremental results with and without PAS using higher progression-free utility for BSC

ICER (£) incremental (QALYs)	Base-case (no PAS)	Lower axitinib PF utility (no PAS)	Base case (PAS)	Lower axitinib PF utility (PAS)
Prior Cytokine	[redacted]	[redacted]	£65,326	£66,639
Prior Sunitinib	[redacted]	[redacted]	£40,933	£41,363

5.4 Conclusions

The economic model described in the MS is considered by the ERG to meet the NICE reference case[5] and is in-line with the decision problem specified in the scope.

The ERG assessment indicated that the model was generally well presented and reported. Some minor issues were identified regarding the use of standard deviations in the PSA instead of standard errors; whilst these errors have a clear impact on the outcomes of the univariate and probabilistic sensitivity analyses, they do not impact the estimate of the central ICER.

The ERG univariate sensitivity analysis revealed that for the cytokine refractory subgroup, the ICER is extremely sensitive to changes in the HR for the overall survival. At the upper limit of the 95% confidence interval, the ICER would amount to approximately £423,000 (with PAS). This is in sharp contrast with the results for the sunitinib refractory subgroup, where changes to input parameters lead to modest changes in the ICER. This is related to the fact that no measures of uncertainty were provided for the adjustment factor for the BSC arm, which means that this uncertainty was not considered in the univariate (or probabilistic) sensitivity analysis.

The cost-effectiveness results were generally robust under the scenario analyses conducted, though a few scenarios impacted the ICER considerably. For the cytokine refractory subgroup, the most important assumption relates to the extrapolation of the OS in the axitinib arm, when a Gompertz distribution is used rather than a Weibull distribution, the ICER increases sharply. For the sunitinib refractory subgroup, the factor having a significant impact was the approach used to model OS in the BSC arm; using RENCOMP observational data with an indirect comparison led to a substantially higher ICER.

It is important to realize is that many uncertainties related to the health economic evaluation in the sunitinib refractory subgroup have not been quantified, and thus are not represented in the central estimates of the ICER or in the CEACs. As discussed in chapter 4, both the STC approach and the RENCOMP approach to estimating the PFS and OS of the BSC group have the potential for considerable bias, either upwards or downwards. This means that the same is true for the ICERs reported in chapter 5. In addition, the STC approach lacks an estimation of the uncertainty surrounding the point estimates it provides. Again, this also means that the uncertainty around the sunitinib-refractory ICER is most likely severely underestimated.

For the cytokine refractory subgroup most uncertainties have been taken into account, revealing a large uncertainty in the number of QALYs gained, and thus around the ICER.

8 CONCLUSIONS

Overall, the manufacturer's submission is clearly presented and the analyses and underlying assumptions are sound and reasonable. The main issue with this submission is whether a simulated treatment comparison (STC) presents a valid and reliable estimate of the clinical effectiveness of axitinib versus BSC in a sunitinib refractory population. As there is no direct trial evidence it is not possible to compare the results of the STC to any existing evidence so the accuracy and reliability of the results cannot be ascertained. In addition, the uncertainty around the STC results is missing.

The methods for the STC (sunitinib refractory population) follow the recommended method and it appears to be well-conducted and clearly explained, including considerations of the limitations of the method and available trial evidence. However, the STC is basically a comparison of individual treatment arms and is open to considerable bias. There is no way to assess whether or not the final result is biased.

In addition, the reliability of the results of the included studies and the comparability of the trials included are potential issues. These have been adequately reported in the manufacturer's submissions, and have been summarised in this ERG report.

The economic model described in the MS is considered by the ERG to meet the NICE reference case. The model structure was considered to be appropriate and the ERG has no major concerns regarding the selection of data used within the model, beyond the issues relating to the estimation of the treatment effectiveness of axitinib versus BSC.

In conclusion, the ICER for axitinib in a cytokine refractory population is based on a well performed indirect comparison. The results seem reliable, and present an ICER with large confidence intervals due to considerable uncertainty regarding the treatment effect of axitinib on overall survival. The ICER for axitinib in a sunitinib refractory population is based on an STC which did not included the uncertainty surrounding the estimates used in the analyses (indicating that the uncertainty around the ICER is only a lower limit of the true uncertainty), and is basically a comparison of individual treatment arms and is therefore open to considerable bias.

8.1 Implications for research

The NICE scope specified BSC as the only comparator for axitinib in adult patients with advanced renal cell carcinoma (aRCC) after failure of prior treatment with sunitinib or a cytokine. For the cytokine refractory population there is sufficient evidence to allow an indirect comparison between axitinib and BSC. For the sunitinib refractory population, there is not a network to link axitinib with BSC. Therefore a randomised trial comparing axitinib with BSC would be the first research priority. Alternatively, treatments that have been compared with axitinib in a sunitinib refractory population, such as sorafenib and temsirolimus, could also be compared with BSC, to allow an indirect comparison between axitinib and BSC.

Adverse events (comparators)

In utilising the same strategies reported in 10.2 the same limitations applied. Given the CRD advice on not using RCT filters in these cases, the ERG would recommend removing the study designs filters in lines #17-93 (Medline Search) and replacing them with a suitable adverse events filter, a number of which can be found in the ISSG Search Filters Resource.[49]

Non-RCT Evidence (Axitinib)

Adequate searches were carried out on all NICE required databases. ERG noted the same limitations in the line for Axitinib as in earlier searches (see Clinical Effectiveness 10.2). A test by the ERG in Medline and Embase showed that the omission of the alternative name Inlyta was unlikely to have impacted on recall.

	# ▲	Searches	Results	Search Type	Actions
☐	1	(axitinib or ag13736 or ag 13736).mp.	160	Advanced	Display More »
☐	2	(axitinib or ag13736 or ag 13736 or Inlyta).mp.	160	Advanced	Display More »

Remove Selected | Save Selected | Combine selections with: And Or RSS

Save Search History

Cost effectiveness

Limitations

- The ERG noted the redundant use of economics filter on Cochrane searches
- The ERG noted some disparity between the way that renal cell carcinoma was searched for between this and the earlier clinical effectiveness searches (10.2), especially the omission of terms such as hypernephroma$ or nephroid carcinoma$. It was not clear if this had an impact on the overall recall of the strategies
- The ERG noted that the following Emtree and MeSH terms used in the previous clinical effectiveness searches were not used in the cost effectiveness searches: Emtree *interleukin 2/* (Embase) and the MeSH for both *Interferon-Alpha/* and *Interleukin-2/* (Medline and Cochrane). It was unclear if this had any impact on the recall of results.
- The ERG noted the absence of CAS registry numbers.
- The ERG noted the absence of the following brand names:
    - Tivozanib missing: krn951
    - Pazopanib missing: votrient
    - Interferon alpha missing: varients using alfa rather than alpha
    - Everolimus missing: affinitor or xience or zortress
- No host was reported for the Embase and Econlit searches, ERG assumed this to be Ovid as reported for Medline search
- No host was given for the Cochrane Library searches, but given the syntax used and after a brief investigation of the hits per line, the ERG assumed this to be EBM Reviews from Ovid.
- Results were limited to English language only, which may have resulted in the omission of potentially useful papers in other languages.

Measurement and valuation of health effects

Limitations

- Redundant use of HRQL filter on Cochrane searches