26th March 2009

Pfizer Response to the NICE ACD
Sunitinib for the Treatment of Gastrointestinal Stromal Tumours

Sunitinib is both clinically and cost-effective, compared to best supportive care, when used to treat patients with gastrointestinal stromal tumours. It is understood why the Committee is minded not to recommend sunitinib and trust that the provision of the requested analyses will provide sufficient reassurance that sunitinib is affordable to the NHS.

Pfizer's response to the Appraisal Consultation Document is focused upon providing the following additional information:

1) Detailed explanation of the application of the rank preserving structural failure time (RPSFT) model and justification for its use within the base case of our economic analysis
2) An assessment of the impact upon cost effectiveness of including within the sunitinib costs, additional costs associated with continued use of sunitinib after disease progression
3) An analysis of sunitinib costs that is based upon the sunitinib treatment duration in the expanded access protocol A6181036
4) An analysis that censors patients assigned to best supportive care at the point they crossed over to receive sunitinib.
5) Complete updated probabilistic sensitivity analyses for each of the above analyses

For comparison with the base case analysis, Table 1 presents each additional analysis and their associated PSA estimates (Weibull parameters for each analysis can be found in Appendix 1). While we recognise these are important analyses for testing the variability of the ICER for sunitinib vs. BSC; we are concerned that these analyses each have their own limitations and biases and consequently we maintain that the base case we originally submitted reflects appropriately the benefit of using sunitinib to treat patients with GIST. These additional analyses were performed in response to NICE's request.

In addition, sunitinib for the treatment of GIST should be considered adopting the supplementary criteria for evaluating end of life medicines. The reasons for this are provided below related to the criteria established by NICE:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months;
  - Whilst there is little published data examining life expectancy for patients with advanced /metastatic GIST who have failed imatinib therapy, UK clinical expert opinion suggests that this patient group will live for circa 9 months if they do not have access to sunitinib. Of note, the RPSFT analysis estimated survival in the BSC group to be approximately 39 weeks which is similar to clinical expectation.
- 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,
  - The RPSFT analysis of the A6181004 trial suggests that sunitinib offers patients on average an additional 9 months of life compared to BSC, which is the current standard NHS treatment for this patient group.
- No alternative treatment with comparable benefits is available through the NHS,
  - Sunitinib is the only drug licensed for use post-imatinib failure in advanced/ metastatic GIST.
- The treatment is licensed or otherwise indicated, for small patient populations.
  - Estimates vary widely on the incidence of new cases of GIST in the UK, with figures between 200 and 2,000 quoted (NICE, 2004), with an apparent acceptance of an upper limit of 240 (NICE, 2004). Approximately half of new cases of GIST are likely to be metastatic and/or unresectable on first presentation. (NICE 2004)

Table 1: Cost effectiveness analysis for sunitinib vs. best supportive care

Analysis		Time on treatment (months)	Life years	Mean time in PD (years)	QALYs	Drug cost	Total costs	Cost per Life year gained	Cost per QALY gained	Probability sunitinib cost effective at £20,000/QALY WTP		Probability sunitinib cost effective at £30,000/QALY WTP	
										PSA[1]	PSA[2]	PSA[1]	PSA[2]
Base case	Sunitinib	7.3	1.98	1.38	1.23	£12,391	£19,767						
	BSC	0	1.21	1.02	0.73	£0	£6,068						
	Incremental	7.3	0.77	0.36	0.50	£12,391	£13,699	£17,695	**£27,365**	17%	18%	59%	57%
Include costs of sunitinib post progression‡	Sunitinib	7.3	1.98	1.38	1.23	£14,620	£21,996						
	BSC	0	1.21	1.02	0.73	£0	£6,068						
	Incremental	7.3	0.77	0.36	0.50	£14,620	£15,928	£20,574	**£31,817**	8%	7%	43%	42%
Sunitinib PFS and OS, from expanded access trial	Sunitinib	14.8	1.97	0.77	1.32	£26,567	£33,883						
	BSC	0	1.21	1.02	0.73	£0	£6,068						
	Incremental	14.8	0.76	-0.24	0.58	£26,567	£27,815	£36,578	**£47,628**	N/A	0%	N/A	0%
Censoring placebo pts at crossover to Sutent¶	Sunitinib	7.3	1.98	1.38	1.23	£12,391	£19,767						
	BSC	0	2.63	2.43	1.55	£0	£8,037						
	Incremental	7.3	-0.65	-1.05	-0.32	£12,391	£11,730	BSC dominates sunitinib	**BSC dominates sunitinib**	0%	N/A	3%	N/A

[1] Uses Weibull gamma and lambda parameters of BSC applying the hazard ratio from the RPSFT analysis to estimate sunitinib PFS and OS survival

[2] Uses Weibull gamma and lambda parameters of sunitinib applying the hazard ratio from the RPSFT analysis to estimate BSC PFS and OS survival

‡ Assumes additional cost of £2,229 per patient in PD. Calculated as 22% of patients taking sunitinib in PD for 0.42 years and the dose intensity for these patients was 91.2%. The PSA tests the sensitivity of these costs by varying time spent in PD; dose intensity of drug given when in PD and alters the mean time spent in PFS accordingly.

¶ Weibull curves fitted separately to sunitinib and BSC. The hazard ratio for this analysis is 0.824 (95CI 0.454, 1.499) this has been used within the PSA.

1) Detailed explanation of the application of the rank preserving structural failure time model and justification for its use within our economic analysis

As discussed within our submission, the results of the ITT overall survival analysis within our clinical trial are confounded by treatment crossover to sunitinib in the placebo arm. Although a common approach to this problem is to censor placebo arm patients at the point of crossover, we maintain that in this instance, this would be invalid as a) crossover is informative and patients who crossover are unlikely to be comparable to those who do not; b) crossover occurred very early in the time span of the clinical trial; c) a large proportion (84%) of patients crossed over.

Based on these factors, Pfizer identified the RPSFT analysis as the most appropriate statistical analysis for these data. This methodology can correct for time-dependent treatment changes in survival data whilst respecting the randomisation. (Robins et al 2004). The results from applying this methodology to the sunitinib A6181004 trial were presented at ASCO 2008. In preparing for the NICE submission Pfizer sought independent validation concerning the applicability of this method in this setting. Mr Ian White, a biostatistician at the Institute of Public Health, confirmed the appropriateness of RPSFT and provided additional confirmation that Pfizer had carried out the methodology correctly. Mr White provided further guidance concerning the use of the original ITT p value as it was considered scientifically inappropriate to revise the p value based on the data estimated with RPSFT. In addition, Mr White suggested re-censoring of the data which Pfizer also undertook and used within our base case analysis. Contrary to the suggestion we had carried out the methodology incorrectly, Mr White agreed with the methodology performed and Pfizer applied his advice to the analysis to ensure we were consistent with the current thinking.

Consequently, as agreed by both independent statistical expert advice and the Evidence Review Group (ERG), in this situation the correct analytical approach is to use the rank preserving structural failure time (RPSFT) model.

The purpose of this post-hoc analysis was to estimate the 'true' treatment difference in overall survival between sunitinib and placebo, as if the placebo patients had not crossed over on to sunitinib. Unlike simply censoring overall survival at the time of crossover from placebo to sunitinib treatment, the RPSFT analysis relates a patient's observed event time in the placebo arm to an event time that would have been observed if crossover to sunitinib treatment had not occurred, assuming treatment has a multiplicative effect on a patient's lifetime. This approach produces a randomisation-based effect estimator; that is, the treatment estimate is based on the treatment arms as randomised, thus avoiding many of the potential pitfalls and biases introduced with subgroup analyses. Appendix 2 provides further details upon the methods used.

RPSFT methodology is based upon randomisation, and therefore does not change the level of evidence against the null hypothesis. It does change the estimated hazard ratio, bringing it further from the null, consequently the 95% confidence intervals of the revised hazard ratio are wide. When the placebo data is adjusted for crossover using the RPSFT method, this produced a hazard ratio of 0.505 (95% CI: 0.26 to 1.13).

In our base case economic analysis Weibull curves were fitted by linear regression to the observed overall survival curve for sunitinib and independently for the estimated RPSFT curve for BSC. This provided the best fit to both the sunitinib and BSC data, hence we maintain this is the most appropriate base case. In sensitivity analysis we explored fitting Weibull curves to the RPSFT curve for BSC and applying the hazard ratio to this curve to estimate the sunitinib curve. Using the hazard ratio to predict the sunitinib curve, however, gives a poor visual fit between the modelled and empirical sunitinib curve. As an alternative, the ERG fitted Weibull curves to the sunitinib empirical evidence and applied the reciprocal

of the hazard ratio to estimate the BSC curve (Figure 1). With this approach the curve for the BSC data could be considered a poor fit. Using this alternative approach results in a cost per QALY estimate of £25,783 and is 58% cost effective assuming a willingness to pay of £30,000 per QALY (see Table 2).

Figure 1: Overall survival Kaplan-Meier data

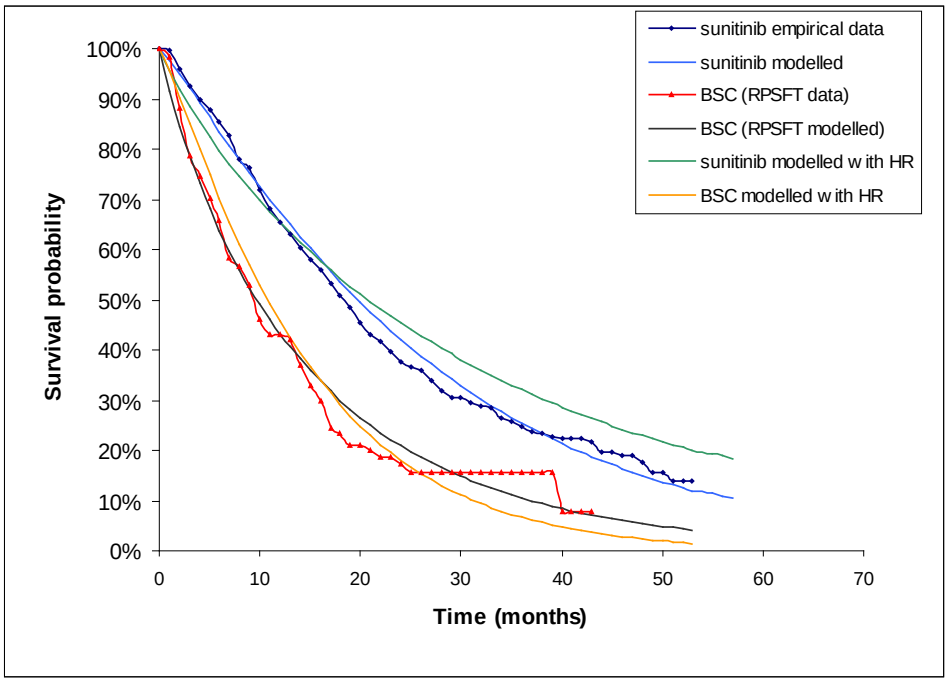


Table 2: Cost effectiveness analysis sunitinib vs. BSC when BSC OS is fitted using hazard ratio.

	Sunitinib	BSC	Sunitinib vs. BSC
Time on treatment (months)	7.3	0.00	7.3
Life years	1.98	1.15	0.83
Mean time in PD (years)	1.38	0.96	0.42
QALYs	1.23	0.70	0.53
Drug cost	£12,391	0	£12,391
Total costs	£19,767	£6,001	£13,767
ICERs			
Cost/life years gained			£16,547
COST/QALY			**£25,783**
Probability sunitinib cost effective at £30,000/QALY WTP			58%
Probability sunitinib cost effective at £20,000/QALY WTP			26%

2) An assessment of the impact upon cost effectiveness of including within the sunitinib costs, additional costs associated with continued use of sunitinib after disease progression

In our base case economic analysis, patients stop sunitinib treatment at disease progression and therefore, sunitinib costs are only incurred for the progression-free phase of the model. Although 22% of sunitinib patients within the clinical trial remained on treatment beyond disease progression, there is insufficient evidence to understand whether this would happen within clinical practice. However, for the purpose of informing NICE decision making, we present estimations for the impact of including additional costs with the continued use of sunitinib after disease progression.

In assessing the cost of using sunitinib beyond disease progression, we have used the formula as suggested by the ERG (page 87 of the ERG report); however when recreating this formula within our model, we noted the cost per patient of sunitinib in PD differed from that reported by ERG (£2,229 calculated by Pfizer, £2,237 calculated by ERG). As we wanted to test individual parameters within this formula, we have used the value we calculated.

Within sensitivity analysis we altered the following parameters within the ERG formula:

- Proportion of patients continuing to receive sunitinib in PD
- Mean duration of patients who took sunitinib in PD
- Dose intensity of sunitinib while taken in PD

The cost effectiveness results generated by including the additional costs associated with sunitinib after disease progression are detailed in Table 1, and the sensitivity analysis results are shown in Table 3.

Table 3: Sensitivity analysis exploring the impact on cost effectiveness of sunitinib vs. BSC including the costs of sunitinib in PD

Parameter	Formula parameters remaining constant	Sensitivity analysis	ICER (£/QALY)
Base case	n/a	n/a	£27,365
Inclusion of costs of sunitinib in PD; RPSFT method for OS	22% patients 0.42 years 91.2% dose intensity	n/a	£31,817
Percentage of patients receiving sunitinib in PD; RPSFT method for OS	0.42 years 91.2% dose intensity	Double % patients reported in trial = 44% Half % patients reported in trial = 11%	£36,270 £29,951
Time spent receiving sunitinib in PD for patients randomised to sunitinib; RPSFT method for OS	22% patients 91.2% dose intensity	Double time reported in trial = 0.84years Half time reported in trial = 0.21	£36,602 £29,599
Dose intensity while receiving sunitinib in PD; RPSFT method for OS	22% patients 0.42 years	Dose intensity = 100% Dose intensity = 50%	£32,247 £29,806
Varying all three parameters simultaneously	n/a	Upper bounds tested for all parameters: 44% patients, 0.84 years; 100% dose intensity Low bounds tested for all parameters: 11% patients, 0.21 years, 50% dose intensity	£46,752 £27,978

3) An analysis of sunitinib costs based upon sunitinib treatment duration in the expanded access protocol A6181036

Pfizer recognises the concern raised by NICE with regards to a difference in the median TTP in the sunitinib arm derived from RCT and EAP. The rationale for these differences is unclear and may be a consequence of a number of factors as listed in Table 4.

Table 4: Study design differences between RCT and EAP

RCT:	EAP:
Double-blind RCT Mature data, 30% of patients were still alive at the data cut-off point.	Open-label Immature results, 50% of patients were still alive at the data cut-off point.
The aim of the study is to test the null hypothesis that the true median TTP is 4 months versus the alternative hypothesis that the true median TTP is at least 6 months.	No formal hypothesis testing was planned. The number of patients to be enrolled was not predetermined and no inferential analyses were planned due to the nature of this study.
Patient population was strictly defined and assessed through inclusion/exclusion criteria.	The EAP included patients who were ineligible for participation in sunitinib RCTs and could not obtain sunitinib. As such, the patient population in the EAP could potentially vary from that of the RCT. Country-specific protocol amendments were allowed and patients as young as 12 years were permitted to enrol. Any patients with potential to derive clinical benefit from treatment with sunitinib were enrolled as judged by the investigator
Tumour measurement/assessment was scheduled every 6 weeks and followed RECIST criteria.	Tumour measurements/assessments were performed as per local standard of care, which may vary significantly between countries. Patient assessment was also carried out on a less frequent basis.
A strict study protocol was applied throughout the study with regard to determining disease progression.	Treatment was continued for as long as there was evidence of disease control in the judgment of the investigator.

We understand the Committee's rationale for exploring this data, as observational studies such as the EAP can reflect "real-world" effectiveness of the intervention. However, the results should be considered with caution as they are susceptible to bias due to internal and external validity issues such as the lack of control of confounding factors. Hence, it could be argued that the phase III RCT A6181004 study is the best evidence to support the use of sunitinib in patients with advanced GIST.

The EAP results support the pivotal RCT and provide more insight into the efficacy and safety of sunitinib. The use of this data within the economic analysis is problematic and leads to a bias against sunitinib for the following reasons:

- There is no control arm in the EAP hence the relative benefit of sunitinib versus BSC is unknown for the EAP and may differ to that from the RCT
- At the time of enrolment the positive results from the A6181004 study had been publicly presented. Therefore sunitinib may have been chosen by patients or

physicians with this knowledge, coupled with a more 'relaxed' inclusion criteria, potentially leading to a different patient population

- At the time of data cut-off 50% of patients were still alive in the EAP compared to 30% in the RCT. Thus, although the overall survival results appear comparable between the RCT and the EAP, the EAP results could be an underestimate of overall survival

In order to incorporate the EAP data into our economic analysis it is believed greater uncertainty has been introduced into the cost-effectiveness results. Within this analysis (presented in Table 1), we have used the EAP data to model sunitinib time to tumour progression and overall survival (see Figure 2) and the PFS data and RPSFT analysed data to model the BSC comparator. In the sensitivity analysis presented in Table 5, we have used alternative methods for fitting the four sets of curves to explore the sensitivity of the methods.

Figure 2: Kaplan-Meier estimates of time to tumour progression and overall survival observed for sunitinib in the expanded access protocol

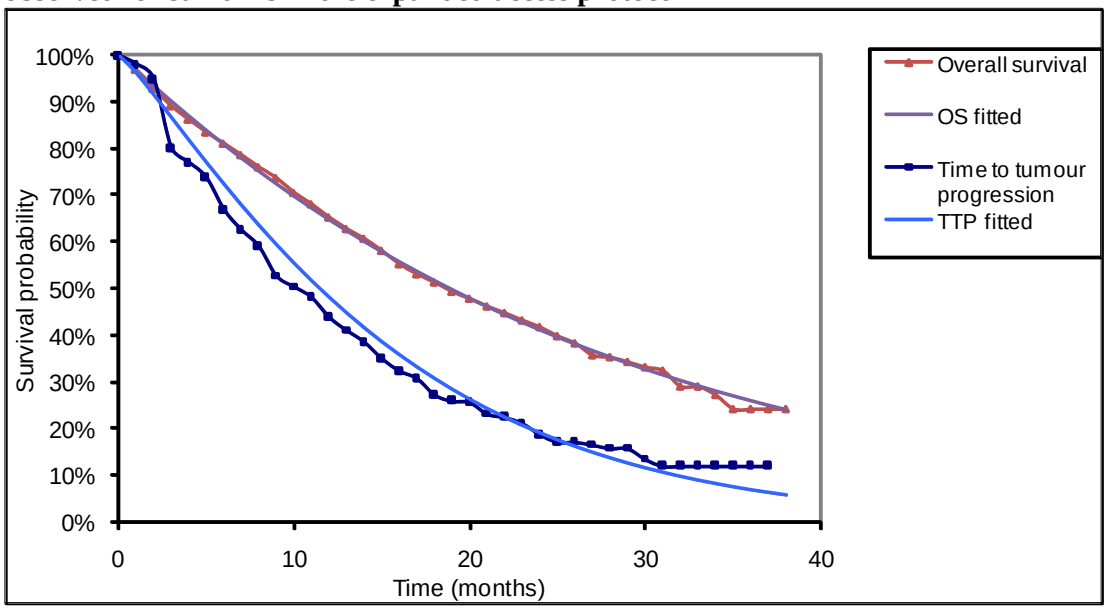


Table 5: Sensitivity analysis to explore using the EAP data

Parameter	Base case	Sensitivity analysis	ICER (£/QALY)
Base case	n/a	n/a	£27,365
Survival curves for PFS and overall survival	Sunitinib and BSC curves fitted independently	n/a	£47,628
	Sunitinib and BSC curves fitted independently	BSC PFS curve fitted using HR; OS independently	£53,002
		BSC PFS fitted independently; OS curve fitted using HR	£43,201
		BSC PFS and OS fitted using HR	£47,547
Source of data	Sunitinib TTP and OS data from EAP; BSC data from clinical trial	Sunitinib TTP from EAP; OS from clinical trial	£47,102
		Sunitinib TTP from EAP; OS from clinical trial – BSC OS from HR	£44,699

4) An analysis that censors patients assigned to best supportive care at the point they crossed over to receive sunitinib

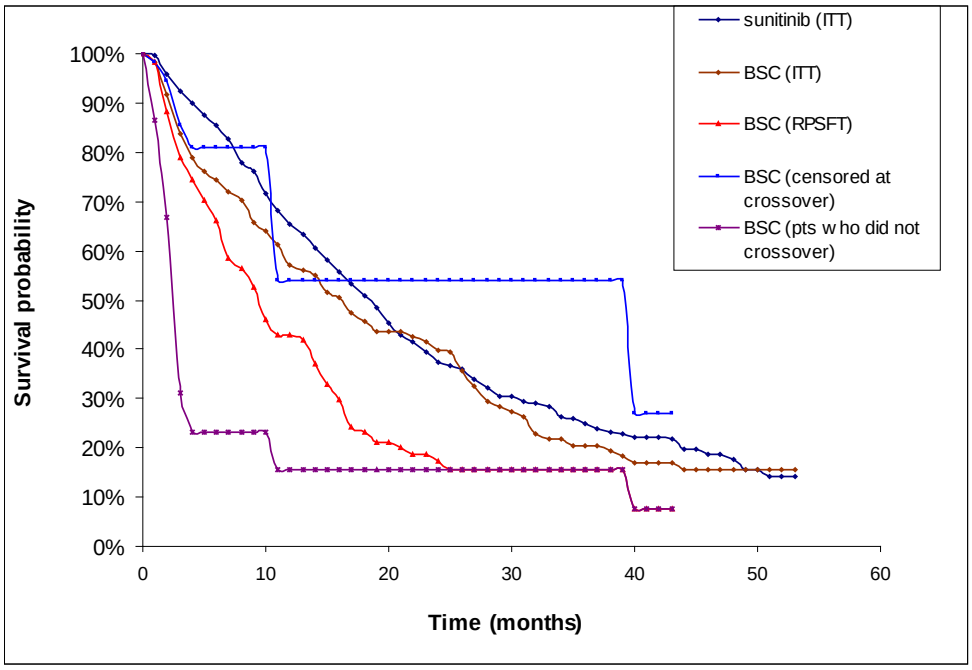
As the ERG has indicated, patient cross over is a recurring problem in cost-effectiveness assessments of cancer drugs and that analysis based on the unadjusted ITT data would underestimate the true relative effect of sunitinib. Traditional approaches to correct for crossover involve censoring placebo patients at the point of crossover so that the overall survival is based on only placebo data. However, as previously discussed, this method has a number of issues associated with it, namely, a) crossover is informative and patients who crossover are unlikely to be comparable to those that do not; b) crossover occurred very early in the time span of the clinical trial and c) a large proportion (84%) of patients crossed over from placebo to sunitinib treatment.

Following the recommendation of the independent DMC after the interim analysis of trial A6181004, Pfizer decided that the trial be completely unblinded and patients who were randomised to placebo, who had not already crossed over to sunitinib at tumour progression prior to interim analysis, were able to crossover to sunitinib treatment even in the absence of tumour progression; this crossover was not mandated but left to the discretion of the investigator. Hence, the decision to cross a patient over to sunitinib treatment was an informed decision. Patients who were not crossed over to sunitinib treatment may have been too ill to benefit from crossing over or they may have had indolent disease and may not have needed to crossover. Either way, data from patients who crossed over and patients who did not crossover cannot be considered representative of the entire placebo randomised population with both subgroups being a biased sample. Any estimation based on censoring the data of patients who crossed over (at the time of crossover) and relying solely on the longer-term follow-up of patients who did not crossover is fraught with bias.

For exploratory purposes we have summarised time to death and produced Kaplan-Meier curves for both an analysis that censors patients at crossover and for the subgroup of 15 of patients who did not crossover from placebo to sunitinib treatment. Median overall survival for BSC when patients are censored for crossover is 156 weeks vs. 72.7 weeks for sunitinib (hazard ratio 0.824, 95% CI: 0.454, 1.499). Median overall survival for the BSC patients who did not crossover is 9.7 weeks.

Figure 3 presents these Kaplan-Meier overall survival curves for comparison with the placebo overall survival ITT curve and the placebo curve generated using the RPSFT methodology. The Kaplan-Meier plot for patients censored at crossover results in a vast overestimation of BSC benefit as the majority of patients crossed over and, as crossover occurred early, only 13 patients had died before censoring. Conversely, the Kaplan-Meier plot for the few patients who did not crossover shows a rapid decline in survival and may provide a worse case scenario for patients with GIST treated with placebo (see Appendix 3 for a comparison of the Kaplan-Meier estimates used in these two analyses). Given the relative small numbers contributing to these two subgroups we urge caution when interpreting this data and are not confident that these present a valid assessment of placebo survival that can be used within economic modelling for decision making.

Figure 3: Exploratory summary of overall survival data for BSC



While these analyses present a very biased estimate of survival for BSC, we have used these values in sensitivity analysis to test our base case results. Using the overall survival data from the analysis in which BSC patients were censored at crossover results in BSC dominating sunitinib that is sunitinib costs more but produces less benefit (see Table 1).
When the overall survival data for the patients who did not cross over is used instead of the RPSFT derived overall survival data the cost per QALY for sunitinib compared to BSC is £20,618 (see Table 6)

Table 6: Cost-effectiveness of sunitinib vs. BSC; BSC OS data derived from patients who did not crossover to sunitinib treatment

	Sunitinib	BSC	Sunitinib vs. BSC
Time on treatment (months)	7.3	0.00	7.3
Life years	1.98	0.86	1.12
Mean time in PD (years)	1.38	0.68	0.70
QALYs	1.23	0.54	0.70
Drug cost	£12,391	0	£12,391
Total costs	£19,767	£5,355	£14,412
ICERs			
Cost/life years gained			£12,916
COST/QALY			**£20,618**
Probability sunitinib cost effective at £30,000/QALY WTP			92%
Probability sunitinib cost effective at £20,000/QALY WTP			46%

5) Complete updated probabilistic sensitivity analyses for each of the above analyses

For ease of reference, complete probabilistic sensitivity analyses (PSAs) have been presented when describing the results for each requested analysis. The probabilistic variables have been corrected for errors and omissions identified by the ERG. Where applicable, probabilistic analyses have been conducted using two alternative ways of estimating efficacy; a) Weibull parameters for BSC are calculated and HR applied to estimate sunitinib survival curves and b) Weibull parameters for sunitinib are calculated and the HR is applied to estimate BSC survival curves.
We have used both approaches as in some instances the Weibull parameters for BSC remain unchanged even though sunitinib survival has changed. Therefore, to run the PSA using Weibull parameters would not investigate the sensitivity within the parameter changes.

Conclusion
In conclusion, although it is understood why the Committee were minded not to recommend sunitinib we trust that in providing the requested analyses the Committee has sufficient reassurance to recommend sunitinib for the treatment of GIST.

While the additional analyses demonstrate the variability of the ICER, the RPSFT methodology is the correct analytical approach to use in interpreting the clinical evidence for sunitinib. Pfizer's use of this methodology has been guided by independent statistical expert advice and the additional detail presented within this document should reassure the Committee that the methods have been used correctly.

The extra analyses presented within this ACD response support the conclusions of Pfizer's original submission that sunitinib is both clinically effective and cost-effective for the NHS compared to best supportive care, when used to treat patients with gastrointestinal stromal tumours.

Appendix 1: Weibull parameters estimated and used within the economic analyses

As within our original submission the two parameters of the Weibull distribution, In(λ) and γ, for baseline PFS and separately for OS were drawn from bivariate normal distributions, using the method of Cholesky matrix decomposition. The variance-covariance matrices used in the matrix decomposition were estimated from linear regression of (In(-InS(t)) against In(t), where S(t) is the survival function at time t.

The calculated Weibull parameters for each new analysis and the hazard ratios applied in PSA are shown in Table 1.

Table 1: Base case parameters of Weibull distribution used within each analysis

Analysis	**Treatment**	**PFS**		**OS**		**Hazard ratio used in PSA**	
		λ	γ	λ	γ	PFS	OS
Base case analysis	Sunitinib	0.171	0.912	0.024	1.133	0.333	0.505
	BSC	0.303	1.360	0.090	0.899		
EAP analysis	Sunitinib	0.039	1.178	0.304	1.030	0.330	0.505
	BSC	0.303	1.360	0.090	0.899		
Censoring placebo pts at crossover to Sutent	Sunitinib	0.171	0.912	0.024	1.133	0.333	0.824
	BSC	0.303	1.360	0.045	0.832		
Analysis of placebo patients who did not crossover	Sunitinib	0.171	0.912	0.024	1.133	0.333	0.505
	BSC	0.303	1.360	0.487	0.433		

Appendix 2: Rationale for use of RPSFT model

Pfizer recognises the concerns raised by NICE with regards to appropriateness of the rank preserving structural failure time model (RPSFT) as a method for dealing with confounding in the OS analysis caused by crossover in the A6181004 trial.

To address NICE's concern over the RPSFT method that was used in Pfizer's submission, the remainder of this document will provide a detailed explanation of the application and methods of the RPSFT model, which was used to derive the overall survival of GIST patients assigned to best supportive care plus placebo while accounting for crossover. We will also provide an explanation as to why the conventional methods, whereby the placebo data is censored at the point of crossover, were not utilised within our submission.

1. RANK-PRESERVING STRUCTURAL FAILURE TIME (RPSFT) MODEL

Introduction

This analysis was performed as a post-hoc analysis reported in the clinical study report of Study A6181004. The analysis was based on Robins and Tsiatis' rank preserving structural failure time models (RPSFT) for survival outcomes which relate a patient's observed event time in the placebo arm to an event time that would have been observed if no crossover to sunitinib treatment had been administered, assuming treatment has a multiplicative effect on a patient's lifetime (accelerated failure time model). Robins and Tsiatis' approach produces a randomisation-based effect estimator; that is, the treatment estimate is based on the treatment arms as randomised, thus avoiding many of the potential pitfalls and biases introduced with subgroup analyses. The RPSFT method has been applied to HIV, coronary heart disease, and other disease areas for estimating the causal effect of the treatment in a randomised trial with a survival outcome and time-dependent treatment switches in the past decade. It hence has been considered as the best available method to account for crossover (Robins et al. 1994; White et al. 1997; Korhonen et al. 1999; White et al. 2003).

Background

In the phase 3 double blinded, placebo controlled randomised trial of sunitinib in imatinib-resistant and intolerant GIST patients (A6181004), the patients in the placebo arm were allowed to cross over to sunitinib treatment at tumour progression due to ethical concerns. Because of the crossover in the placebo arm, the 'naïve' intention-to-treat (ITT) analysis can only estimate the benefit of starting sunitinib immediately compared to deferring the start of sunitinib. However, the relevant question to be answered from a decision-making perspective is the benefit of starting sunitinib compared to not starting sunitinib. Some methods have been proposed for correcting the confounding by treatment crossover. The most well known method is the 'as treated' or 'on-treatment' analysis, in which each patient's follow-up time is censored when they stop receiving their randomised treatment. However, the on-treatment estimates are biased if the patients whose follow-up are censored at time *t* are not the random sample of all patients at risk in their randomised groups at time *t*. In the case of the phase III trial, the decision to cross over was left to the discretion of the investigator and so the decision was informed. Factors that may have influenced this decision could have related to a patient having indolent disease (hence potentially not needing crossover) or a patient being too ill to benefit from crossover. Therefore, these would not necessarily be representative of the placebo group as a whole. The only way to avoid selection bias is to use a method based entirely on the comparability of groups as randomised.

The more commonly used analytic approach is an ITT analysis. That is, the treatment effect is estimated based on the treatment group to which a patient was randomised regardless of whether the patient stayed with their assigned treatment. An ITT analysis has both advantages and disadvantages. A point in its favour is that the identity of the two treatment arms is

guaranteed by randomization. If experimental treatment has no causal effect on a patient's survival, the survival distributions of the two randomised groups will remain the same even in the presence of non-random crossover. However, the treatment effect estimated from the ITT analysis cannot measure the true biological effect of treatment in the presence of crossover. The ITT measure of treatment effect would diminish as crossover rate increased. Therefore, it is important to apply methods which can estimate the true treatment effect of the experimental treatment with the presence of non-random crossover.

The purpose of this post-hoc analysis was therefore to estimate the 'true' treatment difference in overall survival between sunitinib and placebo, as if the placebo patients had not been crossed over on to sunitinib. We used the method of Robins and Tsiatis[1a] which is the best method currently available in the literature that can correct for time-dependent treatment changes in survival data while respecting the randomisation.

Robins et al [1a, 2a, 3a, 4a] developed analytical techniques for survival outcomes which relate a patient's observed event time to an event time that would have been observed if no crossover had occurred. This assumes that the treatment has a multiplicative effect on a patient's lifetime and is more commonly referred to as an accelerated failure model. It makes no assumptions about the relationship between treatment changes and prognosis and is a semi-parametric approach. The estimation approach developed by Robins and Tsiatis does not conflict with the standard ITT log-rank test of no treatment effect used in a randomised trial. The rank preserving structural failure time (RPSFT) model [is] based explicitly on the groups as randomised. The RPSFT model produces estimated treatment effect that is of the same sign as the ITT treatment effect, and that is only statistically significant if the ITT analysis is statistically significant. The RPSFT method uses a causal model which involves a parameter to transform a patient's observed event time to the latent event time that would have been observed for the patient if no treatment had been given, and then the parameter is estimated as the value which makes the latent time balance across randomised groups.

From Study A6181004 for each patient *i*, we have observed:

- the time to death (or censoring), T_i,
- whether crossover occurred, Q_{ik} (where $Q_{ik} = 1$ for crossover and $Q_{ik} = 0$ otherwise and k is the time at which crossover occurred) and
- randomised treatment group, R_i (where $R_i = 1$ for sunitinib and $R_i = 0$ for placebo); note that for sunitinib ($R_i = 1$) crossover was not applicable.

Latent Failure Time

In addressing the issue of crossover, we are interested to know the survival differences we would have observed if no patients in the placebo group had crossed-over to sunitinib treatment.

We assume that for each patient there exists a set of latent failure times $\{U_i, g = h\}$ where $g = h$ indicates that patient i followed the treatment history specified by h until failure. h might be the treatment history of no crossover to sunitinib or the treatment history of crossover to sunitinib. For each patient i, we only observe T_i, which in terms of our set of latent failure times can be written,

$$T_i = U_{i,g=\overline{Q_i}(T_i)}. \qquad [1]$$

In particular, we are interested in the failure time if patient i never crossed over to sunitinib. We will designate this as U_i and refer to it as patient i's baseline failure time, since it is the time to failure if patient i always continued their 'baseline treatment' of placebo.

We define the null hypothesis of no treatment effect on survival time as

$$T_i = U_i = U_{i,g=h} \text{ for all } i \text{ and all } h \qquad [2]$$

If the null hypothesis is true (i.e. there is no treatment effect), then patient *i*'s observed survival time would be unchanged, regardless of crossing over to sunitinib.

Following the accelerated failure time model, we assume that each patient's baseline failure time is related to their observable data by

$$U_i = \int_o^{Ti} \exp\{\psi_0 Q_i(u)\} du \qquad [3]$$

where ψ_0 is an unknown parameter (Robins et al[5a] refer to this as a rank preserving structural failure model). Note: ψ_0 is also known as the acceleration factor such that $\psi_0 < 0$ implies that crossing over treatment extends life by a factor $exp(-\psi_0)$ and $\psi_0 > 0$ implies that crossing over treatment decreases life by a factor $exp(-\psi_0)$.

In Study A6181004, crossing over to sunitinib can only occur at discrete times (i.e. at the clinical visits), hence the right hand side of the equation becomes a weighted sum of time spent in either crossover state (not crossed over or crossed over), where the weights are $\exp\{\psi_0 Q_{i,k}\}$.

The following example, demonstrates the calculation of U_i. If patient *i* had a failure time of *59* days, had attended clinical visits at Day *0*, Day *14*, Day *28* and Day *42* and had crossed over to sunitinib at Day *28*, the notation would be:

$T_i = 59$ (time to failure in days),
$Q_{i,42} = \{0, 0, 1, 1\}$ (indicator variables indicating patient *i* had not crossed over at Day *0* and Day *14*, but had crossed over at Day *28* and Day *42*)

For this example, ψ_0 is assumed to be *-0.1*. U_i is calculated as weighted sum of time spent in either crossover state, where the weights are $\exp\{\psi_0 Q_{i,k}\}$., as follows:

U_i = (*59 - 42*) x *exp* (*-0.1* x *1*) +	between death and last visit
(*42 - 28*) x *exp* (*-0.1* x *1*) +	between visit at days *42* and *28*
(*28 – 14*) x *exp* (*-0.1* x *0*) +	between visit at days *28* and *14*
(*14 – 0*) x *exp* (*-0.1* x *0*)	between visit at days *14* and *0*
= *31* x *0.9048 + 28*	
= *56* (days)	

Under the null hypothesis $\psi_0 = 0$, such that equation [3] returns the identity, $U_i \equiv T_i$ regardless of the observed treatment history. In lay-terms, this means that under the null hypothesis, there is no difference between the treatments and hence there is no benefit in patients crossing over from placebo to sunitinib.

Thus the null hypothesis implies that $\psi_0 = 0$ and hence a test of $\psi_0 = 0$ is a valid α-level test of [2].

To understand the implications of $\psi_0 \neq 0$, we consider a failure time *Vi*, where *Vi* is patient *i*'s time to death. If as suggested previously we assume that *Vi* is related to *Ui* by

$$U_i = \int_0^{V_i} \exp\{\psi_0 \times 1\} du = \exp(\psi_0) V_i \qquad [4]$$

$$\exp(-\psi_0)U_i = V_i \qquad [5]$$

Hence,

$$\frac{V_i - U_i}{U_i} = \exp(-\psi_0) - 1 \qquad [6]$$

which is the fractional increase in survival for the *i*th patient if they cross over to sunitinib.

Estimation of ψ_0

In the example above, ψ_0 was assumed to be known and had a value of *-0.1*, however in reality ψ_0 is unknown and needs to be estimated and once estimated needs to be tested to establish if $\psi_0 = 0$.

Marks and Robins[2a, 3a] provide an in depth description of the estimation process, however the text below provides a brief summary.

The assumption is made that U_i (the baseline failure time) while not observable is a baseline characteristic, like age or height, which therefore implies that it is independent from R_i (randomised treatment group). This means that the baseline failure times in both the sunitinib group $R_i = 1$ and the placebo group $R_i = 0$ are identical, such that

$$\Pr[U_i(\psi) \geq x | R_i = 1] = \Pr[U_i(\psi) \geq x | R_i = 0] \qquad [7]$$

We therefore test whether a particular value of ψ equals ψ_0 by seeing whether equation [4] is true.

The testing procedure has two steps. First, using equation [3] and our hypothesized ψ, we compute the baseline failure times $U_i(\psi)$ for each patient. A grid search over possible values is used in order to determine the appropriate point estimate. Secondly, treating $U_i(\psi)$ exactly as though they were observed failure times, we perform a weighted log-rank test of the hypothesis that the baseline survival curves are identical in the two treatment groups.

The weighted log-rank test statistic is calculated as

$$S_R(\psi, w(i)) = \sum_{i=1}^{n} w_i \left\{ R_i - \frac{\sum_{j \in Y_i(\psi)} R_j}{n_i(\psi)} \right\} \qquad [8]$$

where $Y_i(\psi) = \{j : U_j(\psi) \geq U_i(\psi)\}$, the observable risk set for each patient *i* and $n_i(\psi)$ is the observed number of people in the risk set. Since in our analysis we shall set $w_i = 1$ for all *i*, we will drop the weights from our notation

The variance of this statistic, Ω, can be consistently estimated by the usual formulas for the variance of a log-rank test. The *Z* rank test is therefore calculated as:

$$Z_R(\psi) = \frac{S_R(\psi)}{\sqrt{\Omega}} \qquad [9]$$

which has a normal $N(0, 1)$ distribution.

The presence of censoring introduces extra difficulties; we implemented the re-censoring procedure described by Robins and Tsiatis[1a] as not doing so incurs bias[6a]. Finally, we estimated the hazard ratio for starting sunitinib compared to not starting sunitinib by running a

Cox regression on the observed event times in the sunitinib arm and the estimated U_i values in the placebo arm.

Here is an example to illustrate how the latent event time was derived for a real placebo patient. *T = 47.14* weeks was the observed event time. *C = 18.14* weeks was the time of crossover. *U* was the event time that would have been observed if no sunitinib treatment had been given.

$$U = C + exp(\psi)(T - C)$$

where the parameter $\psi = -0.656$ represents the causal effect of having started treatment. ψ was estimated by computing *U* for a range of possible values of ψ and finding the value for which a log rank test of the equality of *U* across the two groups gives a zero test statistic.

$$U = 18.14 + exp\ (-0.656)(47.14 - 18.14)$$
$$= 18.14 + 0.5189 \text{ x } 29.0 = 33.19(\text{weeks})$$

Because this procedure is based on the randomisation, it does not change the level of evidence against the null hypothesis. It does however change the estimated hazard ratio, bringing it further from the null, as would be expected from the fact that crossovers make the overall treatment experience of the two arms more similar. As a result, the 95% confidence interval is wide.

Results

The purpose of this statistical analysis is to show that using appropriate methodology to correct the effects of crossover from placebo to sunitinib produces an overall survival curve that estimates a treatment difference had patients not crossed over.
When the placebo data adjusted for crossover using the rank-preserving structural failure time model (RPSFT) was analysed, this produced a hazard ratio of 0.505 (95% C.I: 0.262 to 1.134). Table 1 presents the OS estimated using the traditional and novel methods.

Table 1: OS estimated using different statistical methods (ITT Population)

	Median (weeks; 95% CI)			
Phase of Study	**Sunitinib**	**Placebo**	**HR (95% CI)**	**P-value**
Blinded phase* based on ITT	Not reached	Not reached	0.491 (0.290-0831)	0.007
Final study (blinded + open label) based on ITT	72.7 (61.3-83.0)	64.9 (45.7-96.0)	0.876 (0.679-1.129)	0.306
Final study (blinded + open label) based on ITT using RPSFT Method	72.7 (61.3-83.0)	39.2 (28.0-54.1)	0.505 (0.262-1.134)**	0.306

*Interim analysis.
** Empirical 95%CI was obtained using bootstrap samples
Source: Demetri et al. 2008 (revised estimates)

The analysis using the RPSFT approach demonstrated an estimated median OS for the placebo group of 39.0 weeks (95% CI 28.0.-54.1) based on the ITT population. This revealed an unbiased sunitinib treatment effect (HR 0.505, 95% CI 0.262-1.134) comparable to that of the interim OS results. The hazard ratio and 95% CI were re-estimated with a re-censoring procedure following the suggestion of an external subject expert.

Figure 1: Final OS estimated without correcting for crossover using naïve ITT method

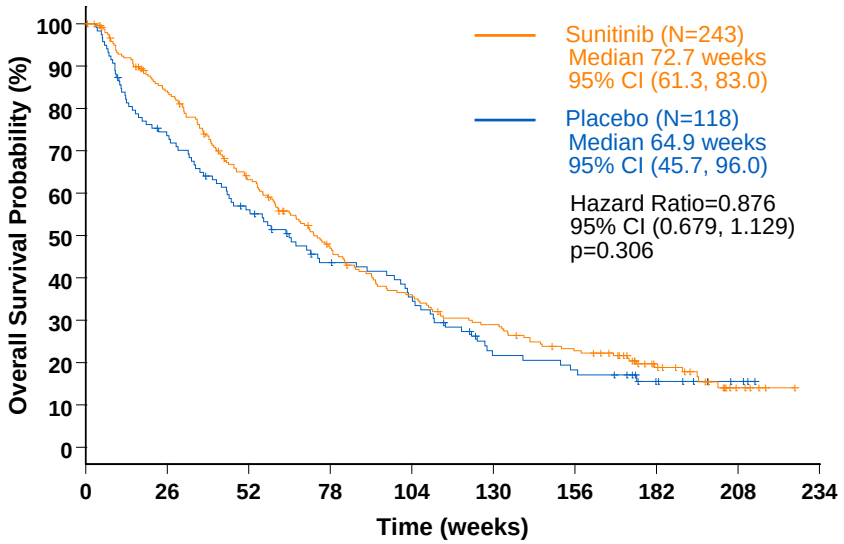


Figure 2 demonstrates the final OS estimated without correcting for crossover using the Kaplan Meier method.

Figure 2: Final OS estimated with correcting for crossover using RPSFT Method

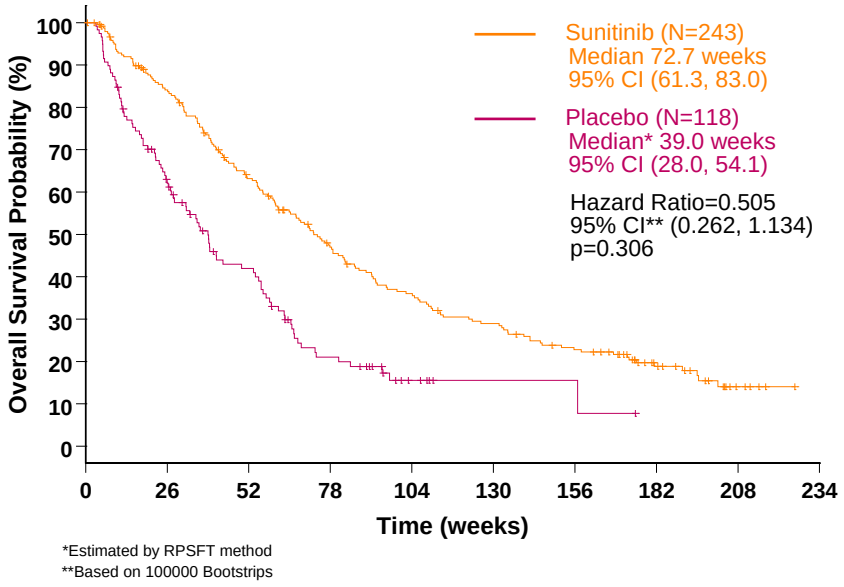
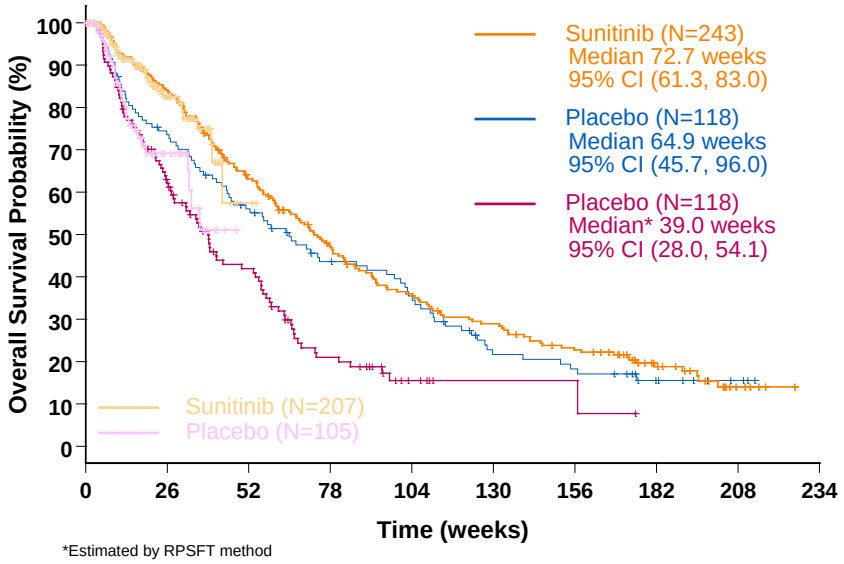


Figure 3: Final OS estimated with correcting for crossover using RPSFT Method (overlapped with interim and naïve ITT results)



There is consistency between the hazard ratio produced using the RPSFT model (HR: 0.505) which adjusted for crossover and the hazard ratio from the interim analysis which was less confounded by crossover (HR: 0.491, 95% C.I: 0.290 to 0.833). This provides validation that the treatment effects observed from the interim analysis would have held true, had placebo patients not been given the ethical choice of crossing over to sunitinib.

Appendix 3: Comparison of the Kaplan Meier estimates of survival between the two censoring analyses

Table 1: Comparison of Kaplan Meier estimates of overall survival between two types of analyses.

Time (weeks)	Survival Probability		Number of Death
	15 placebo patients who did not cross over	All 118 placebo patients with survival times censored at crossover	
0	1.000	1.000	0
3.286	0.933	0.991	1
3.714	0.867	0.981	2
5.000	0.800	0.971	3
5.429	0.733	0.960	4
7.143	0.667	0.945	5
8.571	0.600	0.929	6
9.286	0.533	0.912	7
9.714	0.467	0.895	8
11.286	0.389	0.874	9
11.429	0.311	0.854	10
15.857	0.233	0.811	11
41.714	0.156	0.541	12
156.857	0.078	0.270	13

For 15 placebo patients who did not cross over to sunitinib, the Kaplan Meier estimation of overall survival can be illustrated as in Table 2.

Table 2: Survival estimates for placebo patients who did not cross over to sunitinib

Time (weeks)	Survival Probability	Number of Death	Number of at Risk
0	1.0	0	15
3.286	0.933 (14/15)	1	14
3.714	0.867 (14/15x13/14=13/15)	2	13
5.000	0.800 (14/15x13/14x12/13=12/15)	3	12
5.429	0.733 (14/15x13/14x12/13x11/12=11/15)	4	11
7.143	0.667 (14/15x13/14x12/13x11/12x10/11=10/15)	5	10
8.571	0.600 (14/15x13/14x12/13x11/12x10/11x9/10=9/15)	6	9
9.286	0.533 (14/15x13/14x12/13x11/12x10/11x9/10x8/9=8/15)	7	8
9.714	0.467 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8=7/15)	8	7
10.143+		8	6
11.286	0.389 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8x5/6=7/18)	9	5
11.429	0.311 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8x5/6x4/5=14/45)	10	4
15.857	0.233 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8x5/6x4/5x3/4=7/30)	11	3
41.714	0.156 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8x5/6x4/5x3/4x2/3=7/45)	12	2
156.857	0.078 (14/15x13/14x12/13x11/12x10/11x9/10x8/9x7/8x5/6x4/5x3/4x2/3x1/2=7/90)	13	1
175.286+		13	

+ Censored observation.

For all 118 placebo patients, when survival times are censored at the times of crossover, the Kaplan Meier estimation of overall survival can be illustrated as in Table 3.

Table 3: Survival estimates for all placebo patients when survival time is censored at the time of crossover to sunitinib treatment

Time (weeks)	Survival Probability	Number of Death	Number of at Risk
0	1.00	0	118
+	*10 patients were censored between 0 and 3.286*	0	108
3.286	0.991 (107/108)	1	107
+	*2 patients were censored between 3.286 and 3.714*	1	105
3.714	0.981 (107/108x104/105)	2	104
+	*10 patients were censored between 3.714 and 5.000*	2	94
5.000	0.971 (107/108x104/105x93/94)	3	93
+	*6 patients were censored between 5.000 and 5.429*	3	87
5.429	0.960 (107/108x104/105x93/94x86/87)	4	86
+	*20 patients were censored between 5.429 and 7.143*	4	66
7.143	0.945 (107/108x104/105x93/94x86/87x65/66)	5	65
+	*7 patients were censored between 7.143 and 8.571*	5	58
8.571	0.929 (107/108x104/105x93/94x86/87x65/66x57/58)	6	57
+	*2 patients were censored between 8.571 and 9.286*	6	55
9.286	0.912 (107/108x104/105x93/94x86/87x65/66x57/58x54/55)	7	54
9.286+		7	53
9.714	0.894 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53)	8	52
+	*8 patients were censored between 9.714 and 11.286*	8	44
11.286	0.874 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53x43/44)	9	43
11.286+		9	42
11.429	0.854 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53x42/43x41/42)	10	41
+	*21 patients were censored between 11.429 and 15.857*	10	20
15.857	0.811 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53x42/43x41/42x19/20)	11	19
+	*16 patients were censored between 15.857 and 41.714*	11	3
41.714	0.541 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53x42/43x41/42x19/20x2/3)	12	2
156.857	0.270 (107/108x104/105x93/94x86/87x65/66x57/58x54/55x52/53x42/43x41/42x19/20x2/3x1/2)	13	1
175.286 +		13	

+ Censored observation(s).

References:

Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial. JASA 1994; 89:737-749.

White IR, Walker S, Babiker AG, Darbyshire JH. Impact of treatment changes on the interpretation of the Concorde trial. AIDS 1997; 11:999–1006.

Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC study. Statistics in Medicine 1999; 18: 2879-2897.

White, IR, Carpenter J, Pocock SJ, Henderson RA. Adjusting treatment comparisons to account for non-randomized interventions: an example from an angina trial. Statistics in Medicine 2003; 23:781–793