The clinical and cost-effectiveness of sunitinib for the treatment of gastrointestinal stromal tumours: a critique of the submission from Pfizer

Report commissioned by: NHS R&D HTA Programme

On behalf of: NICE

Produced by: Peninsula Technology Assessment Group (PenTAG)
Peninsula Medical School, Universities of Exeter and Plymouth.

Authors:
Mary Bond, Research Fellow, PenTAG
Martin Hoyle, Research Fellow, PenTAG
Tiffany Moxham, Information Scientist, PenTAG
Mark Napier, Consultant Medical Oncologist, Royal Devon & Exeter Foundation Trust Hospital
Rob Anderson, Senior Lecturer in Health Economics, PenTAG

Correspondence to: Mary Bond

PenTAG
Noy Scott House
Barrack Road
Exeter EX2 5DW
mary.bond@pms.ac.uk

Date completed: 12 January 2009

ABOUT THE PENINSULA TECHNOLOGY ASSESSMENT GROUP (PenTAG)

The Peninsula Technology Assessment Group is part of the Institute of Health Service Research at the Peninsula Medical School. PenTAG was established in 2000 and carries out independent Health Technology Assessments for the UK HTA Programme, systematic reviews and economic analyses for the NICE Centre for Public Health Excellence, and systematic reviews as part of the Cochrane Collaboration Heart Group, as well as for other local and national decision-makers. The group is multi-disciplinary and draws on individuals' backgrounds in public health, health services research, computing and decision analysis, systematic reviewing, statistics and health economics. The Peninsula Medical School is a school within the Universities of Plymouth and Exeter. The Institute of Health Research is made up of discrete but methodologically related research groups, among which Health Technology Assessment is a strong and recurring theme. Projects to date include:

- The Effectiveness And Cost-Effectiveness Of Imatinib (STI 571) In Chronic Myeloid Leukaemia - A Systematic Review (2002)
- Screening For Hepatitis C Among Injecting Drug Users And In Genitourinary Medicine (GUM) Clinics - Systematic Reviews Of Effectiveness, Modelling Study And National Survey Of Current Practice (2002)
- Systematic Review Of Endoscopic Sinus Surgery For Nasal Polyps (2003)
- The Effectiveness And Cost-Effectiveness Of Imatinib For First Line Treatment Of Chronic Myeloid Leukaemia In Chronic Phase (2003)
- The Effectiveness And Cost-Effectiveness Of Microwave And Thermal Balloon Endometrial Ablation For Heavy Menstrual Bleeding - A Systematic Review And Economic Modelling (2004)
- Do The Findings Of Case Series Studies Vary Significantly According To Methodological Characteristics?(2005)
- The Effectiveness And Cost-Effectiveness Of Pimecrolimus And Tacrolimus For Atopic Eczema - A Systematic Review And Economic Modelling (2005)
- The Effectiveness And Cost-effectiveness Of Dual Chamber Pacemakers Compared To Single Chamber Pacemakers For Bradycardia Due To Atrioventricular Block Or Sick Sinus Syndrome - Systematic Review And Economic Evaluation (2005)
- The Effectiveness And Cost-Effectiveness Of Surveillance Of Barrett's Oesophagus: Exploring The Uncertainty (2005)
- The Cost-Effectiveness of testing for hepatitis C (HCV) in former injecting drug users. Systematic Review And Economic Evaluation. (2006)
- The Effectiveness And Cost-Effectiveness Of Cinacalcet for Secondary Hyperparathyroidism in end stage renal disease patients on dialysis. Systematic Review And Economic Evaluation (2007)
- The effectiveness and cost-effectiveness of Carmustine Implants and Temozolomide for the treatment of newly-diagnosed High Grade Glioma**.** Systematic Review And Economic Evaluation (2007)
- The Effectiveness And Cost-Effectiveness of Cardiac Resynchronisation Therapy for Heart Failure. Systematic Review And Economic Evaluation (2007)
- Inhaled Corticosteroids and Long-Acting Beta2-Agonists for The Treatment of Chronic Asthma in Adults and Children Aged 12 Years and Over: a Systematic Review and Economic Analysis (2007)
- Inhaled Corticosteroids and Long-Acting Beta2-Agonists for The Treatment of Chronic Asthma an Children Under the Age of 12 Years: a Systematic Review and Economic Analysis (2007)
- The Effectiveness and Cost-Effectiveness of Cochlear Implants for Severe to Profound Deafness in Children and Adults: A Systematic Review and Economic Model (2009, In Press)
- The Effectiveness and Cost-Effectiveness of Methods of Storing Donated Kidneys from deceased donors: A Systematic Review and Economic Model (2009, in Press)

Source of funding

This report was commissioned by the NHS R&D HTA programme.

Competing interests of authors

None.

Acknowledgements

We would like to thank Joanne Perry (PenTAG) for her administrative support for this project.

Contributions of authors

Mary Bond	Provided overall project management, critiqued the effectiveness evaluation provided by the manufacturer, the formulation of questions for the manufacturer and wrote chapters one, two, three, four and seven of the report
Martin Hoyle	Critiqued the model and economic analysis provided by the manufacturer, formulated questions for the manufacturer and wrote the cost-effectiveness section of the report
Tiffany Moxham	Undertook literature searches for the report, commented on the searches conducted by the manufacturer
Mark Napier	Provided clinical advice throughout the assessment and commented on the draft report.
Rob Anderson	Contributed to the editing and writing of the report and the formulation of questions for the manufacturer

Table of Contents

1. SUMMARY 10
1.1. Scope of the submission 10
1.2. Summary of submitted clinical effectiveness evidence 11
1.3. Summary of submitted cost effectiveness evidence 11
1.4. Commentary on the robustness of submitted evidence 12
1.4.1. Strengths 12
1.4.2. Weaknesses 12
1.4.3. Areas of uncertainty 13
1.5. Key Issues 14
2. BACKGROUND 15
2.1. Critique of manufacturer's description of underlying health problem 15
2.2. Critique of manufacturer's overview of current service provision 15
3. CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM 17
3.1. Population 17
3.2. Intervention 17
3.3. Comparators 17
3.4. Outcomes 18
3.5. Time Frame 18
3.6. Other relevant factors 18
4. CLINICAL EFFECTIVENESS 19
4.1. Critique of manufacturer's approach 19
4.1.1. Description of manufacturers search strategy and comment on whether the search strategy was appropriate. 19
4.1.2. Inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate. 20
4.1.3. Table of identified studies. 20
4.1.4. Details of any relevant studies that were not included in the submission? 23
4.1.5. Description and critique of manufacturers approach to validity assessment 23
4.1.6. Description and critique of manufacturers outcome selection 36
4.1.7. Description and critique of the statistical approach used 37
4.1.8. Summary statement 41
4.2. Summary of submitted evidence 41
4.2.1. Summary of results 41
4.2.2. Critique of submitted evidence syntheses 61
4.2.3. Summary of clinical effectiveness 61
5. ECONOMIC EVALUATION 64
5.1. Overview of manufacturer's economic evaluation 64
5.1.1. Summary of Pfizer's systematic review of cost-effectiveness studies 64
5.1.2. Natural history 66
5.1.3. Treatment effectiveness within submission 67
5.1.4. Health related quality of life 68
5.1.5. Resources and costs 68
5.1.6. Discounting 70
5.1.7. Sensitivity analysis 70
5.1.8. Model validation 70
5.2. Critique of approach used 71
5.2.1. Critical appraisal frameworks 71
5.2.2. Critique of the modelling approach and structure 74
5.2.3. Data inputs 74

5.2.4. Assessment of uncertainty **77**
5.3. Results included in manufacturer's submission **77**
5.4. Comment on validity of results presented with reference to methodology used **81**
5.4.1. Data inputs **81**
5.4.2. Assessment of consistency **89**
5.4.3. Assessment of uncertainty **89**
5.5. Summary of uncertainties and issues **91**
6. ADDITIONAL WORK UNDERTAKEN BY THE ERG 94
7. DISCUSSION 95
7.1. Summary of clinical effectiveness issues **95**
7.2. Summary of cost-effectiveness issues **96**
7.3. Implications for research **96**
8. REFERENCES 97
APPENDIX 1: COMPARISON OF PFIZER MODEL OF SUNITINIB FOR GIST WITH THE ASSESSMENT GROUP MODEL OF SUNITINIB FOR RENAL CELL CARCINOMA FOR MODEL VALIDATION 99
APPENDIX 2: EQ-5D-BASED UTILITY VALUES FROM RCT 101

Tables

Table 1: Studies identified by Pfizer's search strategy 20
Table 2: Summary of ongoing sunitinib phase II/III clinical trial programme for GIST 22
Table 3: Critical appraisal of A6181004 (p32) Demetri et al. 2006/08 24
Table 4: Critical appraisal of A6181036 (Submission p50) Reichardt et al. 2008 (on-going) 30
Table 5: Eastern Cooperative Oncology Group Performance Status, comparing the two included studies' participants. 35
Table 6: The advantages and disadvantages of different analytical approaches to RCTs with crossovers 40
Table 7: Summary of Time to Tumour Progression (ITT) 42
Table 8: Overall Survival estimated using different statistical methods (ITT Population) (p36) 44
Table 9: Summary results for study A6181004 RCT 50
Table 10: Overall Adverse Event Experience during the Blinded Phase (PP population) 51
Table 11: Adverse events that occurred with a frequency of at least 5% greater on sunitinib than on placebo in per-protocol population 52
Table 12: Summary table of results from study A6181036 cohort study 58
Table 13: Most common (≥ 20%) treatment-related non-haematologic adverse events 59
Table 14: Treatment-related haematologic Adverse Events 59
Table 15: Treatment-related AEs related to cardiac function (A6181036) 60
Table 16: Summary results of the main outcome measures in the two included studies 63
Table 17: Critical appraisal checklist based on NICE Reference Case(National Institute for Health and Clinical Excellence 2008) 72
Table 18: Critical appraisal checklist from Drummond and colleagues(Drummond et al. 1997). 72
Table 19: Critical appraisal checklist of Philips et al. (2006)(Philips et al. 2006) for model-based analyses 73
Table 20: Base case results of Pfizer's model using RPSFT method for best supportive care overall survival. 78
Table 21: Results of Pfizer's model using the unadjusted ITT method for best supportive care overall survival. 79
Table 22: Selected sensitivity analyses undertaken by Pfizer † 80
Table 23: Summary of important deterministic sensitivity analyses 92
Table 24: ICER for our (Assessment Group) renal cell carcinoma (RCC) model of sunitinib vs. interferon-□ with one parameter changed at a time 99

Figures

Figure 1: Kaplan-Meier estimates of time to tumour progression (ITT population), interim analysis. 42
Figure 2: Kaplan-Meier Curves of Overall Survival from the Interim Analysis 43
Figure 3: Final Overall Survival estimated with correcting for crossover (RPSFT Method)....45
Figure 4: Final Overall Survival estimated without correcting for crossover (Kaplan-Meier Method)....45
Figure 5: Result of Cox Proportional Hazards Analysis of TTP in various subgroups 48
Figure 6: TTP and overall survival with sunitinib in the ITT Population 56
Figure 7: overall survival for subgroups based on individual baseline factors and prior imatinib treatment history. 57
Figure 8: Pfizer curve fits to progression-free survival Kaplan-Meier data (Submission Figure 13)....84
Figure 9: Pfizer curve fits to overall survival Kaplan-Meier data (Submission Figure 14) 85

Abbreviations and Acronyms

AE	Adverse Events
AIDS	Acquired Immunodeficiency Syndrome
ASCO	American Society of Clinical Oncology
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
CR	Complete Response
CT	Computed Tomography
DPSM	Duration of Performance Status Maintenance
DR	Duration of Response
DSMB	Data Safety Monitoring Board
EAP	Expanded Access Programme
ECOG	Eastern Cooperative Oncology Group
EMEA	European Medicines Evaluation Agency
ESMO	European Society of Medical Oncology
GIST	Gastrointestinal Stromal Tumour
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
HTA	Health Technology Assessment
ITT	Intention To Treat
KM	Kaplan-Meier
MRI	Magnetic Resonance Imaging
NCCN	National Comprehensive Cancer Network
NCI CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
OS	Overall Survival
ORR	Objective Response Rate
PD	Progressive Disease
PDGFRA	Platelet Derived Growth Factor Receptor Alpha
PFS	Progression Free Survival
PP	Per Protocol
PR	Partial Response
QALY	Quality Adjusted Life Year
QoL	Quality-of-Life
RECIST	Response Evaluation Criteria in Solid Tumours

RPSFT	Rank Preserved Structural Failure Time Models
RTK	Receptor Tyrosine Kinase
SBCC	Submission base case 1st cycle costed
SBCF	Submission base case 1st cycle free
SSA	Submission sensitivity analysis
SmPC	Summary of Product Characteristics
TTP	Time to Tumour Progression
TTR	Time to Tumour Response
WHO	World Health Organisation

1. Summary

Text, Tables or Figures shaded in grey have been copied from the submission by Pfizer, hereafter referred to as 'the submission'.

1.1. Scope of the submission

The submission from Pfizer considered the use of sunitinib malate (Sutent®) for the treatment of people with unresectable and/or metastatic gastrointestinal stromal tumours (GIST) after failure of imatinib due to resistance or intolerance.

The comparator in the decision problem was best supportive care which was taken to mean:

treatment to control, prevent and relieve complications and side effects and to improve comfort and quality of life. Within the model it is assumed to include palliative interventions but explicitly excluded the use of active therapy. (Submission p8)

The clinical effectiveness outcomes considered were, overall survival, time to tumour progression, progression free survival, response rates, adverse effects of treatment and health-related quality of life.

The outcomes for the economic analysis were, incremental cost per quality-adjusted life-year, incremental cost per life year gained, resource utilisation and the cost of treating adverse events. The time horizon for the economic analysis was six years and costs were considered from an NHS and personal social services perspective.

In Pfizer's model the first cycle of sunitinib was free for all patients, i.e. there is no cost to the NHS; this reflects a pricing scheme currently being negotiated with the Department for Health.

Subgroup analyses were presented for clinical effectiveness in the following patient groups to show the influence of baseline characteristics:

Study **A6181004** (Demetri et al. 2006/08): Time to tumour progression (TTP) on previous imatinib mesylate therapy (≤ vs. > 6 months); baseline MacGill Pain Questionnaire – present pain intensity (MPQ-PPI) score (0 vs.≥ 1); age (< vs. ≥ 65 years); sex (male vs. female); race (white vs. non-white); Eastern Cooperative Oncology Group (ECOG) performance status (0 vs. 1); and time since initial diagnosis with GIST (< vs. ≥ 6 months).

Study **A6181036** (Reichardt et al. 2008): Age (< vs. ≥ 59 years); ECOG performance status (PS = 0-1vs. 2); prior imatinib dose ≤ (vs. >400mg/day); reason for stopping imatinib treatment (intolerance vs. progressive disease ≤ vs. > 6 months).

No subgroup analyses were conducted for the economic evaluation.

1.2. Summary of submitted clinical effectiveness evidence

The evidence for this submission is based on one RCT (Demetri et al., 2006/08), which compares the effects of sunitinib with placebo for people with unresectable and/or metastatic GIST after failure of imatinib due to resistance or intolerance and with Eastern Cooperative Oncology Group (ECOG) progression status 0-1 (the most physically able), and one, ongoing, non-comparative, cohort study (Reichardt et al., 2008) which gives expanded access to a similar population but with ECOG progression status 0-4.

The RCT was a double-blind, placebo-controlled, parallel-group, multi-centre, phase III clinical trial. The blinded phase became open-label upon disease progression or at the time of interim analysis (54 weeks) when patients were allowed to cross-over from placebo to treatment group.

The results for **overall survival** are similar in both studies; with the RCT reporting results for the sunitinib arm of 73 median weeks (95%CI 61-83) in comparison to 75 (95%CI 68-84) median weeks for the cohort study. However, the results for **time to tumour progression** in the cohort study (median weeks = 41, 95%CI 36-47) are quite different from those of the sunitinib arm of the RCT (median weeks = 29, 95%CI 22-41). These results may be influenced by the different ECOG performance status of the two study populations and a greater median OS for the ECOG grade 0-1 in the cohort study (RCT: 73 weeks (95%CI 61-83), cohort: 88 weeks (95%CI 77-97).

The interim RCT results for **progression free survival** showed that those in the sunitinib group had a significantly better chance of being alive and free from progressive disease than those in the placebo group. Median progression free survival with sunitinib was 24.6 weeks (95% CI:12.1 to 28.4) compared with 6.4 weeks (95% CI:4.4 to 10.0 weeks) on placebo (hazard ratio 0.333, 95% CI:0.238 to 0.467; p<0.001).

1.3. Summary of submitted cost effectiveness evidence

The manufacturer used a Markov model, based on the renal cell carcinoma RCC model developed by PenTAG, to model the cost-effectiveness of sunitinib compared to best supportive care for GIST patients. This had a three state structure; progression free, progressive disease and death.

Pfizer's base case analysis produced an ICER of **£27,365 per QALY** with the first cycle of treatment sunitinib *not costed,* and using effectiveness estimates from their Rank Preserved Structural Failure Time (RPSFT) analysis[a]). When we included the cost of the first cycle of treatment we estimated that the value of the base case ICER was **£32,636 per QALY**, again using RPSFT effectiveness data. Their sensitivity analysis produced a range of ICERs from £15,536 per QALY to £59,002 per QALY.

When a conventional method of unadjusted ITT analysis is used to calculate the base case ICER, values of £93,062 per QALY (first cycle costed) and £77,107 per QALY (first cycle free), are produced. However, this method does not account for the overestimated effectiveness results in the placebo arm due to crossovers; independent, expert statistical opinion favours the RPSFT method.

1.4. Commentary on the robustness of submitted evidence

1.4.1. Strengths

Clinical effectiveness

- The searches are appropriate and include all relevant studies
- The RCT is of high quality

Cost-effectiveness

- The approach taken to modelling is reasonable
- The sources and justification of estimates were also generally reasonable

1.4.2. Weaknesses

- The evidence is based on only one completed and published RCT. The expanded access cohort study is ongoing, is not comparative and only published as an abstract at the time of this report.
- The majority of the control population (84%) in the RCT crossed over to the intervention group. This gave rise to the use of unusual methods of analysis (RPSFT) to correct for the bias this may have introduced. While we believe this to be the correct approach, we have been unable to check that it was applied correctly.

[a] The RPSFT is discussed in Section 4.1.7.1

- In their economic evaluation, Pfizer have presented a miscalculation of cost-effectiveness using the ITT overall survival data for best supportive care (Kaplan-Meier analysis). The stated ICER is £34,649 per QALY when it should have been £93,062 per QALY with sunitinib fully costed (or £77,107 per QALY if the first cycle of treatment is free). (Pfizer corrected this error following questions from us)

- A number of errors and omissions were also made in the probabilistic sensitivity analysis:

    - Pfizer used the standard deviation rather than the standard error for the utilities.

    - In the model, Pfizer assume a standard deviation of 0.02 for progression free survival, whereas the report says 0.20

    - Importantly, Pfizer have not modelled all the uncertainty in the treatment effect for progression free survival and overall survival

    - There are errors in the Cholesky matrix decompositions in modelling the uncertainty of the fit of the Weibull curves for treatment effectiveness in worksheets "PFS", "overall survival_RPSFT analysis" and "overall survival_ITT analysis".

1.4.3. Areas of uncertainty

- Given that there are several major errors in the PSA, and that we do not have the information to fully correct the PSA, the precise degree of uncertainty in the base case ICER is unknown. However, we can say that the uncertainty in the base case ICER (reported as £27,365 per QALY –1st cycle free) is substantial, given the wide (95%) confidence interval for the hazard ratio of overall survival of 0.262-1.234 (using the RPSFT method).

- The use of the RPSFT method of analysis has had a very large impact on cost-effectiveness; the ICER using this method produces (£32,636 per QALY -1st cycle costed) is a great deal less than that based on the unadjusted ITT data analysis (£93,062 per QALY -1st cycle costed). Expert statistical advice from Ian White (MRC Biostatistics Unit, Cambridge), indicates that the RPSFT is the correct method for analysis and that it appears to have been correctly applied.

- However, we caution that the base case ICERs may be slightly too low as Pfizer's calculation does not include the cost of sunitinib in progressive disease for some

patients randomised to sunitinib (54 patients in the sunitinib arm carried on with this treatment after disease progression), and who theoretically may have benefited.

1.5. Key Issues

- The use of the RPSFT method of analysis (instead of the conventional approach of censoring participants at the point of crossover) greatly affects the estimated cost-effectiveness of sunitinib for GIST. However, this is a common analysis issue in trials of cancer drugs that are found to be effective mid-trial, and the use of the RPSFT seems appropriate.

- The lack of costing of sunitinib in progressive disease for patients initially randomised to sunitinib, which does not reflect the treatment of some patients in the RCT (22% continued with sunitinib after disease progression).

- The large amount of uncertainty in the relative treatment effectiveness for overall survival between sunitinib and best supportive care under the RPSFT method.

- Whether to assume that the first cycle of sunitinib is free to the NHS.

- Patients in the expanded access cohort study had a longer median time to tumour progression than those in the RCT.

2. Background

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

In Section 4.1 (Submission p11) Pfizer provided a summary of the characteristics of gastrointestinal stromal tumours (GISTs) and their treatment. Brief evidence was given of the incidence of disease, its aetiology, treatment and prognosis; although a description of best supportive care (BSC) has been omitted. A useful algorithm (Submission Fig 1, p13) of the management of patients with metastic/unresectable GISTs was provided.

2.2. Critique of manufacturer's overview of current service provision

Section 4.5 (p15) of the submission states that there is wide variation in the methods used to manage imatinib resistant/intolerant GIST patients. However, it does not say what these variations might be, other than persisting with imatinib as part of BSC (even when the patient does not respond to it), or increasing the dose of imatinib above 400mg/day, although this is against NICE current guidance (see next page). Pfizer gives no indication of how many patients may currently be being treated in these ways in the UK or quantify the benefits/disbenefits of doing so. The submission states that:

> Some clinicians consider best supportive care as maintaining patients on imatinib 400 mg/d, regardless of resistance, with the aim of maintaining some symptomatic control.

However, our clinical advisor disagrees that many UK clinicians would consider the ongoing use of imatinib, when progressive disease occurs, to be part of BSC. No other means of BSC are described by Pfizer as part of current service provision. NICE's consultees report that there is variation in the types of BSC offered within the NHS (Burnham and Kaestner, NHS organisation consultee statements to NICE). The submission acknowledges that sunitinib is the first choice of second line treatment for imatinib resistant/intolerant GIST, this is also recognized by NICE's consultees. Furthermore, (as Burnham points out) the European Society for Clinical Oncology (ESMO) guidelines suggest that sunitinib is the standard second line therapy when adherence problems have been ruled out and progression to 800mg/day imatinib has been tried first.(Casali et al. 2008)

Overall, the submission broadly agrees with clinical and expert patient opinion put forward by NICE's consultees and our clinical expert. Evidence about the frequency of use of sunitinib is given on p6 of the submission.

The sunitinib dose is 50mg daily for 4 weeks, followed by a 2-week treatment-free period to complete a 6-week cycle.

The duration of treatment depends on the success of treatment and the tolerability of the drug.

The current NICE guidance for the treatment of GIST is documented (Submission p16) (Technology Appraisal no86, NICE 2004, www.nice.org.uk. Accessed 11 December 2008).

"Imatinib treatment at 400 mg/day is recommended as first-line management of people with KIT (CD117)-positive unresectable and/or KIT (CD117)-positive metastatic gastro-intestinal stromal tumours (GISTs).

Continuation with imatinib therapy is recommended only if a response to initial treatment (as defined below) is achieved within 12 weeks.

Responders should be assessed at intervals of approximately 12 weeks thereafter. Continuation of treatment is recommended at 400 mg/day until the tumour ceases to respond, as defined below.

An increase in the dose of imatinib is not recommended for people receiving imatinib who develop progressive disease after initially responding"

3. Critique of manufacturer's definition of decision problem

3.1. Population

The population considered by the submission is people with unresectable and/or metastatic GISTs after failure of imatinib due to resistance or intolerance.

This is an adequate description of the population under consideration, and concurs with that defined in the NICE Scope. [a]

3.2. Intervention

The intervention is sunitinib malate (Sutent®). Sunitinib gained marketing authorisation, for the treatment of gastrointestinal stromal tumour (GIST) after the failure of imatinib mesylate (Glivec®) treatment due to resistance or intolerance, in June 2006.

> The sunitinib dose is 50mg daily for 4 weeks, followed by a 2-week treatment-free period to complete a 6-week cycle. The duration of treatment depends on the success of treatment and the tolerability of the drug. (Submission p6)

3.3. Comparators

The single comparator was best supportive care, which was taken to mean:

> Treatment to control, prevent and relieve complications and side effects and to improve comfort and quality of life. Within the model it is assumed to include palliative interventions but explicitly *excluded the use of active therapy*. (Submission Table 2, p8, emphasis added)

However, "palliative treatment" with imatinib is used in the economic model in sensitivity analysis of BSC for those with progressive disease.

This terminology for BSC is different from that used in the Scope from NICE, which describes best supportive care as possibly comprising "radiofrequency ablation of the tumour, stenting, embolisation and laser endoscopy". However, our clinical expert considers that Pfizer's description of BSC fits well with clinical practice, with the exception of the use of imatinib

[a] Hereafter this will be referred to as 'the Scope'.

therapy in the face of disease progression as part of BSC. NICE's consultees say that BSC is the only alternative to sunitinib in this population but do not give details of what BSC includes (Kaestner, NHS Organisation consultee statement to NICE).

3.4. Outcomes

The outcomes considered for assessing **clinical effectiveness** were, overall survival (overall survival), time to tumour progression (TTP)[a], progression free survival (PFS), overall confirmed objective response rate (ORR), time to tumour response (TTR), duration of response (DR), duration of performance status maintenance (DPSM), adverse effects of treatment and health-related quality of life, (these outcome measures are defined on p35 of this report).

The outcomes for the **economic analysis** were, incremental cost per quality-adjusted life year, incremental cost per life-year gained, resource utilisation and the cost of treating adverse events. There was no discussion of appropriate ways for measuring these outcomes in the decision problem section. However, these are the appropriate outcomes for this assessment.

3.5. Time Frame

The time horizon for the economic analysis was six years. Our clinical expert agrees that this is an appropriate time frame, with all GIST patients expected to be dead within it.

3.6. Other relevant factors

The submission states that dosing of participants was in accordance with the marketing authorisation for sunitinib in the trials which parameterised the model (Submission p23). In the event of an adverse reaction to sunitinib, the dose was reduced; in the first instance to 37.3mg/day, and if necessary to 25mg/day.

[a] TTP was not specified in the Scope as an outcome measure; PFS is more usually the primary outcome measure in cancer studies.

4. Clinical effectiveness

4.1. Critique of manufacturer's approach

4.1.1. Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

Manufacturer searches were performed in the following databases:

- EMBASE 1996- 2008 Search Date: 16 September 2008
- MEDLINE 1996-2008 Search Date: 16 September 2008
- MEDLINE In-Process 24/07/08-16/09/08 Search Date: 16 September 2008
- The Cochrane Database of Systematic Reviews: Version: 2008 Issue 3
- ASCO Website 2000-2008 Hand Search Date: 16 September 2008
- Documentum (Pfizer Clinical Trials Database) Search Date: 23 September 2008

Separate search strategies were provided for EMBASE, Medline, Medline in-process, and the Cochrane Database of Systematic Reviews (CDSR) by the manufacturer. EMBASE, Medline, Medline in-process database searches are based on a conjunction of terms identifying the GIST population and terms identifying sunitinib as an intervention. For each term, a combination of thesaurus headings (where possible) and free-text search-words was used. For CDSR searches only the GIST population was searched for. No comparators or outcomes were specified to limit the searches in any of these databases.

The EMBASE and Medline searches included a study design filter to limit hits to clinical trials, meta-analysis and reviews. Cochrane and Medline in-process searches did not include any study design filters. There were no additional filters applied to any of the databases.

The manufacturer states that for cohort studies, search results were viewed (scanned) prior to a filter being added.

All the combination of terms within the search strategies to define the GIST population and/or the intervention and resources used were appropriate, replicable, and the resulting hits appear correct given the search date and database/interface used. However the clinical trials, meta-analysis, and review filter used is limited in identifying all controlled trials . It is also unclear why filters were added if the manufacturer was looking at the complete set of results prior to adding the filters. The ERG re-ran the base search strategy adding a more extensive RCT (increased sensitivity) filter of the additional studies found and none were considered to meet all inclusion criteria. The ERG also checked for on-going trials in the

Meta Register of Controlled Trials and in the ClinicalTrials.gov online database as a standard check. No additional trials were found.

4.1.2. Inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

The submission included the following kinds of studies of clinical effectiveness:

> Phase III randomised controlled trials were included if they compared sunitinib with existing standard of care (best supportive care) for the treatment of unresectable or metastatic GIST failing on imatinib 400 mg/d. Primary outcomes of interest were time to tumour progression (TTP). Secondary outcomes were identified as progression free survival (PFS), overall survival (overall survival), objective response rates (ORR), adverse events and health-related quality-of-life (QoL). Studies were excluded if they did not report either of the primary outcomes. Use of data from phase II studies and from non-randomised studies was only considered where there was insufficient evidence from good quality phase III trials. Reports of any studies not available in English were excluded as the time scale of the review precluded time for translation. (Submission p21)

These inclusion and exclusion criteria are appropriate. However, the submission does not explain the process used in study selection (e.g. how many people were involved in reviewing abstracts and titles? How were differences in opinion resolved? What was the process of selection or rejection of retrieved papers?)

4.1.3. Table of identified studies.

Three studies, listed in Table 1, were identified by Pfizer's search strategy.

Table 1: Studies identified by Pfizer's search strategy

Published randomised controlled trials
Phase II studies Sunitinib has been studied in patients with GIST in one phase II study, the results of which were presented at 2 international conferences:
Demetri GD, Desai J, Fletcher JA, Morgan JA, Fletcher CDM, Kazanovicz A, et al. SU11248, A multi-targeted tyrosine kinase inhibitor, can overcome imatinib resistance caused by diverse genomic mechanisms in patients with metastatic gastrointestinal stromal tumour (GIST). 40th Annual Meeting of American Society of Clinical Oncology 2004, Abstract 3001.

Maki RG, Fletcher A, Heinrich MC et al. SU11248 in patients with imatinib-resistant GIST: results form a continuation trial. Oral presentation at: 41st Annual Meeting of American Society of Clinical Oncology 2005, Abstract 9011.
Phase III studies
One pivotal, phase III study was undertaken in patients with GIST and has been the subject of a number of presentations at various meetings. For the purpose of clarity we have listed the key published paper describing the interim results of the trial and the most recent analysis of this data, inclusive of survival data, from ASCO 2008. A6181004
Demetri GD. Oosterom A, Garrett CR, Blackstein ME, Shah MH, Verweij J et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. The Lancet 2006; 368 (9544):1329-1338
Demetri GD, Huang X, Garett CR, Schöffski P, Blackstein ME, Shah MH et al. Novel Statistical Analysis of Long-term Survival to Account for Crossover in a Phase III Trial of Sunitinib versus Placebo in Advanced GIST after Imatinib Failure. The 44th Annual Meeting of the American Society of Clinical Oncology 2008, Abstract 35113
Non-RCT studies [cohort]
Initial results from the ongoing worldwide, phase III, open label treatment use trial in patients with metastatic and/or unresectable GIST were presented at ASCO 2008: A6181036 Reichardt P, Kang YK, Ruka W et al. Detailed analysis of survival and safety with sunitinib in a worldwide treatment use trial of patients with advanced GIST. The 44th Annual Meeting of the American Society of Clinical Oncology 2008, Abstract 10548 (Submission p19)

Three studies, which are part of Pfizer's sunitinib phase II/III clinical trial programme in GIST, are shown below in Table 2. Although trial A6181047 was completed in April 2008, no results have been reported in the submission.

Table 2: Summary of ongoing sunitinib phase II/III clinical trial programme for GIST

Efficacy and Safety Studies	Study start Study end	Design	Treatment duration Follow-up
Phase II Studies A6181047	28/09/2005 10/04/2008	Open-label, uncontrolled, multi-centre study assessing the use of a continuous daily dose of sunitinib. Patients are randomised to a morning or evening continuous daily dose	No. of days on treatment; Median, (range): 319 (17-654) Follow-up: 28 days after the last dose of the study drug
Phase III Studies A6181036 Worldwide treatment use trial	Dec 2007 - ongoing	An open label [cohort] treatment use study designed to permit access to sunitinib prior to regulatory approval and also provide real world efficacy and safety data.	Latest update reported median follow up of 51 weeks (Reichardt et al. 2008)
A6181112	Recruiting	Phases 3b study of patients with GIST who have had progressive disease while on 400mg imatinib. Patients will be randomly assigned to either sunitinib 37.5mg daily or imatinib 800 mg daily.	Patients treated until progressive disease, withdrawal from study or survival until 2 years after final patient recruited (Submission p20)

The submission did not report which studies were excluded or give reasons for exclusions. The QUOROM flow chart in Appendix 2 Fig 1 (Submission p114) does not cross-reference with the above Tables very well and it is not clear what the nature of the 40 full papers reviewed was, as no details are given.

Ultimately two studies were included in the systematic review of clinical effectiveness; these were **A6181004 an RCT**, published as:

Demetri GD. Oosterom A, Garrett CR, Blackstein ME, Shah MH, Verweij J et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *The Lancet* 2006; 368 (9544):1329-1338

Demetri GD, Huang X, Garett CR, Schöffski P, Blackstein ME, Shah MH et al. Novel Statistical Analysis of Long-term Survival to Account for Crossover in a Phase III Trial of Sunitinib versus Placebo in Advanced GIST after Imatinib Failure. *The 44th Annual Meeting of the American Society of Clinical Oncology* 2008, Abstract 35113

And **A6181036 an extended access cohort study**, only published as a conference abstract:

Reichardt P, Kang YK, Ruka W et al. Detailed analysis of survival and safety with sunitinib in a worldwide treatment use trial of patients with advanced GIST. *The 44th Annual Meeting of the American Society of Clinical Oncology* 2008, Abstract 10548

4.1.4. Details of any relevant studies that were not included in the submission?

No relevant studies were found that were not included in the submission.

4.1.5. Description and critique of manufacturers approach to validity assessment

Details of Pfizer's critical appraisal of study A6181004 (RCT), alongside our critique, can be seen in Table 3 below. These are followed by our critique of their critical appraisal of study A6181036 (cohort) in Table 4.

Table 3: Critical appraisal of A6181004 (p32) Demetri et al. 2006/08

Question	Pfizer response	ERG comment
How was allocation concealed?	The clinical site staff were provided patient identifiers, demographic information, and stratification variables only.	This description by Pfizer indicates that allocation concealment may have been very poor. However, Demetri 2006 indicates that it was the clinical site staff who provided this information to a centralised randomisation system who assigned a unique patient identifier. This would provide adequate allocation concealment.
What randomisation technique was used?	Randomisation was done centrally with an interactive voice response system. The centralised randomisation system assigned unique numbers to each patient and provided treatment group information. Patients were stratified by best outcome of previous imatinib treatment (disease progression within 6 months vs. disease progression beyond 6 months of treatment initiation or intolerance to imatinib) and baseline McGill Pain Questionnaire score (0 vs. 1 or more). The 2:1 randomisation was used to minimise the number of patients treated with placebo.	This is an adequate system for randomisation.
Was a justification of the sample size provided?	Yes A total of 281 patients with disease progression were estimated to be needed to detect 50% improvement (Hazard Ratio 0.67) using a two-sided, unstratified log-rank test with an overall two-sided significance level of 0.05 and power of 0.90. It was estimated that 357 patients	Our calculations indicate that this is an adequate sample size to detect the level of improvement specified. The hazard ratio was based on TTP.

Question	Pfizer response	ERG comment
	(238 in sunitinib arm and 119 in placebo arm) would need to be enrolled in order to observe 281 patients with progressive disease by the end of the minimum follow-up period.	
Was follow-up adequate?	Yes 10 December 2003 - 01 January 2005. At the data cut-off (January 2005), patients had been followed-up for maximum of 54 weeks.	Yes, this was adequate as the median time to tumour progression in the sunitinib arm was 27.3 weeks, 95% CI 16.0-32.1. The trial recruitment period was the same as the length of follow up.
Were the individuals undertaking the outcomes assessment aware of allocation?	A6181004 trial consisted of two phases, (1) double blinded and (2) open-label. Patients experiencing disease progression were unblinded, and patients who had been receiving placebo crossed over to open-label treatment with sunitinib; patients who had been receiving sunitinib during the blinded phase study continued to do so after unblinding if, in the opinion of the investigator, there was sufficient evidence of clinical benefit. Treatment was unblinded at the recommendation of the Independent Data and Safety Monitoring Board, and all patients were allowed to cross over to open-label sunitinib.	This response doesn't really answer the question. However, Demetri 2006 clarifies that assessors were blinded until the point of the interim analysis at 54 weeks from the start of the trial (except for those patients whose tumours progressed).
Was the design parallel-group or crossover? Indicate for each crossover trial whether a carry-over effect is likely.	A6181004 trial consisted of two phases, (1) double blinded parallel group and (2) open-label crossover phases (See above).	This is an accurate description of the trial design. A "carry over" effect is likely as members of the unblinded placebo group crossed over to sunitinib at 54 weeks from the start of the trial, possibly masking any longer term benefits of sunitinib in the open-label phase in ITT analysis.

Question	Pfizer response	ERG comment
Was the RCT conducted in the UK (or were one or more centres of the multinational RCT located in the UK)? If not, where was the RCT conducted, and is clinical practice likely to differ from UK practice?	A6181004 trial was a multicentre RCT that enrolled 56 centres from 11 countries, including four UK centres.	Our clinical expert advice is that as all the centres were in the developed world it is unlikely that their clinical practice would differ from that of the UK. However, only twelve participants (4%), from four centres, were from the UK. The participating centres were in: USA, Italy, Australia, UK, France, Singapore, Spain, Canada, Netherlands, Belgium and Switzerland.
How do the included in the RCT participants compare with patients who are likely to receive the intervention in the UK? Consider factors known to affect outcomes in the main indication, such as demographics, epidemiology, disease severity, setting.	There were four UK centres participated in A6181004 trial. There is no evidence to suggest that patients treated in the UK would differ from the overall patient population considered within the A6181004 trial, or importantly that UK patients would respond to sunitinib treatment in a different manner from that observed in the A6181004 trial.	Our clinical expert agrees that the demographic, epidemiologic and setting information presented broadly reflects patient profiles in the UK who are likely to receive this treatment. With the exception that >80% of participants had previously had imatinib at > 400 mg/day. However, the inclusion criteria confined the disease severity of participants to grades 0 or 1 of the ECOG performance status guide; i.e. the most physically able GIST patients. Thus people who were more physically disabled by GIST were excluded. See Table 5 below for a copy of the table, the number of participants in each category and how this compares with study A6181036.
For pharmaceuticals, what dosage regimens were used in the RCT? Are they within those detailed in the Summary of Product	The dose of sunitinib was 50 mg orally once daily as a single agent for 4 consecutive weeks followed by a 2-week rest period to form a complete cycle of 6 weeks. Sunitinib was taken orally in the morning with a glass of water without regard to meals	This is an accurate description of the dosage regimens in the trial and in the summary of Product Characteristics. Patients were monitored for toxicity from sunitinib: Intrapatient dose reduction by one,

Question	Pfizer response	ERG comment
Characteristics?	beginning on Day 1 of the study. This dosing regimen is detailed in the Summary of Product Characteristics (SmPC) (http://emc.medicines.org.uk (See Appendix 1, Section 10.1).	and if needed two, dose levels (to 37.5 mg/day and then to 25 mg/day) were required depending on the type and severity of toxicity encountered. Overall, 62 (31%) vs.12 (12%) experienced a delay or change in dosing in sunitinib and placebo groups respectively. (Pfizer response to ERG questions)
Were the study groups comparable?	In A6181004 trial, study groups were comparable; all characteristics were well balanced between sunitinib and placebo groups (e.g. age, sex, baseline ECOG performance status).	The study groups were well balanced for demographic and ECOG performance status and exposure to imatinib. However, they did differ slightly in their best response to imatinib, with 24.6% of the intervention group showing a partial response and 34.3% of the control group showing a partial response. (Correspondingly, 42.0% of those on sunitinib and 34.3% of those on placebo had stable disease). (Submission Table 3, p26) However, these differences are not statistically significantly different (X^2=2.43, p=0.3) Nevertheless, our clinical expert's advice is that the two groups are comparable in respect to their likely response to sunitinib.
Were the statistical analyses used appropriate?	Yes Kaplan-Meier methods and log-rank tests.	Yes, as reported (Submission p29): Time to tumour progression in each group was assessed using Kaplan-Meier methods and compared with the log-rank test (primary efficacy analysis). A stratified log-rank test and Cox regression models were

Question	Pfizer response	ERG comment
		used to explore the potential effects of the stratification factors and patients' baseline characteristics on the primary endpoint (primary efficacy sub-analyses). Other time-to-event data, including progression-free survival and overall survival, were assessed with Kaplan-Meier methods and log-rank tests. The proportion of patients who achieved an objective tumour response was calculated for each arm and compared by means of a χ^2 test. To explore potential confounding influence of crossover, a post-hoc analysis of overall survival was recently published (Demetri et al, 2008). This analysis was performed using rank preserved structural failure time model (RPSFT) method (Robins & Tsiatis 1991). The RPSFT method estimates the true treatment effect, even in the presence of non-random non-compliance, i.e. the effect that would be realised if all individuals complied with the treatment protocol to which they were assigned, while preserving the unbiased test of the null hypothesis available.
		We have investigated the appropriateness of the RPSFT as a method for dealing with bias caused by crossovers in RCTs. Our expert statistical advice is that this is the most appropriate way of analysing this kind of data. (personal communication Dr Ian White, University of Cambridge, Nov. 27th

Question	Pfizer response	ERG comment
		2008). See Section 4.1.7 for further comment.
Was an intention-to-treat analysis undertaken?	Yes	Yes, as reported on (Submission p29): Analysed study populations included intention-to-treat (ITT; all patients randomised to treatment), modified ITT (all ITT patients with disease progression on imatinib confirmed by central radiology laboratory), and per-protocol (all patients who received at least one dose of assigned study treatment). ITT data are presented for efficacy and per-protocol data for safety; modified ITT data are discussed where relevant. Protocol-defined interim analyses of efficacy and safety were planned after 141 and 211 patients had documented progressive disease. The nominal levels of significance for the interim analyses were determined using the Lan-DeMets procedure with an O'Brien-Fleming stopping boundary (Lan & DeMets, 1983).
Were there any confounding factors that may attenuate the interpretation of the results of the RCT(s)?	A6181004 trial was a crossover study, 84% of patients randomised to receive placebo crossed over to sunitinib arm. In total, 247 (out of 312) patients ultimately received open-label sunitinib. The crossover design has a potential to give a biased estimates of treatment effect for overall survival data.	We agree that the unblinding of patients and assessors may have allowed loss of randomisation, potentially confounding the outcomes.

Source: Demetri et al. 2006; 2008; A6181004 trial report

Table 4: Critical appraisal of A6181036 (Submission p50) Reichardt et al. 2008 (on-going)

Question	Pfizer comment	ERG response
Did the study address a clearly focused issue?	Yes	The objective of the study is to allow access to sunitinib for GIST in patients who might benefit and who would otherwise not have access to sunitinib due to other trials' inclusion criteria, unavailability of trials or lack of regulatory approval where they live. This study aims to obtain broad safety and efficacy data from a large GIST population. This study is still ongoing.
Did the authors use an appropriate method to answer their question?	Yes expanded access Programme to facilitate early access to sunitinib.	Details of this study are only available as a conference abstract, and at www.clinicaltrials.gov/ct2/show/record/NCT00094029?id=A6181036&rank=1 (accessed on 14/11/2008) which does not specify a research question or hypothesis. As the objective of the study is to increase access to sunitinib, this pragmatic open-label, single group design is an appropriate design.
Was the cohort recruited in an acceptable way?	Yes. The cohort in the study A6181036 was representative of a defined population. The study included patients who might benefit from this therapy, but who are ineligible for sunitinib clinical trials because of pre-specified entry criteria, or for whom there are no GIST trials available in a particular country in which regulatory approval has not yet been granted.	No information is provided about how the cohort was recruited.
Was the exposure	Yes Study A6181036 used objective	Yes, the dosing regimen is strictly controlled in repeated six week cycles (four weeks on

Question	Pfizer comment	ERG response
accurately measured to minimize bias?	measurements and clear inclusion/exclusion criteria. These measure are reflect the clinical practice.	two weeks off). Dosing can be reduced if side effects became distressing. Treatment is continued as long as there is evidence of disease control in the judgement of the investigator. Tumour measurements/assessments are performed as per local standard of care. Safety and tolerability are assessed by monitoring adverse events and laboratory abnormalities, and by physical examination. (Submission p50)
Was the outcome accurately measured to minimize bias?	Yes Study A6181036 prospectively assessed objective measurements: AEs, ORR, TTP and overall survival. All tumour measurements/assessments were in the ITT population to minimise bias.	Yes, ITT analysis was used to measure clinical outcomes. (Submission p52)
Have the authors identified all important confounding factors?	Yes Study A6181036 identified all potential confounding factors that might affect the clinical outcome such as age, ECOG PS, and prior imatinib dosage.	Factors that could cause variations in effectiveness are listed as baseline characteristics (Submission p49). Some of these were explored in subgroup analyses (age, ECOG PS, prior imatinib dose and reason for stopping imatinib treatment) (Submission p54).
Was the follow up of subjects complete enough?	Yes. Study A6181036 is currently ongoing. As of December 2007, 1,126 patients were enrolled in the study and 1,117 patients comprised the ITT population. Treatment is continued for as long as there is evidence of disease control in the judgment of the investigator.	This is an ongoing study that is still recruiting. It began in September 2004 and is scheduled to end in December 2009. By December 2007 1117 participants had been analysed by ITT for TTP and overall survival. (Submission p48)

Question	Pfizer comment	ERG response
Was the follow up of subjects long enough?	Survival is monitored for up to 2 years after the last dose of sunitinib.	Safety outcomes will be measured for five years and efficacy outcomes for two years after the last dose of sunitinib according to the submission (Submission p48) and study information on www.clinicaltrials.gov/ct2/show/record/NCT00094029?id=A6181036&rank=1 (accessed on 14/11/2008). As the median survival of patients on sunitinib in Demetri et al. 2006/08 was 72 weeks, this follow-up period is long enough.
What are the results of this study?	Based on results from this treatment-use trial, sunitinib appears to be generally well tolerated in patients with imatinib-resistant or-intolerant advanced GIST who were ineligible for other sunitinib clinical trials. The safety profile observed in this study was similar to that seen with sunitinib in a prior phase III GIST study, with most AEs mild to moderate in severity. Sunitinib was effective in the treatment of patients with advanced GIST after imatinib failure, corroborating previous studies. The median estimated TTP and overall survival from this ongoing study are 41 and 75 weeks, respectively.	At the time of writing the submission, the ITT population had been followed up for a median of 51 weeks (range, 0.1-159) N=1117 (Submission p52). **Safety** Toxicity was graded according to the National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI CTCAE version 3.0). 1=mild, 2=moderate, 3=severe, 4=life threatening/disabling, 5=death related to AE. Fatigue (42%), diarrhoea (39%), and nausea (28%) were the most commonly reported treatment-related non-haematologic (Table 14) (Submission p54) These were mainly grade 1 or 2 in severity. Fatigue (8%), hand–foot syndrome (8%), hypertension (5%), and diarrhoea (5%) were the most commonly reported treatment-related non-haematologic grade 3/4 AEs. Treatment-related hypothyroidism (all grades) was reported in 10% of patients. Treatment-related haematologic AEs

Question	Pfizer comment	ERG response
		included thrombocytopaenia (19%), neutropaenia (18%), and anaemia (14%; Table 15) (Submission p55). **Efficacy** The median estimated TTP was 41 weeks (95% CI: 36–47, Figure 9A). Five hundred and sixty-four patients (50%) in the ITT population were alive at the time of data cut-off. The median estimated overall survival was 75 weeks (95% CI: 68–84, Figure 9B). (Submission p53) Median overall survival of the subgroup analysis suggested that age (age <59 years, 85 weeks vs. age ≥59 years, 65 weeks), ECOG PS (ECOG PS=0–1, 88 weeks vs. ECOG PS=2, 27 weeks), and prior imatinib dosage (imatinib ≤ 400 mg/day, 90 weeks vs. > 400 mg/day, 70 weeks) may be important prognostic factors affecting the clinical outcome in this patient population, but further studies are required to confirm this. (Submission p54)
How precise are the results? How precise is the estimate of the risk***?***	The ITT population was followed up for a median of 51 weeks (range, 0.1–159). The median estimated TTP was 41 weeks (95% CI: 36–47). 564 patients (50%) in the ITT population were alive at time of data cut-off. The median estimated overall survival was 75 weeks (95% CI: 68–84).	The results and estimates of risk are reasonably precise, giving medians and 95% confidence intervals.
Were the results valid?	Yes. Median sunitinib overall survival of 75 weeks was consistent to that	With the limited information available from the study abstract and the submission it is not possible to accurately say if the results

Question	Pfizer comment	ERG response
	seen with a pivotal phase III RCT (74 weeks, see Section 6). The safety profile observed in this study was similar to that seen with sunitinib in a pivotal phase III GIST study (see Section 6.7), with most AEs mild to moderate in severity.	are valid. Although this is not an RCT, the population is reasonably large (N=1117) which adds weight to the results. The results for overall survival do appear similar to those of RCT trial A6181004: Results of 73 (95%CI 61-83) median weeks are quoted in Table 7 (Submission p38), for the RCT in comparison to 75 (95%CI 68-84) median weeks for this cohort study. However, the results for TTP in this cohort study (median weeks = 41 95%CI 36-47) are quite different from the RCT (median weeks = 29 95%CI 22-41). (Submission p39) The details in the submission of the inclusion/exclusion criteria for the RCT are more extensive than those of the cohort study (see Submission pages 24 and 48) as are the details of baseline characteristics (see Submission pages 25 and 49). Both analyses were ITT. The inclusion criteria for study A6181036 show that there was no restriction on ECOG performance status and consequently there are a higher proportion of less well participants in this study. See Table 5 below for details.
Can the results be applied to the local population?	There is no evidence to suggest that patients treated in the UK would differ from the overall patient population considered within the A6181036 trial, or importantly that UK patients would respond to sunitinib treatment in a different manner from that observed in the A6181036 trial.	The inclusion criteria for this study and the baseline characteristics reflect the patient population in the UK who would be eligible for this treatment.

Question	Pfizer comment	ERG response
Do the results of this study fit with other available evidence?	Yes (See Section 6 for details of the pivotal phase III RCT)	The only other available evidence is the RCT A6181004 considered above. Broadly, the evidence for overall survival fits well with this trial, although the TTP results differ. Please see the comments for the second to last question for details.

Table 5: Eastern Cooperative Oncology Group Performance Status, comparing the two included studies' participants.

Grade	ECOG category	A6181004 RCT N=312 N (%)	A6181036 cohort study N=1117 N (%)
0	Fully active, able to carry on all pre-disease performance without restriction	140 (45)	420 (38)
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work	168 (54)	515 (46)
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours	4 (1)	134 (12)
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours		>2 38 (3)
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair		
5	Dead		Missing 10 (1)

Source: Eastern Cooperative Oncology Group, Robert Comis M.D., Group Chair. (Oken et al. 1982)

4.1.6. Description and critique of manufacturers outcome selection

A6181004 RCT

In study A6181004, the primary efficacy endpoint was Time to Tumour Progression (TTP).

Secondary endpoints included overall survival (overall survival), progression-free survival (PFS), overall confirmed objective response rate (ORR), time to tumour response (TTR), duration of response (DR), and duration of performance status maintenance (DPSM) (time from date of randomisation to the last time the performance status was no worse than at baseline or to death from cancer).[a]

Safety and tolerability were assessed by analysis of adverse events. Patient reported outcomes were also assessed and included patient reported pain intensity and general quality of life (QoL) assessments. (Submission p28)

These outcome measures are appropriate and include all necessary indicators of effectiveness. Table 11 on page 53 of the submission reports the adverse events considered by Pfizer.

A6181036 cohort study

The primary outcome measure used in study A6181036 was safety and tolerability; this was measured over five years.

The secondary outcome measures were time TTP, ORR and overall survival. They were measured over a two year period. (Submission p48)

This study has used fewer outcome measures. The absence of PFS, TRR, DR and DPSM mean that potentially important information is not being collected. The difference between PFS and TTP is that PFS includes censoring due to death for any reason.

---

[a] Definitions of outcome measures in trial A6181004:
TTP: Time from randomisation to first documentation of objective tumour progression.
PFS: Time from randomisation to first documentation of objective tumour progression or to death due to any cause (on treatment or within 28 days of last dose).
overall survival: Time from date of randomisation to date of death due to any cause. In the absence of confirmation of death, survival time was censored at the last date the patient was known to be alive.
ORR: Proportion of patients with confirmed complete (CR) or confirmed partial (PR) response according to RECIST, relative to the total population of randomised patients. Confirmed responses were those that persisted on repeat imaging study ≥ 4 weeks after initial documentation of response.
TTR: Time from date of randomisation to first documentation of objective tumour response that was subsequently confirmed. TTR was only calculated for the subgroup of patients with a confirmed objective tumour response.
Source: Demetri et al. 2006

4.1.7. Description and critique of the statistical approach used

4.1.7.1. A6181004 RCT

Analysed study populations included intention-to-treat (ITT; all patients randomised to treatment), modified ITT (all ITT patients with disease progression on imatinib confirmed by central radiology laboratory), and per-protocol (all patients who received at least one dose of assigned study treatment). ITT data are presented for efficacy and per-protocol data for safety; modified ITT data are discussed where relevant. Protocol-defined interim analyses[a] of efficacy and safety were planned after 141 and 211 patients had documented progressive disease. The nominal levels of significance for the interim analyses were determined using the Lan-DeMets procedure with an O'Brien-Fleming stopping boundary (Lan & DeMets, 1983).

Time to tumour progression in each group was assessed using Kaplan-Meier methods and compared with the log-rank test (primary efficacy analysis). A stratified log-rank test and Cox regression models were used to explore the potential effects of the stratification factors and patients' baseline characteristics on the primary endpoint (primary efficacy sub-analyses).

Other time-to-event data, including progression-free survival and overall survival, were assessed with Kaplan-Meier methods and log-rank tests. The proportion of patients who achieved an objective tumour response was calculated for each arm and compared by means of a χ^2 test.

To explore potential confounding influence of crossover, a *post-hoc* analysis of overall survival was recently published (Demetri et al, 2008). This analysis was performed using rank preserved structural failure time model (RPSFT) method (Robins & Tsiatis 1991). The RPSFT method estimates the true treatment effect, even in the presence of non-random non-compliance, i.e. the effect that would be realised if all individuals complied with the treatment protocol to which they were assigned, while preserving the unbiased test of the null hypothesis available from the ITT analysis. (Submission p29)

General approach

Pfizer have analysed appropriate study populations (ITT for efficacy and per protocol for safety). The timing of the interim analyses were correctly determined *a priori*, and time-to-event data (TTP, overall survival and PFS) were properly analysed using Kaplan-Meier

[a] The 'interim analysis' refers to data from the blinded phase of the study only.

methods and log-rank tests. Categorical data (ORR) were appropriately analysed with a χ^2 test. The potentially confounding effects of baseline characteristics on the primary end-point were correctly explored with a stratified log-rank test and Cox regression models.

Patient crossover issues

A recognised problem in the analysis of data from an RCT occurs when a trial becomes "unblinded", and a proportion of participants cross over from the control to the intervention group (i.e. they switch from the treatment to which they were originally randomised). If conventional Kaplan-Meier analysis of ITT comparison groups is used, this may reduce the likelihood of detecting benefit from the intervention, as any gains from the intervention during the unblinded phase will also be experienced by some of those originally allocated to placebo.

This problem is often dealt with by censoring the data of the control group at the point at which they crossover. Pfizer have decided not to take this route because:

> .. this is invalid because patients who cross over are unlikely to be comparable to those who do not. (Submission p31)

We agree with this assertion and statistical expert advice (personal communication from Ian White, MRC Biostatistics Unit, University of Cambridge, 27th November 2008), confirms that in this situation the correct analytical approach is to use the RPSFT method used by Pfizer, as outlined below. It should be noted that this method was used post-hoc when the Kaplan-Meier ITT analysis had failed to show a benefit from sunitinib. In addition, although White had some oversight of the methods and commented that the results 'look about right', he did not actually conduct the analysis (and received no remuneration for his advice to Pfizer). Therefore, although it is the correct approach, we cannot be certain that the methods were applied correctly.

In particular the RPSFT method:

- Estimates the times of death of patients randomised to placebo as if they had not crossed over to receive the intervention
- Is based on the ITT population
- Is a non-parametric model that produces a randomisation-based effect estimator

Rationale for using Rank Preserved Structural Failure Time (RPSFT) model

Because of treatment crossovers in the placebo arm, the intention-to-treat analysis estimates the benefit of starting sunitinib immediately compared to deferring the start of sunitinib. However, the relevant parameter from a decision-making perspective is the benefit of starting sunitinib compared to not starting sunitinib. We therefore corrected for treatment crossovers in the placebo arm.

A common approach to this problem is to censor placebo arm patients at the point of crossover, but this is invalid because patients who cross over are unlikely to be comparable to those who do not. The only way to avoid selection bias is to use a method based entirely on the comparability of groups as randomised (White 2005b).

We used the method of Robins and Tsiatis (1991) which is the only method currently available in the literature that can correct for time-dependent treatment changes in survival data while respecting the randomisation. This method is based on the accelerated failure time model $U = T_{start} + \exp(\psi)(T-T_{start})$ where T is the observed event time, T_{start} is the time of starting treatment, U is the event time that would have been observed if no treatment had been given, and the parameter ψ represents the causal effect of having started treatment. ψ is 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. 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 values in the placebo arm.

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.

The initial Hazard Ratios and 95% Confidence Intervals for all analyses utilizing the RPSFT method are derived from the Cox regression analysis as presented in the 2008 publication (Demetri et al, 2008). Advice received since the publication is that because this procedure is based on the randomisation, it does not change the level of evidence against the null hypothesis and therefore a different analytical approach needs to be used. Adopting this results in a wider 95% confidence interval and for transparency we have therefore also presented revised estimates from our updated analysis. It should also be noted that after review by an independent statistician Pfizer was made aware of a number of methodological

issues with the original RPSFT analysis, we therefore took the opportunity of the availability of the final data to re-conduct the analysis. This updated analysis has been externally reviewed and approved. (Submission p31)

However, Pfizer report three different values for the numbers of participants who crossed over from the placebo arm to sunitinib; on p27, Table 4 says that 59 (56%) patients taking the placebo had entered the open label phase, but p36 of the submission says that 99 (94%) patients randomised to placebo crossed over and p38 of the submission states that over 80% crossed over to sunitinib. These different values for crossing over were explained by Pfizer, in response to our question, as follows:

The data cut-off and interim analysis (termination of blinded phase) of the study A 6181004 was performed in January 2005 at which point there were 59 patients crossed over the sunitinib arm from placebo arm. However, patient enrolment continued until May 2005 at which point there were 243 patients in the sunitinib arm and 118 patients in the placebo arm. Of the 188 patients in the placebo arm 99 patients (84%) had subsequently received open label sunitinib. Therefore, statement in the submission referring to 99 patients (over 80%) crossed over to sunitinib (page 36) is the correct value and have been used in the economic model.

Table 6 below highlights some of the possible approaches to analysing RCT data when participants have crossed over from placebo to intervention.

Table 6: The advantages and disadvantages of different analytical approaches to RCTs with crossovers

Analytical approach	Advantages	Disadvantages
Double blind phase only: ITT with data censoring at crossover	Randomised	Ignors longer follow-up data from open label phase
ITT of all data ignoring the crossover	ITT	Gives a biased estimate
RPSFT: ITT analysis accounting for crossover effects	Analyses by ITT but accounts for the effects of crossovers	Not in common usage

A6181036 cohort study, treatment-use trial

The number of patients to be enrolled was not predetermined and no inferential analyses were planned due to the nature of this study. Descriptive statistics (mean, standard deviation,

median, minimum and maximum values) are utilized to summarize all continuous data. All categorical data are summarized using frequencies and percentages.

The study population for updated efficacy and safety analyses includes all patients enrolled in the study receiving at least one dose of sunitinib (ITT population). TTP and overall survival are estimated using the product-limit method of Kaplan and Meier. (Submission p50)

These statistical methods are appropriate for the design of this study.

4.1.8. Summary statement

The submission contains all the relevant studies and the relevant data within those studies. The submitted evidence also adequately reflects the decision problem defined in the submission.

4.2. Summary of submitted evidence

4.2.1. Summary of results

4.2.1.1. Study A6181004 (Phase III RCT)

Primary endpoint results – time to tumour progression (TTP)

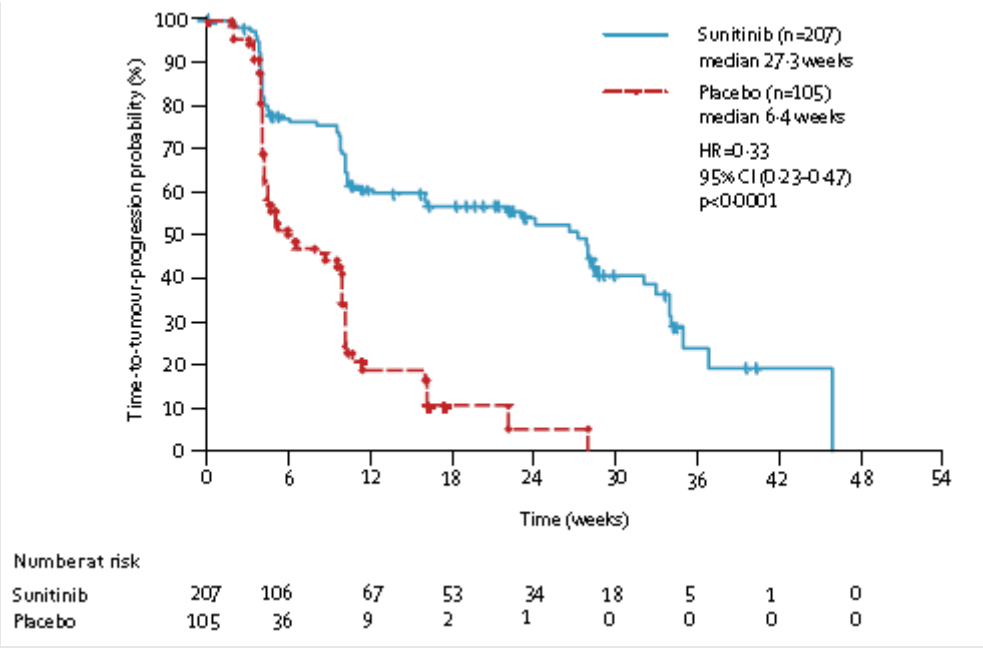
Time to Tumour Progression, defined as the time from the first dose of study drug to first documentation of progressive disease. Disease progression was assessed by using RECIST[a] criteria and evaluated by the investigators as well as an independent, third-party radiology laboratory.

The blinded phase of the A6181004 trial was terminated early (January 2005) when a planned interim analysis revealed significantly longer TTP in patients treated with sunitinib compared with patients treated with placebo. Overall, 82 (40%) of sunitinib treated patients and 67 (64%) of placebo treated patients, had disease progression at the time of analysis. (Submission p35)

---

[a] Response Evaluation Criteria in Solid Tumours (RECIST):
CR (complete response) = disappearance of all target lesions
PR (partial response) = 30% decrease in the sum of the longest diameter of target lesions
PD (progressive disease) = 20% increase in the sum of the longest diameter of target lesions
SD (stable disease) = small changes that do not meet above criteria

Figure 1: Kaplan-Meier estimates of time to tumour progression (ITT population), interim analysis.



Source: Demetri et al. 2006

Median time to tumour progression for the ITT population, the primary study endpoint, was more than four times as long with sunitinib (27.3 weeks, 95% CI 16.0–32.1) as with placebo treatment (6.4 weeks 4.4–10.0; HR 0.33, 95% CI 0.23–0.47; p<0.0001) on the basis of central radiology laboratory assessment (Figure 1, Table 7). A clear difference between the treatment groups was noted around week 4. The greater TTP obtained with sunitinib compared with placebo was confirmed by the stratified analysis when controlling for stratification factors (HR 0.33, 95% CI 0.22–0.46; p<0.0001). (Submission p36)

It is not clear what the stratification factors were, but it seems likely that they were the baseline characteristics shown in Figure 5, p39.

Table 7: Summary of Time to Tumour Progression (ITT)

Phase of Study	Median (weeks; 95% CI)		HR (95% CI)	P-value
	Sunitinib	Placebo		
Blinded phase	27.3 (16.0–32.1)	6.4 (4.4–10.0)	0.33 (0.23–0.47)	<0.0001
Entire study (blinded + open label)	28.6 (22.0-41.0)	N/A	N/A	N/A

N/A-Not applicable

Source: Demetri et al. 2006; 2008

The median TTP for the 99 patients randomised to receive placebo who crossed over to sunitinib treatment was 28.6 weeks (95% CI 22.0-41.0) and similar to that observed during the blinded phase in the sunitinib group. The summary of TTP results are shown in Table 6 above. (Submission p36)

The results from the primary outcome measure show a clear improvement in TTP from sunitinib when compared to placebo for imatinib resistant/intolerant GIST patients.

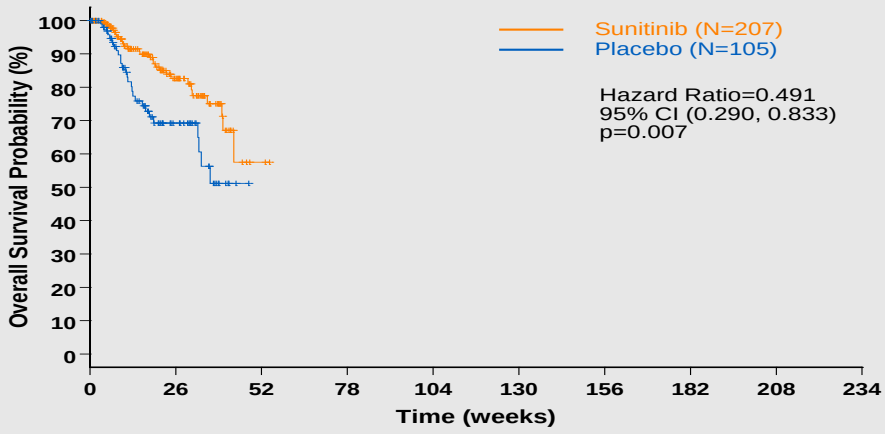
Secondary endpoint results

Overall Survival (overall survival)

Submission p37

Interim analysis (blinded phase) demonstrated that overall survival obtained with initial sunitinib treatment was better that that obtained with placebo (HR 0.49, 95% CI 0.29–0.83; p=0.007, Figure 2). Since more than half the patients in the sunitinib group were still alive at the time of the interim analysis, a median overall survival value could not be calculated (Table 8).

Figure 2: Kaplan-Meier Curves of Overall Survival from the Interim Analysis



Source: Demetri et al. 2006 (Submission p37)

Although the crossover situation was relatively simple (placebo control; only crossover in one direction; no other treatments available, etc.), the conventional statistical methods still could not provide an unbiased estimation of the treatment effect. In this situation, the RPSFT

approach becomes the most appropriate choice to estimate the treatment effect of overall survival when it is diluted or confounded by the crossover in a trial like this. (Submission p37)

Table 8: Overall Survival estimated using different statistical methods (ITT Population) (p36)

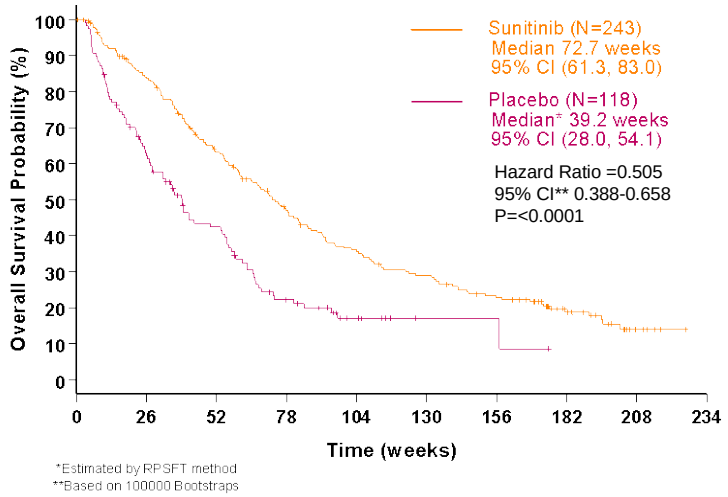
Phase of Study	Median (weeks; 95% CI)		HR (95% CI)	P-value
	Sunitinib	Placebo		
Blinded phase* Kaplan-Meier Method	Not reached	Not reached	0.491 (0.290-0831)	0.007
Entire study (blinded + open label) RPSFT Method	72.7 (61.3-83.0)	39.2 (28.0-54.1)	0.505 (0.388-0.658)	<0.001
Entire study (blinded + open label) Kaplan-Meier Method	72.7 (61.3-83.0)	64.9 (45.7-96.0)	0.876 (0.679-1.129)	0.306

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

The analysis using RPSFT approach demonstrated an estimated median overall survival for the placebo group of 39.2 weeks (95% CI 28.0.-54.1) based on the ITT population. This revealed a statistically significant sunitinib treatment effect (HR 0.505, 95% CI 0.388-0.658, p<0.0001) comparable to that of the interim overall survival results (**Error! Reference source not found.** and Figure 3). The re-calculated confidence intervals for the HR from the RPSFT approach, following external recommendation, are 0.262-1.134. (Submission p39)

It should be noted that the ITT analysis using Kaplan-Meier methods show no statistically significant overall survival benefit from sunitinib for the entire study analysis (see Figure 4). Given the high numbers crossing over to sunitinib, we agree that this method has problems with masking the possible benefits of the intervention and that the RPSFT is the most appropriate method in these circumstances.

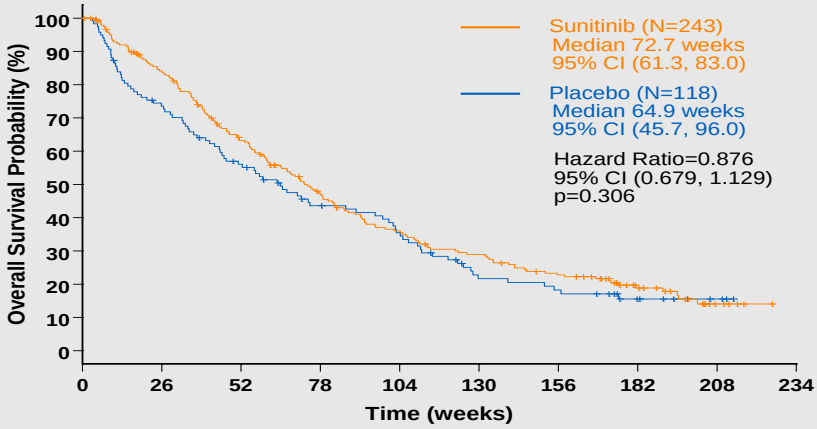
Figure 3: Final Overall Survival estimated with correcting for crossover (RPSFT Method)



Source: Demetri et al. 2008 (revised estimates, for details see Section 6.3.5)

Figure 4 demonstrates the final overall survival estimated without correcting for crossover using the Kaplan-Meier method. (Submission p39)

Figure 4: Final Overall Survival estimated without correcting for crossover (Kaplan-Meier Method)



Source: Demetri et al. 2008 (Submission p39)

The analysis of overall survival using the Kaplan-Meier method includes data collected after unblinding of individual patients randomised to placebo who experienced disease

progression and crossed over into open-label treatment, the analysis is not fully blinded and includes potential influence of open-label sunitinib treatment on patients randomised to placebo.

The median sunitinib overall survival of 74 weeks was consistent with that of currently ongoing expanded-access trial A6181036 (see section 6.2.1). The latest update (median follow up of 51 weeks) of this trial demonstrated overall survival survival of 75 weeks (95% CI: 68–84) in patients treated with sunitinib (Reichardt et al. 2008). (Submission p40)

However, this is not the case if participants with ECOG performance status of >2 are excluded.

Progression-Free Survival (PFS)

Eighty-nine (43%) patients treated with sunitinib and 70 (67%) patients treated with placebo had disease progression, or were dead at the time of this interim analysis.

There was a statistically significant difference in median PFS with sunitinib (24.6 weeks 95% CI: 12.1 to 28.3 weeks) compared with 6.4 weeks (95% CI: 4.4 to 10.0 weeks) on placebo (HR 0.333, 95% CI: 0.238 to 0.467; p< 0.001). The risk of disease progression, or death in the sunitinib group was approximately 33% of that in the placebo group. These results are consistent with the analysis of TTP in this study.

Overall Confirmed Objective Response Rate (ORR)

Fourteen (6.8%) patients on sunitinib and no (0%) patients on placebo responded (PR) indicating a significantly higher response rate on sunitinib (treatment difference: 6.8%; 95% CI: 3.34 - 10.18; p=0.006).

Time to Tumour Response (TTR)

TTR was only determined for patients with a confirmed objective response (CR or PR); no patients on placebo had a response. Fourteen patients (6.8%) patients on sunitinib had experienced a response at the time of this interim analysis. The median TTR on sunitinib was ten weeks (95% CI: 9.7 to 16.1 weeks).

Duration of Response (DR)

DR was only determined for patients with an objective response (complete response or partial response) who subsequently experienced disease progression or death. Of the 14 sunitinib patients with a confirmed objective response, only three of these patients (21%) had

subsequently experienced progression at the time of this analysis. The observed DRs were 29.9, 23.3, and 15.9 weeks.

Duration of Performance Status Maintenance (DPSM)

Seventy-nine (38%) sunitinib treated patients and 38 (36%) placebo treated patients, had experienced performance status worsening at the time of analysis. The median DPSM was 18.9 (95% CI: 12.1 - 33.9 weeks) vs. 16.1 weeks (95% CI: 6.1 weeks to (upper limit could not be calculated because the data were not mature).

EQ-5D Health State Profile

EQ-5D health state profile was analysed to compare the difference in treatments between patients who answered "no problem" and who answered "at least some problem". Compliance was generally high, with > 75% expected patients completing questionnaire at each time point.

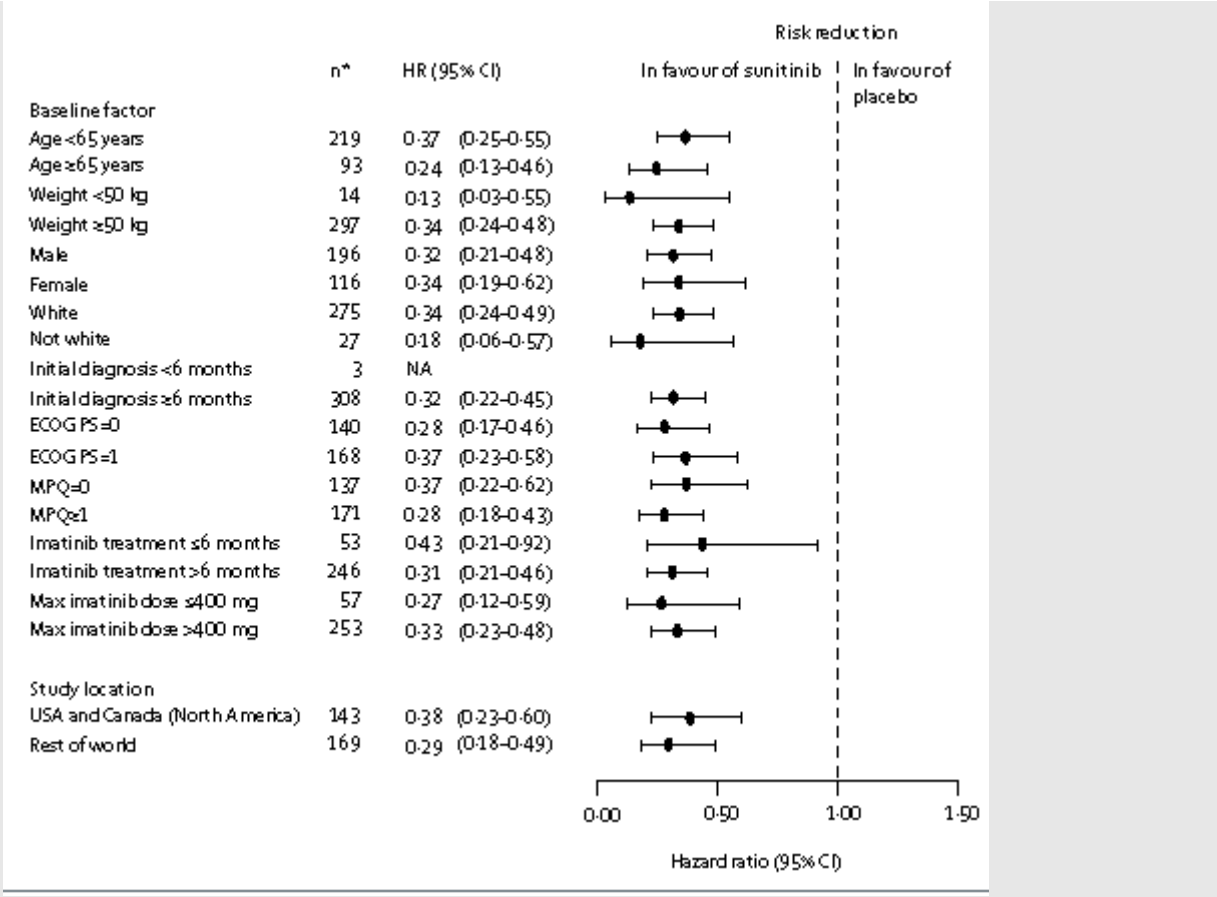
The numbers and percents of patients who reported “no problem,” “some problem,” or “extreme problem,” and the number and percentage of patients reporting at least some problem (some or extreme problem), were measured by time point, dimension of the EQ-5D descriptive system, and treatment arm. The statistical significance tests of between-treatment differences did not show significant differences. Utility values were calculated for the cost effectiveness analysis (Section 7).

Subgroup analysis

The influence of baseline factors on the treatment effect in different populations was analysed by using a Cox proportional hazards model including TTP on previous imatinib mesylate therapy (≤ vs. > 6 months), baseline MPQ-PPI[a] score (0 vs. ≥ 1), age (< vs. ≥ 65 years), sex (male vs. female), race (white vs. non-white), ECOG performance status (0 vs. 1), and time since initial diagnosis with GIST (< vs.≥ 6 months), controlling for each factor 1 at a time (Figure 5). (Submission p41)

[a] MPQ-PPI McGill Pain Questionnaire- present pain intensity

Figure 5: Result of Cox Proportional Hazards Analysis of TTP in various subgroups



Source: Demetri et al. 2006; A6181004 trial report

In the primary analysis, the treatment effect was statistically significant ($p < 0.001$) overall and when controlling for each individual baseline factor. The only baseline factors considered of interest to pursue in the multivariate model (i.e., with $p < 0.10$) were TTP on previous imatinib mesylate therapy (≤ vs. > 6 months; hazard ratio: 1.657; 95% CI: 1.109 to 2.475; $p = 0.014$), indicating longer TTP for patients who had previously experienced longer (> 6 months) disease control on imatinib mesylate, gender (male vs. female; HR: 1.457; 95% CI: 1.029 to 2.064; $p = 0.034$), indicating longer TTP for female patients, and weight (HR: 0.991; 95% CI: 0.983 to 1.000; $p = 0.041$) indicating a trend for longer TTP in patients with lower body weight. The results for the treatment effect were similar in the Modified ITT and Per Protocol populations. The Cox proportional hazards models were repeated, with treatment-by-baseline factor interactions included, and the hazard ratios and 95% CIs are presented in Figure 5 above. For all subgroups, the hazard ratio was significantly less than 1.000 (i.e., the 95% CI did not overlap 1.000), indicating that all subgroups, defined on the basis of baseline factors, benefited from sunitinib. (Submission p42)

The subgroup analysis shows that all the subgroups benefited from sunitinib. Multivariate analysis also indicated that women, people who had experienced longer disease control with imatinib (>6 months) and those who weighed <50 kg, had significantly longer TTP.

In response to our question to Pfizer about the numbers of people who were randomised to sunitinib and who continued to take it after disease progression, we received the following response:

At the time of database cut off for the interim report (January 1, 2005), 19 of 51 sunitinib patients (N=207) with a disease progression crossed over to open-label treatment (Please see Table 4, page 27 of the submission). Subsequently, 152 patients who had been randomised to receive sunitinib treatment (N=243) received open-label sunitinib treatment.

Overall of the 243 patients originally randomised to sunitinib, 54 continued to take sunitinib in progressive disease.

This means that 22% of those with progressive disease who were in the sunitinib group continued to take the drug. We are unable to say whether this would have had any effect on the outcome measures and are unaware of any evidence from RCTs that suggests that sunitinib has a palliative effect in progressive disease or slows the rate of progression. However, Pfizer's search strategy would not have picked up observational studies that may have addressed these issues.

The summary table of outcomes shown below (Table 9) clearly shows the benefits of sunitinib over BSC for the outcomes of time to tumour progression, overall survival and progression free survival.

Table 9: Summary results for study A6181004 RCT

Outcomes		Sunitinib n=207				Placebo n=105						
	Analysis	n	%	Median time in weeks	95% CI	n	%	Median time in weeks	95% CI	Hazard ratio	95% CI	p
Time to tumour progression												
Number with progression [blinded phase interim analysis]	KM ITT	82	40	27.3	[16.0-32.1]	67	64	6.4	[4.4-10.0]	0.33 or 0.346*	[0.23-0.47]	0.0001
Entire study blinded and x overs [started with placebo and crossed over to sunitinib]		99		28.6	[22.0-41.0]							
Overall survival												
Blinded phase interim analysis	KM ITT	207				105				0.49	[0.29-0.831]	0.007
Entire study blinded and x overs	RPSFT ITT	243		72.7	[61.3-83.0]	118		39.2	[28.0-54.1]	0.505	[0.388-0.658] OR [0.262-1.134] *	<0.001
Entire study blinded and x overs	KM ITT	243		72.7	[61.3-83.0]	118		64.9	[45.7-96.0]	0.876	[0.679-1.129]	0.306
Six month survival		168	81			68	65					
Progression free survival				24.6	[12.1-28.3]			6.4	[4.4-10.0]	0.333	[0.238-0.467]	<0.001
No. with disease progression or dead at interim ananlysis [54 weeks]		89	43			70	67					
Overall confirmed objective response rate		14	6.8			0	0					0.006
Time to tumour response	RPSFT ITT			10	[9.7-16.1]	0	0					
Duration of response[of those who went onto progression]		3.00	21	29.9, 23.3, 15.9								
Duration of performance status maintenance		79	38	18.9	[12.1-33.9]	38	36	16.1	[6.1-not calculated]			

* Both values given in the submission, KM: Kaplan-Meier, RPSFT: Rank Preserved Structural Failure Time, ITT: intention to treat

Safety results for A6181004 RCT

In the A6181004 trial, the safety was analysed in the per protocol (PP) population, which consists of all patients who received at least 1 capsule of study medication with treatment assignments designated according to actual study treatment received. On sunitinib, the PP population included 202 of 207 randomised patients (98%; 4 randomised patients had been treated but did not have drug dosing data available and are excluded from the PP population; 1 patient randomised to sunitinib but did not receive treatment), and, on placebo, the PP population included 102 of 105 randomised patients (97%). The overall adverse experience during the blinded phase is summarised by treatment arm in Table 10.

Definitions of mild, moderate and severe adverse events were not provided.

Table 10: Overall Adverse Event Experience during the Blinded Phase (PP population)

Category	**Sunitinib (N=202)**	**Placebo (N=102)**
Patients with ≥ 1 adverse event, n (%)	190 (94)	99 (97)
Patients with ≥ 1 serious adverse event, n (%)	70 (35)	25 (25)
Patients with ≥ 1 treatment-related adverse event, n (%)	168 (83)	60 (59)
Patients with≥ 1 treatment -related serious adverse event, n (%)	40 (20)	5 (5)
Patients who discontinued due to adverse events, n (%)	19 (9)	8 (8)
Patients who died within 28 days of treatment, n (%)	13 (6)	8 (8)

Source: A6181004 trial report

Treatment-related adverse events, serious adverse events, and treatment-related serious adverse events appeared to be more common on the sunitinib arm. Nineteen (9%) patients in the sunitinib group and eight (8%) in the placebo groups discontinued treatment because of adverse events.

Table 11 presents a summary of adverse events that occurred with a frequency of at least 5% greater on sunitinib than on placebo in the per-protocol population. (Submission p43)

Table 11: Adverse events that occurred with a frequency of at least 5% greater on sunitinib than on placebo in per-protocol population

	Sunitinib (n=202)			Placebo (n=102)		
	Grade 1/2	Grade 3	Grade 4	Grade 1/2	Grade 3	Grade 4
Non-haematological*						
Fatigue	58 (29%)	10 (5%)	0 (0%)	20 (20%)	2 (2%)	0 (0%)
Diarrhoea	52 (26%)	7 (3%)	0 (0%)	8 (8%)	0 (0%)	0 (0%)
Skin discolouration	50 (25%)	0 (0%)	0 (0%)	6 (6%)	0 (0%)	0 (0%)
Nausea	47 (23%)	1 (1%)	0 (0%)	10 (10%)	1 (1%)	0 (0%)
Anorexia	38 (19%)	0 (0%)	0 (0%)	5 (5%)	1 (1%)	0 (0%)
Dysgeusia	36 (18%)	0 (0%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Stomatitis	30 (15%)	1 (1%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Vomiting	30 (15%)	1 (1%)	0 (0%)	5 (5%)	1 (1%)	0 (0%)
Hand-foot syndrome	19 (9%)	9 (4%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Rash	24 (12%)	2 (1%)	0 (0%)	5 (5%)	0 (0%)	0 (0%)
Asthenia	18 (9%)	6 (3%)	0 (0%)	2 (2%)	2 (2%)	0 (0%)
Mucosal inflammation	24 (12%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Dyspepsia	22 (11%)	1 (1%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)
Hypertension	15 (8%)	6 (3%)	0 (0%)	4 (4%)	0 (0%)	0 (0%)
Epistaxis	14 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Hair-colour changes	14 (7%)	0 (0%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)
Dry mouth	13 (6%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)
Glossodynia	11 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Haematological						
Anaemia†	117 (58%)	7 (4%)	0 (0%)	59 (58%)	2 (2%)	0 (0%)
Leucopenia	104 (52%)	7 (4%)	0 (0%)	5 (5%)	0 (0%)	0 (0%)
Neutropenia	86 (43%)	17 (8%)	3 (2%)	4 (4%)	0 (0%)	0 (0%)
Lymphopenia	80 (40%)	18 (9%)	1 (1%)	31 (30%)	2 (2%)	1 (1%)
Thrombocytopenia	72 (36%)	8 (4%)	1 (1%)	4 (4%)	0 (0%)	0 (0%)

Data are number (%). *Treatment-related. †Anaemia was included in the table, despite a difference of less than 5% between the treatment groups, because of its frequency and clinical relevance in GIST.

Source: Demetri et al. 2006 (Submission p44)

An explanation of the grading system in the above table was not provided.

Adverse events were generally mild to moderate in intensity and easily managed by dose reduction, dose interruption, or standard supportive medical treatments. Overall numbers of events of any grade for the most common treatment-related adverse event, fatigue, were 68 (34%) for sunitinib and 22 (22%) for placebo (Table 9). The incidence of grade 3 fatigue was similar between the treatment groups; there were no cases of grade 4 fatigue. Other serious treatment-related non-haematological adverse events that seemed to be experienced more frequently on sunitinib treatment included hand-foot syndrome, diarrhoea, and hypertension;

serious haematological adverse events also seemed to be more frequent with sunitinib than with placebo.

Patients Completing the Study

Three hundred and twelve patients were enrolled into the study. 207 (66%) patients were randomised to sunitinib and 105 (34%) to placebo. At the cut-off for analysis 134 (65%) of the sunitinib treated patients compared with 34 (32%) of the placebo-treated patients were ongoing in double-blind treatment. Nineteen (9%) sunitinib treated patients and 59 (56%) placebo treated patients went on to receive open-label sunitinib.

Adverse Events Resulting in Dose Reduction

Blinded phase: Sixty two (31%) patients treated with sunitinib reported 149 adverse events that resulted in either a delay or change in dose. Twelve (12%) patients treated with placebo reported 16 adverse events that resulted in either a delay, or change in dose. Adverse events that most commonly led to dose reductions, or delays for sunitinib treated patients included gastrointestinal disorders (11%) and blood and lymphatic disorders (6%). No commonly occurring adverse event led to dose reduction or treatment delay in patients treated with placebo.

Adverse Events Resulting in Discontinuation of Study Treatment

Blinded phase: Nineteen (9%) patients treated with sunitinib and eight (8%) patients treated with placebo discontinued treatment due to adverse events. Eleven (5%) and two (2%) were considered as treatment related.

Crossover phase: Seventeen patients (22%) discontinued because of adverse events during the crossover phase, including disease progression (9 patients, 12%) and fatigue (2 patients, 3%). Five patients (6%) discontinued because of adverse events of which at least one was considered to be related to the study treatment; 12 patients (15%) discontinued because of adverse events that were solely related to the study disease.

Treatment-Related Adverse Events

Blinded phase: Fatigue, diarrhoea, skin discolouration, nausea, anorexia, stomatitis, asthenia, constipation, dysgeusia, vomiting, palmar-plantar erythrodysesthesia syndrome, rash, anaemia, mucosal inflammation, dyspepsia, and hypertension occurred 5% more frequently with sunitinib compared with placebo. No adverse events occurred more frequently with placebo.

Crossover phase: The most common adverse events during the crossover phase included constitutional symptoms (fatigue, asthenia, anorexia, and pyrexia), gastrointestinal symptoms (abdominal pain, nausea, dyspepsia, vomiting, diarrhoea, and stomatitis), and skin and subcutaneous disorders (skin discoloration). Apparently more common during the crossover phase than the blinded phase were oedema peripheral, disease progression, and dyspnoea; the increase in these events may reflect the more advanced disease state of patients in the crossover population. (Patients were required to experience disease progression during the blinded treatment phase before crossing over to open-label treatment.)

Severe Treatment-Related Adverse Events (Grade 3 and 4)

Blinded phase: The most frequently reported sunitinib treatment-related severe adverse events included fatigue, diarrhoea, palmar-plantar erythrodysesthesia syndrome and anaemia. Only two kinds of severe treatment-related adverse events occurred more frequently with placebo, abdominal pain and disease progression. Seventeen (8%) patients treated with sunitinib versus one patient (1%) treated with placebo experienced treatment-related adverse events, of a maximum severity (grade 4).

Crossover phase: One patient (1%) experienced a grade 5 (fatal) adverse event. Three patients (4%) experienced adverse events with a maximum severity of grade 4; the grade 4 events were fatigue, leucopoenia, thrombocytopenia, perirectal abscess, mucosal inflammation, and transient ischemic attack (each 1 patient, 1%). Treatment-related grade 3 events experienced by 2 or more patients during the crossover phase were fatigue (7 patients, 9%); neutropoenia (5 patients, 6%); nausea (4 patients, 5%); anorexia and palmar-plantar erythrodysesthesia syndrome (each 3 patients, 4%); and leucopoenia, stomatitis, and abdominal pain (each 2 patients, 3%).

Serious Treatment-Related Adverse Events

Blinded phase: Forty (20%) patients treated with sunitinib and five (5%) patients treated with placebo reported serious treatment related adverse events. Treatment-related serious adverse events that were experienced by patients treated with sunitinib include anaemia (2%), abdominal pain and tumour haemorrhage (2%), thrombocytopenia (1%), diarrhoea (1%), pulmonary embolism (1%), neutropoenia (1%), melaena (1%), nausea (1%), vomiting (1%) and pyrexia (1%). Treatment-related serious adverse events that were experienced by patients on placebo were vomiting (2%) and fatigue (2%).

Crossover phase: Thirty one patients (40%) experienced serious adverse events during the crossover phase, including 16 patients (21%) who experienced treatment-related serious adverse events. The only treatment-related serious adverse events that were experienced by

more than 1 patient during the crossover phase were nausea and vomiting (each 2 patients, 1%).

Deaths

Blinded phase: Twenty-three (11%) patients treated with sunitinib and 11 patients (11%) treated with placebo died during the blinded phase, or after discontinuing the randomised double blind treatment phase without crossing over to receive open-label sunitinib.

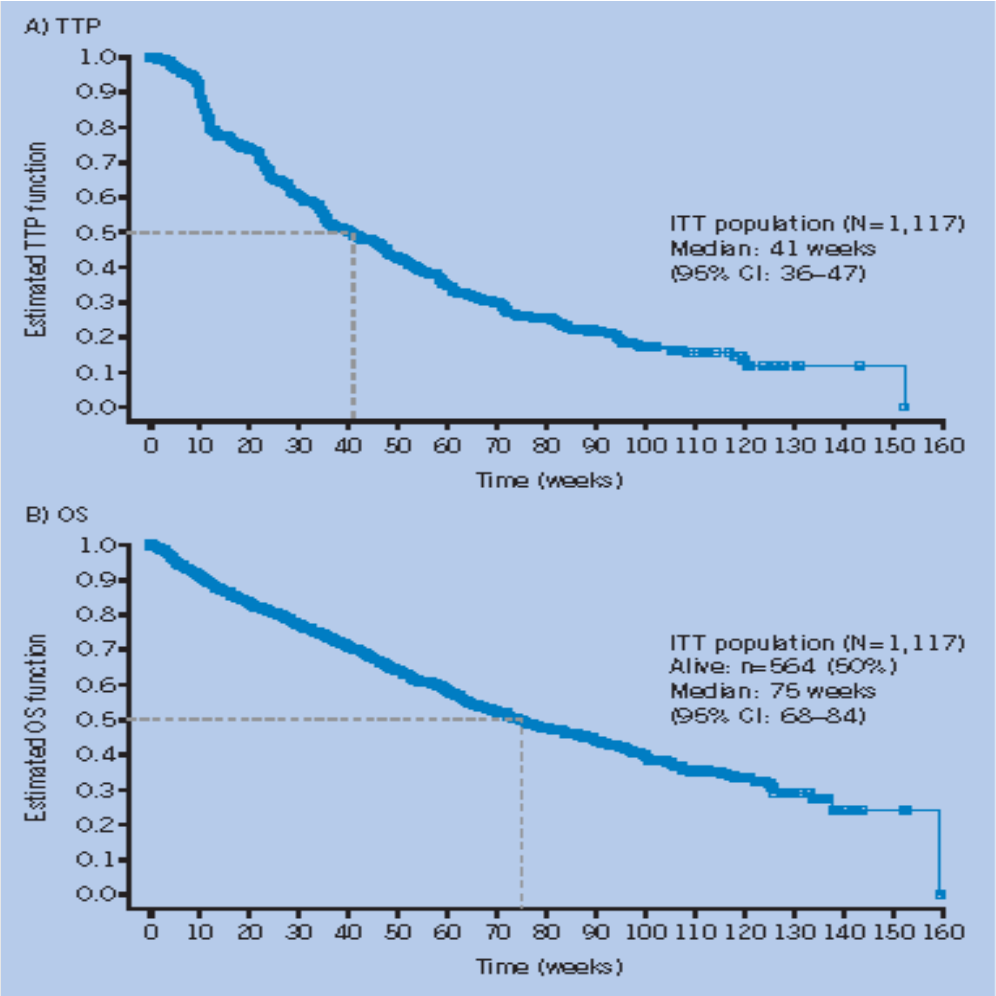
Crossover phase: Eighteen patients (23%) died on study during the crossover phase. Seventeen on-study deaths (22% overall) were secondary to progressive disease, and 1 (1%) was related only to the study drug. Four patients (5%) died during the follow-up period (i.e., more than 28 days after their last treatment); all deaths during the follow-up period were secondary to progressive disease. (Submission p46)

Most adverse events in this study were mild to moderate. However, in the blinded phase of the study, 40 (20%) people taking sunitinib compared to five (5%) people taking placebo had at least 1 severe treatment-related adverse event. Nevertheless, most people continued with their treatment as only 19 (9%) of those in the sunitinib group left the study due to adverse events, compared to eight (8%) in the placebo group. Nevertheless, 62 (31%) people with adverse symptoms had a delay or reduction in their sunitinib dose. Thirty six people discontinued treatment with sunitinib, due to adverse events, over the whole study period. The proportion of deaths during the blinded phase was similar for both groups (sunitinib n=23 (11%), placebo n=11 (11%)).

4.2.1.2. Study A6181036 cohort study

The ITT population was followed up for a median of 51 weeks (range, 0.1–159). Figure 6 below show TTP and overall survival in the ITT population

Figure 6: TTP and overall survival with sunitinib in the ITT Population

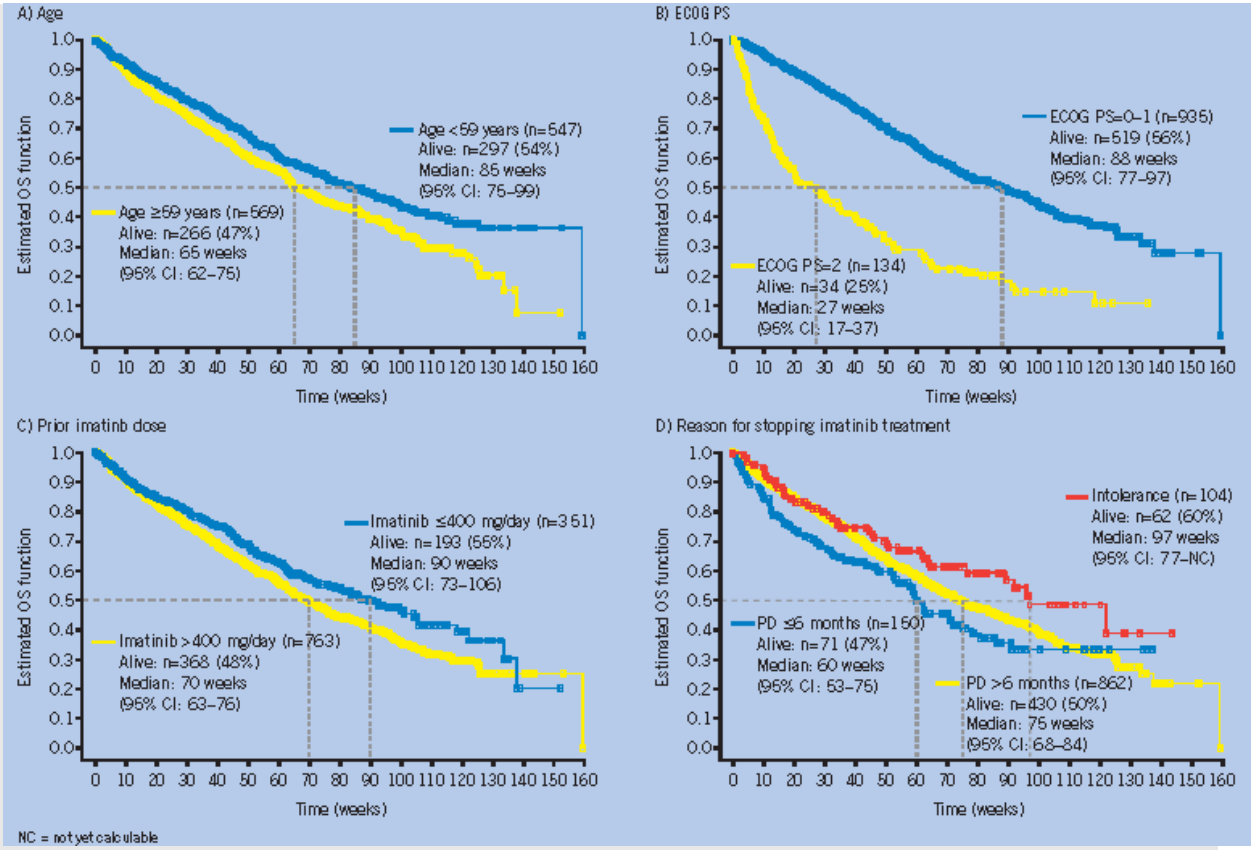


Source: Reichardt et al. 2008 (Submission p53)

The median estimated TTP was 41 weeks (95% CI: 36–47, Figure 6A). Five hundred and sixty-four patients (50%) in the ITT population were alive at the time of data cut-off. The median estimated overall survival was 75 weeks (95% CI: 68–84, Figure 6B).

Figure 7 below compares survival data for subgroups based on individual baseline factors and prior imatinib treatment history.

Figure 7: overall survival for subgroups based on individual baseline factors and prior imatinib treatment history.



Source: Reichardt et al. 2008 (Submission p54)

Median overall survival of the subgroup analysis suggested that age (age <59 years, 85 weeks vs. age ≥59 years, 65 weeks), ECOG PS (ECOG PS=0–1, 88 weeks vs. ECOG PS=2, 27 weeks), and prior imatinib dosage (imatinib ≤ 400 mg/day, 90 weeks vs. > 400 mg/day, 70 weeks) may be important prognostic factors affecting the clinical outcome in this patient population, but further studies are required to confirm this.

The overall results of the ongoing study A6181036, demonstrated that sunitinib is effective in the treatment of patients with advanced GIST after imatinib failure with median estimated TTP and overall survival of 41 and 75 weeks, respectively.

The characteristics listed above as possibly predicting response to sunitinib are different from those found in the RCT (which are, gender, weight and time on imatinib before PD). Unsurprisingly, from the analysis of ECOG PS it can be seen that those with a performance status of 2 had much shorter median overall survival than those with a performance status of 0-1 (0-1= 88 weeks (95%C 77-97), 2= 27 weeks (95%CI 17-37). In comparison, the

participants of the RCT, who all had ECOG PS = 0-1, had a median overall survival of 73 weeks (95%CI 62-83).

The summary table below (Table 12) shows that, despite the different patient populations, the results for overall survival are similar to those of RCT trial A6181004. Results (reported as median weeks) for the RCT, of 73 weeks (95%CI 61-83) (Submission Table 7, p38) and (Submission Fig 7, p39), in comparison to 75 weeks (95%CI 68-84) for this cohort study. (Submission Fig 9, p53)

However, the results for TTP in the cohort study (median weeks = 41 95%CI 36-47) are quite different from the RCT (median weeks = 29, 95%CI 22-41). (Submission p39)

Table 12: Summary table of results from study A6181036 cohort study

	Sunitinib n=1117 N alive at follow up	%	Median time in weeks	95% CI	Analysis
Time to tumour progression	1117	100	41	36-47	KM ITT
Overall survival	564	50	75	[68-84]	KM ITT
Follow-up was for a median of 51 weeks (range, 0.1-159)					
	Overall survival by subgroup				
Age	N alive at follow up	%	Median time in weeks	95% CI	Analysis
<59 years n=547	297	54	85	[75-99]	KM ITT
≥59 years n=569	266	47	65	[62-75]	KM ITT
[ECOG PS]					
PS=0-1 n=519	519	56	88	[77-97]	KM ITT
PS=2 n=134	34	25	27	[17-37]	KM ITT
Prior imatinib dose					
≤ 400 mg/day n=351	193	55	90	[73-106]	KM ITT
> 400 mg/day n=763	368	48	70	[63-76]	KM ITT
Reason for stopping imatinib treatment					
intolerance n=104	62	60	97	[77-not calculated]	KM ITT
PD >6 months n=862	430	50	75	[68-84]	KM ITT
PD ≤6 months n=150	71	47	60	[53-75]	KM ITT

KM: Kaplan-Meier, ITT: intention to treat

Safety Results – A6181036

The majority of adverse events observed in this study were mild to moderate. Fatigue (42%), diarrhoea (39%), and nausea (28%) were the most commonly reported treatment-related

non-haematologic AEs (Table 13: Most common (≥ 20%) treatment-related non-haematologic adverse events).

Table 13: Most common (≥ 20%) treatment-related non-haematologic adverse events

AE	Sunitinib (N=1,117)			
	Grades 1–2 n (%)	Grade 3 n (%)	Grade 4 n (%)	Total* n (%)
Fatigue	374 (33)	88 (8)	3 (0.3)	465 (42)
Diarrhoea	383 (34)	55 (5)	0 (0)	439 (39)
Nausea	291 (26)	23 (2)	0 (0)	315 (28)
Hand–foot syndrome	210 (19)	86 (8)	2 (0.2)	298 (27)
Anorexia	230 (21)	22 (2)	1 (0.1)	253 (23)
Mucosal inflammation	229 (21)	20 (2)	1 (0.1)	250 (22)
Stomatitis	228 (20)	20 (2)	1 (0.1)	249 (22)
Hypertension	188 (17)	58 (5)	2 (0.2)	248 (22)
Vomiting	210 (19)	25 (2)	2 (0.2)	237 (21)

*Twenty-three grade 5 events deemed to be treatment-related have occurred in the study, including one case of diarrhoea and one of nausea

Source: Reichardt et al. 2008 (Submission p54)

These were mainly grade 1 or 2 in severity. Fatigue (8%), hand–foot syndrome (8%), hypertension (5%), and diarrhoea (5%) were the most commonly reported treatment-related non-haematologic grade 3/4 AEs.

Treatment-related hypothyroidism (all grades) was reported in 10% of patients. Treatment-related haematologic AEs included thrombocytopaenia (19%), neutropaenia (18%), and anaemia (14%; Table 14).

Table 14: Treatment-related haematologic Adverse Events

AE	Sunitinib (N=1,117)			
	Grades 1–2 n (%)	Grade 3 n (%)	Grade 4 n (%)	Total n (%)

Thrombocytopaenia	156 (14)	44 (4)	13 (1)	213 (19)
Neutropaenia	119 (11)	76 (7)	6 (1)	201 (18)
Anaemia	108 (10)	37 (3)	14 (1)	159 (14)

Source: Reichardt et al. 2008 (Submission p55)

Most events were grade 1 or 2. Febrile neutropaenia was reported in only three patients. Treatment-related AEs related to cardiac function included heart failure, congestive heart failure, myocardial infarction, reduced ejection fraction, and pulmonary oedema (all ≤ 0.6%; Table 15).

Table 15: Treatment-related AEs related to cardiac function (A6181036)

AE	Sunitinib (N=1,117)				
	Grades 1–2 n (%)	Grade 3 n (%)	Grade 4 n (%)	Grade 5 n (%)	Total n (%)
Heart failure*	1 (0.1)	4 (0.4)	0 (0)	2 (0.2)	7 (0.6)
Congestive heart failure	2 (0.2)	2 (0.2)	2 (0.2)	0 (0)	6 (0.5)
Myocardial infarction	0 (0)	0 (0)	1 (0.1)	1 (0.1)	2 (0.2)
Ejection fraction†	2 (0.2)	1 (0.1)	0 (0)	0 (0)	3 (0.3)
Pulmonary oedema	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0.1)

*Includes acute heart failure. †Includes reduced ejection fraction.

Source: Reichardt et al. 2008 (Submission p55)

Overall, the safety profile observed in the study A6181036 was acceptable with mild to moderate AEs and similar to that seen with sunitinib in a pivotal phase III RCT.

The most common adverse events in this study were similar to those of A6181004, fatigue, diarrhoea and nausea. The majority of these cases were mild to moderate. The numbers of people who needed to reduce their dose of sunitinib due to adverse events is not reported.

Twenty three people died due to treatment related reasons during the follow up period (median 51 weeks, range 0.1-159).

4.2.2. Critique of submitted evidence syntheses

No evidence syntheses were presented in the clinical effectiveness section of the submission.

4.2.3. Summary of clinical effectiveness

The submission contains all the relevant studies and the relevant data within those studies. The submitted evidence also adequately reflects the decision problem defined in the submission.

The submission from Pfizer included two studies; a good quality RCT, A6181004, n=312 (Demetri et al. 2006;Demetri et al. 2008) and an ongoing, cohort study A6181036, n=1117 (Reichardt et al. 2008, abstract only).

In order to deal with potential bias in the results caused by patients crossing over from placebo to sunitinib after unblinding, the RCT's time dependent outcomes were analysed using RPSFT methods. These methods are more appropriate than censoring the data at the primary endpoint, as they allow analysis for a longer follow up period by estimating the differences between the groups as if the participants in the placebo group had not crossed over to sunitinib; i.e. by ITT.

The population of the RCT was restricted to those in ECOG performance status grades O-1, whilst that of the cohort study was inclusive of grades 0-4.

The results for overall survival are similar in these studies; with the RCT reporting median survival of 73 weeks (95%CI 61-83) in comparison to 75 weeks (95%CI 68-84) for the cohort study. However, there was greater median OS for the physically fitter patients with ECOG grad 0-1 in the cohort study (cohort: 88 weeks (95%CI 77-97), than those in the RCT: 73 weeks (95%CI 61-83). The results for median TTP in the cohort study (41 weeks 95%CI 36-47) are quite different from the RCT (29 weeks 95%CI 22-41), (Submission p39). These results may be influenced by the different ECOG performance status of the two study populations (see Table 5), and a greater median OS for the ECOG grade 0-1 in the cohort study (RCT: 73 weeks (95%CI 61-83), cohort study: 88 weeks (95%CI 77-97).

Time to event study data from the RCT were analysed both conventionally using the unadjusted interim ITT data, and more unusually, using the RPSFT model to adjust for the bias introduced in the ITT analysis from participants in the placebo group crossing over to receive sunitinib. Independent expert statistical advice is that the RPSFT approach is appropriate in these circumstances.

In the subgroup analyses of the RCT, baseline characteristics were explored using Cox proportional hazard models. In the analysis of the primary outcome of TTP, the treatment effect remained statistically significant overall and when each factor was controlled for individually ($p<0.001$). In the cohort study, the subgroup analysis indicated that age, ECOG performance status and prior imatinib dosage may be prognostic factors for response to sunitinib. Further research is needed to determine which factors predict a positive response to sunitinib.

In the RCT 168 (83%) of participants in the sunitinib group had an adverse event of any severity, compared to 60 (59%) in the placebo group. In both studies most adverse events were mild to moderate. Overall, the most commonly occurring symptoms were fatigue (RCT: 68 (43%); cohort study: 465 (42%)) and diarrhoea (RCT: 59 (29%); cohort study: 439 (39%)).

The main clinical results of TTP and overall survival have been summarised in Table 16 overleaf.

Table 16: Summary results of the main outcome measures in the two included studies

		Sunitinib				Placebo						
Time to tumour progression	**Analysis**	**n**	**%**	**Median time in weeks**	**95% CI**	**n**	**%**	**Median time in weeks**	**95% CI**	**Hazard ratio**	**95% CI**	**p**
A6181004 number with progression [blinded phase interim analysis]	KM ITT	82	40	**27**	**16-32**	67	64	6.4	4.4-10.0	0.33 or [0.346]	0.23-0.47	0.0001
A6181004 entire study blinded and x over phases [started with placebo and crossed over to sunitinib]		99		**29**	**22-41**							
A6181036		1117	100	**41**	**36-47**							
Overall survival												
A6181004 blinded phase interim analysis	KM ITT	207				105				0.49	0.29-0.831	0.007
A6181004 entire study blinded and x over phases	RPSFT ITT	243		**73**	**61-83**	118		**39**	**28-54**	0.505	0.388-0.658 OR 0.262-1.134 both given in submission	<0.001
entire study blinded and x over phases	KM ITT	243		**73**	**61-83**	118		**65**	**46-96**	0.876	0.679-1.129	0.306
A6181036		564	50	**75**	**68-84**							

KM: Kaplan-Meier, ITT: intention to treat

5. Economic evaluation

In this chapter, we assess the cost-effectiveness analysis submitted by Pfizer. Overall, we found Pfizer's economic model and evaluation to be good, with the following three exceptions.

- First, Pfizer originally made a serious miscalculation in the cost-effectiveness estimate based on the unadjusted ITT overall survival data for BSC: as a result they quoted £34,649 per QALY, whereas the correct value is £77,107 per QALY if the first cycle of treatment is free and £93,062 per QALY if the first cycle is costed. In their response to our questions, Pfizer acknowledged their error.

- Second, they did not include the acquisition cost of sunitinib in progressive disease for patients originally randomised to sunitinib (which was a feature of the RCT from which effectiveness estimates have been used).

- Third, there are several serious errors in the probabilistic sensitivity analysis (PSA) (Section 5.4.3.2).

In this chapter, we start with a summary of the systemic review of cost-effectiveness studies presented by Pfizer and the methods used in the economic evaluation (Section 5.1). Then we present a critique of the methods they used (Section 5.2). This is followed by a description of Pfizer's results (Section 5.3) and our comment on their validity (Section 5.4). Finally there is a summary of the uncertainties surrounding the economic evaluation (Section 5.5).

5.1. Overview of manufacturer's economic evaluation

5.1.1. Summary of Pfizer's systematic review of cost-effectiveness studies

5.1.1.1. Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

Searches were performed in the following databases from "inception" to 16 September 2008 (except where stated):

MEDLINE

MEDLINE In-Process July-September 2008

EMBASE

NHS EED

HTA database of the Centre for Review and Dissemination

ECCO 14, ASCO, and IAPOR: Conference Proceedings (via Internet)

Separate search strategies were provided for each of the databases in the manufacturer's submission (Appendix 3/in answer to initial questions). All database searches are based on thesaurus (were possible) and free-text words for the GIST population combined with terms for finding economic evaluations. None of the database searches use a filter specifically for finding individual utility or quality of life information. There are no additional limits or filters on any of the search strategies. No specific details of terms used within the conference proceeding search are provided. The databases and search strategies reported are appropriate for identifying specific economic evaluations for GIST but not for additional parameters.

5.1.1.2. Search results

Pfizer identified two cost-effectiveness studies (Chabot et al. 2008;Contreras-Hernandez et al. 2008), these are described on pages 58-9 of their submission. Both studies are available in full paper form, and both use clinical effectiveness parameters derived from the interim analysis of the RCT included in their clinical effectiveness systematic review (Demetri et al. 2006); whereas the model supplied by Pfizer to NICE uses longer follow-up efficacy data from the same trial. Given that the cost-effectiveness estimates are highly sensitive to the clinical effectiveness of sunitinib versus BSC, this limits the comparability of these studies to Pfizer's model. Comparability is also limited because both cost-effectiveness studies were not performed from a UK cost perspective. The included model-based studies were:

Chabot et al. (2008) modelled the cost-effectiveness of sunitinib versus BSC in GIST patients intolerant or resistant to imatinib, from the perspective of the Canadian healthcare system. The ICER was Can$79,884 per QALY.

Contreras-Hernandez et al. (2008)(Contreras-Hernandez et al. 2008) modelled second-line treatment of sunitinib and imatinib in patients with advanced GIST, from the perspective of the Mexican healthcare system. The ICER for sunitinib versus BSC was US$46,108 per life year.

Our search for published economic evaluations of sunitinib for GIST identified three studies: the same two studies as above, plus a model-based analysis that Pfizer submitted to the

Scottish Medicines Consortium (2006) (SMC). The analysis for the SMC considered sunitinib compared to BSC in patients with malignant GIST who are intolerant to or unresponsive to imatinib or who develop progressive GIST after treatment with imatinib. Resource use was estimated from the Scottish treatment guidelines, trial resource use and expert opinion. We do not know whether the first cycle of sunitinib was assumed to be free of charge. Utility values were taken from the RCT of sunitinib versus BSC (Demetri et al. 2006). The ICER was £65,000 per QALY. It is not made clear, but we assume that the efficacy data was taken from the interim analysis, reported in Demetri et al (2006).

In contrast, the base case ICER for the economic analysis presented by Pfizer to NICE was £27,365 per QALY, which assumes that the first cycle of sunitinib is free to the NHS. We assume that the large difference in the ICER compared with the model presented to the SMC was due to the fact that the SMC model used interim efficacy data, whereas the NICE model used much more mature efficacy data. In addition, the RPSFT method was used in Pfizer's model for NICE, but we assume that this method was not used in the SMC model (this method of analysis is discussed in Section 5.4.1.1). In Section 5.3, we stress that the RPSFT method yields a far lower ICER than the unadjusted ITT data.

Pfizer's economic model submitted to NICE

We now turn to the economic evaluation that Pfizer presented to NICE. Pfizer reports cost per QALY estimates for sunitinib versus BSC for patients with unresectable and/or metastatic GISTs after failure of imatinib due to resistance or intolerance. The model was built in Microsoft Excel. It is described in detail on p59-81 of Pfizer's report. Here, we summarise the main features of the model.

5.1.2. Natural history

Pfizer's model is based on the simple Markov model of metastatic renal cell carcinoma (RCC) which we (PenTAG), submitted to NICE in the assessment of four drugs for renal cell carcinoma (RCC) (Submission p63). Pfizer's model is therefore based on three discrete health states: progression-free survival (PFS), progressive disease (PD) and death. All patients enter the model in PFS, having failed imatinib therapy. They remain in PFS until they experience disease progression or death. Once patients enter PD, they remain there until death.

For sunitinib and for BSC separately, a Weibull curve describes the number of patients alive over time (overall survival), and another Weibull curve describes the number of patients in PFS over time (Submission p64-5 Figs. 13 - 14). Fitting of these curves to trial data is

described in the following section. For each treatment, the number of patients in PD at any time is calculated as the number alive minus the number in PFS at that time.

The time horizon of the model is 6 years, and the model cycle length is 6 weeks (to reflect the duration of a cycle of treatment with sunitinib). A half-cycle correction is modelled.

5.1.3. Treatment effectiveness within submission

Treatment effectiveness is taken exclusively from the RCT of sunitinib versus BSC (Demetri et al. 2006).

In all analyses, Pfizer consider PFS for sunitinib and BSC and overall survival for sunitinib based on the ITT data. For the base case Pfizer adjust the ITT overall survival for BSC using the RPSFT method (see Section 4.1.7.1 of this report). In a sensitivity analysis, they use the unadjusted ITT overall survival for BSC.

Effectiveness data were not taken from the interim analysis of the RCT reported in Demetri et al (2006)(Demetri et al. 2006). Instead, longer follow-up survival data from the RCT was used. These data were similar, but not exactly the same, as that reported in abstract form(Demetri et al. 2008). Follow up for the interim overall survival data is about 1 year, whereas follow up is about 4.5 years for the mature data used in Pfizer's model.

Weibull curves were fitted to the Kaplan-Meier data by regression, (see p64 of the Submission). Pfizer considered two methods for fitting the PFS and overall survival curves for sunitinib;

A. Weibull curves for sunitinib and BSC were fitted independently.

B. The Weibull curve for BSC was fitted, and the Weibull curve for sunitinib was calculated by multiplying the Weibull parameter λ for the BSC curve by the hazard ratio.

Pfizer found that Method B did not give a good fit to the sunitinib Kaplan-Meier curve (Figures 13 and 14, Submission p64-65). Therefore, they used Method A in the base case analysis, and Method B in sensitivity analyses.

Pfizer state that they did not perform subgroup analyses because clinical effectiveness was not presented by subgroup in the RCT (Submission p61). However, we note that hazard ratios for TTP are presented by several prognostic factors in Figure 3 of Demetri et al (2006) reproduced on p42 of the Submission.

5.1.4. Health related quality of life

Health state utilities were taken from the main RCT, in which the EQ-5D questionnaire was used. During PFS, the average utility of patients treated with sunitinib was 0.731, and for BSC it was 0.781. In PD in the model, the utility for both sunitinib and BSC patients was set equal, at 0.577. The utility of 0.731 for patients on sunitinib in PFS was calculated as a weighted average of the utility at the end of each four week treatment period of sunitinib, and the utility at the end of the two week rest period. The utility of 0.577 for PD was calculated from the average EQ-5D scores at confirmation of disease progression in the two arms of the RCT.

Pfizer state that they did not explicitly model disutility due to adverse events because this would require numerous assumptions, and because such disutility is already implicit in the lower utility value for sunitinib compared to patients in PFS (Submission p74-5).

5.1.5. Resources and costs

The model uses costs based on the NHS and PSS perspective. Costs included were drug cost, disease management costs (such as appointments, scans and tests), some adverse event costs and palliative care costs associated with death.

5.1.5.1. Drug costs

In the model, patients treated with sunitinib are assumed to take one tablet per day for four weeks, followed by two weeks without sunitinib. The cost of imatinib (used in sensitivity analysis) and sunitinib are taken from the British National Formulary (BNF 2008). Sunitinib is available in two pack sizes: 28 and 30 capsules. A 28-capsule pack of 50mg tablets costs £3,138.80 and a 30-capsule pack costs £3,363 (the same cost per tablet)(BNF 2008). Pfizer assume no drug costs for patients in the BSC arm.

The relative dose intensity of sunitinib is an important model parameter. It is defined as the average cost of sunitinib allowing for the actual treatment interruptions and dose reductions experienced in the RCT divided by the cost of sunitinib given no treatment interruptions or dose reductions. Pfizer assume a relative dose intensity for sunitinib of 88.6%, which they state was measured in the RCT (Submission p75). In a sensitivity analysis, they assume a dose intensity of 100%, because of the uncertainty of dose intensity in clinical practice (Submission p67).

In the base case, Pfizer assume that the first 6-week cycle of sunitinib is free to the NHS. In reply to our questions on their report, Pfizer state that the Department of Health have agreed that the first cycle of sunitinib would be free for GIST patients.

In the model, patients stop sunitinib treatment at disease progression. However, in the RCT, sunitinib dosing is continued or modified based on clinical and laboratory findings and continued during disease progression for some patients. Dose adjustments are recommended to manage grade 3 or 4 toxicity related to sunitinib.

5.1.5.2. Disease management costs

Resource use was not measured directly in the RCT (Submission p75). Instead, Pfizer consulted a panel of four consultant oncologists currently treating GIST patients in the UK, to estimate the following resource use (Submission p66). For patients in PFS, irrespective of treatment, they modelled resource use associated with outpatient monitoring, scans and tests. They assumed one outpatient appointment per month, at £112 per visit, one CT scan every 3-months, at £141 per scan, and standard blood tests once per month, at £3 per test. The associated costs per resource use item were assumed to be the same as in the renal cell carcinoma (RCC) model developed by us, the ERG, and were taken from the NHS Reference Costs. The resultant cost of medical management in PFS over a 6-week cycle was £224.

In the base case, Pfizer assumed the same resource use in PD as for PFS.

In a sensitivity analysis, Pfizer assumed the same resource use as we previously assumed for patients treated for RCC (Thompson-Coon et al. 2008). Specifically, for PFS, we assumed the same resource use as described above for patients treated with sunitinib, but a different resource use for patients on BSC: 1 GP visit per month and 1 CT scan every 6 months. This gives a total cost of £83 per 6-week cycle for patients on BSC in PFS. For patients on sunitinib and BSC in PD, we assumed 1 GP visit per month, 1.5 community nurse visits per month, morphine sulphate daily, and no test or outpatient visits, giving a cost of £435 per patient per 6-week cycle (Thompson-Coon et al. 2008). Under these assumptions, the ICER is £29,000 in Pfizer's GIST model (Submission p84).

In another sensitivity analysis, Pfizer assumed that 70% of patients receive imatinib in PD. The cost of imatinib therapy was correctly assumed to be £2,246 per cycle (BNF 2008).

Pfizer assumed a cost of death of £3,923, averaged over hospital and hospice stays, based on an estimate from the literature(Coyle et al. 1999). Pfizer assumed no cost of death in a sensitivity analysis.

5.1.5.3. Adverse events costs

In the base case, Pfizer included only costs associated with Grade 3 or 4 adverse events, since they expected these to incur additional NHS costs. The incidence of Grade 3 or 4 adverse events were taken from the RCT (Demetri et al. 2006). Pfizer assumed that only two adverse events would incur a treatment cost: hypertension and hypothyroidism. They assumed that treatment for hypertension would continue for the duration of PFS and would comprise two GP visits per year (£36 per visit), two district nurse visits per year (£25 per visit), and medication for hypertension (£246 per year), giving a total cost of £367 per year. The results reported in Table 2 p22 and Table 27 p 83 of the Submission, imply these costs were only applied to the sunitinib patients.

Although not an adverse event reported in the RCT, Pfizer's clinical panel suggested that sunitinib treated patients would be monitored for hypothyroidism and where necessary given a thyroxin. They assumed that 10% of patients would receive a thyroxin at £2.28 per cycle, based on the price of Levothyroxine.

All costs are inflated to 2007/8 prices.

5.1.6. Discounting

Future costs and benefits were discounted at 3.5% as specified in the NICE reference case (National Institute for Health and Clinical Excellence 2008).

5.1.7. Sensitivity analysis

One-way sensitivity analyses and probabilistic sensitivity analyses are reported.

5.1.8. Model validation

On p80-81 of their Submission, Pfizer state how they tested the internal validity of their model. Nonetheless, as reported in Section 5, we found one very important logical error in their model. Pfizer originally quoted cost-effectiveness using the unadjusted ITT overall survival data for BSC as £34,649 per QALY, whereas the correct value is £77,107 per QALY.

Pfizer have since acknowledged this error and provide a corrected figure of £77,107 per QALY in their response to our questions.

Pfizer's GIST model is based on the structure of the renal call carcinoma (RCC) model, developed by the present ERG, PenTAG (Submission p63). We therefore tried to re-create the results of Pfizer's GIST model using our RCC model, by setting the parameter values in our RCC model to those used in the GIST model. This was useful for two purposes. First, to check the internal consistency and logic of Pfizer's model. Second, to account for the difference in cost-effectiveness of sunitinib versus interferon-α for RCC, £71,500 per QALY using our model, and the cost-effectiveness of sunitinib versus BSC for GIST, £27,400 per QALY using Pfizer's model.

We were able to recreate all Pfizer's GIST model outputs (Table 20), including the base case deterministic ICER of £27,365 per QALY, almost exactly (£27,355 per QALY) by changing several of the parameters in our RCC model, see Appendix 1. This is powerful evidence that there are now no logical errors or internal inconsistencies in the base case calculated from Pfizer's GIST model. The large difference in cost-effectiveness between sunitinib versus interferon-α for RCC and sunitinib versus BSC for GIST was almost completely due to differences in the clinical effectiveness (survival curves) between RCC and GIST. Differences in the other parameters, such as costs and utilities, explained almost none of the difference in cost-effectiveness (Appendix 1).

5.2. Critique of approach used

In this section, we comment on Pfizer's approach / methodology. First, we consider the model against checklists of good practice. Then we critically appraise the model structure and data. In Section 5.4, we comment on Pfizer's results with reference to their methods.

5.2.1. Critical appraisal frameworks

We considered Pfizer's economic evaluation against the following widely-used study quality checklists: NICE Reference Case(National Institute for Health and Clinical Excellence 2008)(Table 17), Drummond et al. (1997)(Drummond et al. 1997) (Table 18), and Philips et al. (2006)(Philips et al. 2006) for decision model-based economic evaluations (Table 19).

Table 17: Critical appraisal checklist based on NICE Reference Case(National Institute for Health and Clinical Excellence 2008)

NICE reference case requirement		Critical Appraisal	Reviewer comment
Defining the decision problem	The scope developed by the Institute	✓	
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	✓	Comparator is BSC.
Perspective on costs	NHS and PSS	✓	
Perspective on outcomes	All health effects on individuals	✓	Disutility of adverse events not modelled directly, but indirectly via lower utility in PFS for sunitinib compared to BSC.
Type of economic evaluation	Cost-effectiveness analysis	✓	
Synthesis of evidence on outcomes	Based on a systematic review	✓	Demetri et al (2006) RCT of sunitinib *v.* BSC
Measure of health benefits	QALYs	✓	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	✓	EQ-5D survey during RCT
Source of preference data for valuation of changes in HRQL	Representative sample of the public	✓	EQ-5D survey
Discount rate	3.5% pa for costs and health effects	✓	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓	

Table 18: Critical appraisal checklist from Drummond and colleagues(Drummond et al. 1997).

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	✓	-
Is there a clear description of alternatives (i.e. who did what to whom, where, and how often)?	✓	Sunitinib *v.* BSC for patients with unresectable and/or metastatic GISTs after failure of imatinib due to resistance or intolerance
Has the correct patient group / population of interest been clearly stated?	✓	No patient subgroups.
Is the correct comparator used?	✓	BSC
Is the study type reasonable?	✓	Markov cost-utility model
Is the perspective of the analysis clearly stated?	✓	UK NHS & PSS
Is the perspective employed appropriate?	✓	-

Item	Critical Appraisal	Reviewer Comment
Is effectiveness of the intervention established?	✓	Quality of single RCT(Demetri et al. 2006) is good. Sunitinib clearly improves TTP compared to BSC. However, overall survival data of BSC is compromised by substantial post-treatment crossover from BSC arm to sunitinib.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	✓	6-year time horizon. After 6 years, virtually all modelled patients are dead. Hence the time horizon is effectively life time, and appropriate.
Are the costs and consequences consistent with the perspective employed?	✓	All costs from UK NHS & PSS perspective.
Is differential timing considered?	✓	-
Is incremental analysis performed?	✓	-
Is sensitivity analysis undertaken and presented clearly?	✓	Univariate and probabilistic sensitivity analyses clearly presented, but serious errors with PSA.

Table 19: Critical appraisal checklist of Philips et al. (2006)(Philips et al. 2006) for model-based analyses

Dimension of quality			Comments
Structure			
S1	Statement of decision problem/objective	✓	Sunitinib *v.* BSC for patients with unresectable and/or metastatic GISTs after failure of imatinib due to resistance or intolerance.
S2	Statement of scope/perspective	✓	NHS and PSS perspective. Cost and benefit inputs are consistent with the perspective. Scope of model stated.
S3	Rationale for structure	✓	Cohort model is appropriate.
S4	Structural assumptions	✓	Model assumptions are mostly explained clearly in the report. Overall, we are satisfied with the structural assumptions. Weibull functions were fitted to PFS and overall survival from the single RCT.
S5	Strategies / comparators	✓	See S1.
S6	Model type	✓	Cohort model is appropriate.
S7	Time horizon	✓	The model time horizon is 6 years, which is long enough, since by then great majority of patients are modelled to have died. Since overall survival is fairly mature, therefore little extrapolation is necessary.
S8	Disease states / pathways	✓	The disease states: PFS, PD, death are commonly used for terminal cancers.
S9	Cycle length	✓	6 weeks is appropriate.
Data			
D1	Data identification	✓	Data identification methods are well described.
D2	Pre-model data analysis	?	RPSFT method described only in outline.
D2a	Baseline data	✓	Baseline data from the single RCT, which is appropriate.
D2b	Treatment effects	?	Base case treatment effect estimated using RPSFT method. We understand that this method is appropriate, however, we have no guarantee that it has been implemented correctly.

Dimension of quality			Comments
D2c	Quality of life weights (utilities)	?	From EQ-5D in the RCT. We believe that the utility value for PD, 0.577, is uncertain (Section 5.4.1.5).
D3	Data incorporation	✓	Data incorporated in the model is referenced and generally well described. Data incorporation is transparent. For the PSA, the choice of distribution for each parameter has been described and justified.
D4	Assessment of uncertainty	X	Not all types of uncertainty have been addressed, and there are errors in the PSA, see Section 5.4.3.2.
D4a	Methodological	✓	Single type of model, which is adequate.
D4b	Structural	?	Further sensitivity analyses for assumption for time in PD in placebo arm would have been welcome.
D4c	Heterogeneity	✓	No patient subgroups, as appropriate.
D4d	Parameter	✓	Probabilistic and univariate sensitivity analyses performed.
Consistency			
C1	Internal consistency	?	There is evidence of tests on the mathematical logic of the model. However, there was a serious error in the calculation of the ICER for the unadjusted ITT population.
C2	External consistency	?	The results of Pfizer's model are not reconciled with the other models for sunitinib for GIST. However, this is difficult, because the other models use less mature clinical data, and are non-UK.

✓ indicates 'clear', X indicates 'concerns',? indicates 'some concerns'

5.2.2. Critique of the modelling approach and structure

The structure of Pfizer's cohort-based cost-effectiveness model is simple, appropriate and widely used for terminal cancers. The use of the PFS and PD health states is appropriate and consistent with the clinical outcomes in oncology trials.

A 6-year time horizon was used in the model. As very few patients are predicted to survive 6 years after starting treatment, the time horizon is effectively lifetime and is appropriate. The model cycle is 6 weeks, and a half-cycle model correction was applied, which is appropriate.

5.2.3. Data inputs

In this section, we consider the data used in the cost-effectiveness model.

5.2.3.1. Patient group

The modelled patient population is appropriate and is taken directly from the RCT (Demetri et al. 2006); they are people with unresectable and/or metastatic GISTs after the failure of imatinib. However, the extent to which patients from any RCT are representative of patients in routine practice is always uncertain. In particular, we do not know the characteristics of GIST patients in the UK who fail on imatinib. One characteristic that will probably be different

is the dose of imatinib on which patients fail to respond. In England, due to existing NICE Guidance no GIST patients should exceed a dose of 400 mg/day, whereas in the RCT less than 20% of patients had failed on doses this low.

5.2.3.2. Clinical effectiveness data

We agree with Pfizer's approach to using clinical effectiveness data in the cost-effectiveness model. In particular, we agree with their use of the RPSFT method, and with their method of fitting survival curves. In Section 5.4.1.1 we discuss this further, and we give further sensitivity analyses, e.g. use of clinical effectiveness data from the expanded access trial (Reichardt et al. 2008).

5.2.3.3. Drug costs

We have two reservations about Pfizer's method of modelling the cost of acquisition of sunitinib.

First, in all analyses, Pfizer assume that the first 6-week cycle of sunitinib is free to the NHS. In reply to our questions on their submission, Pfizer say that the Department of Health have agreed that the first cycle of sunitinib would be free for GIST patients. Nonetheless, we present all cost-effectiveness results below on two bases: first cycle is free and not free to the NHS.

Second, in Pfizer's model, patients stop sunitinib treatment at disease progression. However, in the RCT, at the time of disease progression, patients on sunitinib were given the opportunity to continue treatment with sunitinib at the investigator's discretion. At disease progression, patients on placebo were also given the opportunity to cross over to open-label sunitinib (Demetri et al. 2006). Fifty four (22%) of the 243 patients originally randomised to receive sunitinib continued with this treatment at the time of disease progression (source: comment from Pfizer in response to question from us). In Section 5.4.1.2 below, we update Pfizer's model to include the cost of sunitinib in PD.

Given that sunitinib is taken orally, Pfizer reasonably assume no drug administration costs.

5.2.3.4. Disease management costs

The model is based on the NHS and PSS perspective, which is appropriate (National Institute for Health and Clinical Excellence 2008). We do not expect the use of sunitinib to incur substantial costs or savings to other government departments.

As stated in Section 5.1.5, in a sensitivity analysis, Pfizer assume that 70% of patients would receive imatinib in PD (both after sunitinib and BSC). Under this assumption, the base case ICER becomes £32,200 per QALY assuming the first cycle of sunitinib is free and £37,500 per QALY assuming the first cycle of sunitinib is not free. However, we consider this analysis to be inappropriate for the following reasons. First, we have no evidence that such high proportions of patients took imatinib in PD the RCT (Demetri et al. 2006), and indeed Pfizer acknowledge this in their reply to our questions on their analyses (quote from Pfizer: "In the base case economic analysis, imatinib 400mg/d was excluded as it would introduce uncertainty into the model because the survival benefit conferred by its use and the proportion of patients who receive it for maintenance within the trial are unknown"). Second, this is not a NICE-approved use of imatinib in the NHS. Third, it seems unlikely that patients would be treated with imatinib after sunitinib, because patients in the RCT were resistant or intolerant to imatinib.

All non-drug costs are inflated to 2007/8, which is appropriate.

5.2.3.5. Adverse event costs

We believe that Pfizer's approach to modelling the costs of adverse events is appropriate. They have taken the incidence of the adverse event of hypertension from the RCT, and modelled only Grade 3 and 4 events. The costs of treating adverse events are consistent with those assumed by us previously, the assessment group for drugs for RCC Thompson-Coon et al. 2008. Given that the patients in the expanded access trial were, on average, less fit than patients in the RCT, higher rates of adverse events may be experienced in clinical practice.

5.2.3.6. Health-related quality of life

As stated in Section 5.1.4, health state utilities were taken from the RCT (Demetri et al. 2006), in which the EQ-5D questionnaire was used. This is clearly the most appropriate source of utility values for this appraisal.

In their reply to our question, Pfizer state that the PD utility of 0.577 was calculated from the average EQ-5D score measured *at confirmation of disease progression* in the two arms of the RCT. This therefore represents the utility at the time the patient progresses. However, for the assumption for the utility of patients in PD in the economic model, we require the average utility over the *whole* period that patients are in PD. This is likely to be lower than 0.577 (the utility of patients at the time of disease progression), assuming that utility

decreases from the time of progression until the time of death. We discuss the implications for cost-effectiveness in Section 5.4.1.5.

However, in response to our questions, we obtained full EQ-5D data for a range of disease states, including for PD in both trial arms. This data showed that , in fact, the EQ-5D utility during PD was 0.74 (SD 0.207) for those in the sunitinib trial arm, and 0.692 (SD 0.337) for those in the placebo trial arm (these data were from 237 and 306 data points respectively, rather than measured once at time of disease progression). Ultimately, the utility value for PD is not a major determinant of the ICET, and higher values also decrease the ICER. (see Section 5.4.1.5).++++

Although Pfizer did not explicitly model disutility due to adverse events (Submission p74-5), we are satisfied that such disutility is implicit in the utilities in PFS, and is manifest by the fact that in PFS, the utility for patients on sunitinib is lower than for patients on BSC.

5.2.4. Assessment of uncertainty

5.2.4.1. Probabilistic sensitivity analysis

The following parameters were varied stochastically in the probabilistic sensitivity analysis: utility for PFS, management costs and clinical effectiveness related to regression fit (Submission p78-9). Importantly, not all the uncertainty in clinical effectiveness is modelled, and the implications for this are discussed in Section 5.4.3.2. Utilities were modelled as beta distributions, which is appropriate, and medical management costs were modelled by Gamma distributions, which is also appropriate. The parameters of the Weibull distributions were assumed bivariate normal, which is appropriate.

Pfizer present a cost-effectiveness acceptability curve (CEAC) on Figure 16, p83 of their submission.

5.3. Results included in manufacturer's submission

Here, we present a summary of the results of Pfizer's model. See p81-84 of Pfizer's submission for further details. The acquisition cost of sunitinib is by far the largest single cost (Table 20, Table 21). The ICER using the RPSFT method for BSC overall survival, at £27,365 per QALY (1st cycle sunitinib free) is far lower than using the unadjusted ITT method for BSC overall survival, at £77,107 per QALY. This is because patients in the BSC arm are

assumed to spend far less time in PD under the RPSFT method (1.02 vs. 1.60 years), and hence far lower total QALYs under the RPSFT method (0.73 versus 1.07).

Table 20: Base case results of Pfizer's model using RPSFT method for best supportive care overall survival.

	Sunitinib		BSC (overall survival RPSFT)	Sunitinib vs. BSC (overall survival RPSFT)	
	(1st cycle not free)	(1st cycle free)		(1st cycle not free)	(1st cycle free)
Time on treatment (months)	7.3		n/a	7.3	
Life years	1.98		1.21	0.77	
Mean time in PD (years)	1.38		1.02	0.36	
QALYs	1.23		0.73	0.50	
Drug cost	£15,030†	£12,391	£0	£15,030†	£12,391
Monitoring	£799		£249	£551	
Blood tests	£22		£7	£15	
CT scans	£336		£105	£232	
Adverse events	£11		£0	£11	
BSC in PD	£2,692		£1,985	£708	
Death	£3,515		£3,724	-£208	
Total costs	£22,406†	£19,767	£6,070§	£16,337†	£13,699
ICERs					
Cost / Life years gained				£21,103†	£17,695
Cost/QALY				**£32,636**†	**£27,365**
Probability sunitinib cost effective at £30,000 / QALY WTP				Unknown¶	Approximately 50%¶
Probability sunitinib cost effective at £20,000 / QALY WTP				Unknown¶	Unknown¶

§ Incorrectly quoted as £6,315 by Pfizer.
¶ Figures unknown because of several important errors in PSA (see Section 5.4.3.2).
† Calculated by ERG. All other figures calculated by Pfizer and checked by the ERG.

Table 21: Results of Pfizer's model using the unadjusted ITT method for best supportive care overall survival.

	Sunitinib		BSC (overall survival unadjusted ITT)	Sunitinib vs. BSC (overall survival unadjusted ITT)	
	(1st cycle not free)	(1st cycle free)		(1st cycle not free)	(1st cycle free)
Time on treatment (months)	7.3		n/a	7.3	
Life years	1.98		1.79	0.19	
Mean time in PD (years)	1.38		1.60	-0.22	
QALYs	1.23		1.07	0.17	
Drug cost	£15,030†	£12,391	£0	£15,030†	£12,391
Monitoring	£799		£249	£551	
Blood tests	£22		£7	£15	
CT scans	£336		£105	£232	
Adverse events	£11		£0	£11	
BSC in PD	£2,692		£1,985	-£422	
Death	£3,515		£3,724	-£28	
Total costs	£22,406†	£19,767	£7,017	£15,388†	£12,750
ICERs					
Cost / Life years gained				£79,669†	£66,010
Cost/QALY				**£93,062**†	**£77,107**
Probability sunitinib cost effective at £30,000 / QALY WTP				Unknown¶	Unknown¶
Probability sunitinib cost effective at £20,000 / QALY WTP				Unknown¶	Unknown¶

¶ Figures unknown because of several important errors in PSA (see Section 5.4.3.2).
† Calculated by ERG. All other figures calculated by Pfizer and approved by ERG.

Given several important errors in the PSA, uncertainty in the ICERs is largely unknown (Section 5.4.3.2).

Pfizer also performed several sensitivity analyses (Submission p83-4). All analyses are based on the RPSFT method to allow for crossovers from placebo to sunitinib, and all analyses assume the first cycle of sunitinib is free, unless otherwise stated. We display Pfizer's analyses that we consider to be important in Table 22 below, plus some additional related analyses of our own.

When the sunitinib PFS and/or overall survival curves are fitted by applying the appropriate hazard ratio to the BSC Weibull curve, the ICER falls markedly (Table 22).

Table 22: Selected sensitivity analyses undertaken by Pfizer †

	Base case	Sensitivity analysis	ICER (Cost/QALY) Sunitinib vs. BSC
Base case	n/a	n/a	**£27,365 (1st cycle sunitinib free)** **£32,636(1st cycle sunitinib not free)** §
Costs			
Medical management costs using ERG RCC assumptions	£224 for both treatments in PFS and in PD	£81 for BSC in PFS, £224 for sunitinib in PFS, £435 for both treatments in PD	£29,033
Clinical effectiveness		Sunitinib PFS curve fitted using hazard ratio	£19,434
	Sunitinib curves fitted independently	Sunitinib overall survival curve fitted using hazard ratio	£22,003
Survival curves for PFS and overall survival		Sunitinib PFS and overall survival curves fitted using hazard ratios	£15,536
		BSC PFS curve fitted using hazard ratio	£27,396 §
	BSC curves fitted independently	BSC overall survival curve fitted using hazard ratio	£25,783 §

	Base case	Sensitivity analysis	ICER (Cost/QALY) Sunitinib vs. BSC
		BSC PFS and overall survival curves fitted using hazard ratios	£25,811 §

† All analyses are based on the RPSFT method to allow for cross-overs from placebo to sunitinib, and all analyses assume the first cycle of sunitinib is free, unless otherwise stated.

§ Calculated by us, the ERG

5.4. Comment on validity of results presented with reference to methodology used

5.4.1. Data inputs

5.4.1.1. Clinical effectiveness data

The assumptions used to model PFS and overall survival of BSC and sunitinib are very important in determining cost-effectiveness (Table 22). Indeed, this is apparent when we reconcile the cost-effectiveness of sunitinib versus BSC for GIST with the cost-effectiveness of sunitinib versus interferon-α for RCC (Section 5.1.8).

The overall survival data used in the model are mature, with only 14% of modelled patients still alive in the sunitinib treatment arm, and 8% in the BSC arm at the maximum follow-up time of the RCT (4.7 years). Hence, relatively little extrapolation of the overall survival curves are needed, which increases our confidence in the cost-effectiveness results. By contrast, the **published** overall survival data is far less mature, with about 60% of patients still alive in the sunitinib treatment arm, and about 50% in the placebo arm (Demetri et al. 2006).

The interim analysis overall survival Kaplan-Meier data are published (Demetri et al. 2006), but the final overall survival Kaplan-Meier data, a crucial model input, is not. Therefore, there is less assurance that this data, provided by Pfizer, is correct. Nonetheless, we can compare the median overall survival assumed in the model with the values quoted in the abstract which reports the analysis of the more mature data (Demetri et al. 2008). These values are similar, but not exactly the same: sunitinib overall survival: 78 weeks model versus 74.7 weeks abstract, BSC unadjusted ITT: 69 weeks model versus 64.9 weeks abstract, BSC RPSFT method: 41 weeks model versus 36.0 weeks abstract. Furthermore, the hazard

ratios differ slightly between those used in the model and quoted in the abstract: hazard ratio unadjusted ITT: 0.876 model versus 0.82 abstract, hazard ratio RPSFT: 0.505 model versus 0.46 abstract. In their reply to our questions, Pfizer explained that the values quoted in the abstract were based on the "interim" data (although presumably not the "interim" data in Demetri et al 2006, and the values used in the model were "the final re-calculated analyses after the closure of database" (Pfizer additional comments to ERG). Therefore, we assume that the data used in the model are more up to date than that in the abstract.

The PFS Kaplan-Meier data used in the model are virtually identical to that published in the main RCT paper (Demetri et al. 2006). Furthermore, the PFS hazard ratio of 0.33 used in the model agrees with the value quoted in the RCT paper.

Given that cost-effectiveness is strongly influenced by the assumption for BSC overall survival, we undertook the following sensitivity analysis. We assumed that the beneficial effect of sunitinib in terms of survival stops when treatment with sunitinib stops, i.e. the average time patients spend in PD is equal for patients in the sunitinib and BSC arms. We implemented this in Pfizer's model by setting the QALYs in PD equal for the two treatments and the costs in PD equal. We ignored the very slight discounting error associated with this method, namely that the method does not allow for the fact that QALYs in PD are on average incurred later in the sunitinib arm than in the BSC arm. In this case, the ICER is £56,200 per QALY assuming the first cycle of sunitinib is costed and £47,100 per QALY assuming it is free.

RPSFT method

Patient cross-over is a recurring problem in cost-effectiveness assessment of cancer drugs. The rank-preserved structural failure time RPSFT method, which adjusts for patient cross-over, has been discussed in the previous chapter (Section 4.1.7.1). Analyses based on the unadjusted ITT data would almost certainly underestimate the true relative treatment effect of sunitinib. Therefore, we recognise the need to correct for patient crossovers, and we support Pfizer's attempt by using the RPSFT method. Our analysis highlighted a crucial feature of this health technology assessment, namely that cost-effectiveness is radically improved when the RPSFT method is adopted. Using the unadjusted ITT data for overall survival for BSC, the ICER is £93,062 per QALY (1st cycle sunitinib not free) and £77,107 per QALY (1st cycle sunitinib free). This falls dramatically to £32,636 per QALY (1st cycle sunitinib not free) and £27,365 per QALY (1st cycle sunitinib free) under the RPSFT method (Table 22).

Therefore, it is crucial that we have confidence that the RPSFT method has been correctly performed. From the Web of Knowledge database, we identified 68 papers that cite the

original statistics paper that describes the RPSFT method(Robins & Tsiatis 1991). None of these papers are of cost-effectiveness studies. This suggests that the method has rarely, if ever, been used in cost-effectiveness models. But of course, this in itself does not mean that the method is inappropriate in this instance.

As mentioned in the previous chapter, whilst an unpaid, independent statistician, Ian White (MRC Biostatistics Unit, Cambridge), who has published on methods of adjusting for patient cross-over (White 2005a;White 2006), has endorsed the use of the method in this application, he did not perform the calculations. Therefore, we cannot be completely certain that the method has been correctly implemented. Furthermore, the RPSFT analysis was unplanned (Submission p29). We do, however, have some weak evidence to suggest that the method has been applied correctly: the mean overall survival hazard ratio under the RPSFTM of 0.505 as estimated by Pfizer is similar to the mean overall survival hazard ratio of 0.49 for the interim ITT data, before patient cross-over. But of course the interim analysis is based on far less mature data than that used in the final analysis (on which the RPSFT method is based).

On the other hand, we have a reason to question whether the method has been correctly implemented by Pfizer. Ian White (MRC Biostatistics Unit, Cambridge), advised Pfizer that the 95% confidence interval of the overall survival hazard ratio of 0.388 – 0.658 (mean 0.505) as originally calculated by Pfizer, was incorrect. Instead, Ian White states that the confidence interval should be 0.262 – 1.134 (Submission p39).

Curve fitting

Given that cost-effectiveness is heavily influenced by clinical effectiveness, it is important that the Weibull survival curves have been fitted correctly to the Kaplan-Meier data. In short, we are satisfied that this is the case.

The Weibull curves were fitted by linear regression to one data point per month to improve the fit to the actual data by preventing the first few data points in the trial data, at times less than one month, from dominating the fit (Figure 8, Figure 9).

We found similar ICERs when we used an alternative method of fitting Weibull curves to the Kaplan-Meier data by minimising the sums of squares. Fitting PFS data by minimising the sums of squares gives an ICER of £27,600 per QALY assuming the first cycle of sunitinib is free, and fitting overall survival data by sums of squares gives an ICER of £27,000 per QALY. These figures are very similar to Pfizer's base case of £27,400 per QALY assuming the first cycle of sunitinib is free.

The modelled curve for PFS for sunitinib strongly influences cost-effectiveness (Table 22), because the mean PFS for sunitinib determines the total cost of acquisition of sunitinib, which is the largest cost in the analysis. Cost-effectiveness is far less sensitive to PFS for BSC, because patients on BSC incur no large costs in this health state.

In the base case, Pfizer fit curves separately to sunitinib and to BSC, and in sensitivity analyses, they fit curves to BSC, and then estimate the sunitinib curves by applying the appropriate hazard ratio (for PFS or overall survival) to the BSC curves. We agree with Pfizer that their base case methodology is preferable, because under the alternative method, the sunitinib curves are visually poor fits to the Kaplan-Meier data.

Figure 8: Pfizer curve fits to progression-free survival Kaplan-Meier data (Submission Figure 13)

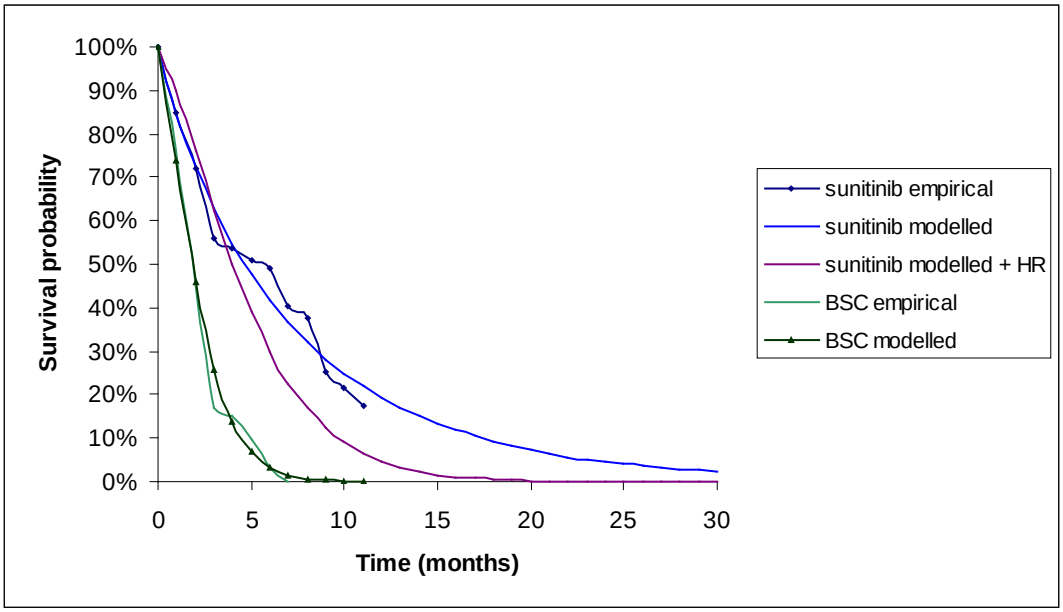
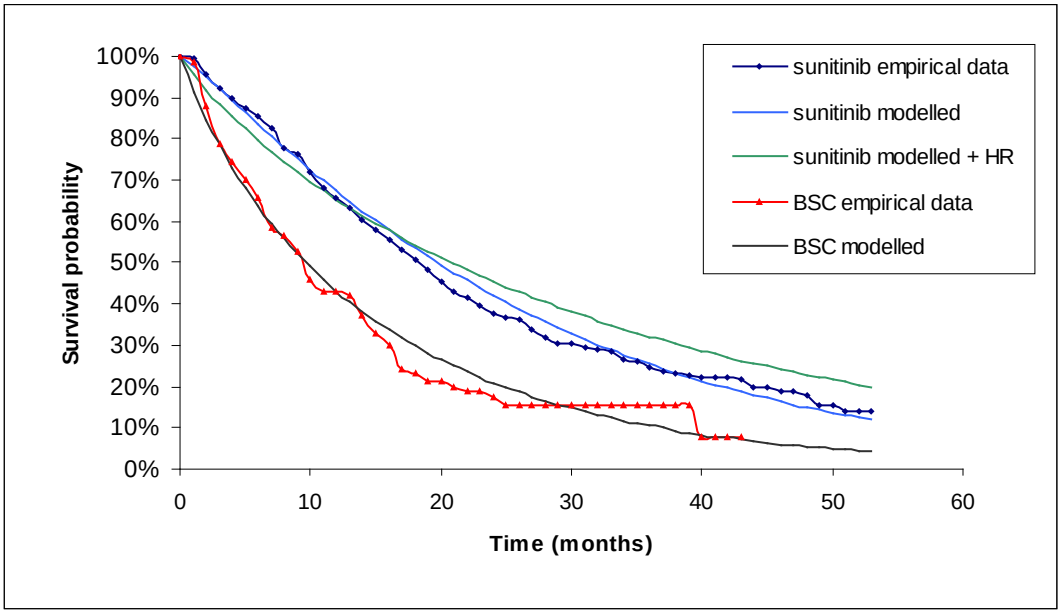


Figure 9: Pfizer curve fits to overall survival Kaplan-Meier data (Submission Figure 14)



Cohort study A6181004

Next, we consider the clinical effectiveness of sunitinib for GIST as measured in the expanded access cohort study (Reichardt et al. 2008) in the context of cost-effectiveness. Pfizer did not consider this data for assessing cost-effectiveness. This is an ongoing worldwide, phase III, multicentre, open-label treatment-use trial that aims to provide access to sunitinib to GIST patients who might benefit from this therapy, but who are ineligible for sunitinib clinical trials because of pre-specified entry criteria, or for whom there are no GIST trials available in a particular country in which regulatory approval has not yet been granted (see previous chapter and Submission p48). As in the RCT, patients had failed prior treatment with imatinib, defined either as progression of disease or as significant toxicity that precluded further treatment with imatinib. The regime of administration of sunitinib was the same as in the RCT. Pfizer state that there is no evidence to suggest that patients treated in the UK in clinical practice would differ from the overall patient population considered within the expanded access trial (Submission p52). In fact this population is more representative of UK GIST patients than those in the RCT, as participants were included with all ECOG grades of performance status and not just 0 and 1 as in the RCT (i.e. the most able see Table 5). Therefore, the patient population is appropriate for the cost-effectiveness model.

Median PFS and overall survival for sunitinib quoted in Pfizer's submission p51 are slightly different that quoted in Reichardt et al (2008): median PFS is 41 weeks in Pfizer's submission versus 37 weeks in Reichardt et al (2008), median overall survival is 75 weeks in Pfizer's submission versus 73 weeks in Reichardt et al (2008). We assume that the values

quoted by Pfizer are more up to date than those quoted in Reichardt et al (2008), and therefore we consider these now. The median overall survival of 75 weeks is virtually identical to the value of 73 weeks assumed in Pfizer's model (although the model, being based on the Demetri RCT data, is restricted to patients with a baseline ECOG performance status of 0 or 1). However, the median PFS of 41 weeks is substantially more than the value of 23 weeks assumed in Pfizer's model.

When we adjust the PFS curve for sunitinib to give a median of 41 weeks, by changing the Weibull parameter λ from 0.171 to 0.090, whilst leaving the parameter γ constant at 0.91, the base case ICER increases from £27,400 to £46,300 assuming the first cycle of sunitinib is free, and from £32,600 to £50,900 when the first cycle of sunitinib is not free. In response to our questions, Pfizer told us that the dose intensity in the expanded access trial has not been calculated. Therefore, we have assumed that the dose intensity of sunitinib in the expanded access trial was equal to the value of 88.6% in the RCT. Also, in this analysis, we assume the same PFS BSC survival curve from the RCT. However, this is probably unimportant because, as noted above, cost-effectiveness is insensitive to this model input. The ICERs increase substantially when we use PFS for sunitinib from the expanded access trial because patients typically stay in PFS for longer in the expanded access trial compared to the RCT, thereby incurring greater sunitinib acquisition costs. We do not know why patients typically remain in PFS longer in the expanded access trial.

5.4.1.2. Drug costs

The cost of acquisition of sunitinib, at £27,200 per patient per year, is by far the largest cost in the model. In all their analyses, Pfizer assume that the first 6-week cycle of sunitinib is free to the NHS. In reply to our questions on their submission, Pfizer state that the Department of Health have agreed that the first cycle of sunitinib would be free for GIST patients. However, as sensitivity analyses, we calculated that Pfizer's base case ICER of £27,365 per QALY increases to £32,636 per QALY when we assume that the first cycle of sunitinib is not free. NB (both results based on the RPSFT method for estimating PFS and OS with BSC). Assuming the unadjusted ITT analysis, Pfizer's ICER is £77,107 per QALY (1st cycle free) or an estimated £93,062 per QALY (1st cycle not free).

Treatment with sunitinib was modelled to continue until the occurrence of either disease progression or unacceptable side effects. The efficacy data from the RCT therefore reflects treatment interruptions and dosage reductions. For consistency of costs and clinical effects, the cost of sunitinib is modelled to be reduced by the dose intensity. The unpublished dose intensity of 88.6%, which Pfizer state is taken directly from the RCT (Submission p75), is

consistent with the value of 86.4% which Pfizer used in their assessment of the cost-effectiveness of sunitinib for RCC Thompson-Coon et al 2008. Although the dose intensity is held constant in the PSA, we appreciate that it may be difficult to model uncertainty in this important parameter. Nonetheless given that Pfizer have not incorporated uncertainty in the dose intensity, this means that uncertainty in cost-effectiveness is underestimated.

In Pfizer's model, patients stop sunitinib treatment at disease progression. However, in the RCT, at the time of disease progression, patients on sunitinib were given the opportunity to continue treatment with sunitinib at the investigator's discretion. At disease progression, patients on placebo were also given the opportunity to cross over to open-label sunitinib (Demetri et al. 2006). 54 (22%) of the 243 patients originally randomised to receive sunitinib continued with sunitinib treatment at the time of progression (comment from Pfizer in response to question from us). Although we do not know whether sunitinib affects overall survival if taken after disease progression, for consistency between the overall survival of patients in the sunitinib arm of the RCT and the costs of sunitinib, we should model the sunitinib acquisition costs associated with the 22% of patients in the sunitinib arm who continued with sunitinib treatment once they have progressed. We believe that the only way to know whether sunitinib treatment in PD improves overall survival would have been to include a third treatment arm in the RCT which would have been identical to the sunitinib treatment arm in the RCT, except that sunitinib would not have been given in PD.

On average, the 54 patients taking sunitinib in PD took sunitinib for 0.42 years in PD, and the dose intensity for these patients was 91.2% in PD (data from Pfizer). Then the per patient cost of sunitinib in PD is the proportion of patients originally randomised to sunitinib who took sunitinib in PD x annual per patient cost of sunitinib acquisition x mean time sunitinib took sunitinib in PD x dose intensity of these patients x cost discount factor to mean time of sunitinib use in PD = 22% x £27,203 x 0.42 years x 91.2% x 1/(1.035)^(0.42/2 + 0.61) years = £2,237 (where mean time sunitinib patients spent in PFS = 0.61 years). Including this cost in Pfizer's model, under the RPSFT method, gives an ICER of £31,800 per QALY assuming the first cycle of sunitinib is free, and £37,100 per QALY assuming the first cycle of sunitinib is not free. Using the unadjusted ITT analysis, the ICER is of £90,500 per QALY assuming the first cycle of sunitinib is free, and £106,500 per QALY assuming the first cycle of sunitinib is not free.

5.4.1.3. Disease management costs

As noted in the previous Section 5.2.3.4, per-patient non-drugs costs are far smaller than drug costs.

As stated in Section 5.1.5, based on advice from their panel of four specialists in the treatment of GIST, for the base case, Pfizer assume that patients in PFS would have one outpatient appointment every month at £112 per visit, one CT scan every 3 months at a cost of £141 per scan and standard blood tests once every month at a cost of £3 per test. This gives the total cost of medical management in PFS over a 6-week cycle of £224. Also for the base case, Pfizer assume that the resource use for disease monitoring and management in PD is exactly as for PFS.

As a sensitivity analysis, Pfizer assume the same medical management resource use as we assumed for RCC, see Section 5.1.5.

5.4.1.4. Adverse event costs

We believe that the costs of treating adverse events have been appropriately incorporated in the model. This cost accounts for only a tiny fraction of the total mean per patient costs.

5.4.1.5. Health-related quality of life

Pfizer state that the utility of progression free survival in the sunitinib arm was 0.731, and for BSC it was 0.781. In PD, the utility for both sunitinib and BSC was equal, at 0.577. We agree with Pfizer's method of estimating the utility of 0.731 for patients on sunitinib in PFS as an average of the value at the end of the four week treatment period of sunitinib, and the utility at the end of the two week rest period. Pfizer state that the PD utility of 0.577 was calculated from the average EQ-5D score, over both treatment arms, measured *at the termination of the double-blind phase* of the RCT. These utility values are published in a cost-effectiveness study of sunitinib for GIST (Chabot et al. 2008), which increases our confidence in their accuracy.

We asked Pfizer for further information to explain their calculation of the utility of 0.577 for patients in PD. We are unable to reconcile the data they supplied to the value of 0.577, because they quote values of 0.740 for patients taking sunitinib and 0.692 for patients on BSC (Appendix 2). We also note that the value of 0.740 for patients taking sunitinib in PD seems inconsistent with the utility of 0.731 for patients taking sunitinib in PFS, since we would expect the value in PD to be lower than the value in PFS, given that the health of patients is expected to deteriorate as the disease worsens. Nonetheless, assuming these utility values for PD, the base case ICER (assuming the RPSFT method) falls to £22,500 per QALY assuming the first cycle of sunitinib is free and £26,800 per QALY assuming the first cycle of sunitinib is costed.

Fortunately, the ICER is reasonably insensitive to the utility in PD: a utility of 0.50 gives an ICER of £29,000 per QALY, and a utility of 0.40 gives an ICER of £31,400 per QALY assuming the first cycle of sunitinib is free.

The ICER is also rather insensitive to the utility in PFS: reducing the sunitinib and BSC utilities by 0.10 gives an ICER of £29,800 per QALY, and increasing the utilities by 0.10 gives an ICER of £25,300 per QALY.

5.4.2. Assessment of consistency

5.4.2.1. Internal consistency

We have thoroughly checked the mathematics, statistics, internal logic and implementation of the model in Excel, as well as the cost-effectiveness results presented by Pfizer. As stated at the beginning of this chapter, we discovered the following important logical error in the economic model. Pfizer originally quoted the cost-effectiveness using the unadjusted ITT overall survival data for BSC as £34,649 per QALY (assuming the first cycle of sunitinib is free), whereas the correct value is £77,107 per QALY. In their response to questions from us, Pfizer acknowledged this error.

5.4.2.2. External consistency

We agree with Pfizer's assertion that it is difficult to compare the results of their model with the two published models of the cost-effectiveness of sunitinib for GIST(Chabot et al. 2008;Contreras-Hernandez et al. 2008), because these two models use the interim effectiveness data, and because they both consider a non-UK cost perspective.

5.4.3. Assessment of uncertainty

5.4.3.1. One-way sensitivity analyses

Pfizer present one-way sensitivity analyses (see Table 22). In additional, we present several further analyses, see Section 5.5.

5.4.3.2. Probabilistic sensitivity analysis

We found several serious errors with the PSA;

- Importantly, Pfizer have modelled only one aspect of uncertainty in the treatment effect on PFS and overall survival. They have modelled uncertainty in the Weibull parameters only due to uncertainty in fitting the Weibull curves to the deterministic Kaplan-Meier curves. They have not allowed for uncertainty in the treatment effect due to the finite sample size of the patient population in the RCT. This uncertainty is commonly modelled by allowing for the uncertainty in the hazard ratio. However, given that Pfizer have, quite legitimately, not used the hazard ratios in their deterministic base case, this uncertainty is missing. Given that the hazard ratio for overall survival based on the RPSFT method is highly uncertain, 0.505 (0.262 - 1.134), a very important source of uncertainty is missing in the model.

  For example, suppose we use the Weibull curve fit for overall survival as calculated by Pfizer, and suppose we assume a hazard ratio of 1, which falls well within the 95% confidence interval of (0.262 – 1.134) under the RPSFT method, then this yields a very higher ICER of £230,000 per QALY assuming the first cycle of sunitinib is free and £279,000 per QALY assuming the first cycle of sunitinib is costed. In this case, the per patient QALYs for sunitinib are only slightly greater, by 0.05, than for BSC, because patients are in PFS for longer on sunitinib, where they enjoy a slightly better quality of life. Furthermore, assuming the upper 95% confidence interval hazard ratio of 1.134, patients accrue fewer QALYs with sunitinib compared to BSC, and so sunitinib is dominated by BSC. Now, assuming the lower 95% confidence interval hazard ratio of 0.262, the ICER falls to £17,900 assuming the first cycle of sunitinib is free and £21,100 per QALY assuming the first cycle of sunitinib is costed.

- There are several errors in the fit to the Weibull curves in the modelling of uncertainty in treatment effectiveness. Technically, there are errors in the Cholesky matrix decompositions. [a] When we correct these errors, cost-effectiveness becomes slightly more certain.

- For the utilities, importantly, Pfizer have used the standard deviation, when they should have used the standard error, over-estimating the uncertainty of the utility values.

---

[a] In worksheet “PFS”, the reference in cell AA43 should be raised to the power of 2, and cell AL43 should reference cell AM36 to the power of 2. These errors are repeated in worksheets “overall survival_RPSFT analysis” and “overall survival_ITT analysis”.

- In the model Pfizer assume a standard deviation of 0.02 for the utility in PFS, whereas 0.20 is quoted in their submission (p74) and in the cost-effectiveness analysis of Chabot et al (2008)(Chabot et al. 2008).

- Even using Pfizer's incorrect assumptions above, we produced a very different CEAC when we ran the PSA using all of Pfizer's base case assumptions. Specifically, we found much more uncertainty in cost-effectiveness than is suggested by Pfizer's CEAC. We suspect, but are not certain, that the CEAC displayed in Pfizer's submission may have been produced with no allowance for uncertainty in the utilities.

Given these serious errors, we suggest that the estimate of uncertainty in cost-effectiveness calculated by Pfizer, in particular as described by the CEAC, should be ignored.

We can, however, say that **the uncertainty in the base case ICERs of £27,400 per QALY (first cycle sunitinib free) and £32,600 per QALY (first cycle not free) is substantial,** given the very wide confidence interval (0.262 - 1.134) in the overall survival hazard ratio under the RPSFT method.

5.5. Summary of uncertainties and issues

We suggest that there are three crucial issues in the economic evaluation: (1) the use of the RPSFT method to adjust for patient cross-over from BSC to sunitinib, (2) the costing of sunitinib acquisition in progressive disease for patients initially randomised to sunitinib, and (3) the large uncertainty in the relative treatment effectiveness for overall survival between sunitinib and BSC under the RPSFT method.

1) First, Pfizer's base case ICERs of £27,365 per QALY (first cycle sunitinib free) and £32,636 per QALY (first cycle not free), which depend on the RPSFT method to adjust for patient cross-over, are far lower than the ICERs of £77,100 per QALY (first cycle sunitinib free) and £93,100 per QALY (first cycle not free) based on the unadjusted ITT data (Table 23). Therefore, it is essential that the RPSFT method is appropriate and that it has been correctly implemented.

An independent statistician has advised us that the RPSFT method is appropriate in this case. However, we cannot be certain that the method has been implemented correctly. If the method has been implemented correctly, then we endorse Pfizer's use of the RPSFT method.

2) The second crucial issue concerns the use of sunitinib in progressive disease (PD). Although in the RCT, 22% of patients initially randomised to sunitinib continued to take sunitinib in PD, in their model Pfizer assumed no drug cost of sunitinib for these patients in PD. When we allow for this extra cost, the ICERs increase (Table 23).

3) The third crucial issue concerns the large uncertainty in the relative treatment effectiveness for overall survival between sunitinib and BSC. We cannot quantify precisely the uncertainty in the base case ICERs due to several serious errors in the PSA (Section 5.4.3.2). However, the uncertainty will be substantial given the very wide 95% confidence interval, (0.262 – 1.134), in the overall survival hazard ratio under the RPSFT method. In particular, assuming a hazard ratio of 1, which falls well within this confidence interval, yields a very higher ICER of £230,000 per QALY, assuming the first cycle of sunitinib is free.

We present several important sensitivity analyses in Table 23 below. We suggest that the base case ICERs of £27,400 per QALY (first cycle sunitinib free) and £32,600 per QALY (first cycle not free), should be considered in conjunction with these sensitivity analyses. The ICERs for the sensitivity analyses should be considered with reference to the full explanations of the analyses given earlier in this chapter.

According to Pfizer, the budget impact of funding sunitinib for GIST for the NHS in England and Wales is small: £247,000 in 2009 rising to £1,123,000 in 2013 (assuming first cycle of sunitinib is free) (see Submission p87).

Table 23: Summary of important deterministic sensitivity analyses

Parameter	Pfizer base case assumption	Sensitivity analysis	ICER (£/QALY)		Details
			1st cycle sunitinib not free	1st cycle sunitinib free	
Pfizer base case (RPSFT method)	n/a	n/a	£32,600	£27,400§	Table 22, page 80, Section 5.4.1.2, page 86
BSC overall survival	RPSFT method	Unadjusted ITT	£93,100	£77,100§	Table 22, page 80, Section 5.4.1.2, page 86
Sunitinib treatment in PD for patients randomised to	0% patients	22% patients	£37,100	£31,800	

Parameter	Pfizer base case assumption	Sensitivity analysis	ICER (£/QALY)		Details
			1st cycle sunitinib not free	1st cycle sunitinib free	
sunitinib, RPSFT method					Section 5.4.1.2, page 86
Sunitinib treatment in PD for patients randomised to sunitinib, unadjusted ITT population	0% patients	22% patients	£106,500	£90,500	
Beneficial effect of sunitinib when treatment with sunitinib stops (Mean time patients in progressive disease, PD)	Beneficial effect of sunitinib continues in PD. (Mean time PD sunitinib > mean time PD BSC)	Beneficial effect of sunitinib ends in PD. (Mean time PD sunitinib = mean time PD BSC)	£56,200	£47,100	Section 5.4.1.1, page 81
Sunitinib PFS, RPSFT method	From RCT (Demetri et al. 2006)	From expanded access trial (Reichardt et al. 2008)	£50,900	£46,300	Section 5.4.1.1, page 81
Utility in progressive disease	0.577	< 0.577	ICER increases by a few £thousand		Section 5.4.1.5, page 88

§ ICERs calculated by Pfizer and endorsed by ERG. All other ICERs calculated by ERG only.

6. Additional work undertaken by the ERG

Below, we present a list of the additional sensitivity analyses that we have performed. The results of the most important of these analyses are summarised in the previous Section, Section 5.5;

- All ICERs presented by Pfizer assume the first cycle of sunitinib is free. In addition, we present ICERs assuming the first cycle is not free.

- We assumed that the beneficial effect of sunitinib in terms of survival stops when treatment with sunitinib stops, i.e. the average time patients spend in progressive disease (PD) is equal for patients in the sunitinib and BSC arms. See Section 5.4.1.1, p81.

- Pfizer fitted Weibull curves to the Kaplan-Meier data by regression. We used an alternative method of minimising the sums of squares. See Section 5.4.1.1, p81.

- The median time sunitinib patients spent in PFS in the RCT was 23 weeks. In a sensitivity analysis, we assumed the median time of 41 weeks, from the experience of the expanded access trial. See Section 5.4.1.1, p81.

- Although 22% of patients initially randomised to sunitinib continued to take sunitinib in progressive disease (PD) in the RCT, Pfizer assumed no cost of sunitinib acquisition for these patients in PD. We modelled this extra cost, see Section 5.4.1.2, p86.

- The utility of patients in PD in the economic model should represent the average utility over the whole period that patients are in PD. It might be thought that this is likely be lower than 0.577 (the utility of patients at the time of disease progression), the value used by Pfizer in their model. However, full utility data supplied by Pfizer (see Appendix 2) indicates that the EQ-5D utility of living with progressive disease is actually closer to 0.7). Section 5.4.1.5, p88.

- We found several serious errors in the PSA but were unable to fully correct them. However, we calculated an estimated ICER assuming the upper and lower confidence intervals for the hazard ratio for overall survival under the RPSFT method. This showed that the ICER under the RPSFT method is very uncertain.

7. Discussion

7.1. Summary of clinical effectiveness issues

- The literature search strategy was appropriate, replicable, and the hits appear correct in relation to the search date and databases/interfaces used. We are confident that there are no relevant and good quality studies which have not been presented in the submission.

- In the included studies sunitinib was taken in accordance with the manufacturer's dosing regime; i.e. 50 mg/day for four weeks followed by two weeks treatment free to give a six week cycle. In the event of toxicity (grade 3 or 4) the dosage was reduced in 12.5 mg/day steps to 37.5 mg/day then 25mg/day if necessary, but not below this level.

- The submission from Pfizer included one good quality RCT (A6181004) and one, ongoing, cohort study to extend access to sunitinib (A6181036). The *overall* survival results for sunitinib patients in both these studies were similar (RCT: overall survival = 73 median weeks (95% CI 61-83) in comparison to 75 median weeks (95% CI 68-84) for the cohort study). However, a greater median OS for the ECOG grad 0-1 in the cohort study was found compared to the RCT (RCT: 73 weeks (95%CI 61-83), cohort study: 88 weeks (95%CI 77-97). The results for time to tumour progression in sunitinib patients in the cohort study were quite different (RCT: median weeks = 29, 95% CI 22-41, cohort study: median weeks = 41, 95% CI 36-47). These results may be influenced by the different ECOG performance status of the two study populations.

- In the RCT 168 (83%) of participants in the intervention group had an adverse event of any severity compared to 60 (59%) in the control group. In both studies most adverse events were mild to moderate. Overall, the most commonly occurring symptoms were fatigue (RCT: 68 (43%); cohort study: 465 (42%)) and diarrhoea (RCT: 59 (29%); cohort study: 439 (39%)).

- In order to deal with potential bias in the results caused by patients crossing over from placebo to sunitinib after unblinding, the RCT's time dependent outcomes were analysed using RPSFT methods. These methods are more appropriate than censoring the data at the primary endpoint, as they allow analysis for a longer

follow up period by estimating the differences between the groups as if the participants in the placebo group had not crossed over to sunitinib; i.e. by ITT.

- The population of the RCT was restricted to those in ECOG performance status grades O-1, whilst that of the cohort study was inclusive of grades 0-4.

7.2. Summary of cost-effectiveness issues

- Use of the RPSFT method to adjust for patient cross-over from BSC to sunitinib. In particular, under this method, the improved survival for patients randomised to sunitinib is assumed to continue in PD.
- No costing of sunitinib acquisition in progressive disease for patients initially randomised to sunitinib (which does not reflect the treatment of some patients in the RCT).
- The large uncertainty in the relative treatment effectiveness for overall survival between sunitinib and BSC under the RPSFT method.
- Whether to assume the first cycle of sunitinib is free to the NHS.
- Patients on sunitinib spent longer in PFS in the expanded access trial compared to the RCT.

7.3. Implications for research

- The cohort study showed response to sunitinib according to ECOG performance status. This evidence should be explored further with an RCT.
- Further RCTs are needed to determine whether there are any other potential prognostic factors for response to sunitinib; including whether people with a Kit exon 9 mutation have longer progression free survival than those with a Kit exon 11 mutation.
- Pfizer report that 22% of participants in the RCT randomised to sunitinib continued to take it when in progressive disease. Future RCTs of sunitinib for GIST should include measures of palliative or other effects in progressive disease to support this use.

8. References

BNF. British National Formulary (BNF) 56. Online edition . 2008. 19-3-2008.

Casali, P. G., Jost, L., Reichardt, P., Schlemmer, M., Blay, J. Y., & On behalf of the ESMO Guidelines Working Group 2008, "Gastrointestinal stromal tumors: ESMO Clinical Recommendations for diagnosis, treatment and follow-up", *Annals of Oncology*, vol. 19, no. suppl_2, p. ii35-ii38.

Chabot, I., LeLorier, J., & Blackstein, M. E. 2008, "The challenge of conducting pharmacoeconomic evaluations in oncology using crossover trials: the example of sunitinib for gastrointestinal stromal tumour", *Eur J Cancer*, vol. 44, no. 7, pp. 972-977.

Contreras-Hernandez, I., Mould-Quevedo, J. F., Silva, A., Salinas-Escudero, G., Villasis-Keever, M. A., Granados-Garcia, V., vila-Loaiza, G., Petersen, J. A., & Garduno-Espinosa, J. 2008, "A pharmaco-economic analysis of second-line treatment with imatinib or sunitinib in patients with advanced gastrointestinal stromal tumours", *Br J Cancer*, vol. 98, no. 11, pp. 1762-1768.

Coyle, D., Small, N., Ashworth, A., Hennessy, S., Jenkins-Clarke, S., Mannion, R., Rice, N., & Ahmedzai, S. Costs of palliative care in the community, in hospitals and in hospices in the UK. Critical Reviews in Oncology:Hematology 32, 71-85. 1999.

Demetri, G. D., Huang, X., Garrett, C. R., Schoffski, P., Blackstein, M. E., Shah, M. H., & et al. 2006, "Novel Statistical Analysis of Long-term Survival to Account for Cross-Over in a Phase III Trial of Sunitinib versus Placebo in Advanced GIST after Imatinib Failure", in *The 44th Annual Meeting of the American Society of Clinical Oncology*.

Demetri, G. D., van Oosterom, A. T., Garrett, C. R., Blackstein, M. E., Shah, M. H., Verweij, J., McArthur, G., Judson, I. R., Heinrich, M. C., Morgan, J. A., Desai, J., Fletcher, C. D., George, S., Bello, C. L., Huang, X., Baum, C. M., & Casali, P. G. 2006, "Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial", *Lancet*, vol. 368, no. 9544, pp. 1329-1338.

Drummond, M., O'Brien, B., Stoddart, G., & Torrance, G. 1997, *Methods for the economic evaluation of health care programmes* Oxford University Press, Oxford, UK.

National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. 2008. London, NICE.

Oken, M. M., Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., & Carbone, P. P. 1982, "Toxicity and response criteria of the Eastern Cooperative Oncology Group", *American Journal of Clinical Oncology.5(6):649-55*.

Philips, Z., Bojke, L., Sculpher, M., Claxton, K., & Golder, S. 2006, "Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment: A Review and Consolidation of Quality Assessment", *Pharmacoeconomics*, vol. 24, no. 4, pp. 355-371.

Reichardt, P., Kang, Y. K., & Ruka, W. "Detailed analysis of survival and safety with sunitinib in a worldwide treatment use trial of patients with advanced GIST", in *The 44th Annual Meeting of the American Society of Clinical Oncology: abstract*.

Robins, J. M. & Tsiatis, A. A. 1991, "Correcting for non-compliance in randomized trials using rank preserving structural failure time models", *Communications in Statistics - Theory and Methods*, vol. 20, no. 8, pp. 2609-2631.

Thompson-Coon, J., Hoyle, M., Green, C., Liu, Z., Welch, K., Moxham, T., & Stein, K. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systemtatic review and economic evaluation. NICE . 2008. 19-12-2008. http://www.nice.org.uk/guidance/index.jsp?action=download&o=41488

White, I. R. 2005b, "Uses and limitations of randomization-based efficacy estimators", *Stat Methods Med Res*, vol. 14, no. 4, pp. 327-347.

White, I. R. 2005a, "Uses and limitations of randomization-based efficacy estimators", *Statistical Methods in Medical Research*, vol. 14, no. 4, pp. 327-347.

White, I. R. 2006, "Estimating treatment effects in randomized trials with treatment switching", *Statistics in Medicine*, vol. 25, no. 9, pp. 1619-1622.

Appendix 1: Comparison of Pfizer model of sunitinib for GIST with the Assessment Group model of sunitinib for renal cell carcinoma for model validation

Pfizer's GIST model is based on the renal call carcinoma (RCC) model written by the ERG. In order to validate their model we tried to re-create the results of Pfizer's GIST model using our RCC model by setting the parameter values in our RCC model to those used in the GIST model, see Table 24 below.

For discussion of the results in the table, see Section 5.1.8.

Table 24: ICER for our (Assessment Group) renal cell carcinoma (RCC) model of sunitinib vs. interferon-□ with one parameter changed at a time

Parameters are changed to values used in Pfizer's GIST model.

Parameter	Assessment Group RCC model base case value	Pfizer GIST model base case value	ICER (£/QALY) (Assessment Group RCC model using GIST parameter value)	% increase in ICER from RCC base case
Assessment Group RCC base case	n/a	n/a	**£71,462**	n/a
Survival curves				
PFS Weibull sunitinib	$\lambda = 0.055$, $\gamma = 1.004$	$\lambda = 0.171$, $\gamma = 0.912$ (shorter-tailed)	£35,954§	-50%
PFS Weibull comparator	$\lambda = 0.132$, $\gamma = 1.004$	$\lambda = 0.303$, $\gamma = 1.360$ (shorter-tailed)	£69,274	-3%
PFS Weibull sunitinib & comparator	above	above	£36,442	-49%
Overall survival Weibull sunitinib	$\lambda = 0.007$, $\gamma = 1.447$	$\lambda = 0.024$, $\gamma = 1.133$ (shorter-tailed)	£87,336†	22%
Overall survival Weibull comparator	$\lambda = 0.011$, $\gamma = 1.447$	$\lambda = 0.024$, $\gamma = 1.133$ (shorter-tailed)	£45,501††	-36%
Overall survival	above	above	£50,856	-29%

Parameter	Assessment Group RCC model base case value	Pfizer GIST model base case value	ICER (£/QALY) (Assessment Group RCC model using GIST parameter value)	% increase in ICER from RCC base case
Weibull sunitinib & comparator All survival curves above	above	above	£26,244	-63%
		Utilities		
Utilities	0.78 PFS, 0.70 PD for both sunitinib and interferon-α	0.731 PFS sunitinib, 0.781 PFS BSC. 0.577 PD for both sunitinib and BSC	£78,069	9%
		Costs		
First cycle sunitinib	Normal price	Free	£65,362	-9%
Sunitinib dose intensity	86.4%	88.6%	£73,447	3%
Cost of death	Not included	£3,923 per patient	£71,294	0%
Cost comparator drug and drug administration	Interferon-α: £813 / 6 weeks, £112 / 6 weeks administration	BSC: £0 acquisition and administration	£78,226	9%
PD medical management cost per patient	£435 / 6 weeks	£224 / 6 weeks	£72,594	2%
Adverse event cost per patient	Sunitinib £88, Interferon-α £4	Sunitinib £11, BSC £0	£71,295	0%
Time horizon	10 years	6 years	£73,053	2%
All parameters above			**£27,355**	-62%
Pfizer GIST model base case	n/a	n/a	**£27,365**	-62%

§ ICER decreases greatly because sunitinib GIST PFS curve is much shorter-tailed than RCC curve, hence lower cost of sunitinib incurred.

† ICER increases because sunitinib GIST overall survival curve is shorter-tailed than renal cell carcinoma curve, hence fewer QALYs accumulated in sunitinib arm.

†† ICER decreases greatly because BSC GIST overall survival curve is much shorter-tailed than RCC interferon curve, hence fewer QALYs accumulated in comparator arm.

Appendix 2: EQ-5D-based utility values from RCT

The table below, concerning EQ-5D-based utility values from the RCT, was provided by Pfizer. We are unable to calculate Pfizer's assumption for utility in progressive disease, 0.577, from this data see Section 5.4.1.5

	Sunitinib trial arm			Placebo trial arm		
	No. of data points (%*)	From: No. of patients (%*)	Mean EQ-5D utility score (SD)	No. of data points (%*)	From: No. of patients (%*)	Mean EQ-5D utility score (SD)
Double blind phase:	1067 (99.2)	200 (100)	0.744 (0.234)	324 (99.1)	98 (99.0)	0.750 (0.269)
During PFS	933 (99.4)	195 (100)	0.750 (0.235)	258 (98.9)	97 (99.0)	0.753 (0.255)
During PD	198 (97.1)	132 (97.1)	0.728 (0.209)	115 (99.1)	74 (98.7)	0.724 (0.320)
AFTER double blind phase:	40 (97.6)	18 (100)	0.805 (0.189)	192 (97.5)	54 (100)	0.673 (0.345)
During PFS	21 (95.5)	17 (100)	0.799 (0.207)	173 (97.2)	52 (100)	0.675 (0.348)
During PD	24 (100)	9 (100)	0.827 (0.162)	30 (100)	22 (100)	0.694 (0.275)
COMBINED all data:	1107 (99.1)	200 (100)	0.746 (0.233)	516 (98.5)	100 (99.0)	0.722 (0.302)
During PFS	936 (99.1)	195 (100)	0.751 (0.235)	263 (98.9)	97 (99.0)	0.748 (0.260)
During PD	237 (97.1)	135 (97.8)	0.740 (0.207)	306 (98.1)	86 (98.9)	0.692 (0.337)

PFS = Progression-free survival; PD = Progressive disease.
* % of patients, as a percentage of all patients/or potential datapoints