Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours

A Single Technology Appraisal

Produced by	Peninsula Technology Assessment Group (PenTAG) University of Exeter Medical School South Cloisters St Luke's Campus Heavitree Road Exeter EX1 2LU
Authors	**Tracey Jones-Hughes**,[1] *Research Fellow* **James Dunham**,[1] *Graduate Research Assistant* **Sophie Robinson**,[1] *Information Scientist* **Mark Napier**,[2] *Consultant Medical Oncologist* **Martin Hoyle**,[1] *Associate Professor* [1] Peninsula Technology Assessment Group (PenTAG), Exeter, UK [2] Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
Correspondence to	Tracey Jones-Hughes, South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU

Date completed	31/05/2017
Source of funding	This report was commissioned by the NIHR Systematic Reviews Programme as project number 16/51/15.
Declared competing interests of the authors	None
Acknowledgments	We acknowledge the excellent administrative support of Sue Whiffin and Jenny Lowe (both of University of Exeter).
Rider on responsibility for document	The views expressed in this report are those of the authors and not necessarily those of the NIHR SR Programme. Any errors are the responsibility of the authors.
This report should be referenced as follows	Jones-Hughes T, Dunham J, Robinson S, Napier M, Hoyle M. Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2017.

Contributions of authors

Tracey Jones-Hughes	Led the critique of the company's decision problem and clinical effectiveness evidence. Wrote the Summary, Background, Decision problem, Clinical effectiveness and Overall conclusions. Compiled the report. Provided overall project management.
James Dunham	Contributed to the critique of the economic model and contributed to writing the Cost-effectiveness and End-of-life sections.
Sophie Robinson	Led the critique of the company's literature searching for this submission. Wrote the review of the literature searches for the report. Contributed to the writing and editing of the report.
Mark Napier	Provided clinical advice on soft tissue sarcoma and its management within the NHS. Reviewed and revised a draft version of the report.
Martin Hoyle	Contributed to writing the Cost-effectiveness and End-of-life sections. Contributed to the critique of the economic model and is the guarantor of the report.

Contents

List of tables

List of figures

Abbreviations

AESI	Adverse Events of Special Interest
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BSA	Body Surface Area
BSC	Best Supportive Care
CDF	Cancer Drug Fund
CEA	Cost Effectiveness Analysis
CEAC	Cost Effectiveness Acceptability Curve
CRD	University of York Centre for Reviews and Dissemination
CI	Confidence Interval
CrI	Credible Interval
CRUK	Cancer Research UK
ECOG	Eastern Cooperative Oncology Group
EED	National Health Service Economic Evaluations Database
eMIT	Electronic Market Information Tool
EPAR	European Public Assessment Report
EQ5D	EuroQol-5 Dimensions
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
FDA	Federal Drug Administration
GIST	Gastrointestinal stromal tumour
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
ICER	Incremental Cost-Effectiveness Ratio
ICC	Interstitial cells of Cajal
IPCW	Inverse Probability of Censoring Weights
IPE	Iterative Parameter Estimation
ITT	Intent-to-treat
KM	Kaplan Meier
LYG	Life-Years Gained
MedDRA	Medical Dictionary for Regulatory Activities
NICE DSU	National Institute for Health and Care Excellence Decision Support Unit
ORR	Objective Response Rate
OS	Overall Survival
OSA	One-way Sensitivity Analysis
PFS	Progression Free Survival
PPS	Post-Progression Survival
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-Adjusted Life Year

RCT	Randomised Controlled Trial
RECIST	Response Evaluation Criteria in Solid Tumours
RPSFT	Rank Preserving Structural Failure Time method (RPSFT)
SAE	Serious Adverse Event
SPC	Summary of Product Characteristics
TEAE	Treatment-Emergent Adverse Event
TSD	Technical Support Document
WHO	World Health Organisation

1 Summary

1.1 Critique of the decision problem in the company submission

The company defined the population as patients with unresectable or metastatic gastrointestinal stromal tumours (GIST) whose disease has progressed on, or who are intolerant to, previous treatment with imatinib and sunitinib. This definition agrees with the population described in the NICE scope.[1]

The intervention in the decision problem was regorafenib, and the comparator was best supportive care (BSC), as in the NICE Scope. The outcomes in the company submission also match those in the Scope.

Although the NICE scope did not consider any subgroups, preplanned investigations by the company include geographic region, prior line of treatment, age, sex, baseline BMI, duration of imatinib treatment, ECOG performance status, and mutational status

1.2 Summary of clinical effectiveness evidence submitted by the company

The primary focus of the company's submission was the GRID study, which was a phase 3 randomised controlled trial. The GRID study was double-blind and multi-centre (57 centres, 17 countries).

Patients were randomised to regorafenib + best supportive care (N=133) or to placebo + best supportive care (N=66). Baseline characteristics were reported as being balanced between arms, however, there was a slight imbalance where 67% of participants receiving regorafenib and 83% receiving placebo had >18 months of previous imatinib therapy.

At the June 2015 cut-off, fifty eight participants in the placebo arm (87.9%) had crossed over to the regorafenib arm.

Outcome results were as follows:

Progression-free survival

The regorafenib group was assessed via blinded review to be superior to the placebo group (147 days [4.8 months] vs 28 days [0.9 months]), with the risk of progression or death in the regorafenib arm lower than in the placebo arm (Hazard ratio [HR] 0.27, 95% CI 0.19-0.39; p<0.000001).

Secondary endpoints

The uncorrected median OS for the regorafenib and placebo arms was 17.4 months. Following adjustment for crossover, median OS was shown to be longer in the regorafenib

group (529 days) than in the placebo group (338 days [p = 0.00095] using the Iterative Parameter Estimation method; 361 days [p = 0.00286] using the Rank-Preserving Structural Failure Time method). The estimated corrected hazard ratio of regorafenib to placebo using the RPSFT and IPE correction methods were 0.616 (95% CI 0.435 - 0.871) and 0.586 (95% CI 0.417 - 0.824), respectively.

Other secondary outcomes were reported as follows (Source: Bayer submission, section 1.3, p20):

- Median time to progression (TTP) was significantly longer in the regorafenib arm than in the placebo arm (5.4 months [165 days] versus 0.9 months [28 days], HR 0.248, 95% CI 0.170–0.364; p<0.000001).
- Tumour Response Rate, showed no statistically significant difference between arms despite the higher trend in the regorafenib group (4.5%) compared to the placebo group (1.5%).
- Disease Control Rate (DCR) was significantly higher in the regorafenib group (52.6%) vs. the placebo group (9.1%) (one-sided p<0.000001)
- For HRQoL, there was no statistically significant difference between patients receiving regorafenib and patients receiving placebo.

As noted above, no statistically significant difference was evident between treatment groups for tumour response rate. However, the company highlight that within-tumour necrosis promotes disease stabilisation without reduction in size, which is an observed effect of kinase inhibitors in TKI-resistant disease.

Subgroups

Bayer found regorafenib to be effective across all subgroups for progression-free survival except for the small subset of patients (n=22) with duration of imatinib treatment of less than 6 months.

Overall survival for subgroups was presented as uncorrected for crossover and corrected via RPSFT and IPE. The results were similar for the main OS results, however, confidence intervals were wide and indicating heterogeneity and a lack of statistical significance. Bayer do point out that the low number of events within subgroups will contribute to this.

Adverse events

During the double-blind study phase of the GRID study, drug-related adverse events were reported in 132 (100%) patients in the regorafenib group and 61 (92%) patients in the placebo group. Treatment discontinuations due to regorafenib-related events were reported

for 16.8% of patients and distributed across all system organ classes. Five deaths were reported as related to regorafenib treatment by investigators (cardiac arrest, acute hepatic failure, acute kidney injury, colonic perforation, and thromboembolic event).

The most serious adverse drug reactions in patients receiving regorafenib were haemorrhage, severe liver injury, and gastrointestinal perforation and the most common adverse events included hand-foot skin reaction (HFSR), hypertension, diarrhoea, mucositis and fatigue

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

The methods for the searches and systematic review were adequate and well described, therefore, the ERG concluded that the company did not miss any evidence.

The primary focus of the company's submission was the CRC study. This was an appropriately-designed double-blinded, multi-centre RCT. The treatment arms were balanced and patients were representative of the UK population.

The crossover of 87.9% of placebo-treated patients to open-label regorafenib following disease progression may cause the OS to be overestimated, assuming regorafenib provides a clinical benefit for this outcome. Therefore Bayer applied two correction methods, which have been assessed as appropriate by the ERG, resulting in a statistically significant difference for OS in favour of regorafenib.

1.4 Summary of cost-effectiveness evidence submitted by the company

So far, we have received a total of three versions of Bayer's economic model and cost-effectiveness results.

We received Bayer's economic model and full report on 21st March 2017.

On 25th April 2017, after an earlier request for clarification from us, we received a second version of Bayer's economic model and cost-effectiveness results. This included some updated OS data, as discussed in Section 5.3.6, p74.

On 16th May 2017, in response to another request for clarification from us, we received a third version of Bayer's economic model and cost-effectiveness results. In addition to the updated OS data, this also included some updated data on treatment duration of regorafenib as discussed in Section 5.3.8.1, p102.

1.4.1 Company's systematic review of economic evaluations

Bayer conducted a systematic literature review of economic and cost-effectiveness studies. They considered only one study to be relevant, an analysis for the relevant patient population in England. The base case ICER for regorafenib vs placebo was £34,420 - £40,188 per QALY according to the method of adjustment for treatment switching.

1.4.2 Company's submitted economic evaluation

1.4.2.1 Methods

The company presented a model-based economic evaluation to address the decision problem.

Bayer submitted a partitioned survival model with three independent health states; progression-free survival (PFS), post-progression survival (PPS), and Death. Patients enter the model upon treatment commencing for either regorafenib or the comparator, best supportive care (BSC). The model uses a 28-day cycle length and a time horizon of 40 years. A half-cycle correction is applied. Outputs of the model (costs, life years and quality-adjusted life years [QALYs]) were discounted at 3.5% per annum.

Health state utility values in the base case were estimated using EQ-5D measurements from patients in the GRID trial. Paired samples and repeated measures methods were used to estimate the values, with paired comparisons preferred by Bayer. Bayer's base case HSUVs are independent of treatment group. Bayer extensively examine the effect of different HSUV estimates in their scenario analyses. The impact of adverse events on health-related quality of life was also directly modelled for the treatment groups.

Costs were modelled from the NHS and Personal Social Services perspective. Bayer's base case includes options for costing the drug at list price, as well as offering a confidential patient access scheme (PAS) applied to the cost of regorafenib.

Bayer's method of modelling the treatment duration of regorafenib changed substantially from the time of their original report submission to the time of our report submission. Regorafenib treatment in the regorafenib arm of GRID was continued after disease progression. However, instead, Bayer modelled regorafenib treatment only up to progression, as they claimed this would be as in clinical practice in England & Wales, citing surveys of physicians. In response to our question for clarification, they completely changed their method of modelling treatment duration. In particular, they now model treatment with regorafenib for the entire duration as experienced in GRID RCT. We agree with this approach.

Other resource costs for regorafenib and placebo patients were identified through using clinician surveys conducted by Bayer. This included one-off costs, such as end-of-life costs, as well as health state costs, which consisted of outpatient monitoring visits, regular tests and medication for pain management. A variety of sources were used to estimate unit costs, including: Published studies, PSSRU Unit 2015, NHS reference costs 2015/16, 2016/17 National Tariff, and the Drug tariff 01/2017. The costs of adverse events were also modelled, although they were negligible. Univariate and probabilistic sensitivity analyses were conducted to explore uncertainty in the incremental cost-effectiveness ratio (ICER) and to identify parameters to which the model was sensitive. Scenario analyses to examine the model's sensitivity to structural assumptions were also conducted.

1.4.2.2 Clinical outcomes in model

Treatment effectiveness was estimated using the GRID trial. The economic model considered progression-free survival and overall survival. In their base case, Bayer assume the lognormal distribution for PFS, which we consider reasonable.

87.9% of patients in the placebo arm crossed over to the regorafenib arm after disease progression. This introduces the possibility of overestimating OS in the placebo arm and hence confounding the cost-effectiveness estimates. Bayer considered three crossover correction methods; Iterative Parameter Estimation (IPE), Rank Preserving Structural Failure Time method (RPSFT), and Inverse Probability of Censoring Weights (IPCW). The IPCW method was rejected due to the high proportion of placebo patients crossing over, and we consider this reasonable. In their base case, Bayer assume the IPE method and we also consider this reasonable. The cost-effectiveness of regorafenib is extremely sensitive to the adjustment for treatment switching, specifically, Bayer's base case ICER of £38,000 per QALY assuming the PAS increases to over £100,000 per QALY based on the unadjusted ITT OS data.

In their original report, Bayer presented OS data with a cut-off date of June 2015. In our clarification letter, we ask Bayer whether they could provide us with more mature data, given that the existing data is now about two years out of date, and that a reasonable amount of extrapolation is required. In response, on 25th April 2017, we received OS data from Bayer with cut-off of 2017. Bayer also included an updated version of their economic results.

Despite the fact that the Kaplan-Meier graph for the placebo arm changed only very slightly from the 2015 to the 2017 data cut-off, Bayer estimate a shorter OS for placebo after correction for cross-over using the 2017 data, compared to the 2015 data. Specifically and importantly, the estimated mean OS in the placebo arm decreases by 24%. Bayer justify this as follows: "*This is a result of the greater follow-up time allowing for a longer potential*

censoring date within the crossover adjustment calculation" (Bayer response to clarification, p11). This reduction in mean OS substantially improves the cost-effectiveness of regorafenib. For example, assuming the PAS, the ICER for regorafenib vs. BSC decreases from £49,000 to £38,000 per QALY.

However, we have several important concerns with the switching adjustment applied to the 2017 data. Given these concerns, we use the 2015 data-cut for OS in our base case.

We turn now to the extrapolation of OS. Two consultant oncologists, who specialised in the disease area, validated the fittings of various parametric models, on behalf of Bayer. They argued that the loglogistic, Weibull and Gompertz models all looked clinically plausible. However, in their base case, Bayer chose the log-logistic distribution for OS based on the accuracy of the fit the data from GRID.

We surveyed the literature for studies that could help to inform the extrapolation of OS. We found just one relevant study, which suggested, if anything, a reasonably long tail for OS. However, we caution not to rely solely on this study to inform extrapolation, due to limitations in comparability with the GRID study. On balance, in our base case, we model OS as the average of the shorter tailed Weibull and longer tailed log-logistic distributions.

Bayer do not explicitly model background general population mortality. In our base case, we include this additional mortality.

1.4.2.3 End of Life criteria

We agree with Bayer that regorafenib meets the End of Life criteria.

1.4.3 Results

In Bayer's base case analysis (without/with PAS), treatment with regorafenib resulted in 1.7333 QALYs at a cost of [redacted]/£47,249, while treatment with the placebo resulted in 0.761 QALYs at a cost of £10,395. The QALY differential was 0.971 and the cost differential was £[redacted]/£36,854. The corresponding ICERs per QALY were £[redacted]/£37,941.

Regorafenib was predicted to result in QALY gains in both PFS and OS, with the benefits roughly similar in both health states. The overall QALY gain depends heavily on the treatment switching adjustments.

Drug acquisition costs were by far the largest cost in the regorafenib arm at £[redacted]/£[redacted] which was also the incremental cost as the placebo arm had zero drug costs. Other cost differentials were much smaller; the next largest incremental cost was +£[redacted] for monitoring costs in the regorafenib arm. Remaining costs were very similar between the two treatment arms.

In the probabilistic sensitivity analysis, the ICERs per QALY were similar to the deterministic case at £[redacted]/£38,494 without and with the PAS. Both costs and QALYs were very similar to the base case. At a willingness to pay threshold of £50,000 per QALY, regorafenib had a [redacted]/82% chance of being cost-effective.

Univariate sensitivity analyses were also carried out, indicating that results were sensitive to a number of parameters. Regorafenib drug costs and utility discount rates were the most impactful parameters, with HSUVs and cost discount rates also being significant.

Bayer also carried scenario analyses looking at assumptions for: OS extrapolation, treatment switching, resource use, and utility elicitation method. The most impactful of these were the choice of OS extrapolation, and the method of treatment switching adjustment.

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

The derivation of the PenTAG base case is shown in Table 1 below.

Total uncertainty in the cost-effectiveness of regorafenib versus BSC is high due to:

- Substantial uncertainty in the adjustment for widespread treatment switching on diseases progression, from BSC to regorafenib.
- Important uncertainty in the extrapolation of OS.

In key plausible scenario analyses we suggest alternative plausible methods of extrapolating OS, and of modelling costs and QALYs only whilst patients are in PFS.

Table 1. Derivation of PenTAG base case ICERs Regorafenib vs. BSC (£ per QALY)

				Regorafenib price	
				PAS	List
	Bayer base case			**£38,000**	[redacted]
	PenTAG assumption	**Bayer assumption**			
1	OS from 2015 data-cut	OS from 2017 data-cut	(Section 5.3.6.2, p79)	£49,000	[redacted]
2	Include general mortality from UK population	Do not including general mortality from UK population	(Section 5.3.6.3, p87)	£41,000	[redacted]

3	OS average of Log-logistic / Weibull	OS average Log-logistic	(Section 5.3.6.3, p87)	£41,000	[redacted]
4	Utilities decrease with age	Utilities independent of age	(Section 5.3.7, p95)	£39,000	[redacted]
1 + 2				£52,000	[redacted]
1 + 3				£52,000	[redacted]
2 + 3				£43,000	[redacted]
1 + 2 + 3				£55,000	[redacted]
		ICER		**£56,000**	[redacted]
1+2+3+4	**PenTAG base case**	**Uncertainty**		**High, mostly due to switching adjustment, but also extrapolation.**	

Key: ICER, incremental cost-effectiveness ratio; OS, overall survival; PAS, patient access scheme; PFS, progression-free survival; QALYs, quality-adjusted life year(s); Dark shading indicates ICER > £50,000 per QALY.

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

1.6.1 Strengths

- Double-blind multi-centre randomised trial.
- The population recruited to the GRID study was representative of the typical UK population.
- Bayer's analysis has been clearly described.
- There were no noteworthy wiring errors in the economic model.

1.6.2 Weaknesses and areas of uncertainty

- Whilst Bayer were responsive to our questions of clarification, they sent us a total of three version of their economic model.
- The substantial amount of treatment switching introduces a great deal of uncertainty in the estimated cost-effectiveness.
- We are not convinced that Bayer have correctly adjusted for treatment switching in the most recent, 2017, data cut.

- Extrapolation of OS also introduces appreciable uncertainty in the estimated cost-effectiveness.

2 Background

2.1 Critique of company's description of underlying health problem

Gastrointestinal stromal tumours (GIST) are a rare type of soft tissue sarcoma that develops in the connective tissues of the digestive system, commonly (~60%) in the wall of the stomach. However, they may originate elsewhere such as the small intestine (~30%) or oesophagus and, very rarely, outside the gastrointestinal tract.[2]

GIST are the most common mesenchymal neoplasms of the GI tract, but actually represent less than 1% of tumours in this region. [3]

For many people with GIST, the c-kit oncogene which is found in all cells in the body has undergone a mutation. This oncogene directs the cell to produce the KIT protein, subsequently causing the cell to replicate. Within the interstitial cells of Cajal (ICCs), the c-kit gene is inactive unless there is a need for more cells. However, in most GISTs, there may be an inherited mutation of the c-kit gene leading to a high cell division rate.

A less frequent mutation also exists, known as PDGFRA, which causes the cell to overproduce a different protein (also called PDGFRA), but which has the same effect as KIT.

The majority of GISTs will have one or other of the mutations, but not both. There is also a small population of patients who have neither of these changes.

Bayer describe GIST in more detail as follows (Source: Bayer submission, section 3, p38):

> *Gastro-intestinal stromal tumours (GISTs) are rare connective tissue tumours that show a differentiation profile similar to the interstitial cells of Cajal involved in the regulation of the digestive system.[…] Pathologically, most of GISTs are caused due to oncogenic mutations in either KIT or PDGFRA (23). The majority of the cases (75% to 80%) have KIT mutations that typically affect the juxtamembrane domain encoded by exon 11, while 5% to 8% GISTs have PDGFRA mutation and 12% to 15% have KIT and PDGFRA wild-type mutations (23).*

The ERG believes the description given by the company is appropriate.

2.1.1 Epidemiology

According to Amelio et al. 2014, UK estimates of GIST annual incidence range from 1.32–1.50 per 100,000 population, which is equivalent to approximately 800–900 new cases each year.[4] No UK prevalence has been reported, however, reports from western Sweden estimate prevalence at 12.9 per 100,000.[4]

As mentioned above, most GISTs are sporadic and occur because of a c-kit or PDGFRA oncogene mutation encouraging the GIST cells to grow and multiply. There are also a number of rarer types of GIST some of which may be due to an inherited gene mutation:[5]

- Wild-type GIST: A type of GIST that is not caused by a known cell mutation
- Paediatric GIST: A GIST affecting children and young adults. Paediatric GIST is very rare
- Syndromic GIST: A type of GIST linked to Carney's Triad Syndrome and Carney-Stratakis Syndrome and Neurofibromatosis.
- Familial GIST: A rare inherited form of GIST

Risk factors include age and sex, as GIST most often occurs in people older than 50 and is slightly more common in men than in women.[6]

2.1.2 Diagnosis

The symptoms of GIST can vary according to size and location of the tumour. Initial diagnosis, following clinical examination may be via a diagnostic scan and biopsy. [5]

With regard to metastatic GIST, Bayer list the typical symptoms below (Source: Bayer submission, section 3.2, p39):

> *Metastatic GIST is a terminal disease for which patients may experience general systemic symptoms such as fever, nausea, abdominal discomfort and weight loss as well as psychological distress and functional impairments (5)*

Other symptoms may include fatigue, blood in stools or vomit and anaemia.

2.1.3 Prognosis and burden of disease

The overall 5-year survival rate for people with GIST has been reported as 76%.[6] However, this was estimated from data collected between 2003 and 2009 from the American Cancer Society and both diagnosis and available treatments have improved since then. The estimate reduces to 74% if the cancer has spread locally, and falls to 48% for distant metastasise.[6] Whereas, if the cancer is contained within the original organ, the 5-year survival rate is improved at 91%.

The most reliable prognostic factors for GIST are considered to be:[7]

- The size of the primary tumour,
- The mitotic index i.e., the ratio between the number of cells in a population undergoing mitosis to the number of cells in a population not undergoing mitosis.

- The location of the primary lesion, with small bowel and rectal primary GIST less favourable than gastric GISTs.
- PDGFRA mutations which are most commonly associated with gastric primary lesions have a more favourable prognosis.
- Histologic type may also impact prognosis, with spindle cell displaying a higher five-year survival rate than epithelioid or mixed histology. However, in contrast, others report a prognostic influence of the degree of cellularity but not histologic subtype.

The company submission provides the following details on prognosis for people with GIST: (Source: Bayer submission, Section 3, p38)

> *For people with GIST, the prognosis depends mainly on whether the tumour is resectable. Size, location, and stage of tumour at initial diagnosis are also important factors for the prognosis of the tumour (26).*
>
> *Surgery represents the cornerstone treatment of localised GISTs (26). Complete removal of GIST is potentially curative, especially when it is small in size and the risk classification is low. However, the risk of relapse after surgery can be substantial, as defined by available risk classifications […]*

2.2 Critique of company's overview of current service provision

2.2.1 Current UK GIST treatment pathway

A summary of treatment options for GIST are as follows:[8]

- Localized, smaller (resectable) tumours - surgery is the main treatment and for tumours that are small and are not growing quickly, this is often the only treatment needed. Recurrence is more likely if the tumour is larger, did not start in the stomach, or if the cancer cells have a high mitotic rate. In this case, an adjuvant treatment with imatinib may be recommended for a minimum of a year post-surgery. For tumours that are highly likely to come back, many doctors now recommend giving patients at least 3 years of imatinib.
- Localized, larger (marginally resectable) tumours - may require more extensive surgery leading to further health problems later on. Therefore, once a biopsy confirms the tumour is a GIST, treatment with imatinib is usually commenced and continues at least until the tumour stops shrinking. At this point, surgery may be possible. If the tumour is still too large for surgery, imatinib may be continued, followed by sunitinib if the first-line treatment is no longer effective. If sunitinib is no longer working, the targeted drug regorafenib may help some patients.

- Unresectable tumours and metastases - imatinib is usually the preferred first treatment option. It is continued as long as the tumour has a stable response. If the tumour progresses, it may respond to increasing the dose of imatinib. If the tumour continues to grow or the side effects from imatinib are too severe, a switch to sunitinib may be helpful. If sunitinib is no longer working, regorafenib may help some patients as a third-line treatment. If the tumour shrinks enough with targeted therapy, surgery may then be an option for some patients. This might be followed by more targeted therapy if it is still effective. If the cancer has spread to only 1 or 2 sites in the abdomen (such as the liver), the surgeon may advise removing the main tumour and trying to remove these other tumours as well. Usually this should be considered only for tumours that are slow growing or those causing local complications such as uncontrollable bleeding. Other options to treat cancers that have spread to the liver include ablation and embolization. These treatments may include radiofrequency ablation (RFA; using electric currents to heat the tumour), or ethanol ablation (injecting concentrated alcohol into the tumour). Cancers that are no longer responding to the targeted drugs discussed above can be hard to treat. Some doctors may recommend trying other targeted drugs, such as sorafenib (Nexavar®), dasatinib (Sprycel®), or nilotinib (Tasigna®), although it's not yet clear how helpful these drugs are.
- Recurrent tumours - Treatment options for GISTs that recur after treatment depend on the location and extent of the recurrence. For most recurrences, treatment with imatinib is probably the best way to shrink any tumours, as long as it is still effective and the patient can tolerate taking it. If the starting dose of imatinib does not work, the dose can be increased. Another option is to try sunitinib or regorafenib. If the cancer comes back as a single, well defined tumour, removing or destroying the tumour may be an option. Doctors are still not certain if removing GISTs that come back after treatment helps people live longer. Some studies have found that it does, but other studies disagree.

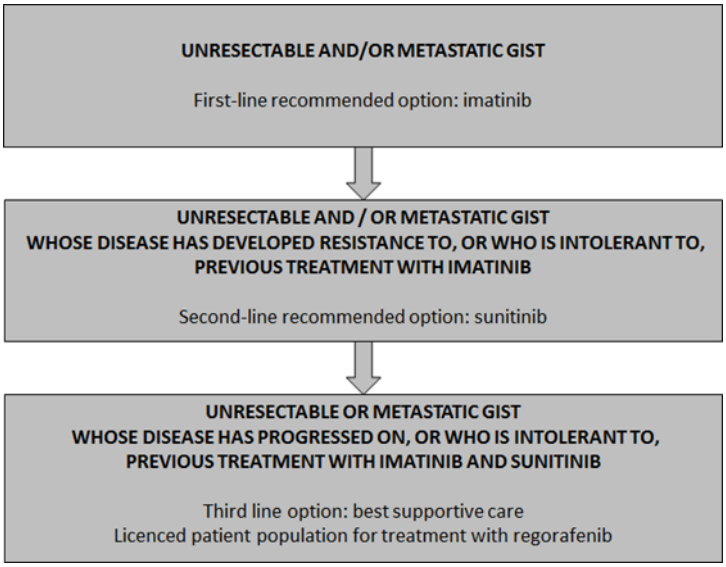
Best supportive care is provided to patients who fail to respond to imatinib and sunitinib. Although there is no strict definition, this generally involves care to prevent or treat the symptoms of GIST, side effects caused by treatment, and psychological, social, and spiritual problems related to a disease or its treatment. Radiation therapy is sometimes given as supportive care to relieve pain in patients with large tumours that have spread.[9]

Bayer report that pain management may be administered as follows: (Source: Bayer submission, Section 3, p39)

According to a survey conducted in 2013 and involving physicians from England and Wales, pain management treatments were confirmed to comprise co-codamol, tramadol, paracetamol, morphine sulphate and dexamethasone.

Within the UK, the clinical pathway falls under the NICE pathway for stomach cancer, as shown in Figure 1, which also includes the proposed position for regorafenib: (Source: Bayer Submission, Section 3, p40)

Figure 1. UK Clinical pathway for GIST



Source: Bayer submission, Section 3, p40, Figure 1

2.2.2 Anticipated place of regorafenib in clinical practice

In England, there are no other lines of therapy recommended by NICE for the treatment of patients with unresectable or metastatic GIST whose disease has progressed upon treatment with sunitinib. Therefore, Bayer anticipates that regorafenib will be an option for this population of approximately 60 new patients per year. The ERG considers this an appropriate figure, given approximately half of new cases of GIST are likely to be metastatic and/or unresectable on first presentation and will initially be treated with imatinib/sunitinib.[10]

3 Critique of company's definition of decision problem

The company presented their decision problem within the Executive Summary chapter, under the subheading 'statement of the decision problem' (Bayer submission, Section 1.1, p. 16). A summary table of the NICE Scope, the company's decision problem and the ERG's critique is presented below (Table 2).[1] Clearly, Bayer's definition of the decision problem is closely aligned with the NICE Scope.

Table 2. Summary table of decision problem critique

Decision problem	NICE Scope	Company's decision problem	ERG notes
Population	Patients with unresectable or metastatic gastrointestinal stromal tumours (GIST) whose disease has progressed on, or who are intolerant to, previous treatment with imatinib and sunitinib.	As per Scope	No comments
Intervention	Regorafenib	As per Scope	No comments.
Comparator	Best supportive care (BSC)	As per Scope	No comments
Outcome	The outcome measures to be considered include: • Overall survival • Progression-free survival • Adverse events of treatment • Health-related quality of life	As per Scope	The company include additional secondary outcomes • time to progression, • tumour response • objective response rate • disease control rate • duration of response

Source: NICE Scope[1] and Bayer submission, Table 1, p. 16–17

3.1 Population

The defined population in the company's submission (patients with unresectable or metastatic gastrointestinal stromal tumours (GIST) whose disease has progressed on, or who are intolerant to, previous treatment with imatinib and sunitinib), agrees with the population specified in the NICE Scope. Inclusion criteria also require an ECOG performance status of 0 or 1 which is standard for RCTs. [11]

3.2 Intervention

The company's decision problem specified the intervention as 'regorafenib', which matches the NICE Scope.[1]

Regorafenib is a multitargeted tyrosine kinase inhibitor with antiangiogenic activity. It has inhibitory action against several tyrosine kinases, including KIT, PDGFRA, bFGFR, VEGFR1-3, TIE2, RET, BRAF and BRAF V600E.[12]

Regorafenib (Stivarga®, Bayer) is approved for the treatment of adult patients with unresectable or metastatic gastrointestinal stromal tumors (GIST) who progressed on or are intolerant to prior treatment with imatinib and sunitinib. The marketing authorisation is presented by the company as follows (Source: Bayer submission, Section 2.2, p29):

> *Initial marketing authorisation for regorafenib (Stivarga®) was received on June 27th, 2013 for the treatment of metastatic colorectal cancers who have been previously treated with, or are not considered candidates for, available therapies.*
>
> *On June 26th, 2014 the CHMP released its positive opinion on the extension of indication for regorafenib in the treatment of adult patients with unresectable or metastatic gastrointestinal stromal tumours (GIST) who progressed on or are intolerant to prior treatment with imatinib and sunitinib.*
>
> *Treatment with regorafenib (Stivarga®) for patients with unresectable or metastatic GIST whose disease has progressed on, or who are intolerant to, previous treatment with imatinib and sunitinib has been funded through the Cancer Drug Fund (CDF) since 2013.*

The recommended dose on the marketing authorisation is 160 mg once daily for three weeks followed by one week off therapy. The clinical evidence supplied by Bayer is in agreement with this.[11]

3.3 Comparators

The only comparator listed in the NICE scope and in the company submission is best supportive care (BSC).[1] Best supportive care is defined by the company as (Source: Bayer submission, Section 1, p13):

> *…any method to preserve the comfort and dignity of the patient, excluding disease-specific antineoplastic therapy, radiation therapy, or surgical intervention (8).*
>
> *[…]According to two surveys, conducted in 2013 and 2016 and involving physicians from England and Wales, pain management treatments were confirmed to comprise co-codamol, tramadol, paracetamol, morphine sulphate and dexamethasone.*

Similarly, the GRID study includes placebo+BSC (blind) for comparator, with BSC defined in study protocol as follows:[11]

> *any method to preserve the comfort and dignity of the patients, and excludes any disease-specific anti-neoplastic therapy such as any kinase inhibitor, chemotherapy, radiation therapy, or surgical intervention.*

Chemotherapy is also listed as an exclusion to BSC in the GRID study.

3.4 Outcomes

The outcomes in the company submission comply with the scope (Source: Bayer submission, Section 4.3, p68):

- Overall survival - Assessment of survival status was performed every 3 months.
- Progression-free survival (primary endpoint) - PFS was assessed by central radiology reviewers who were masked to assignment and data from patients. Two readers reviewed the images. Tumour assessments were made at baseline, then every 4 weeks for the first 3 months, every 6 weeks for the months 4 to 6, and every 8 weeks thereafter until the end of study drug administration.
- Adverse events of treatment - Investigators rated severity of adverse events according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0) [NCI CTCAE V4.0].
- Health-related quality of life - Health-related quality of life questionnaires (EORTC QLQ-C30 and EuroQoL EQ-5D) were routinely completed by patients.

Other outcomes included in the study protocol include;

- time to progression,

- tumour response
- objective response rate
- disease control rate
- duration of response

4 Clinical effectiveness

4.1 Critique of the methods of review(s)

4.1.1 Searches

Bayer presented a literature search protocol to support its review of clinical effectiveness. This protocol included systematic searches of key biomedical databases using a literature search strategy, searching of conference websites and a search of clinical trials.gov. The literature search was last updated in December 2016.

The bibliographic database searching used a search strategy that took the following form:

1. (controlled index terms for gastrointestinal tumous and various sub-types of gastrointestinal tumour including gastrointestinal stromal tumour) OR
2. (free-text terms for gastrointestinal tumour and various related terms) AND
3. (controlled index terms for regorafenib or drug therapy or palliative therapy) OR
4. (free-text terms for regorafenib or drug therapy or palliative therapy) AND
5. (a range of search terms for study design (RCTs, clinical trials, controlled studies, comparative studies and prospective studies) NOT
6. (a range of search terms to exclude case studies, conference abstracts and letters) AND
7. (limited to 2000 onwards and humans).

The search strategy was applied in the following bibliographic databases: Medline-in-Process (PubMed), Medline and Embase (Elsevier at Embase.com) and The Cochrane Library.

The following conference websites were searched: American Society of Clinical Oncology (ASCO) in 2016 (month not stated) and European Society for Medical Oncology (ESMO) in 2016 (month not stated). Finally, clinicaltrials.gov was searched for relevant, unpublished studies (no date for this search is stated).

The literature searching for clinical effectiveness studies is reasonably well conducted and reported. However there are a few concerns.

- We do not have access to Embase.com so are unable to test the searches but the value of searching Medline and Embase simultaneously with one strategy is debatable since these databases use different indexing terms (Emtree for Embase and MeSH for Medline).

- The filter used to limit to RCTs is not the Cochrane search filter or any other validated filter that we recognize. It is unclear why a validated search filter was not used to limit to RCTs.
- Systematic reviews and meta-analyses were not searched for.
- The search for the intervention could have included further terms such as Stivarga and BAY 73-4506.
- The report describes 'hand-searching' clinicaltrials.gov, it is unclear what this entails and no further information is given.

There is insufficient information about the screening methods used for the review. Bayer have provided further details about their methods for screening in clarification but it is still not clear whether full text studies were double screened.

4.1.1.1 Quality of Life

Bayer presented a literature search protocol to support its review of health-related quality-of-life studies. This protocol included systematic searches of key biomedical databases using a literature search strategy and searching of conference websites. The literature search was last updated in December 2016.

The bibliographic database searching used a search strategy that took the following form:

1. (controlled index terms for gastrointestinal tumour and various sub-types of gastrointestinal tumour including gastrointestinal stromal tumour) OR
2. (free-text terms for gastrointestinal tumour and various related terms) AND
3. (a range of search terms for health utilities and quality of life) NOT
4. (a range of search terms to exclude conference abstracts) AND
5. (limited to 2000 onwards, English language and humans).

The search strategy was applied in the following bibliographic databases: Medline-in-Process (PubMed), Medline and Embase (Elsevier at Embase.com), EconLIT and NHS EED (The Cochrane Library).

The following conference websites were searched: American Society of Clinical Oncology (ASCO) in 2016 (month not stated), European Society for Medical Oncology (ESMO) in 2016 (month not stated), International Society for Pharmacoeconomics and Outcomes Research in 2016 (month not stated) and International Society for Quality of Life Research (ISOQoL) (month not stated).

The literature searching for health-related quality-of-life studies is reasonably well conducted and reported. However there are a few concerns.

- We do not have access to Embase.com so are unable to test the searches but the value of searching Medline and Embase simultaneously with one strategy is debatable since these databases use different indexing terms (Emtree for Embase and MeSH for Medline).
- It is not clear why NHS EED was included in the update search in December 2016 when this database has not been updated since April 2015.

There is insufficient information about the screening methods used for the review Bayer have provided further details about their methods for screening in clarification but it is still not clear whether full text studies were double screened.

4.1.1.2 Adverse events

Bayer did not undertake separate literature searches to identify studies reporting adverse events.

4.1.2 Inclusion criteria

Bayer's inclusion criteria in the search strategy are given below (Table 3) with an additional column added to the right of the table, taken from the Scope for reference and comparison. Comments about the differences in inclusion criteria are outlined below the table.[1]

Table 3. Scope of the literature review: PICOS criteria for study inclusion

Criteria	From Bayer Definition	From Scope
Population	Adult patients with metastatic, advanced, or unresectable GIST. Including 3rd line or later patients.	People with unresectable or metastatic gastrointestinal stromal tumours whose disease has progressed on, or who are intolerant to, previous treatment with imatinib and sunitinib
Interventions/ comparators	Regorafenib studies vs. placebo or BSC	Best supportive care
Outcomes	Efficacy outomes e.g. progression-free survival (PFS), overall survival (OS), Time to progression (TTP), disease control rate (DCR), response rate (ORR), duration of response (DOR). Safety outcomes e.g. adverse events Health-related Quality of life (HRQoL)	The outcome measures to be considered include: • Overall survival • Progression-free survival • Adverse effects of treatment • Health-related quality of life
Study Design	Randomised control trials (of any blinding status); non-randomised,	

Criteria	From Bayer Definition	From Scope
	controlled studies; uncontrolled single-arm trials; Cohort studies	

Key: GIST, gastrointestinal stromal tumour; BSC, best supportive care; PFS, progression-free survival; OS, overall survival; TTP, time to progression; DCR, disease control rate; ORR, overall response rate; DOR, duration of response; HRQoL, health-related quality of life.
Source: Bayer submission, Table 8, pp. 50–51 and NICE Scope[1]

4.1.2.1 Population

The population defined by Bayer differs slightly to the scope in that 3rd line or later patients are specified, whereas the population in the scope are intolerant to previous treatment with imatinib and sunitinib.[1] However, since 3rd line patients are likely to have received imatinib and sunitinib, the ERG believes the populations are essentially the same.

4.1.2.2 Interventions/comparators

The NICE Scope lists only best supportive care, whereas Bayer specify placebo or best supportive care. The use of a placebo would be necessary in blinded trials and is, therefore, appropriate.

4.1.2.3 Outcomes

The outcomes listed by Bayer include all those specified in the NICE Scope.

4.1.2.4 Study design

Bayer include several types of study design, including RCTs. Although the NICE Scope did not restrict study design, the NICE reference case guide to the methods of technology appraisal 2013 (Chapter 5.2.3)[13] recommends studies should be restricted to RCTs and when they are not available, non RCTs.

4.1.2.5 Study selection

The process for study selection as described by Bayer is standard for systematic reviews.

From 3,764 unique citations identified, 3173 were excluded and 591 were taken to full-text screening.

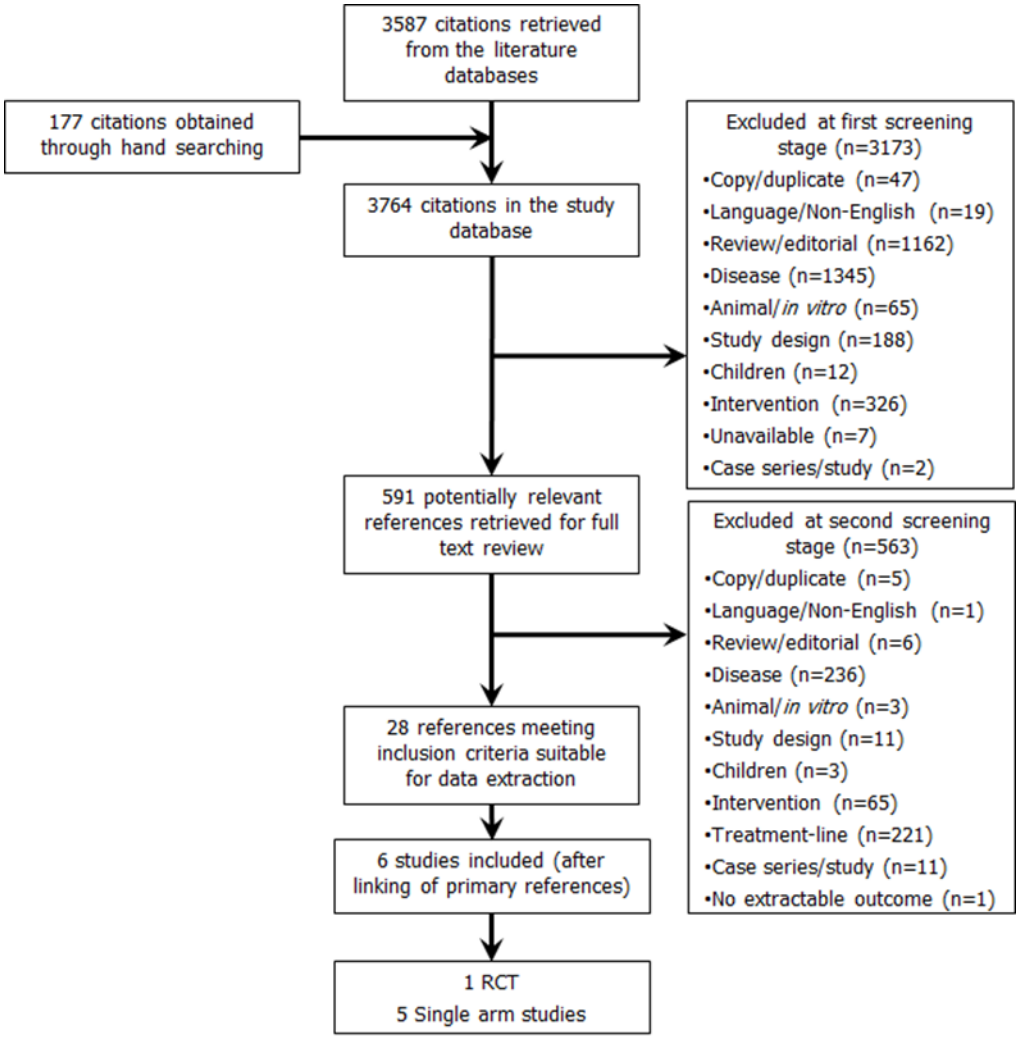
A further, 563 studies were excluded leaving the following (Source: Bayer submission, Section 4.1, p52):

Of relevance to the decision problem in this submission, 28 publications concerned the use of regorafenib. These publications related to 6 studies: one randomised controlled trial (RCT), and 5 single-arm studies. The single-arm studies included limited information and

patient numbers. This section further focuses on the identified RCT, the optimum design for assessing the benefits of treatments in oncology.

The PRISMA diagram reported in Bayer's submission is copied below (Figure 2).

Figure 2. PRISMA study flow diagram



Source: Bayer submission, Section 4.1, p52.

4.1.3 Critique of data extraction

The data extraction process is briefly explained for the one included study.[11] It is unclear if this was performed or checked independently by two researchers.

4.1.4 Quality assessment

Details of the company's critical appraisal of the GRID study,[11] alongside our critique, can be seen in Table 4.

Table 4. Critical appraisal of GRID study

Critical appraisal criterion	Bayer's Assessment	ERG Comment
Was randomisation carried out appropriately?	Yes Randomisation was performed via an interactive voice response system (IVRS). Investigators received the randomisation number for each participant through the IVRS and study drug supply was also managed via IVRS. Computer-generated randomisation lists were prepared by Bayer (pre-allocated block design, block size 12). Randomisation was stratified by treatment line (3rd vs. 4th line therapy or beyond) and geographical region (Asia vs. rest of the world).(Source: Bayer submission, Section 3, p58)	Block randomization with stratification is an appropriate method to ensure populations for the two treatments are approximately equal in size and balanced.
Was the concealment of treatment allocation adequate?	Yes Investigators received the randomisation number for each participant through the IVRS and study drug supply was also managed via IVRS. Regorafenib and placebo were identical in appearance.	The ERG agree that the method of allocation concealment is adequate.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes Demographics and baseline disease characteristics were comparable between the regorafenib and the placebo groups	The groups are generally balanced. However, the placebo group had a slightly larger population receiving >18 months of imatinib therapy (regorafnib 67% vs. placebo 83%).
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes Investigators received the randomisation number for each participant through the IVRS and study drug supply was also managed via IVRS. All patients, investigators, and the study sponsor were masked to treatment assignment through the use of the unique drug pack numbers preprinted onto each bottle, which was assigned to the patient by the IVRS. Regorafenib and placebo were identical in appearance in order to preserve blinding. Assessment of the primary endpoint (PFS) was carried out by central radiology reviewers who were masked to assignment and data from patients.	The ERG agree that the methods of blinding are adequate.
Were there any	No	The treatment duration was longer in the regorafenib arm, hence, the higher

Critical appraisal criterion	Bayer's Assessment	ERG Comment
unexpected imbalances in drop-outs between groups?	A higher number of patients withdrew from double-blind treatment in the regorafenib arm of the study (38%) than in patients receiving placebo (11%). This was mainly due radiological progression.	number of withdrawals due to radiological progression.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No Results of all pre-specified outcomes are reported in full.	The outcome measures listed in the protocol for the trial correspond with the outcome measures reported.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes/Yes/Yes The primary analysis was performed in the ITT population using radiological assessments taken during the double-blind phase of the study only. This was appropriate. Missing or unevaluable tumour assessments were not used in the calculation of derived efficacy variables unless a new lesion occurred, or the lesions that were evaluated already showed progressive disease (PD). No imputation was performed for missing lesion assessment and tumour response. For example, if a patient missed a scan visit and PD was documented at the next available scan visit, the actual visit date of the first documented PD was used. If a date was incomplete, such as only the year and month were available, day 15 of the month was used for the calculation.	Yes, we agree the main analysis adopts 'intention to treat' principles. The methods for dealing with missing data in this population appear to be standard.

Key: IVRS, interactive voice response system; IWRS, interactive web response system;
Source: Bayer submission, Appendix 3, p 36, Table 14

4.1.5 Evidence synthesis

From the searches, only one RCT was identified. Therefore synthesis of the evidence was not required.

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

4.2.1 Methods

The single RCT (study name GRID; main publication by Demetri et al. 2013) identified was presented in detail within the submission.[11] No additional relevant studies were identified by the ERG.

4.2.1.1 Study objective

The objectives are reported in the company submission as follows (Source: Bayer submission, Section 4.3, p51):

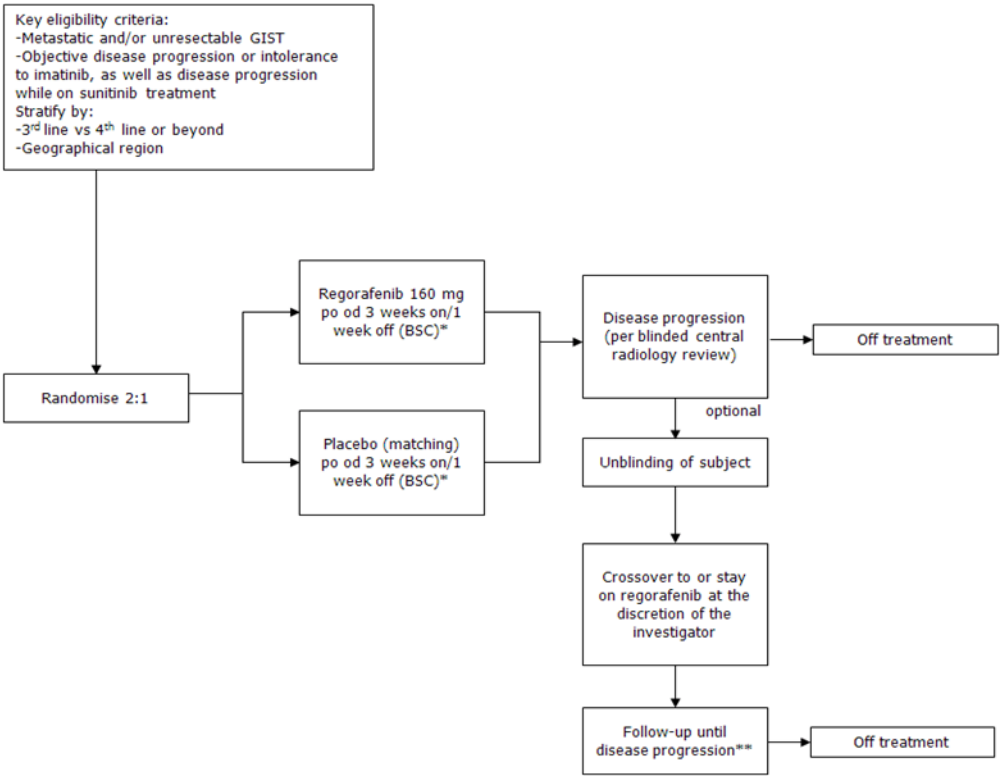
> *The primary objective of the GRID study was to compare regorafenib and placebo treatment in terms of progression-free survival (PFS) in patients with metastatic and/or unresectable GIST who have progressed after therapy with at least imatinib and sunitinib.*
>
> *Secondary objectives included evaluation of overall survival (OS), time to progression (TTP), disease control rate (DCR), tumour response rate (RR), duration of response (DOR), and safety of regorafenib. Health-related quality of life, pharmacokinetics, secondary PFS during open label treatment, and biomarker analysis were exploratory objectives within the study.*

The objectives correspond to the outcome measures detailed in the NICE Scope.[1]

4.2.1.2 Study design and treatment

The GRID study was a multicentre (57 centres; 17 countries), randomised, blinded, phase 3 trial investigating the efficacy of regorafenib for patients with GIST who have previously been treated with imatinib and sunitinib.[11] The overall trial design is displayed in Figure 3.

Figure 3. GRID trial design



Key: BSC, Best Supportive Care; GIST, Gastrointestinal Stromal Tumour; po, per os
Notes: ** Patients could continue treatment with regorafenib even after 1st progression (for regorafenib patients) or 2nd progression (for cross over patients)
Source: Bayer submission, Section 4.3, p58

As shown in Figure 3, participants receive either regorafenib or placebo once daily for the first 3 weeks of each 4-week cycle. Regorafenib was administered as 4 x 40 mg tablets, with a matching placebo for the control arm and both were stored in identical containers. Patients continued to receive treatment until disease progression, clinical progression, toxicity or consent withdrawal.

The intervention and control arms also included BSC, which is defined by the company as follows (Source: Bayer submission, Section 4.3, p 61):

> *BSC was defined as any method to preserve the comfort and dignity of the patient, and included any concomitant medications or treatments: antibiotics, analgesics, radiation therapy for pain control (limited to bone metastases), corticosteroids, transfusions, psychotherapy, growth factors, palliative surgery, or any other symptomatic therapy necessary to provide BSC, except other investigational anti-tumour agents or anti-neoplastic chemo/hormonal/immune/radio-therapy.*

Concomitant medication was permitted at the discretion of the principal investigator and in accordance with the protocol.

A pre-specified schedule was followed with regard to dose modification for unacceptable toxic effects, hand-foot skin reaction and hypertension. A maximum of two dose-reductions due to toxicity were permitted (from 160 mg to 120 mg to 80 mg) (Table 5). A subsequent dose re-escalation was allowed, subject to resolution of toxicities (Table 6).

Table 5. Regorafenib dose levels

Dose level	Dose	Form
Dose level 0 (standard dose)	160mg po od	4 tablets of regorafenib, 40mg/tablet, or 4 matching placebo tablet
Dose level -1	120mg po od	3 tablets of regorafenib, 40mg/tablet, or 3 matching placebo tablet
Dose level -2	80mg po od	2 tablets of regorafenib, 40mg/tablet, or 2 matching placebo tablet

Source: Bayer submission, Section 4.3, p62

Table 6. Dose modification for toxicities related to study drug (except hand-foot skin reaction and hypertension)[a]

NCI-CTCAE v4.0	Dose Interruption	Dose Modification	Dose for Subsequent Cycles
Grade 0-2	Treat on time	No change	No change
Grade 3[b]	Delay until < grade 2[b]	Reduce 1 dose level	If toxicity remains <grade 2, dose re-escalation can be considered at the discretion of the treating investigator. If dose is re-escalated and toxicity (≥ grade 3) recurs, institute permanent dose reduction
Grade 4	Delay until < grade 2[b]	Reduce by 1 dose level. Permanent discontinuation can be considered at treating investigator's discretion.	

Notes: a, excludes alopecia, non-refractory nausea/vomiting, non-refractory hypersensitivity and asymptomatic laboratory abnormalities; b, If no recovery after a 4-week delay, treatment will be permanently discontinued

Source: Bayer submission, Section 4.3, p62

With regard to hand-foot skin reaction, dose modification was adjusted according to skin toxicity grade with supportive measures. According to the level of toxicity or the number of occurrences, treatment may be discontinued or re-escalated.

[redacted]

Randomisation

Randomisation and allocation was performed with stratification by treatment line (3rd vs. 4th line therapy or beyond) and geographical region (Asia vs. rest of the world) as follows: (Source: Bayer submission, Section 4.3, p58):

Randomisation was performed via an interactive voice response system (IVRS). Investigators received the randomisation number for each participant through the IVRS and study drug supply was also managed via IVRS. Computer-generated randomisation lists were prepared by Bayer (pre-allocated block design, block size 12).

With regard to stratification, overstratification can lead to loss of information, but unstratified analyses are not appropriate when there is heterogeneity between strata. Given the variables used for stratification are considered prognostic indicators, this suggests that the stratified analyses are appropriate

Study duration

Patients continued masked study treatment until disease progression, unacceptable toxicity or withdrawal of patient from the study.

Participants receiving placebo were given the option to cross-over to regorafenib if they experienced disease progression. For participants on regorafenib, open-label regorafenib was offered upon progression, if this was considered clinically beneficial.

Blinding

Treatment allocation was masked for patients, investigators and the study sponsor. This was achieved with the appearance of regorafenib and placebo being identical and unique pack numbers pre-printed onto bottles. Central radiology reviewers were blinded for PFS assessment.

Inclusion/exclusion

Table 7 gives a summary of the inclusion/exclusion criteria for the GRID trial. Those listed are in keeping with the NICE Scope.[1, 11]

Table 7. Eligibility criteria

Key inclusion criteria	Key exclusion criteria
• At least 18 years of age • Histologically confirmed metastatic and/or unresectable GIST in people who have experienced disease progression or intolerance to imatinib, as well as disease progression while on sunitinib. • At least one measurable lesion with CT or MRI (according to RECIST, version 1.1). A lesion in a previously irradiated area was eligible as long as there was objective evidence of progression of the lesion prior to study enrolment. • An ECOG PS score of 0-1 at study entry • Adequate haematological, hepatic, cardiac, and renal function. • Resolution of all toxic effects of previous therapy to grade 1 or lower (excluding alopecia, anaemia, and hypothyroidism).	• Prior treatment with regorafenib, or any VEGFR inhibitor except sunitinib. • Use of any approved tyrosine kinase inhibitors or investigational agents within 1 week or a minimum 5 half-lives of the agent, whichever is shorter, prior to receiving study drug. • Previous or concurrent cancer that is distinct in primary site or histology from GIST within 5 years prior to randomisation EXCEPT for curatively treated cervical cancer in situ, non-melanoma skin cancer, and superficial bladder tumours • Congestive heart failure ≥ New York Heart Association (NYHA) class 2. • Unstable angina, new-onset angina, myocardial infarction less than 6 months before start of study drug. • Cardiac arrhythmias requiring anti-arrhythmic therapy (beta blockers or digoxin are permitted). • Uncontrolled hypertension • Pheochromocytoma. • Arterial or venous thrombotic or embolic events such as cerebrovascular accident (including transient ischaemic attacks), deep vein thrombosis, or pulmonary embolism within the 6 months before start of study drug. • Ongoing infection > grade 2 National Cancer Institute-Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 4.0. • Symptomatic metastatic brain or meningeal tumours unless the patient is > 6 months from definitive therapy, has a negative imaging study within 4 weeks of study entry, and is clinically stable with respect to the tumour at the time of study entry.

Key: ECOG PS, Eastern Cooperative Oncology Group performance status; RECIST, Response Evaluation Criteria in Solid Tumors
Source: Bayer submission, Section 4.3, p59

Location

The multi-centre GRID study was conducted in 57 sites across 17 countries including Asia, China, Japan, Singapore, South Korea, Austria, Belgium, Canada, Netherlands, Poland,

Spain, United Kingdom, United States Finland, France, Germany, Israel and Italy. The proportion of patients based in Europe was 58%.

Study endpoint

The study endpoints and definitions are presented in Table 8

Table 8. Study endpoints

End point	Timing of assessment	Definition
	Primary end point	
Progression free survival (PFS)	At baseline, then every 4 weeks for the first 3 months, every 6 weeks for the months 4 to 6, and every 8 weeks thereafter until the end of study drug administration.	The date of randomisation to the date of first observed radiological progression according to blinded central radiology review, or death due to any cause, if death occurred before progression. The actual date of radiological assessment was used as the date of progression. Patients without tumour progression or death at the time of analysis were censored at their last date of radiological tumour assessment.
	Secondary end points	
Overall survival (OS)	Every 3 months	The date of randomisation until the date of death due to any cause. If a patient was alive at the date of database cut-off, they were censored at this point. All patients were followed for survival until death was documented, except for those who specifically withdrew consent to follow-up.
Time to progression (TTP)	As for PFS	The date of randomisation until the date of radiological progression. Patients without tumour progression at the time of analysis were censored at their last date of radiological tumour assessment. The date of progression was the date of first observation of progression.
Tumour response rate (ORR)	As for PFS	The proportion of patients with the best overall tumour response of partial response (PR) or complete response (CR) according to RECIST version 1.1 criteria that is achieved during treatment or within 30 days after termination of study medication.
Disease control rate (DCR)	As for PFS	The rate of complete response or partial response plus stable disease lasting for at least 12 weeks.

End point	Timing of assessment	Definition
Duration of response (DOR)	As for PFS	The number of days from the date of first documented objective response of PR or CR, whichever is noted earlier, to first disease progression or death before progression. Patients without progression or death before progression at the time of analysis were censored at the date of their last tumour assessment.
Safety	Days 1 and 15 of each treatment cycle for the first six cycles. Cardiac function was assessed at screening, day 1 of the first two treatment cycles (and subsequent cycles at the discretion of the investigator), and at treatment end.	Investigators rated severity of adverse events according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0) [NCI CTCAE V4.0].
Exploratory endpoints		
Health-related quality of life (HRQoL)	At baseline (Day 1 of Cycle 1), on day 1 of cycles 2-4, and day 1 of every other cycle thereafter and within 14 days at the end of treatment.	Health-related quality of life questionnaires (EORTC QLQ-C30 and EuroQoL EQ-5D) were routinely completed by patients.
Pharmacokinetics	Day 15 of cycles 1 and 2	Only performed in patients from selected sites.
Biomarker evaluation	At screening, on day 1 and day 15 of cycle 1, day 15 of subsequent cycles, and at the end of treatment)	Including tumour genotype for mutational status of target oncogene.
Secondary PFS during open label treatment	Only investigator assessments were made during the open-label period.	The time from first progression until second progression or death, whatever came first, during or after open-label treatment with regorafenib per investigator assessment

Key: CR, Complete response; ORR, objective response rate; OS, overall survival; PFS, progression free survival; PR, partial response
Source: Bayer submission, Section 4.3, Table 15, pp 68-70.

Tumour response was based on Response Evaluation Criteria In Solid Tumours (RECIST) (v1.1), with the following modifications: *no lymph nodes and no bone lesions were chosen as target lesions, and PET scan was not considered acceptable for radiological evaluation*. (Source: Bayer submission, Section 4.3, p66)

Furthermore, progression was defined as a growing new tumour nodule within a pre-existing tumour mass expanding on at least two sequential images or must be at least 2 cm in size and a new active lesion.

In order to minimise bias, PFS was assessed by blinded central radiology reviewers. Each image was reviewed by two readers, with adjudication by another radiology reviewer with discordant results.

These endpoints agree with the publication [11] and the protocol for the trial. The ERG considers them appropriate for a study investigating GIST.

4.2.1.3 Statistical analysis

The company state that their primary hypothesis is as follows (Source: Bayer submission, Section 4.4.3, p76):

> *The null hypothesis that both treatment arms have the same PFS distribution was tested against the alternative hypothesis that the distribution of PFS times in the regorafenib arm is different from the control arm according to the Lehmann alternative.*

This statement is in keeping with the study objective of the trial.

4.2.1.3.1 Analysis population

The different populations reported within Bayer's submission for their analyses, along with their definitions are presented in Table 9.

Table 9. Analysis population

Analysis Population	Definition
Intent-to-treat population (ITT)	The full study data set from the GRID study containing data on randomised patients (n= 133 for regorafenib; n=66 for placebo). The ITT population was used for the analysis of the primary efficacy analysis. Subjects in the ITT population were analysed as randomised.
HRQoL evaluable population	Full analysis set patients with evaluable patient reported outcome assessments at baseline and at least one post-baseline assessment. EORTC QLQ-C30 global health status was completed by 183 (92%) patients at baseline, 167 (84%) patients at cycle 2, and 126 (63%) patients at cycle 3.
Safety analysis set	All randomised patients who received at least one dose of study medication (n=132 for regorafenib; n=66 for placebo)[a]
Patient Reported Outcome analysis set (PROAS)	All full analysis set patients with evaluable PRO assessments at baseline and at least one post-baseline assessment (n=123 for regorafenib; n= 62 for placebo).

Notes: a, One patient in the regorafenib group was not treated with study drug
Source: Bayer submission, Table 17, p75

The ITT and safety populations are defined appropriately.

4.2.1.3.2 Determination of sample size

Bayer report that sample size was based on assuming a 100% improvement in PFS for regorafenib, with 199 patients randomised (2:1 regorafenib to placebo), a one-sided alpha of

0.01 and a power of 0.94. It should be noted that the one-tailed test provides more power to detect an effect. However, this test is appropriate since regorafenib is unlikely to be less effective than placebo

As such, the number of events required for final analysis were 144 events, which corresponds to 81 events within the regorafenib group of 133 patients (61%) and 63 events in the placebo group of 66 patients (95%).

Bayer also include the following assumptions of (i) exponential distribution of the PFS event times, (ii) median time of PFS in the control group of [redacted] and [redacted] drop-out rate of patients evaluable for PFS. (Source Bayer submission, Section 4.4, p77)

Missing data

The methods used for handling missing data were as follows: (Source: Bayer submission, Section 4.4, pp 77-78)

> *Missing or not evaluable tumour assessments […] were not used in the calculation of derived efficacy variables unless a new lesion occurred, or the lesions that were evaluated already showed progressive disease (PD). No imputation was performed for missing lesion assessment and tumour response. For example, if a patient missed a scan visit and PD was documented at the next available scan visit, the actual visit date of the first documented PD was used.[…]*

The ERG considers this approach acceptable

4.2.1.3.3 Primary, secondary and tertiary outcomes

Primary outcome – progression-free survival

A stratified log rank test (by prior therapies and geographical region) with a one-sided alpha of 0.01 was used to compare PFS of regorafebin vs. placebo.

Median times to PFS were estimated using the Kaplan-Meier method. Hazard ratios (HR) and 95% confidence intervals (CI) were derived from a Cox proportional hazard model.

Preplanned subgroups for PFS were: (Bayer submission, Section 4.4, p81)

- Line of treaments: 3rd line, 4th line and beyond
- Geographical region,
- Age: <65 years, ≥65 years
- Sex,
- ECOG performance status 0, 1

- Baseline body mass index (BMI) (kgm^{-2}):<25, 25≤BMI<30, 30≤BMI
- Duration of imatinib treatment (months): <6, ≥6<18, ≥18
- Mutational status: initial KIT Exon 11 mutation, initial KIT Exon 9 mutation

Sensitivity analysis included the number of PFS events originally planned in the protocol (no. of events=122), unstratified PFS analyses and PFS analysis according to the assessment of local investigators.

Secondary PFS, assessed during open label treatment, was considered a tertiary outcome.

Secondary outcomes – overall survival

The methods used for TTP and OS analysis were as for PFS. The Cochran-Mantel-Haenszel test was employed for ORR and DCR, whereas DOR received a descriptive analysis. The methods used for adjusting for crossover from placebo to open-label regorafenib are described below: (Source: Bayer submission, Section 4.4, p80)

> *A pre-planned interim analysis of overall survival was done at the time of the final PFS analysis… An updated analysis of OS, was performed as of the cut-off date of 08 June 2015, when approximately 160 deaths had occurred. For the updated analysis of OS, a secondary analysis was performed which applies the Rank Preserving Structural Failure Time (RPSFT) method and the Iterative Parameter Estimate (IPE) method to correct for the effect of cross-over of patients from the placebo treatment to regorafenib treatment on the OS endpoint.*

These methods of adjustment are discussed in more detail in Section 5.3.6.2, p79. However, in our opinion, both the IPE and RPSFT are reasonable candidate adjustment methods.

Tertiary outcomes

Data on HRQoL were obtained via the EORTC QLQ-C30 and EuroQol EQ 5D assessment tools and analysed as described by Bayer: (Source: Bayer submission, Section 4.4, p80)

> *…using an analysis of covariance (ANCOVA) model, comparing the time-adjusted AUCs between the two treatment groups with covariates for baseline HRQoL score and stratification factors. Least-squares mean estimates, standard errors, and 95% confidence intervals (CI) were estimated for each treatment group and for the treatment group difference.*

Exploration of covariates was performed using linear mixed effects models and sensitivity analysis assessed via various imputation methods for missing data.

Descriptive analysis was performed on safety parameters and exploratory endpoints.

Subgroup analysis for OS was as for PFS, with the exception of mutational status. This analysis was also adjusted for crossover using the RPSFT and IPE method (see 5.3.6.2, p79).

Overall, the ERG agrees the statistical analysis were appropriate.

4.2.2 Results

4.2.2.1 Population distribution

Of 199 people recruited, 133 were randomised to receive regorafenib+BSC and 66 to placebo+BSC.

The number of participants evaluable for each of the different populations (ITT, safety and patient reported outcomes), are presented in Table 10.

Table 10. Population distribution for analysis

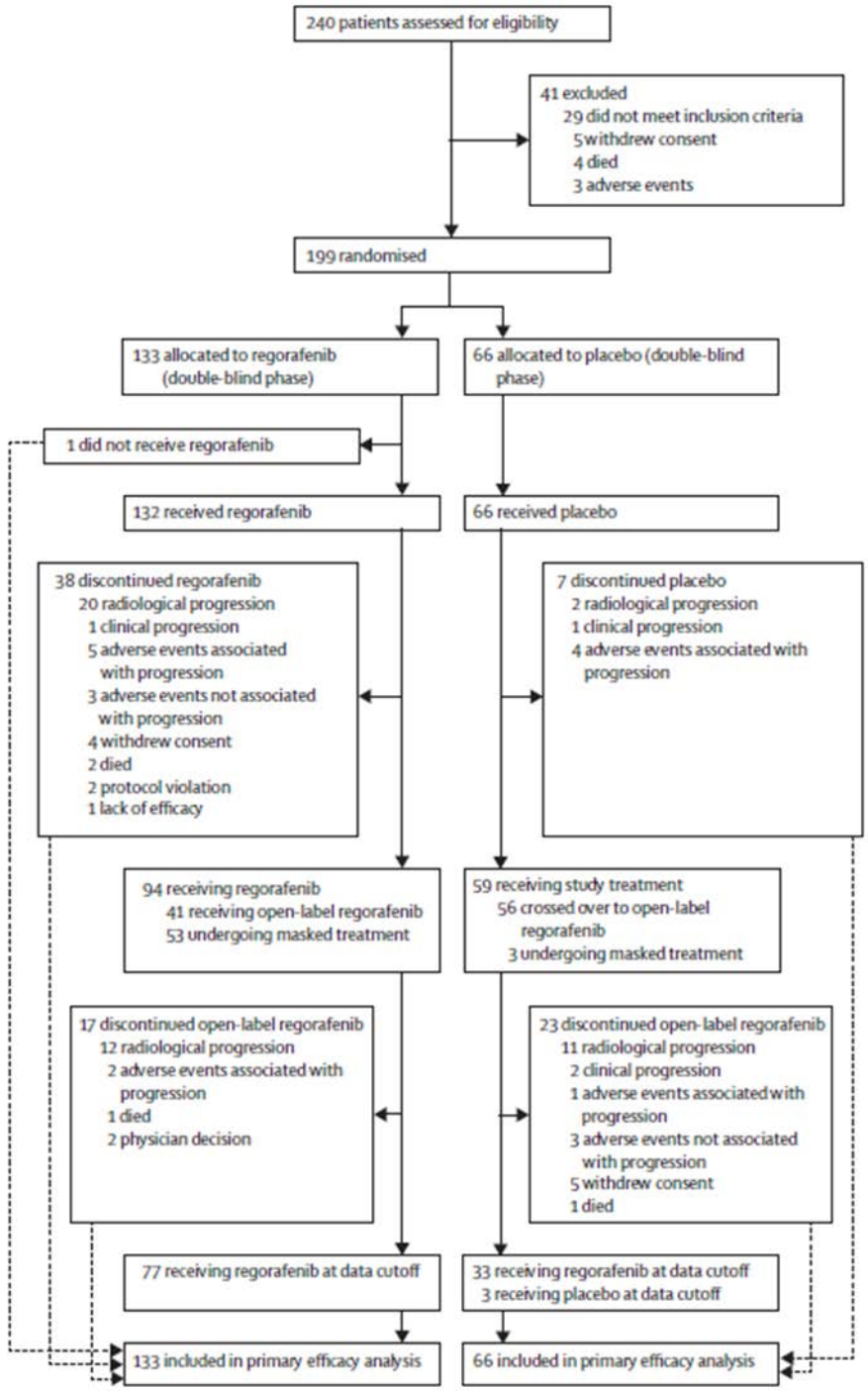
Analysis population	Regorafenib+BSC (n=133)	Placebo+BSC (n=66)
ITT	133 (100%)	66 (100%)
Safety[a]	132 (99.2%)	66 (100%)
Patient Reported Outcomes	123 (92.5%)	62 (93.9%)

Key: ITT, intent-to-treat
Notes: a, One patient in the regorafenib group was not treated with study drug
Source: Bayer submission, Section 4.4, p. 75

4.2.2.2 Participant flow

The participant flow is displayed in Figure 4.

Figure 4. CONSORT diagram for GRID study



Source: Bayer submission, section 4.5, Figure 4, p83

Participants assigned to the regorafenib arm were offered open label regorafenib on progression, if considered appropriate (n=41). For participants receiving placebo, 56 crossed over to regorafenib on experiencing disease progression (follow-up every 6 weeks).

[redacted]

[redacted]The most common reason for termination of study treatment was radiologically confirmed disease progression.

However, the overall treatment duration for the double-blind period for those on regorafenib was a median of 22.9 weeks and a mean of 20.2 weeks. For placebo, the median was only 7.0 weeks and mean 9.1 weeks, hence the difference in patient withdrawal between arms.

4.2.2.3 Baseline characteristics and demographics

Baseline characteristics of the ITT population are summarised in Table 57 (Appendix 1). The demographic characteristics are generally well balanced between those randomised to the regorafenib and placebo groups.

The median age was 60 (range 51-67) and 61 (range 48-66) for regorafenib and placebo, respectively. The proportion of men to women in both groups was 64%:36%.

There was a slight imbalance where 67% of participants receiving regorafinib and 83% receiving placebo had >18 months of previous imatinib therapy.

4.2.2.4 Clinical effectiveness results

The results in the company submission are as of 26 Jan 2012, however, OS was analysed as of 8 June 2015, when approximately 160 deaths had occurred. Following a request to the company, we received updated analyses for OS in 2017, however, at the time no CSR was available to confirm results.

4.2.2.4.1 Primary efficacy analysis – progression-free survival

Progression-free survival is presented in Table 11 and Figure 5 for blinded review:

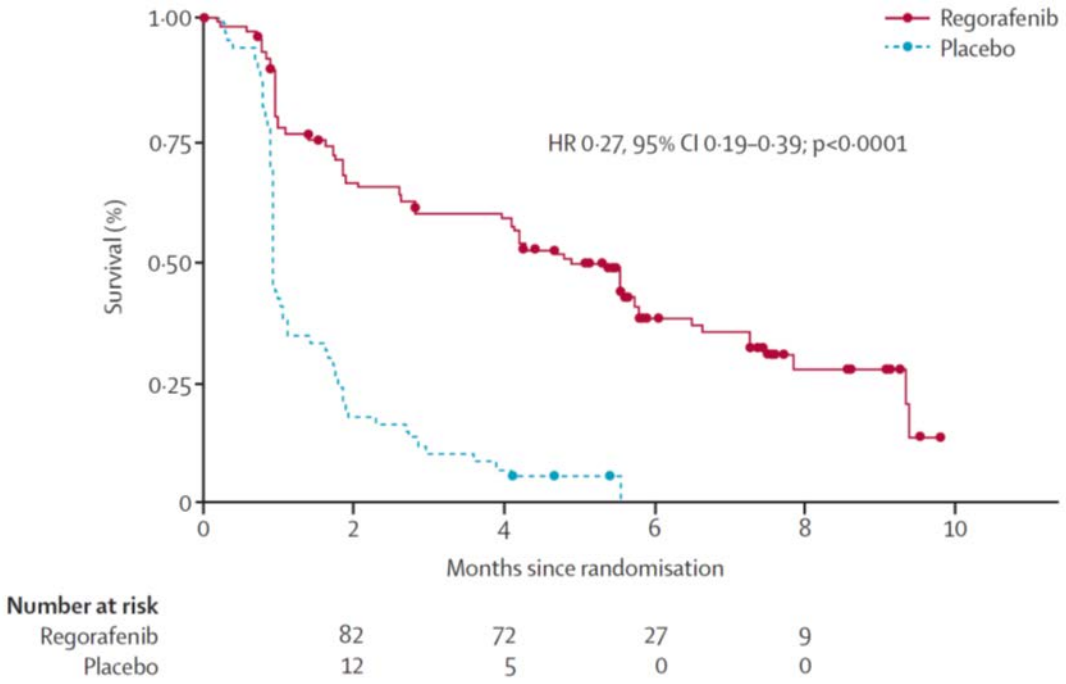
Table 11. Summary of progression-free survival analysis for ITT population

	Regorafenib+BSC (n=133)	Placebo+BSC (n=66)
Blinded review, median PFS, months (IQR)	4.8 (1.4-9.2)	0.9 (0.9-1.8)
Investigator assessment, median PFS, months (IQR)	7.4 (2.7-not calculable)	1.7 (0.9-2.7)
Median PFS, months (IQR		
Blinded review, hazard ratio (95% CI; p value)	0.27 (0.19-0.39; p<0.000001)	
Investigator assessment, hazard ratio (95% CI; p value)	0.22 (0.14-0.35; p<0.0001)	
3 month PFS rate, % (95% CI)	60 (51-68)	11 (3-18)

6 month PFS rate, % (95% CI)	38 (29-48)	0 (0-0)
After 144 events, as specified in the protocol, n (%)	81 (60.9)	63 (95.5)

Source: Bayer submission, Section 4.7, p90

Figure 5. KM estimates of the PFS rate (144 events) during the GRID trial, (central assessment, ITT)



Source: Bayer submission, Section 4.7, p95

The study met the protocol-defined primary endpoint of a one-sided alpha of 0.01. Overall, the results indicate a median PFS for blinded review which is higher in the regorafenib arm than placebo (4.8 months [95% CI: 1.4, 9.2] versus 0.9 months [95% CI: 0.9, 1.8], respectively; HR = 0.27; p <0.000001).

The company state that the sensitivity analyses also showed a statistically significant difference and were consistent with the primary analyses. As with the blinded independent review, the investigator's assessment produced a statistically significant result for PFS in favour of regorafenib. However, the ERG have been unable to locate and verify these results.

4.2.2.4.2 Secondary efficacy analysis

Overall survival

Analysis for OS is displayed in Table 12 and Figure 6 which is unadjusted for crossover (database cut-off 08 June 2015).

Table 12. Summary of overall survival analysis with stratification for ITT population unadjusted for crossover

	Regorafenib+BSC (n=133)	Placebo+BSC (n=66)	Regorafenib+BSC (n=133)	Placebo+BSC (n=66)
	2015 cut-off		2017 cut-off	
Median OS, months	17.4	17.4		
Blinded assessment hazard ratio (95% CI)	0.909 (0.653-1.265)		[redacted]	
Investigator assessment, hazard ratio (95% CI; p value)	0.22 (0.14-0.35; p<0.0001)			
No. events at data cut off 08 June 2015, n (%)	109 (82.0)	53 (80.3)		

Source: Bayer submission, Section 4.7, p91

Figure 6. KM estimates of OS during the GRID trial, (central assessment, ITT; data cut-off June 2015)

Source: Bayer submission, Section 4.7, p96

For the unadjusted analysis, regorafenib shows no benefit for overall survival. However, this includes 56 participants from the placebo arm, who following progression, were allowed to cross over to open-label regorafenib. Therefore, adjustments were performed by the company as shown in Table 13 and Figure 7 to Figure 9. Table 13 and Figure 9 also contain data for the 2017 cut-off.

Table 13. Summary of overall survival analyses with corrected cross-over analyses with stratification

	Data cut-off 2012		Data cut-off 2015		Data cut-off 2017	
	Regorafenib (n=133)	**Placebo (n=66)**	**Regorafenib (n=133)**	**Placebo (n=66)**	**Regorafenib (n=133)**	**Placebo (n=66)**
Median OS, months	NA	NA	17.4	11.9[a], 11.1[b]		
Blinded assessment hazard ratio corrected RPSFT (95% CI)[a]	0.537 (0.286-1.007) p-value 0.024725		0.616 (0.435-0.871; p-value 0.002862)		[redacted] p-value 0.000011[e]	
Blinded assessment hazard ratio corrected IPE (95% CI)[b]	0.565 (0.302-1.055) p-value 0.034931		0.586 (0.417-0.824) p-value 0.000949		[redacted] p-value 0.0000022)[e]	
Number of patients with event, n (%)	29 (21.8%)	17 (25.8%)	[redacted]	[redacted]		
Number of patients censored, n (%)	104 (78.2%)	49 (74.2%)	[redacted]	[redacted]		

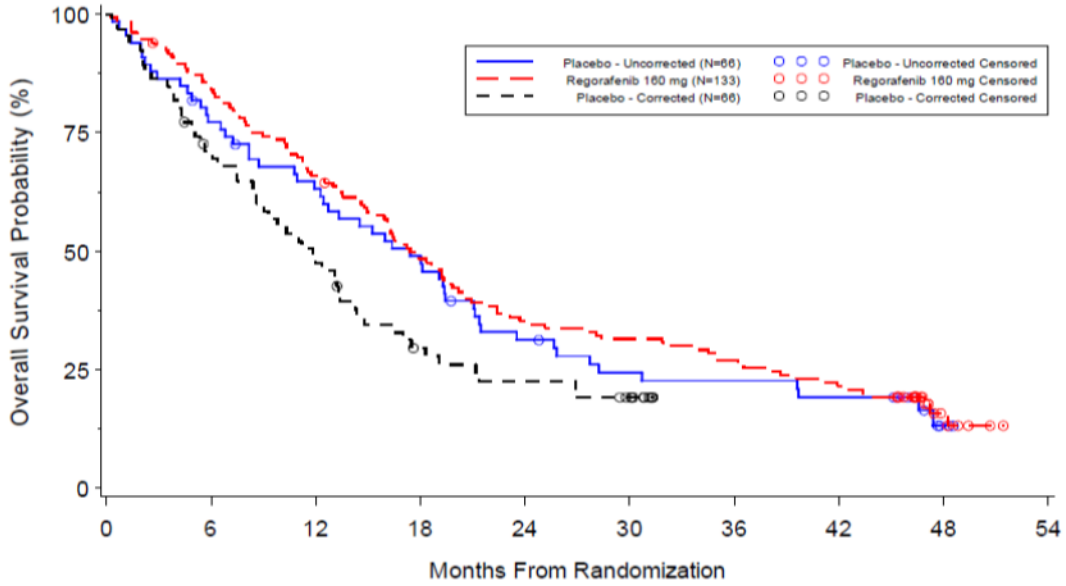
Key: BSC, best supportive care; CI, confidence interval; IPE, iterative parameter estimation; ITT, intention-to-treat; NA, Value cannot be estimated due to censored data; RPSFT, rank preserving structural failure time.

Notes: a, Corrected for the effect of cross-over from the placebo to the regorafenib arm on the OS endpoint by RPSFT method; b, Corrected for the effect of cross-over from the placebo to the regorafenib arm on the OS endpoint by IPE method; c, Using the RPSFT cross-over correction method, the number (%) of patients with an event in the placebo group is 51 (77.3%); d, Using the RPSFT cross-over correction method, the number (%) of patients censored in the placebo group is 15 (22.7%); e,

taken from additional data for stratified analysis supplied by Bayer in response to clarification questions

Source: Bayer submission, section 4.7, p92.

Figure 7. Overall Survival, cross-over correction by RPSFT method (ITT; data cut-off 08 June 2015)

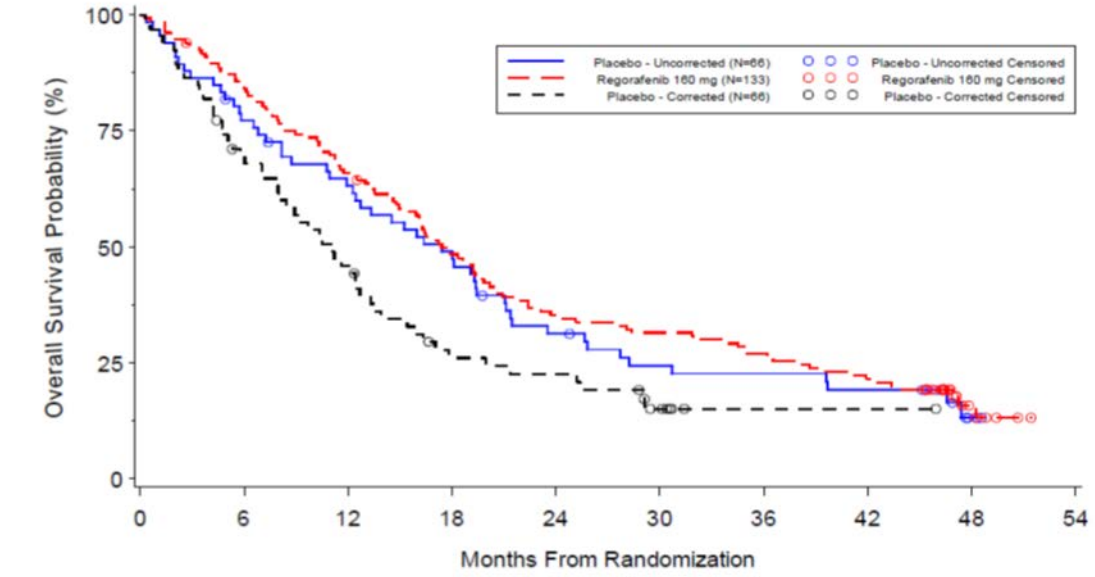


Patients at Risk

Placebo – Uncorrected	50	40	31	19	14	13	11	2
Regorafenib 160 mg	111	87	63	46	41	35	28	6
Placebo – Corrected	45	31	17	13	9	0	0	0

Source: Bayer submission, Section 4.7, p96

Figure 8. Overall Survival, cross-over correction by IPE method (ITT; data cut-off 08 June 2015)



Patients at Risk

Placebo – Uncorrected	50	40	31	19	14	13	11	2
Regorafenib 160 mg	111	87	63	46	41	35	28	6
Placebo – Corrected	44	29	15	13	6	1	1	0

Source: Bayer submission, Section 4.7, p97

Figure 9. Overall Survival, cross-over correction by IPE method (ITT; comparison of 2015 and 2017 data)

Source: Bayer submission, response to clarification

Following adjustment for crossover, both the 2015 and 2017 data, indicate a statistically significant difference in overall survival favouring regorafenib (median OS 17.4 months) over placebo (median OS 11.9 months using RPSFT method or 11.1 months using IPE method).

Time to progression

For the cut-off date of 26th January 2012, 57.1% of participants in the regorafenib group experienced disease progression and 93.9% in the placebo group. Median TTP was reported as 165 days in the regorafenib group and 28 days in the placebo group (HR 0.248, [95% CI: 0.170-0.364, p<0.000001]). Therefore, there is a statistically significant difference between arms, in favour of regorafenib for TTP.

Objective Response Rate, Disease Control Rate and Duration of Response

For ORR, although numerically in favour of regorafenib, there was no statistically significant difference between the two arms: 4.5% with regorafenib (PR n= 6/133) vs. 1.5% with placebo (PR n=1/66) and there were no cases reported of complete response.

[redacted]

The disease control rate (DCR) reflects the percentage of patients with metastatic cancer who have achieved complete response, partial response and stable disease, as opposed to ORR which only includes CR or PR. Stable disease was reported by the company to be 71.4% (95/133 patients) in the regorafenib arm as compared to 33.3% (22/66 patients) in the placebo arm. Therefore, DCR for the regorafenib group was 52.6% (n=70/133) compared with 9.1% (n=6/66) in the placebo group (95% CI: –54.72, –32.49; p<0.0001). Bayer suggest this outcome indicates the clinically meaningful tumour control of regorafenib as a third-line treatment in patients with advanced GIST.

With regard to median duration of response, only one patient in the placebo group reported PR, which was 30 days, whereas the median duration of response for patients receiving regorafenib was 99 days.

Maximum percent reduction in the size of target lesions

[redacted]

4.2.2.4.3 Exploratory endpoints

Secondary PFS (SPFS)

Bayer investigated secondary PFS for participants who crossed over from placebo to regorafenib (n=56; 151 days) and for participants who continued on open label regorafenib,

following progression during the masked period (n=41; 137 days) (Figure 10). Therefore, the company suggest that regorafenib may delay subsequent progression.

Figure 10. KM curves of PFS during treatment with regorafenib by double blind and open label treatment groups

Source: Bayer submission, Section 4.7, p98

Patient reported outcomes

Analysis of HRQoL via the EORTC QLQ-C30 and the EQ5D revealed no statistically significant difference between regorafenib and placebo. Mean changes from baseline were not clinically meaningful (defined as ≤10 points), except for the role function subscale in the regorafenib group.

Mean changes in scores from baseline for EQ-5D index reflected a deterioration in health status for both groups. However, the results for the EQ-VAS appear more variable, with a change from baseline indicating a general reduction in health status for the regorafenib group, but an improvement for the placebo group. However, the company report that analysis of time-adjusted AUC for the EQ-5D index and VAS showed that regorafenib treatment maintained patients' health-related quality of life.

No statistically significant difference in HRQoL was noted in regorafenib-treated patients with dose reduction vs. no dose modification.

Mutational analyses

Mutation data were available for 48% of patients in the GRID study (53% KIT Exon 11; 16% KIT Exon 9; 8% no KIT and no PDGFR mutation).

The company report that both exon 9 mutant and exon 11 mutant subgroups have improved PFS on regorafenib compared to placebo, although this appears to be comparable to the results for the ITT population overall (Table 11):

- KIT Exon 11 (HR of 0.21; 95% CI: 0.10, 0.46)
- KIT Exon 9 (HR of 0.24; 95% CI: 0.07, 0.88)

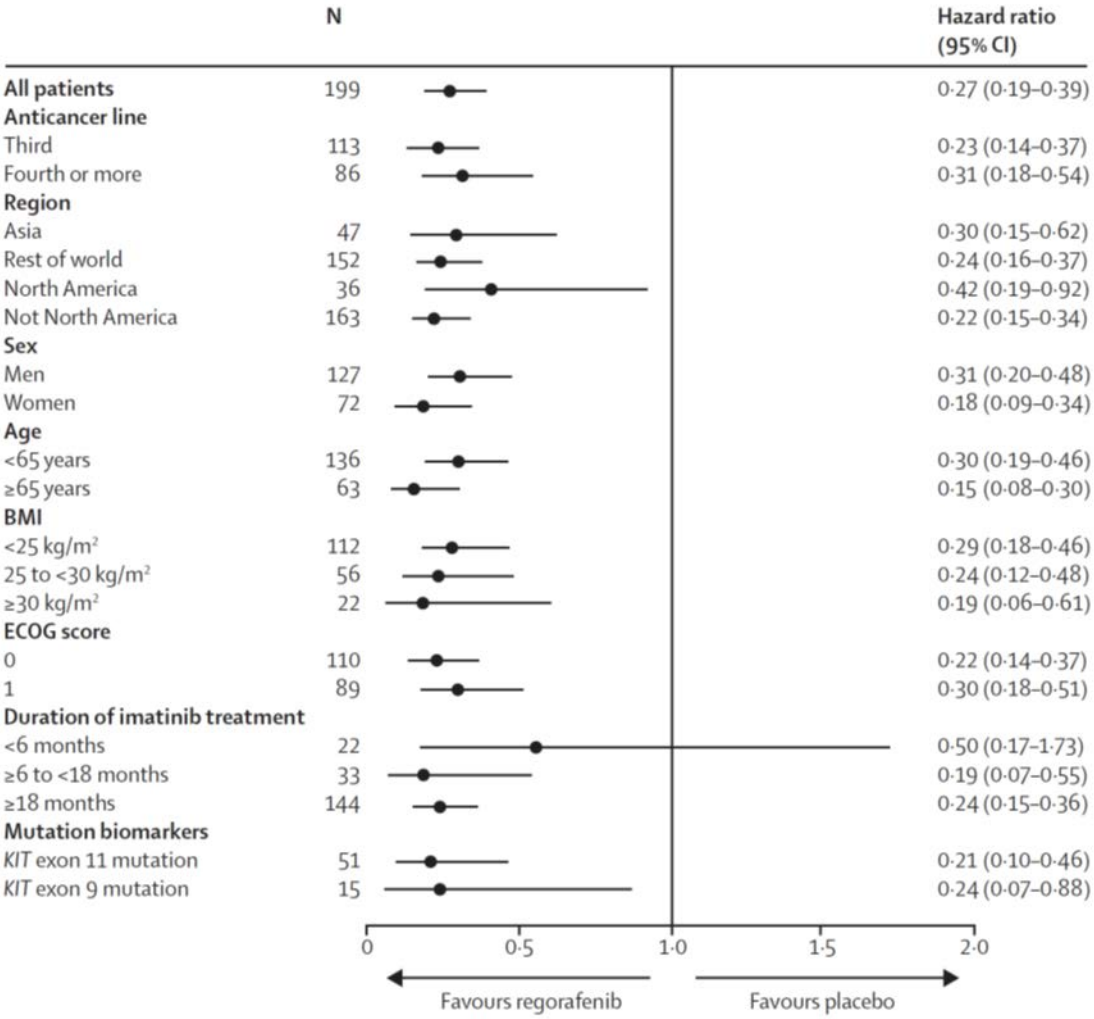
The benefit for other mutations is not reported.

4.2.2.4.4 Subgroup analysis

Progression-free survival

Pre-planned subgroup analysis was performed for PFS as displayed in Figure 11.

Figure 11. Progression-free survival by subgroup



Key: BSC, best supportive care; ECOG, Eastern Cooperative Oncology Group
Source: Bayer submission, Section 4.8, p103

The majority of subgroups in Figure 11 show a statistically significant benefit in PFS for regorafenib. There is little heterogeneity, with similar HRs and generally narrow confidence intervals. The group where this is not the case is for the population who received imatinib for less than 6 months. Bayer suggest this is due to the small sample size of 22.

At the request of some health authorities, Bayer also report:

- there was no correlation between hypertension and length PFS,
- low patient numbers meant conclusions could be drawn on mitotic index and PFS,.
- median PFS times in patients in the regorafenib group who had dose modifications were similar to those in the overall primary analysis .

Overall survival

The subgroup analysis for OS was performed uncorrected and corrected for the effect of crossover using the RPSFT model and the IPE method.

The uncorrected analysis in Figure 12 includes 58 (87.9%) of patients in the placebo + BSC group crossed over to regorafenib treatment. The HRs for most subgroups are close to one, with broad intervals, indicating no statistically significant difference in OS between the two arms. However, as noted by the company, these results should be interpreted with caution due to the low number of events in some subgroups.

Figure 12. OS with regorafenib by double blind and open label treatment groups

Source: Bayer submission, Section 4.8, p106

Figure 13. Overall survival by subgroup, RPSFT correction (data cut-off 08 June 2015)

Source: Bayer submission, Section 4.8, p107

Figure 14. Overall survival by subgroup, IPE correction (data cut-off 08 June 2015)

Source: Bayer submission, Section 4.8, p108

4.2.2.4.5 Adverse events

The GRID study included 198 participants in the safety population, which included 162 in the regorafenib arm and 66 participants in the placebo arm who received at least one dose of regorafenib. The analysis included treatment-emergent adverse events (TEAEs) occurring up to the primary efficacy analysis cut-off date of 26th January 2012.

Secondary analyses included patients who crossed over to regorafenib from placebo (n=132+58) and a subgroup of patients who received regorafenib for over 1 year (n=75).

A summary for all grade adverse events (AEs) is presented in Table 14 which reports the incidences of AEs for > 10 % of people in any treatment arm. The main groups are included, with further detail on individual conditions provided in Appendix 2.

Table 14. Summary of all grade adverse events

	Double-blind treatment (data cut-off 26 January 2012)		Data cut-off 08 June 2015	
	Regorafenib + BSC	Placebo + BSC	Regorafenib-treated at any time during study	Subgroup treated with regorafenib for >1 year
	N=132 n (%)	N=66 n (%)	N=190 n (%)	N=75 n (%)
Any TEAE	[redacted]	[redacted]	[redacted]	[redacted]
Blood and Lymphatics				
Cardiac				
Ear and Labyrinth				
Endocrine				
Gastrointestinal				
General and Administrative Site Conditions				
Hepatobiliary disorders				
Infection and Infestations				
Injury, poisoning and procedural complications				
Investigations				
Metabolism and Nutrition				
Musculoskeletal and Connective Tissue				
Nervous System				
Psychiatric disorders				
Insomnia				
Renal and urinary				
Reproductive system and breast disorders				
Respiratory, Thoracic and Mediastinal				
Skin and subcutaneous tissue				
Vascular				

Key: BSC; Best supportive care; TEAE; Treatment-emergent adverse event; NCI CTCAE, National Cancer Institute Common Terminology Criteria for Adverse Events version 4.0; A patient may have experienced more than one TEAE.

Source: Bayer submission, Section 4.12, p122

In the double-blind study phase, drug-related adverse events were reported by Bayer to be 130 (98%) patients in the regorafenib group and 45 (68%) patients in the placebo group. The most common drug-related AEs were PPES (HFSR), hypertension, fatigue, diarrhoea, and oral mucositis.

Adverse Events of Special Interest

Bayer comment that the toxicity profile of regorafenib is typical for molecule of its type and that as such, events including hypertension, skin (hand-foot syndrome, rash) and gastrointestinal toxicities (diarrhoea, mucositis) are not unexpected. Table 15 displays specific adverse events of interest with regard to regorafenib treatment.

Table 15. Adverse events of special interest during the GRID trial

	Double-blind treatment (data cut-off 26 January 2012)		Data cut-off 08 June 2015	
	Regorafenib + BSC	Placebo + BSC	Regorafenib-treated at any time during study	Subgroup treated with regorafenib for >1 year
	N=132 n (%)	N=66 n (%)	N=190 n (%)	N=75 n (%)
Cardiac	[redacted]	[redacted]	[redacted]	[redacted]
Hepatobiliary disorders			[redacted]	[redacted]
Alanine aminotransferase increased (ALT)			[redacted]	[redacted]
Aspartate aminotransferase increased (AST)			[redacted]	[redacted]
Blood bilirubin increased			[redacted]	[redacted]
Proteinuria			[redacted]	[redacted]
Palmar-Plantar Erythrodysaesthesia Syndrome	[redacted]	[redacted]	[redacted]	[redacted]
Hypertension	[redacted]	[redacted]	[redacted]	[redacted]
Bleeding events			[redacted]	

Source: Bayer submission, Section 4.12, p123

Clearly, the most common adverse event was [redacted]

Serious adverse events

During the double-blind phase, 38 (29%) serious adverse events were reported in the regorafenib group and 14 (21%) in the placebo group. The company report the most common SAEs to be [redacted]

Laboratory parameters

The majority of laboratory abnormalities were grade 1–2, the most common being anaemia for regorafenib-treated patients (144 [77.0%] patients, hyperglycaemia (93.0%), AST increased (67.6%), hypertriglyceridaemia (63.3%), hypoalbuminaemia (62.0%), hypophosphatemia (61.2%), alkaline phosphatase increased (57.4%), and ALT increased (48.9%).

Adverse events leading to withdrawal

Bayer report the following withdrawals due to adverse events during the double-blind phase of the study: (Source, Bayer submission, Section 4.12, p126):

9 patients discontinued due to an AE in the regorafenib-treated group (6.1%) versus 5 patients in the placebo group (7.6%).
[redacted]

Deaths

Bayer report 5 deaths considered to be due to regorafenib (cardiac arrest, acute hepatic failure, acute kidney injury, colonic perforation, and thromboembolic event). Of the patients treated with regorafenib,
[redacted]

Long term safety

With regard to long term safety, Bayer report that: (Source: Bayer submission, section 4.12, p126)

The safety profile of patients on long-term regorafenib treatment (> 1 year; n=75) was comparable with the safety profile of the overall patient population...
[redacted]*….*
[redacted]
For hypothyroidism, the decreasing but not completely absent event rates over time emphasise the label-defined regular monitoring recommendation of thyroid function during regorafenib treatment.

Of note, long-term responders showed around a [redacted] *incidence rate in drug-related grade 3 events as compared to the overall patient population, mainly due to respective increases in grade 3 PPES (HFSR) and hypertension rates. …Treatment discontinuation rates due to regorafenib-related events were comparable between long-term responders and overall patient population* [redacted]

4.2.3 Interpretation

Key efficacy findings from the RCT reported in the submission were as follows:

Progression free survival

Median PFS was more favourable for the regorafenib group than in the placebo group (4.8 months vs. 0.9 months; [HR] 0.27, 95% CI 0.19–0.39; p<0.0001).

The treatment effect of regorafenib was generally consistent across pre-specified subgroups. The company also report the effect is maintained following sensitivity analyses for PFS, however, the ERG have been unable to locate these results.

[redacted]

Overall survival

Prior to adjustment for crossover, median OS time was 529 days (17.4 months) in both treatment groups (HR = 0.909).

Following correction, median OS time was longer in the regorafenib group (529 days or 17.4 months) than in the placebo group (338 days or 11.1 months IPE [p = 0.00095]; 361 days or 11.9 months RPSFT [p = 0.00286]). The estimated corrected hazard ratio of regorafenib to placebo using the RPSFT and IPE correction methods were 0.616 and 0.586, respectively.

Secondary endpoints

Median time to progression (TTP) was significantly longer in the regorafenib arm than in the placebo arm (5.4 months [165 days] versus 0.9 months [28 days], HR 0.248, 95% CI 0.170–0.364; p<0.000001).

Although there was a numerical difference in overall response rate, this was not statistically significant (4.5% vs. 1.5% for the regorafenib and pllacebo group, respectively)

The disease control rate (DCR), which also includes stable disease, was significantly higher in the regorafenib group (52.6%) vs. the placebo group (9.1%) (one-sided p<0.000001).

Adverse events

Common adverse events included hand-foot skin reaction (HFSR), hypertension, diarrhoea, mucositis and fatigue. The most serious adverse drug reactions in patients receiving regorafenib were haemorrhage, severe liver injury, and gastrointestinal perforation. However, in general, treatment with regorafenib was not associated with a substantial reduction in patient reported quality of life compared to placebo

[redacted]

4.2.3.1 Strengths and limitations

Strengths

- Large, prospective, randomised, double-blind, placebo-controlled, multicentre trial.

- A majority of the recruited population were representative of the typical UK patient population
- Subgroup and sensitivity analyses indicate robust results

Limitations

- No ‘active’ comparator due to the lack of approved treatment options available to patients with metastatic or unresectable GIST after they have progressed on imatinib and sunitinib.
- Confounding by crossover of 58 (87.9%) patients from the placebo group to regorafenib treatment upon disease progression.Therefore, two correction methods were used.

5 Cost-effectiveness

5.1 History of Bayer's economic evaluation

So far, we have received a total of three versions of Bayer's economic model and cost-effectiveness results.

We received Bayer's economic model and full report on 21st March 2017.

On 25th April 2017, after an earlier request for clarification from us, we received a second version of Bayer's economic model and cost-effectiveness results. This included some updated OS data, as discussed in Section 5.3.6, p74.

On 16th May 2017, in response to another request for clarification from us, we received a third version of Bayer's economic model and cost-effectiveness results. In addition to the updated OS data, this also included some updated data on treatment duration of regorafenib as discussed in Section 5.3.8.1, p102.

5.2 ERG comment on company's review of cost-effectiveness evidence

5.2.1 Searches

The company conducted a systematic literature review (SLR) of economic and cost-effectiveness studies. The company conducted one primary search in a range of databases indexing published research for cost-effectiveness analyses for treating adults with unresectable and/or metastatic GIST, who have failed to respond to both sunitinib and imatininb. The initial search was from database inception to 21 December 2011, and was then updated 3 times: 21 December 2011 – July 2013, 21 July 2013 – 06 May 2016, and 06 – May to 19 December 2016.

The following electronic databases were searched: MEDLINE, MEDLINE (R) In-Process, EMBASE, EconLIT, and NHS EED. In addition, 3 major conferences were searched for relevant research: American Society of Clinical Oncology, European Society for Medical Oncology, and the International Society for Pharmacoeconomics and Outcomes Research.

5.2.2 Inclusion/exclusion criteria

The company developed a set of inclusion and exclusion criteria which were applied to the search results. The titles and abstracts were independently reviewed by two people and any disparity in decisions whether to include/exclude were reviewed by a third party. The inclusion/exclusion criteria presented by the company are shown below in Table 16.

Table 16. Inclusion/exclusion criteria for review of cost-effectiveness evidence

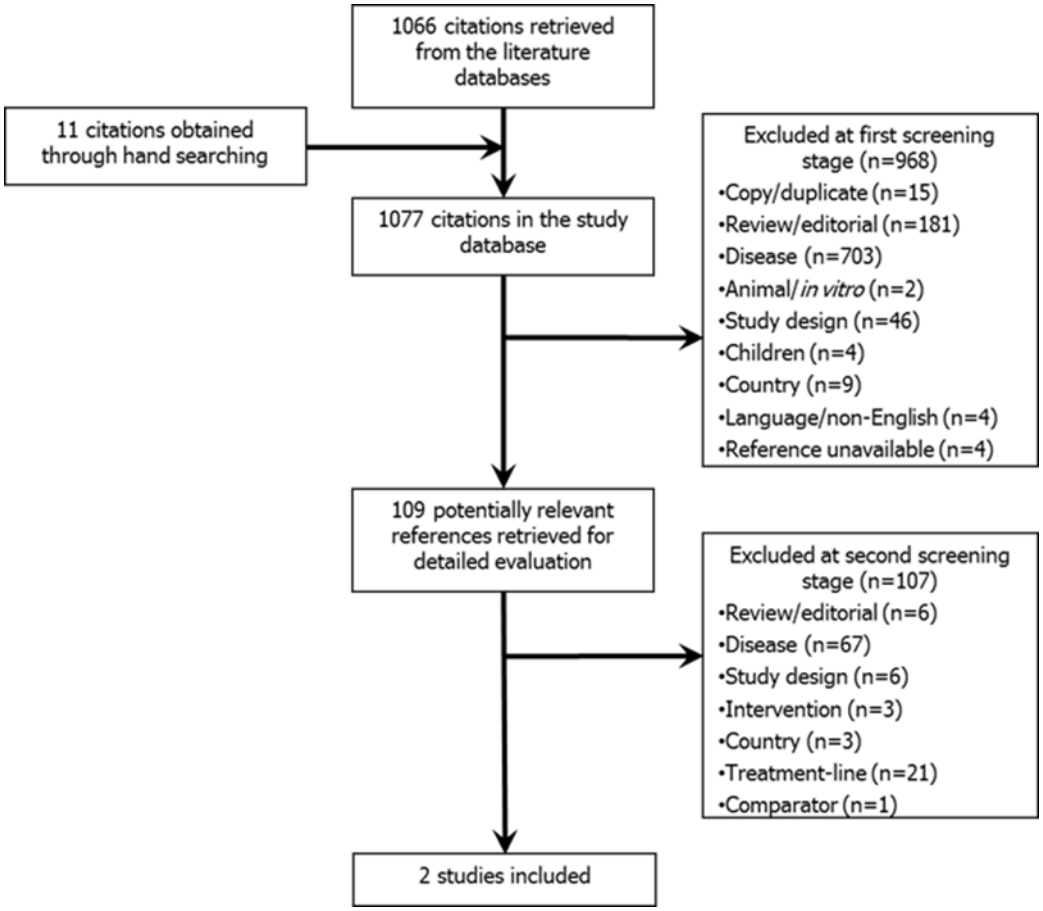
	Inclusion criteria	Exclusion criteria
Study design	Study design appropriate to report the cost of illness and/or resource use for GIST (cost studies analyses, database studies collecting cost or resource use data [including claims databases and hospital records], cross-sectional studies [including surveys] containing cost data, cohort studies containing cost data, longitudinal studies containing cost data, RCT containing piggy-back economic evaluation, cost-effectiveness analyses, cost-utility analyses, cost-benefit analyses, cost-minimisation analyses, budget impact models, cost consequence studies)	Literature and systematic reviews Database studies or epidemiology studies, not collecting cost data RCTs (with no piggy-back economic evaluations) Studies published in non-English language (with/without English abstracts)
Patient population	Studies including adult patients (aged ≥18 years) Studies reporting data in countries of interest (US, Canada, Australia, France, Germany, Italy, Spain, UK, Brazil, Mexico, Japan, China, Korea)	Studies in children or adolescents Studies conducted in animals or in vitro
Disease/ therapy	Studies including patients with metastatic, advanced, and/or unresectable GIST, defined as such using the study author's definition Studies of third-line patients (who have failed two pharmacological therapies). However, as it is was anticipated that studies focused on third-line patients were rare, studies in first- and second-line patients were only excluded at the final stage of the second pass (at the first pass stage there was no exclusion based on therapy line)	Studies that did not include patients with a specific GIST diagnosis (including gastrointestinal leiomyosarcoma that appeared to behave as GIST, soft-tissue sarcoma that appeared to behave as GIST, oesophageal leiomyosarcoma, gastric leiomyoma, gastric leiomyoblastoma, small intestinal leiomyoma and leiomyosarcoma, colonic and rectal leiomyoma and eiomyosarcoma, gastrointestinal autonomic nerve tumour, eiomyoma and leiomyosarcoma of omentum and mesentery, retroperitoneal leiomyosarcoma)
Intervention	Regorafenib	Any other intervention
Comparator	Placebo/BSC	Any other comparator

Key: BSC, best supportive care

5.2.3 Results

Figure 15 shows the PRISMA flow diagram of the included economic studies.

Figure 15. PRISMA flow diagram of economic studies



Two studies were included. Sanz-Granda et al. (2015) is a study which is based in a Spanish healthcare setting, and the company deemed it to not be relevant for England and Wales.[14] Pitcher et al. (2016) is a UK based cost-utility analysis for the relevant patient population in

England.[15] This study utilised a partitioned survival model with 3 states; PFS, PPS, and death. A summary of the included studies is shown in Table 17 .

Table 17. Included studies in cost-effectiveness review

Study	Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
Pitcher, 2016 (UK)[a]	2016	A partitioned survival model was used to model three health states: progression -free, progressed, and dead, over a lifetime horizon.	NA	QALYs using IPE crossover adjustment method: Regorafenib: 1.717 Placebo: 0.969 QALYs using RPSFT crossover adjustment method: Regorafenib: 1.717 Placebo: 1.080	Costs Using IPE crossover adjustment method Regorafenib: £36,258 Placebo: £10,513 Costs using RPSFT crossover adjustment method Regorafenib: £36,258 Placebo: £10,659	**ICERs per QALY gained: For IPE: £34,420 For RPSFT, £40,188**

Key: QALYs, quality-adjusted life years; ICER, incremental cost-effectiveness ratio
Notes: a, Results presented at ISPOR 19th Annual European Congress

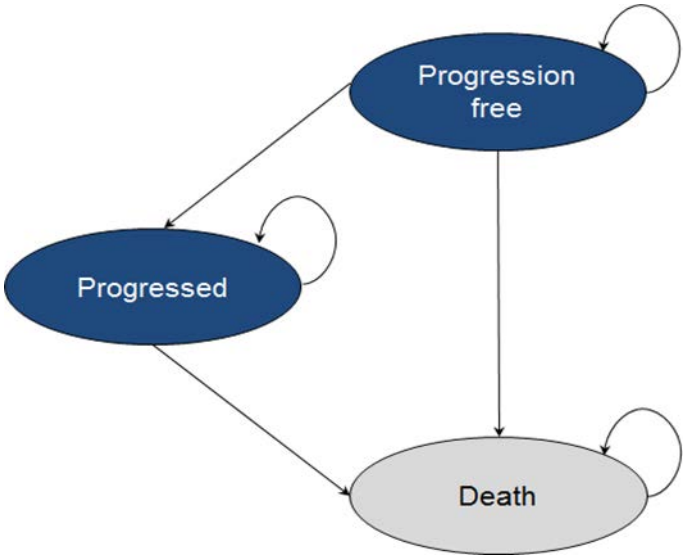
5.3 Summary and critique of company's submitted economic evaluation by the ERG

5.3.1 NICE reference case checklist

5.3.2 Model structure

Bayer submitted a partitioned survival model with three independent health states; PFS, PPS and death (Figure 16). Bayer argues that this structure is commonly used and best reflects the GRID trial as Kaplan-Meier curves for the health states can be used directly.

Figure 16. Bayer's partitioned survival model



Source: Bayer submission, Figure 15, p.147

- Patients start in the PFS state and can remain there, their disease can progress or they can die.
- Patients in the PPS health state can remain there or die.
- Death is the absorbing state.

Patients enter the model upon treatment commencing for either regorafenib or the comparator, BSC. The proportions of patients in each state are calculated as a function of time using parametric extrapolations due to the GRID trial exhibiting significant censoring for both PFS (due to patients dropping out of follow-up) and PPS. The parametric models were fitted to the Kaplan-Meier (K-M) curves from the trial to help inform extrapolation choice. The model uses a 28-day cycle length, corresponding to the proposed regorafenib treatment cycle of 3 weeks on daily treatment followed by 1 week off treatment. A half-cycle correction is applied.

Table 18 (reproduced from Bayer's report) gives a summary of some of Bayer's key modelling assumptions.

Table 18. Model assumptions

Assumption	Reason	Section
Health state assumptions		
Initially all patients begin in the progression free on treatment health state and are assigned progression free disease utility and costs of treatment while on therapy.	This is in line with trial	5.2.2
Patients discontinuing treatment prior to progression are not assigned a cost of active treatment and are assigned progression free utility and other routine costs. Patients can move to the death state based on the OS curve. As there are no cost or outcome implications, the placebo arm does not track patients between on treatment and off treatment states.	This is in line with trial	5.2.2
While in the progressed state, patients are assigned progression state disease utility and costs of disease management. In the progressed state, patients are not assigned costs of regorafenib treatment. Patients can only move from the progressed state to the death state.	Treatment with regorafenib should continue as long as benefit is observed or until unacceptable toxicity occurs	5.2.2
Other assumptions		
Time horizon of 40 years	This should be sufficiently long to capture all the lifetime benefits.	5.2.2
BSC as the only comparator	There are no approved treatments for patients in the given indication for regorafenib.	5.2.4
IPE crossover adjustment	Crossover causes significant bias in the effectiveness estimate if uncorrected. The IPE method provided the least bias for crossover adjustment.	

Log-logistic function used for long term extrapolation of OS	This provided the best statistical fit according to the AIC.	5.3.2
Same utilities used for each treatment arm	No statistically significant treatment effect was found between treatment arms in the utility analyses, therefore the same utilities were applied in both arms.	5.4.1
Resource use based on 2013 physician survey	Physicians were oncologists that had practiced in the area of GIST. The resource use assumptions were then re-evaluated by clinical experts in 2016, and changes to resource use assumptions were explored in scenario analyses.	5.4.1

Source: Bayer submission, Table 61, p.195

5.3.3 Population

The target population is comprised of adults with metastatic and/or unresectable GIST who were previously treated with at least imatinib and sunitinib. Patients enter the model at age 60, the median age from patients in the GRID trial (mean: 58.2 years).

Bayer did not identify any subgroups that would have clinically or economically relevant differences in benefit for regorafenib. We consider this appropriate.

5.3.4 Interventions and comparators

The intervention being investigated is once daily regorafenib at a recommended dosage of 160mg a day in addition to best supportive care (BSC) compared to BSC alone, the "placebo" (Source: Bayer submission, p. 149). Over a 4 week cycle, regorafenib is administered daily for the first 3 weeks, followed by a 1 week break. In the GRID trial, regorafenib could be continued by patients experiencing disease progression based on investigator opinion, and patients on the placebo could also cross over to regorafenib. Despite this, Bayer argue that in accordance with standard practice in England and Wales, regorafenib would only be given to patients whose disease had not progressed in actual practice (Source: Bayer submission, p. 150).

Bayer justify the comparator being solely BSC by referring to physician surveys in 2013 and 2016 in which they found no standard, approved or recommended treatment for patients who had already failed on imatinib and sunitinib. Our clinical expert confirmed that BSC is the sole relevant comparator

5.3.5 Perspective, time horizon and discounting

In the model, the perspective on costs was related to the NHS and Personal Social Services, and direct health effects on patients were considered, in accordance with the NICE reference case.

The time horizon used is 40 years, which Bayer argue is long enough to capture all expected lifetime benefits. In accordance with NICE reference case, benefits and costs are discounted at the standard 3.5 per cent rate. Health effects are measured in quality-adjusted life years (QALYs).

5.3.6 Treatment effectiveness and extrapolation

Treatment effectiveness was estimated using the GRID trial and post-hoc analyses on the data collected.

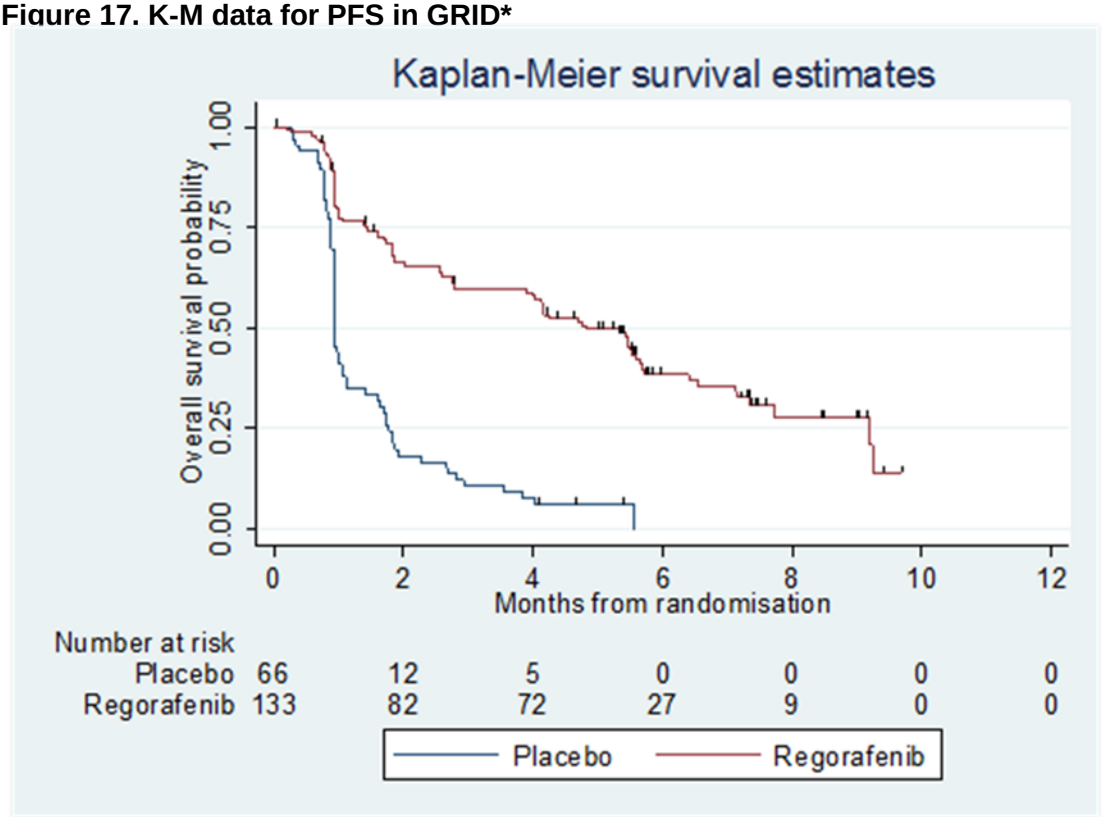
The economic model used the following clinical endpoints:

- Overall survival (OS), the time from entering the model to death from any cause;
- Progression free survival (PFS), the time from entering the model until disease progression (or directly dying);
- Post progression survival, the time from disease progression until death.

In their original report, Bayer presented OS data with a cut-off date of June 2015. In our clarification letter, we ask Bayer whether they could provide us with more mature data, given that the existing data is now about two years out of date, and that a reasonable amount of extrapolation is required. In response, on 25th April 2017, we received OS data from Bayer with cut-off in 2017. Bayer also included an updated version of their economic results.

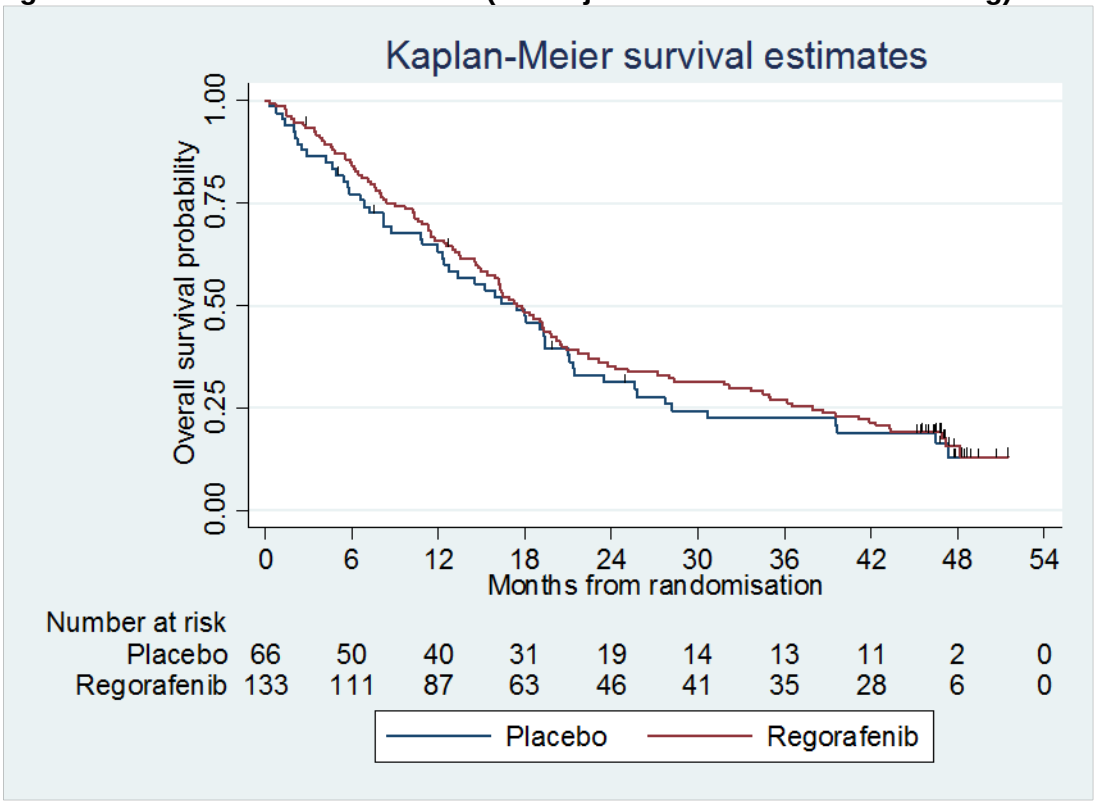
Extrapolation in the model is entirely parametric, as both OS and PFS data from the GRID trial exhibited significant censoring. Figure 17 and Figure 18 below show Kaplan-Meier data for PFS and OS respectively. Bayer have not updated their PFS Kaplan-Meier data, they still use the PFS cut off from back in 26th Jan 2012. This seems curious, because the PFS data is not fully run off. However, given that cost-effectiveness is far less sensitive to PFS than to OS or treatment duration, we pursue this matter no further.

Figure 17. K-M data for PFS in GRID*



Source: Bayer submission, Figure 16, p. 153
*Bayer's Y-axis should read "Progression-free survival".

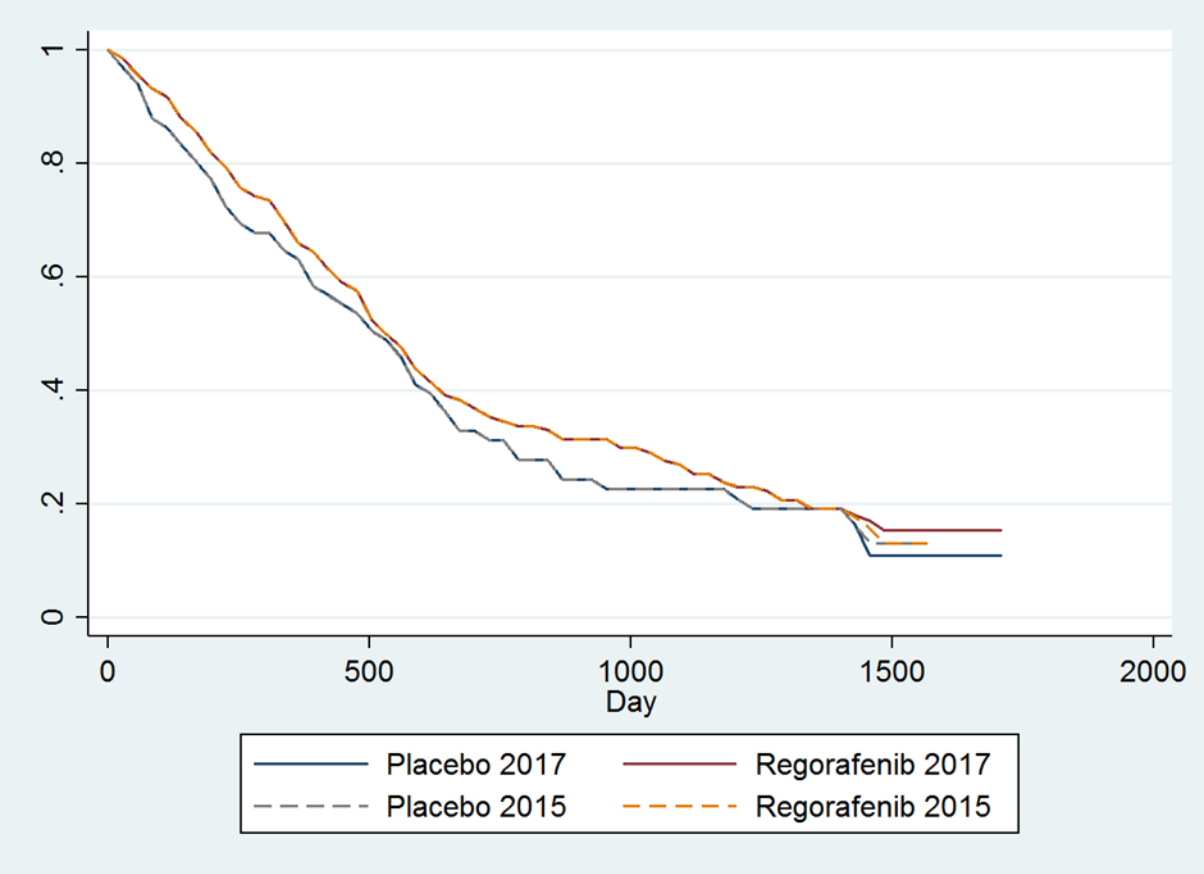
Figure 18. K-M data for OS in GRID (not adjusted for treatment switching)



Source: Bayer submission, Figure 17, p.154

Figure 19 demonstrates that the OS Kaplan Meier curve for the regorafenib arm changes only very slightly using the 2017 data, compared to the 2015 data. The OS is slightly more mature. Similar comments apply to the BSC arm.

Figure 19. Comparison of K-M OS curves not adjusted for treatment switching (2015 vs. 2017 data cut-off)



5.3.6.1 PFS extrapolation

To extrapolate PFS and OS, several parametric models were fitted to the existing GRID trial data:

- Exponential;
- Loglogistic;
- Weibull;
- Lognormal;
- Gompertz.

The parametric models were then assessed for quality of fit to the K-M data visually. Citing the uncertainty of this visual inspection, Bayer also statistically investigated the fits using the Akaike and Bayesian Information Criteria (AIC and BIC). These methods help determine the relative fit of the models by assessing the explanatory power of the model and penalising the number of parameters (to prevent over-fitting), with a lower AIC/BIC value being better. Table 19. AICs and BICs for PFS extrapolation below shows Bayer's AIC and BIC values for

the placebo and intervention for the PFS parametric models, the numbers in bold showing the lowest combined AICs/BICs.

The reason that the AICs/BICs are summed for the two treatments arms is that different parametric models have shapes, which Bayer argue should be avoided (Source: Bayer submission, p.157). Summing the AICs/BICs then gives a single "best" choice for both arms. Bayer therefore chose the lognormal model in the base case and the fit is shown in Figure 20.

We find that the cost-effectiveness of regorafenib is rather insensitive to the choice of distribution function. For example, assuming the shorter-tailed Weibull, Bayer's base ICER assuming the PAS increases only slightly, from £37,900 to £38,800 per QALY.

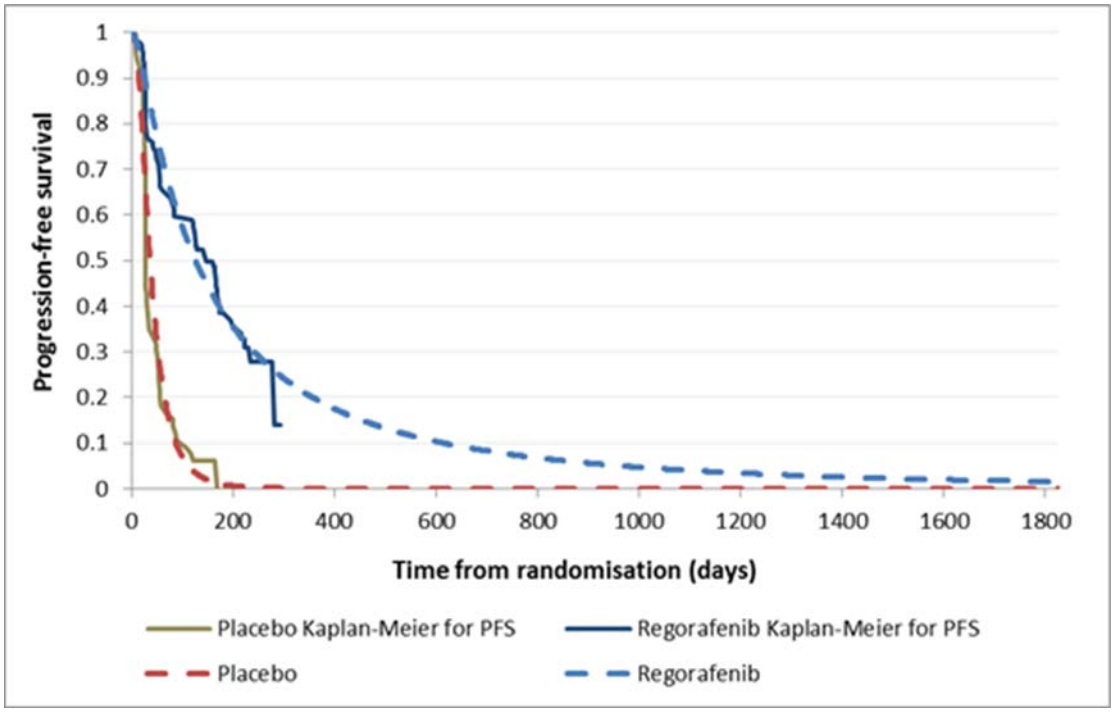
Given this, we accept Bayer's choice of base case, and consider this matter no further.

Table 19. AICs and BICs for PFS extrapolation

Parametric Model	AIC			BIC		
	Placebo	Regorafenib	SUM AIC	Placebo	Regorafenib	SUM BIC
Exponential	170.886	349.477	520.363	173.078	352.368	525.446
Loglogistic	139.045	348.561	487.605	143.424	354.341	497.765
Weibull	162.487	350.95	513.437	162.487	356.731	519.218
Lognormal	142.055	343.396	**485.45**	146.434	349.177	**495.611**
Gompertz	172.009	351.475	523.484	176.388	357.255	533.643

Source: Bayer submission, Figure 31, p. 157

Figure 20. Lognormal model (base case) for PFS compared to GRID PFS K-M data



Source: Bayer submission, Figure 20, p. 158

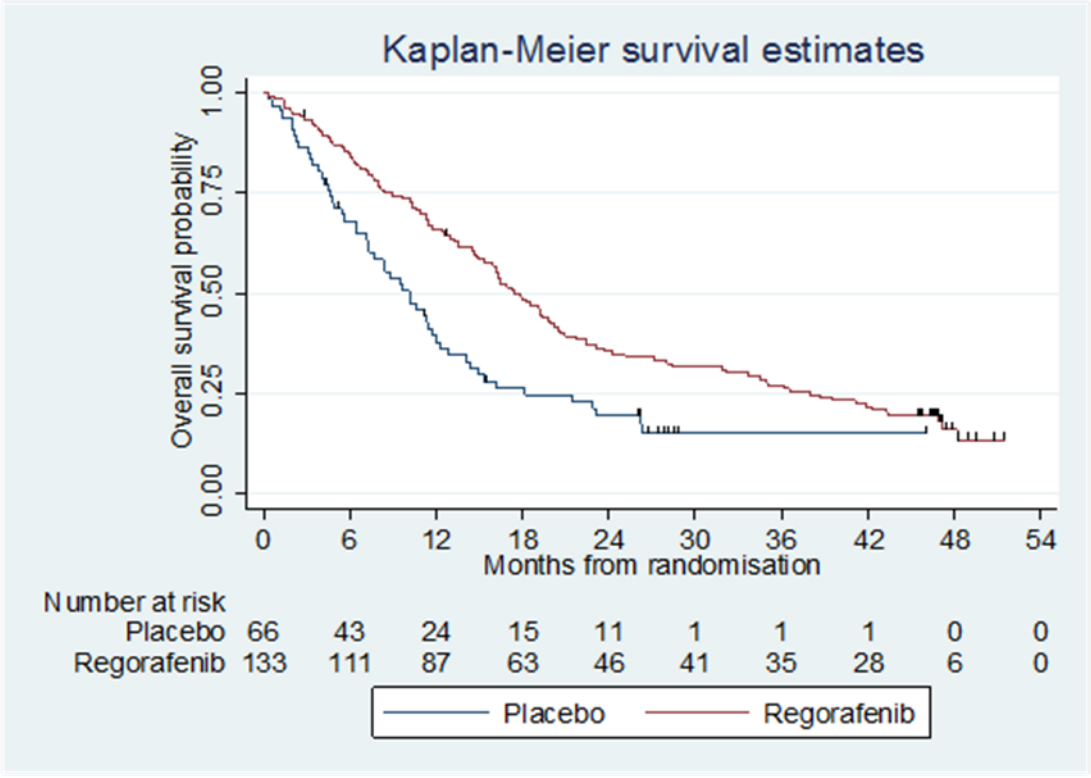
5.3.6.2 OS and crossover adjustments

Due to the crossover design of the GRID trial, estimating OS is more complex than PFS. Cross-over was only permitted after disease progression for the placebo arm, so no adjustment was required for PFS. However, 87.9% (n=58/66) of patients in the placebo arm crossed to the regorafenib arm after disease progression. This introduces the possibility of overestimating OS in the placebo arm if regorafenib gave them benefits in the PPS state and hence confounding the cost-effectiveness estimates.

Three crossover correction methods were considered; Iterative Parameter Estimation (IPE), Rank Preserving Structural Failure Time method (RPSFT), and Inverse Probability of Censoring Weights (IPCW). The aim of these methods is to reconstruct the OS patient level data in the placebo arm as if there had been no crossover in order to get an unbiased estimate of OS in the BSC arm. The IPCW method was discarded due to the high proportion of placebo patients crossing over, which Bayer argue is likely to result in high amounts of bias in treatment effect estimates (Source: Bayer submission, p.152). We agree that the IPCW method can be unreliable if the proportion of patients that switch is high. However, we understand that the method is considered unreliable only if the weights that are applied to the survival data corresponding to the patients that do not switch at very high. Nonetheless, we accept Bayer's justification in rejecting the IPCW method.

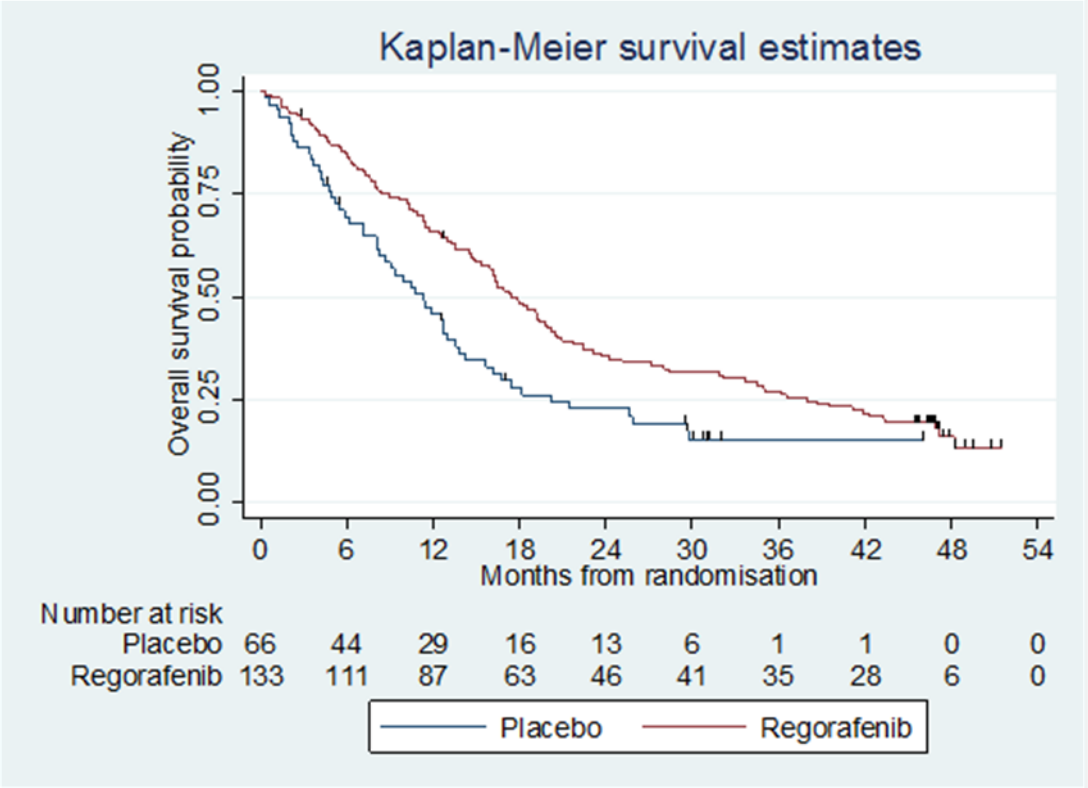
IPE-adjusted and RPSFT-adjusted K-M data for OS are shown in Figure 21 and Figure 22, respectively for both the 2015 data cut-off. Notice that after correction for cross-over, Bayer predict a clear OS benefit of regorafenib versus placebo. Compare this to the unadjusted OS data, in which OS for Regorafenib and placebo were very similar (Figure 19). This alerts us to the fact that the cost-effectiveness of regorafenib is very sensitive to the adjustment for treatment switching. Indeed, without adjusting for treatment switching, allowing for the PAS, Bayer estimate that their base case ICER increases massively, from £38,000 to £149,000 per QALY (Bayer model "Executive Summary" tab, Crossover adjustment method set to "Unadjusted"). We caution that we are not convinced of the accuracy of this figure for two reasons. First, the estimated mean OS for regorafenib changes when we set the adjustment method to "Unadjusted" and second because under the "Unadjusted" method, Bayer's model allows for no cost of regorafenib post-progression in the placebo, whereas we believe it should. However, we can say that the cost-effectiveness of regorafenib is very sensitive to the adjustment for treatment switching.

Figure 21. IPE crossover-adjusted Kaplan-Meier OS data (2015 data cut-off)



Source: Bayer submission, Figure 18, p.155

Figure 22. RPSFT crossover-adjusted Kaplan-Meier OS data (2015 data cut-off)



Source: Bayer submission, Figure 19, p.156

Despite the fact that the Kaplan-Meier graph for the placebo arm changed only slightly from the 2015 to the 2017 data cut-off, Bayer estimate a shorter OS for placebo after correction for cross-over using the 2017 data, compared to the 2015 data (Figure 23). Specifically and importantly, the estimated mean OS in the placebo arm decreases from 1.64 to 1.25 years, a reduction of 24%. Bayer justify this as follows: "*This is a result of the greater follow-up time allowing for a longer potential censoring date within the crossover adjustment calculation*" (Bayer response to clarification, p11). This reduction in mean OS substantially improves the cost-effectiveness of regorafenib. For example, assuming the PAS, the ICER for regorafenib vs. BSC decreases from £49,000 to £38,000 per QALY.

Given the importance of recensoring, we now give a brief explanation of this process. Recensoring involves data being recensored at an earlier time-point to avoid informative censoring and is therefore associated with a loss of longer-term survival information. Some observed events will become censored if the recensoring time is shorter than the counterfactual event time. The time-point at which recensoring occurs is related to the magnitude of the estimated treatment effect; the larger the treatment effect the earlier the recensoring time-point.[16]. Recensoring may lead to biased estimates of the "average" treatment effect in circumstances where proportional treatment effect assumptions do not hold, because longer term data on the effect of treatment may be lost.[16] We understand

that, whilst the NICE TSD recommends recensoring, whether to perform recensoring remains a subject of academic debate. Hence it is probably best to perform the adjustment both with and without recensoring. We further understand that the estimated treatment effect is generally greater when recensoring is performed compared to the analysis without recensoring. **The PenTAG base case employs a similar recensoring approach to Bayer, via the IPE method for treatment switching.**

Figure 23. OS Kaplan-Meier (2015 and 2017 data cut-off comparison)

Source: Bayer response to clarification, Figure 1, p.9

We have several important concerns with the 2017 OS data:

- Only the 2015 data cut appears in Bayer's Clinical Study Report. The 2017 data does not appear in this document. We imagine that the switching adjusted OS data from the 2015 data cut is more likely to be correct, given that it appears in the Clinical Study Report. We have no other means of judging the accuracy of the adjustment for the 2017 data other than Bayer's Addendum.
- Next, we assumed that the maximum follow-up time shown in the Kaplan-Meier graphs for the switching-adjusted placebo OS data would be greater for the 2017 data-cut compared to the 2015 cut, given that the 2017 data is more mature. However, as can be seen in Figure 23, the maximum follow up times are equal, specifically at 1,397 days.

- Next, we remain to be convinced that that a relatively small increase in the maturity of the survival data can results in such a substantial reduction in estimated mean OS for the placebo, of 24%.
- We do not have access to the underlying individual patient data, to enable us to check the switching adjustment.
- Finally, under the RPSFT and IPE methods, by definition, the p-values for the OS HR hazard ratios for the unadjusted (ITT) and switching adjusted data should be identical. However, Bayer quote very different values:
    - 2015 data cut: ITT p value = 0.285777, IPE-adjusted p value = 0.000949, RPSFT-adjusted p value = 0.002862 (Bayer's original report Table 22, p92).
    - 2017 data cut: ITT p value = 0.2298251, IPE-adjusted p value = 0.0000021, RPSFT-adjusted p value = 0.0000071 (Section 4.2.2.4.2, Table 13, p52)

Given all these concerns, we use the 2015 data-cut for OS in our base case.

We understand that the RPSFT method is commonly used in NICE assessments, but the IPE method less so. The IPE method is an extension of the RPSFT method using parametric methods.[16] The same accelerated failure time model is used as for the RPSFT method, but a parametric failure time model is fitted to the original unadjusted ITT data to obtain an initial estimate of the treatment effect. The failure times of switching patients are then re-estimated using this, and this iterative procedure continues until the new estimate is very close to the previous estimate, at which point the process is said to have converged.[16].

The IPE procedure makes similar assumptions to the RPSFTM method – for example the "common treatment effect" assumption. An additional assumption is that survival times follow a parametric distribution, and thus it is important to identify suitable parametric models, which in itself can be problematic.[16] The IPE method is expected to perform similarly, provided a suitable parametric distribution can be identified. Indeed, the results using the IPE and RPSFT methods are similar in our case.

Bayer chose the IPE method for their base case cross-over adjustment method due to Morden et al's study demonstrating this method's efficacy specifically that it performed particularly well in terms of reducing bias in the estimates of the true treatment effect. In line with NICE Decision Support Unit guidance, recensoring was applied to both methods in order to avoid bias.

Hazard ratios for OS for unadjusted, IPE-adjusted, and RPSFT-adjusted models estimated using a Cox model are presented below in Table 20. Figure 24 gives a visual comparison of the 2017 and 2015 OS hazard ratios with the different adjustment methods. Bayer say that their methods allow for recensoring. They further add that the OS HRs corresponding to the 2015 data cut, and reported in the Clinical Study Report, of 0.586 and 0.616 for the IPE and RPSFT methods respectively, were estimated without recensoring.

In our opinion, both the IPE and RPSFT are reasonable candidate adjustment methods. We are not convinced by Bayer's rationale for choosing the IPE method as the base case. It is our understanding that both methods are reasonable candidates. However, Bayer do not say why the IPE is more relevant than the RPSFT method in the specific case of the GRID RCT. Fortunately, **the two methods give reasonably similar estimates of OS for placebo**. Specifically, using the RPSFT method, Bayer's base case ICER under the PAS of £38,000 increases only slightly, to £39,000 per QALY. Therefore, we do not dwell on this issue.

Table 20. OS hazard ratios in for 2015 and 2017 data cut-offs

Crossover Adjustment	2015 cut (no recensoring)	2015 cut (recensoring)	2017 cut (recensoring)
Unadjusted*	0.909	0.909	[redacted]
IPE	0.586	[redacted]	[redacted]
RPSFT	0.616	[redacted]	[redacted]

Figure 24. Visual comparison of OS HRs

At the NICE Decision Problem Meeting on 12th January 2017, we asked Bayer to send us all the data necessary to recreate their adjustment for treatment switching, e.g. the relevant individual patient data from GRID. They replied that they would be very unlikely to send this to us, because it would be against Bayer policy to release such data. Indeed, Bayer have not provided us with the data required for us to check their switching adjustment. Whilst we understand that there may be issues concerning data confidentiality, this does present us with the problem that we are unable to check that the methods have been implemented correctly.

In their original report, Bayer provided some information on the implementation of the IPE and RPSFT methods. They said the methods were implemented in STATA, and the IPE method was implemented using the Weibull parametric failure time model, as in the study by Morden et al [17] similarly, and the RPSFT method was implemented using the logrank test, also recommended by Morden et al[17]. Bayer stated that, in line with the methodological approach recommended by NICE Decision Support Unit[16], recensoring was applied in order to avoid bias for the IPE and RPSFT methods. They noted further that recensoring was not applied for the IPE and RPSFT crossover corrections presented in the GRID clinical study report.

At the clarification stage of this appraisal, given the importance of these methods, we asked Bayer to provide more details on how the implementation of the methods, for example whether the treatment effect of regorafenib was assumed to apply only while regorafenib was being taken, or for the whole period from the start of regorafenib treatment until death.

Bayer responded as follows:

The IPE and RPSFT methods were implemented using Stata 11 and the strbee program developed by White et al. 2002.[18] (http://ageconsearch.umn.edu/bitstream/115957/2/sjart_st0012.pdf), as described by Morden et al. 2011[17] (https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-4). The commands implemented for IPE and RPSFT were as follows (square brackets represent inputs from data):

IPE:

Strbee [treatment], test(weibull) xo0([time to crossover] [crossover flag]) endstudy([study follow-up duration]) ipe

RPSFT:

strbee [treatment], test(logrank) xo0([time to crossover] [crossover flag]) endstudy[study follow-up duration])

A logrank test is implemented for the RPSFT method in order to calculate the test statistic for independence between patients' counterfactual event time and the treatment arm to which they were assigned, as recommended by Morden et al. 2011. For the IPE method, where a likelihood-based analysis is undertaken a Weibull distribution is utilised, also consistent with Morden et al. 2011[17]

Recensoring was implemented directly within the strbee program, using a maximum potential censoring time equal to the duration of study follow up. Recensoring was applied in order to reduce bias from potentially informative censoring as a result of switching (switching itself may potentially be informative if it is related to prognosis). Recensoring is applied in a manner consistent with Morden et al. 2011[17], and discussed further in White et al. 2002 [18].

The entire data for overall survival was used for the crossover adjustment; the assumption is therefore that treatment effect of regorafenib is applied from initiation of treatment until death, regardless of discontinuation. The treatment effect of regorafenib is therefore likely reduced as it will be an average of patients on and off treatment. Only placebo patients who crossover to regorafenib have their survival times adjusted, non-crossers and those in the regorafenib arm are unchanged.

In general, we are satisfied with their response. We do however note the strong assumption in the last paragraph, regarding the assumed duration of the treatment effect of regorafenib.

Given this, we perform a scenario analysis in which we assume that regorafenib improves survival only whilst the patient is taking the drug. In this case, and making a further simplifying assumption that approximately similar proportions of patients are alive on progression in the treatment arms, then to a good degree of accuracy, we model the costs and QALYs only whilst patients are in PFS. We further assume a dose intensity of 87% during PFS, or a mean dose of 139.8mg (compared to the standard dose of 160mg). In this case, Bayer's base case ICERs of £38,000 and [redacted] per QALY increase substantially, to £52,000 and [redacted] per QALY.

5.3.6.3 OS extrapolation

All parametric models for IPE-adjusted placebo and regorafenib OS are shown below for both the 2017 cut-off (Figure 25 and Figure 26).

Figure 25. Parametric models for OS and GRID Kaplan-Meier data, 2017 cut-off (IPE-adjusted placebo)

Source: Bayer response to clarification, Figure 3, p.11

Figure 26. Parametric models for OS and GRID Kaplan-Meier data, 2017 cut-off (Regorafenib arm)

Source: Bayer response to clarification, figure 2, p. 11

As in the extrapolation for the PFS parametric model, Bayer selected the best fit by minimising the sum of the AIC/BIC. The full list of AIC/BICs are shown in the Appendices, for both the 2015 and 2017 cut-offs.

Bayer acknowledge that the lowest AIC/BIC values come from the loglogistic/exponential models. For the 2015 analysis, Bayer settles on the loglogistic model by arguing (Source: Bayer submission, p.160):

> *"…the difference between the BIC values for the exponential and loglogistic models results being smaller compared to the other parametric models. Hence, the loglogistic model was selected for use in the model base case."*

For the 2017 cut-off (base case), Bayer continues to use the log-logistic model for the base case by arguing (Source: Bayer addendum, p. 13):

> *"The loglogistic model gives the minimum AIC for regorafenib OS and for both the RPSFT and IPE methods used in the placebo arm…. Following visual inspection of the parametric functions applied to the Kaplan-Maier curves for the two study arms and analysis of the AIC and BIC, log-logistic was selected as best fitting model."*

Bayer also had the fittings of the 5 parametric models validated by 2 consultant oncologists who specialised in the disease area. They argue that, from a clinical perspective, the

loglogistic, Weibull and Gompertz models all looked clinically plausible (for the 2015 data cut-off). The base case log-logistic model used for the regorafenib arm and the IPE-adjusted placebo arm for OS is shown in Figure 27 (2017 data cut-off).

Bayer also explores using hazard ratios for the regorafenib arm to extrapolate the placebo arm as a sensitivity analysis (rather than extrapolating arms separately). Bayer are unable to reject the proportional hazards assumption, which validates this approach. Bayer settles on using parametric models fitted separately to individual PFS and OS curves for the base case.

Figure 27. Log-logistic models for OS (2017 data cut)

We agree with Bayer that it is good practice to use the same functional form (e.g. log-logistic) for both treatment arms, in accordance with guidance from the NICE Decision Support Unit.[16].

As stated above, Bayer claim that 2 consultant oncologists, who specialise in the disease area believe that, from a clinical perspective, the loglogistic, Weibull and Gompertz models all look clinically plausible for the 2015 data cut-off. Bayer's only justification for choosing the log-logistic for their base case is that it provides the best fit to the trial OS data as measured by AIC / BIC. Whilst we acknowledge that the fit to trial data is a consideration,

we understand that the clinical plausibility of the extrapolations to be critical. The cost-effectiveness of Regorafenib is sensitive to choice of statistical distribution. For example, using the Weibull, Bayer's ICER with the PAS increases from £38,000 to £45,000 per QALY. With the Gompertz, the ICER increases to £47,000 per QALY. Given this, it is worth considering carefully the choice of statistical function.

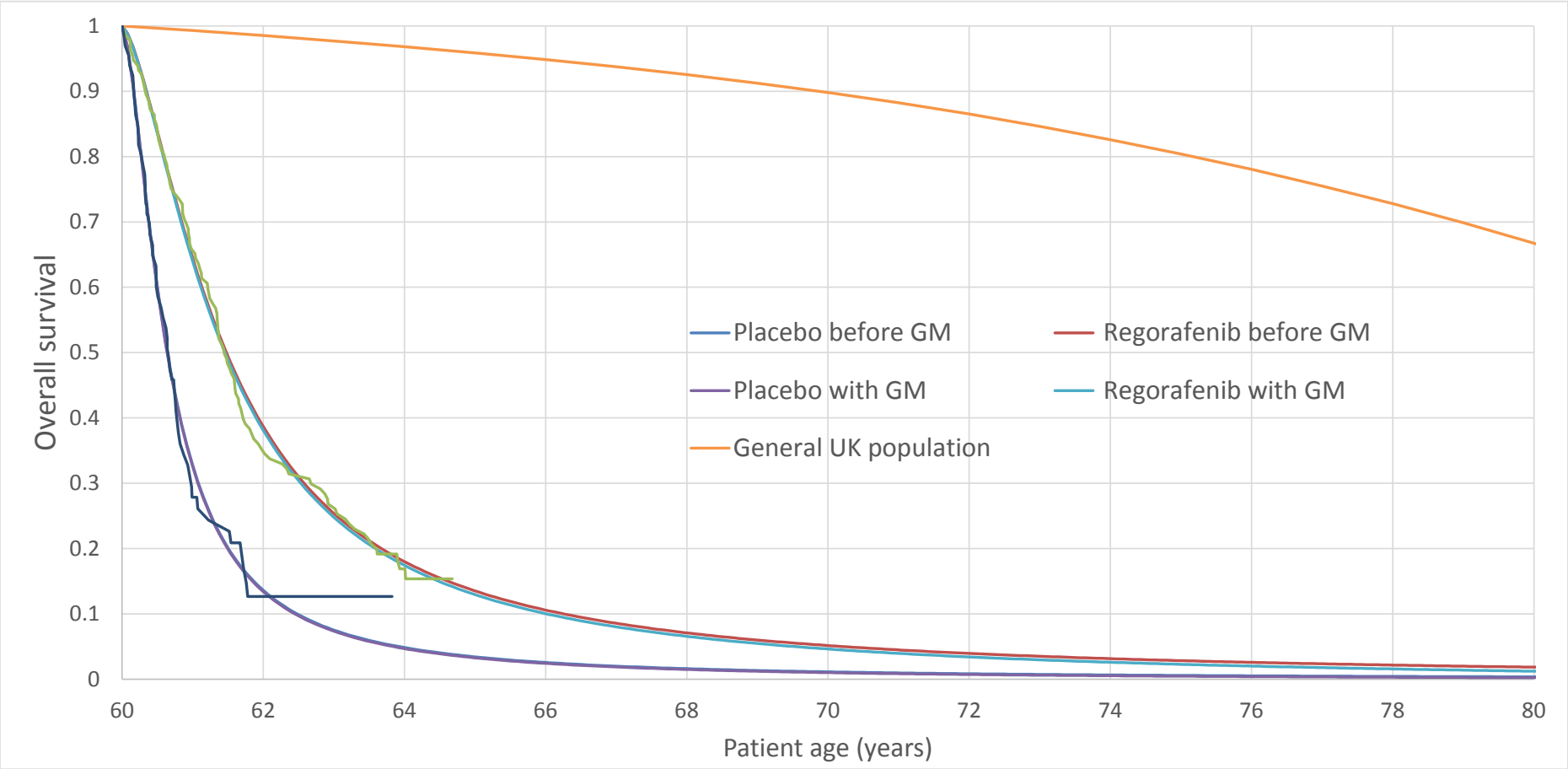
We believe essential to incorporate background mortality. This is because mortality in GRID will be due almost exclusively to causes related to GIST. However, many years later, a much larger proportion of deaths is likely to be due to causes unrelated to GIST, such as heart disease, or diabetes. Bayer's extrapolations make no allowance for this additional mortality.

We have adapted Bayer's model to allow for the extra cause mortality for the general population (Figure 28). Specifically, this change is implemented in worksheet "OS Parametric GRID". Then, the ICERs (£/QALY) increase for log-logistic and Weibull and Gompertz as follows for the example of the PAS:

- £38,000 to £41,000 log-logistic.
- £45,000 to £46,000 Weibull.
- £47,000 to £48,000 Gompertz

The ICERs increase markedly in the case of the log-logistic distribution because this is the longest-tailed distribution, and thus background mortality is more influential as the cohort ages.

Figure 28. OS for regorafenib and placebo with log-logistic extrapolations with and without general background mortality (GM)



Additional searches limited to the previous 10 years were carried out in Medline to identify studies with survival curves for GIST. This was not a systematic review, but we searched for

- Terms for gastrointestinal stromal tumour OR GIST AND
- Terms for survival curve OR Kaplan Meier

Fifty eight papers were identified with potentially relevant data. On screening for the correct population, our search yielded three relevant publications.

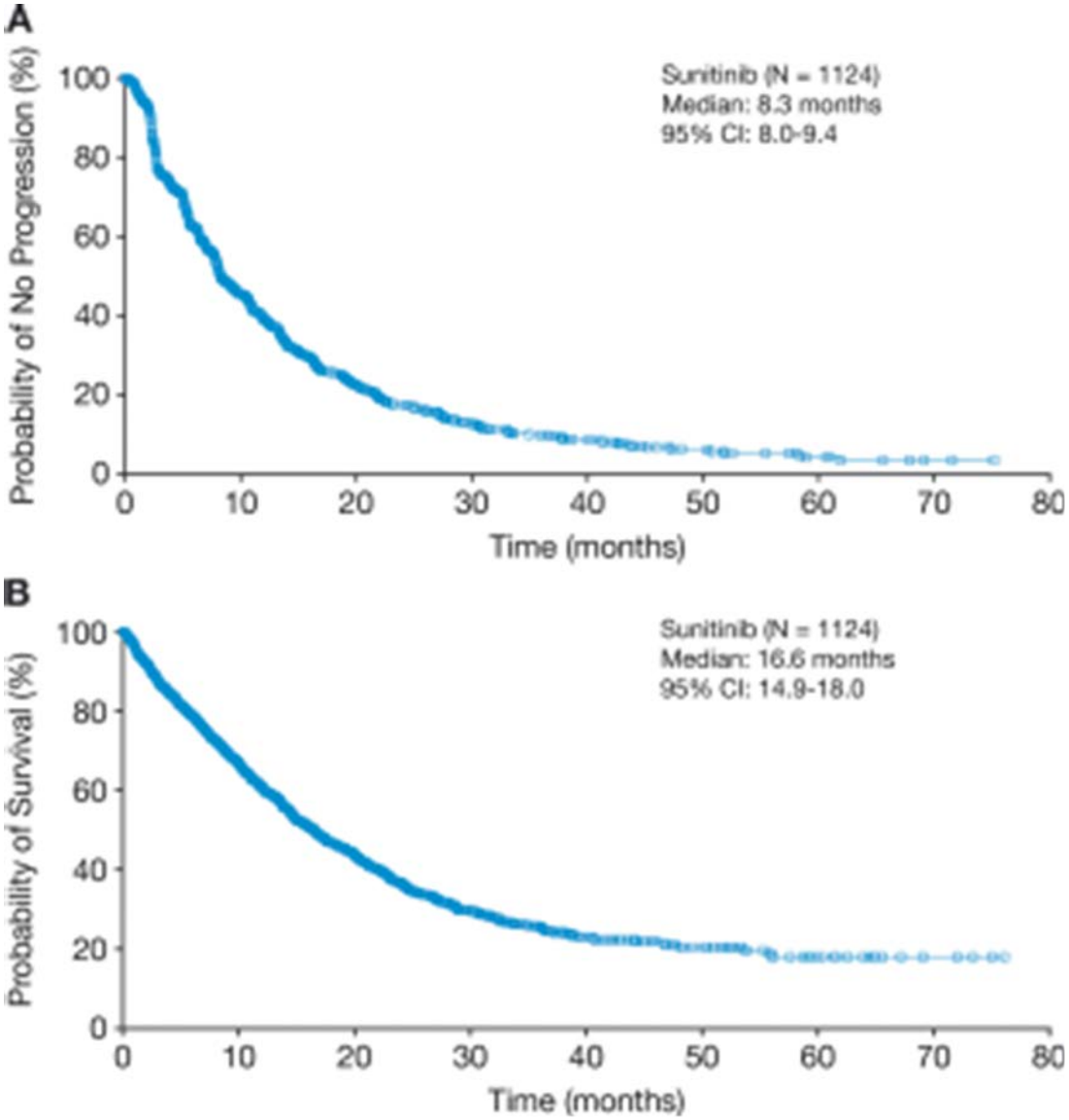
Kang et al (2013)[11] consider a patient population relevant to the current appraisal, namely patients with metastatic or unresectable GIST after failure of imatinib and sunitinib. However, we are unable to use data from this study to inform extrapolation OS in the current HTA because the data is insufficiently mature. Indeed, follow-up in this study was shorter than in GRID.

Yoon et al (2012)[19] consider a patient population less relevant. Whilst patient had failed imatinib, they had not necessary also failed sunitinib. Again, the data from this study is insufficiently mature to guide extrapolation in the current HTA.

The third study, Reichardt et al (2015)[20] is, however, useful because the OS data is slightly more mature than in GRID. Patients had advanced GIST and had previously failed imatinib, not but sunitinib. All 1,124 patients in this large international study took sunitinib. Median patients age was 59, virtually the same as in GRID, at 60 years. 60% of patients were male, again similar to the 64% in GRID. The ECOG distribution was similar compared with that in GRID, with patients typically with a slightly worse ECOG than in GRID.

Median time to progression was substantially longer, at 8.3 months than in the regorafenib arm of GRID (4.8 months). Median OS on sunitinib, at 16.6 months, was however very similar to that of the regorafenib arm of GRID, at 17.3 months (Figure 29).

Figure 29. A: PFS and B: OS from Reichardt et al. (2015) trial of sunitinib



Source: Figure is reproduction of Figure 1 in Reichardt et al. (2015)[20]

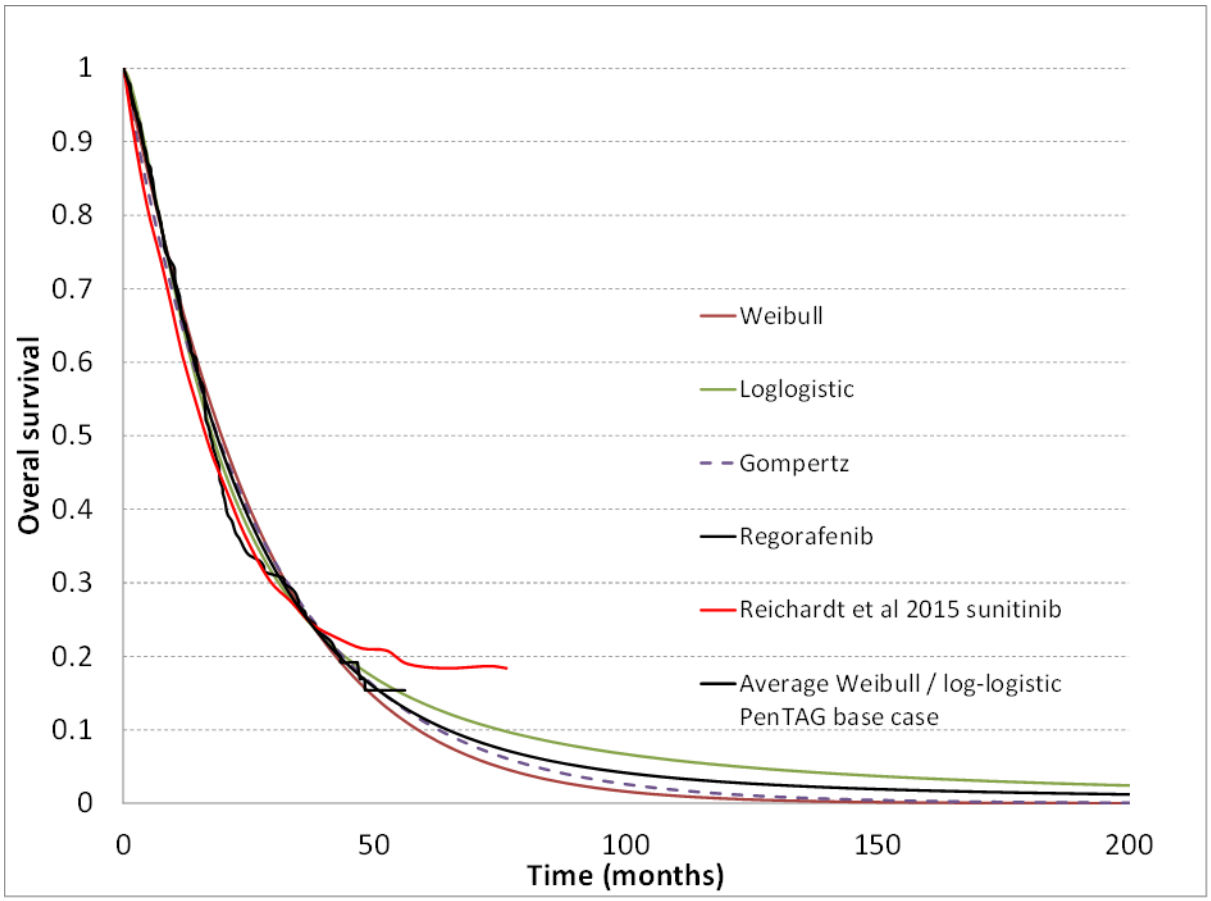
The OS for sunitinib in Reichardt et al. (2015) is slightly more mature than in the regorafenib arm of GRID. Observe also that OS is rather longer-tailed in Reichardt et al. (2015) than in the regorafenib arm of GRID (Figure 30). This might favour the choice of the log-logistic extrapolation over that of the Weibull or Gompertz. However, we caution against relying too much on the data from Reichardt et al. (2015), as:

(a) the uncertainty in the tail of OS in Reichardt et al. (2015) may be large, as the number of patients at risk in the tail might be low (but not reported),

(b) the patients in Reichardt et al. (2015) differed from those in GRID in that they had not previously been treated with suntinib, whereas all patients in GRID had,

(c) the patients in Reichardt et al. (2015) all took sunitinib, verus regorafenib in the regorafenib arm of GRID.

Figure 30. A: OS from Reichardt et al. (2015) trial of sunitinib and PenTAG base case



Source: Figure is reproduction in Reichardt et al. (2015)[20]
Notes: The lognormal fit is not displayed, as it is very similar to the log-logistic, which is shown. Similarly, the exponential fit is not displayed, as it is very similar to the Gompertz, which is shown.

In their base case, Bayer choose the log-logistic distribution to model OS in both treatment arms. As explained above, their two consultant oncologists believed that the log-logistic, Weibull and Gompertz models all look clinically plausible. Bayer's only justification for choosing the log-logistic for their base case is that it provides the best fit to the trial OS data as measured by AIC / BIC. Whilst we acknowledge that the fit to trial data is a consideration, we understand that the clinical plausibility of the extrapolations to be critical.

Having considered everything above, we believe that the evidence in favour of the longer-tailed and shorter-tailed distributions appears evenly balanced. **Therefore, in our base case, we model OS as the average of the Weibull and log-logistic distributions, adjusted for general background mortality (**Figure 30**).** Expressed formally, this is a

form of model averaging, with Bayesian prior weights of 50% applied to the shortest tailed Weibull and longer tailed log-logistic.

In this part of our case, Bayer's base case ICER under the PAS increases from £38,000 to £43,000 per QALY.

We also present Scenario analyses using just the Weibull adjusted for background mortality and Gompertz adjusted for background mortality.

5.3.7 Health related quality of life

Health effects were measured in quality adjusted life-years (QALYs) in accordance with the NICE reference case. Utility estimates were taken directly from the GRID trial using both the EQ-5D and the EORTC QLQ-C30 (Cancer Core Quality of Life Questionnaire) questionnaires, and both were used to estimate health state utility values (HSUVs).

5.3.7.1 EQ-5D

Data were taken from patients who had baseline EQ-5D assessments and at least one post-baseline assessment, and the Patient Reported Outcome Analysis Set (PROAS) was used. Paired-samples comparison and repeated measures analysis was then used to obtain HSUVs.

The paired-samples comparison based on t tests was used to assess intra-patient differences in the EQ-5D at baseline (day 1, cycle 1) and the first post-progression observation (which had to be after the patient knew they had progressed). A total of 77 paired samples were obtained (Table 21). An alternative comparison was also performed where the progression-free state was split into the regorafenib and placebo arms, and the first post-baseline measurement was used in lieu of the first baseline measurement in order to incorporate the treatment effect. Results are shown in Table 22.

Table 21. EQ-5D HSUVs from paired-samples

Health state	Mean utility	Observations, N	SD	SE
Progression-free	0.767	77	0.221	0.025
Progressed	0.647	77	0.343	0.039

Source: Bayer submission, Table 36, p.166

Table 22. EQ-5D HSUVs from paired-samples splitting by treatment in the progression-free state

Health state	Mean utility	Subjects	SD	SE
Progression Free - Placebo	0.583	12	0.341	0.098
Progression Free - Regorafenib	0.702	27	0.281	0.054
Progressed Disease	0.649	49	0.320	0.046

Source: Bayer submission, Table 37, p. 166

Bayer also estimates a linear mixed model with a first-order, autoregressive covariance structure (with subject identity modelled as a random effects) to estimate HSUVs, their repeated measures analysis. Results are shown below in Table 23. Bayer considers this a sensitivity analysis.

Table 23. EQ-5D HSUVs from repeated measures

Health state	Mean utility	SE	95% CI
Progression free	0.743	0.016	0.712, 0.775
Progressed	0.703	0.023	0.657, 0.748

Source: Bayer submission, Table 38, p.167

The repeated measures analysis was also repeated, splitting the progression-free state into the regorafenib and placebo arms (Table 24). This yields a slightly lower HSUV for regorafenib PFS compared to placebo.

Table 24. EQ-5D HSUV from repeated measures and splitting treatment during PFS

Health state	Mean utility	SE	95% CI
Progression Free - Regorafenib	0.741	0.018	0.706, 0.777
Progression Free - Placebo	0.750	0.027	0.698, 0.802
Progressed Disease	0.681	0.023	0.637, 0.725

Source: Bayer submission, Table 39, p. 167

The paired-samples without splitting by pre-progression treatment utility estimates (Table 21**) were used in the base case analysis.** Bayer justifies this by first arguing that the repeated measures analysis is likely to be biased because more measurements were taken for patients in the progression-free state. As utility generally declines over time with age and tumour burden, this could bias estimates. They also note that there were no clinically meaningful differences in EQ-5D between the two treatment arms.

Furthermore, due to the high level of cross-over, the repeated measures analysis would compare non-homogeneous progressed populations; utility observations would be taken for those people in the initial diagnosis of progressed disease and also those under active treatment with regorafenib.

Despite noting that utility often declines with age, Bayer argue that the utility estimates from the GRID trial are constant over time, citing Poole et al (2015)[21] as justification. Bayer do acknowledge that HRQL may decline in the progressed state towards the end of a patient's life, but note that this decrement would apply to both arms and hence no incremental effect would exist, making it reasonable to omit.

Age-related utility decrements were applied to the model for the PenTAG base case. It was assumed that Bayer's baseline utility values incorporated time-invariant characteristics (such as gender), hence the only adjustments needed to be made would be the decrements associated with aging. The values themselves are taken from the Health Survey for England (2012)[22], which give regression coefficients for age and age squared. Therefore, the formula for utility as function of time is:

$$u_{it} = \alpha_i + \beta_{1t}t + \beta_{2t}t^2$$

Where *i* refers to disease state and *t* is time (or age). Since patients enter the model at age 60, the base line utilities values are when *t* = 60. To extrapolate beyond this to *t* + *x*, the equation becomes:

$$u_{it} = \alpha_i + \beta_{1t}(t+x) + \beta_{2t}(t+x)^2 \mid_{t=60}$$

$$= [\alpha_i + \beta_{1t}60 + \beta_{2t}60^2] + \{\beta_{1t}x + 2\beta_{2t}60x + \beta_{2t}x^2\}$$

Where the bracketed term refers to the baseline utilities and term in the curly brackets refers to the added decrement (as β_{1t} and β_{2t} are negative). This has a modest effect on the ICER per QALY, increasing it by around £1,000.

5.3.7.2 EORTC mapping

As with the EQ-5D, paired-samples and repeated measures were used to generate alternative utility estimates. The EORTC QLQ-30 is a commonly used measure of quality of life for cancer patients. Answers were mapped to utilities using the method proposed by Rowen et al[23]. Their mapping algorithm was then applied to the GRID EORTC data to obtain utility estimates. There were 78 paired-samples observations, and the estimates of this method are shown in Table 25. In order to gain a greater number of data points, Bayer used a similar autoregressive covariance structure method as with the EQ-5D, with results shown

below in Table 26. Only patients with non-censored time to progression dates, a baseline assessment and at least one post-baseline assessment were included (n=133). Regardless of the EORTC utility derivation method, the NICE reference case states that EQ-5D results are preferred over other utility measures when they are available, and hence, in the base case, Bayer use the EQ-5D.

Table 25. EORTC mapped utilities from paired-samples

Health state	Mean utility	Observations, N	SD	SE
Progression-free	0.818	78	0.138	0.016
Progressed	0.751	78	0.158	0.018

Source: Bayer submission, Figure 40, p. 169

Table 26. EORTC mapped utilities from repeated measures analysis

Health state	Mean utility	Observations, N	SE	95% CI
Progression free	0.794	320	0.011	0.771, 0.816
Progressed	0.756	128	0.013	0.730, 0.783

Source: Bayer submission, Figure 41, p.169

5.3.7.3 Adverse events

Bayer note that the three most common AEs – hand foot skin reactions (HFSR), diarrhoea, and fatigue – are all easily manageable and their effects on health-related quality of life are negligible. However, they assume that the EQ-5D values obtained from repeated measures, where PFS was split into treatment arms, were inclusive of the treatment-associated adverse events. The results of this analysis are shown in Table 24 (section 5.3.7.1), with the regorafenib arm showing a slightly lower pre-progression utility than the placebo arm.

5.3.7.4 Health-related quality of life studies

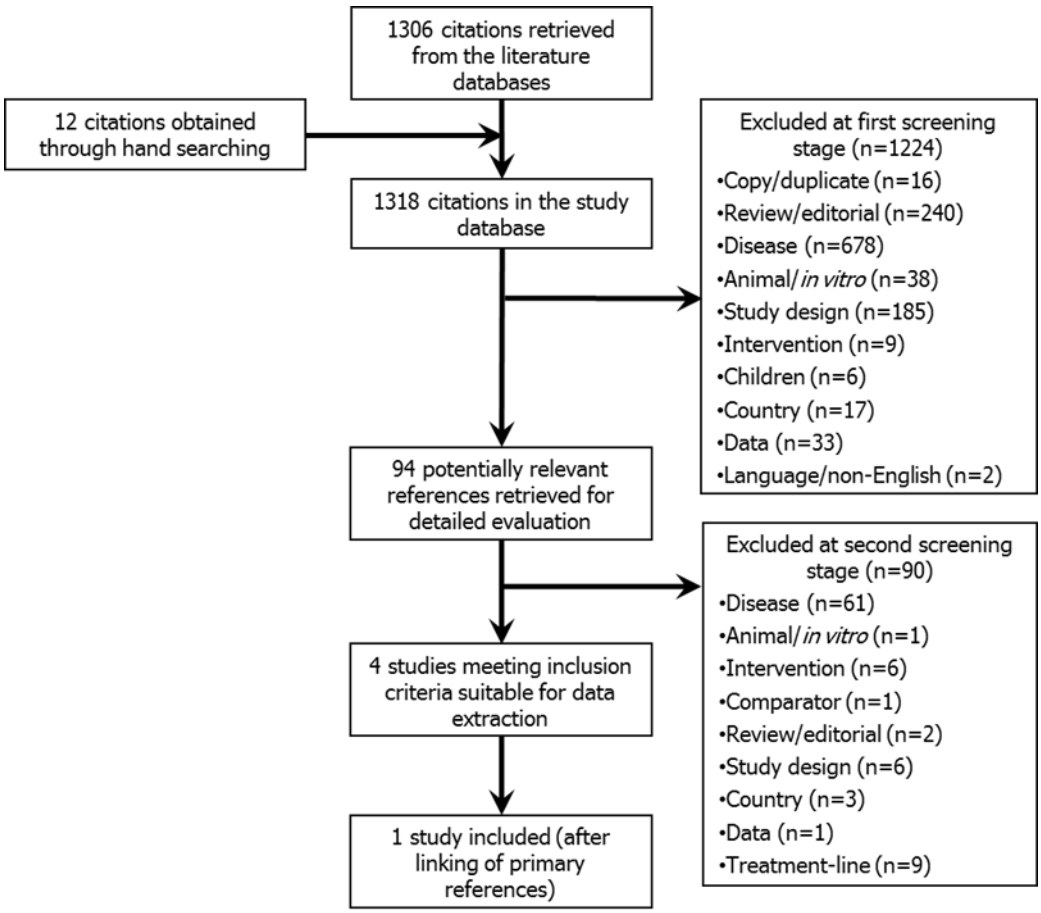
Bayer carried out a full systematic review of published literature to identify health-related QoL studies relevant to the decision problem. The objective was to identify research on utilities associated with GIST and/or studies investigating HRQoL outcomes. The following databases were searched (from inception to 19 December 2016): MEDLINE, MEDLINE (R) In-Process, EMBASE, EconLIT, and NHS EED. The database search was updated 3 times from December 2011 to December 2016. The following conferences were also searched: American Society of Clinical Oncology, European Society for Medical Oncology, International Society for Pharmacoeconomics and Outcomes Research, and the International Society for Quality of Life Research. The set of exclusion/inclusion criteria and the PRISMA flow diagram are shown below in Table 27 and Figure 31.

Table 27. Inclusion/exclusion criteria for cost-effectiveness publications

	Inclusion criteria	Exclusion criteria
Study design	• Study design appropriate to report the HRQoL/utility associated with GIST (patient preference studies, utility mapping studies, cohort studies / longitudinal studies (retrospective), cohort studies / longitudinal studies (prospective), case control studies, cross sectional studies, analysis of hospital records/databases, cost-effectiveness analyses, cost-utility analyses, cost-benefit analyses)	• Literature and systematic reviews • Database studies or epidemiology studies, not collecting utility data • RCTs (with no piggy-back economic evaluations) • Studies published in non-English language (with/without English abstracts)
Patient population	• Studies including adult patients (aged ≥18 years) • Studies reporting data in countries of interest (US, Canada, Australia, France, Germany, Italy, Spain, UK, Brazil, Mexico, Japan, China, Korea)	• Studies in children or adolescents • Studies conducted in animals or *in vitro*
Disease/ therapy	• Studies including patients with metastatic, advanced, and/or unresectable GIST, defined as such using the study author's definition • Studies of third-line patients (who have failed two pharmacological therapies). However, as it is was anticipated that studies focused on third-line patients were rare, studies in first- and second-line patients were only excluded at the final stage of the second pass (at the first pass stage there was no exclusion based on therapy line)	• Studies that did not include patients with a specific GIST diagnosis (including gastrointestinal leiomyosarcoma that appeared to behave as GIST, soft-tissue sarcoma that appeared to behave as GIST, oesophageal leiomyosarcoma, gastric leiomyoma, gastric leiomyoblastoma, small intestinal leiomyoma and leiomyosarcoma, colonic and rectal leiomyoma and eiomyosarcoma, gastrointestinal autonomic nerve tumour, eiomyoma and leiomyosarcoma of omentum and mesentery, retroperitoneal leiomyosarcoma)
Intervention	• Regorafenib	• Any other intervention
Comparator	• Placebo/BSC	• Any other comparator

Source: Bayer submission, Table 42, p.172

Figure 31. PRISMA flow diagram for HRQoL studies



Source: Bayer submission, Figure 24, p.173

After a full text review, only one relevant study was found, the GRID study, shown in Table 28.

Table 28. Summary list of published HRQOL studies

Study	Country	Population	Intervention	Sample size	Elicitation method	Health states	Utility score
GRID **Poole et al. (2015) (40)**	**58 years (SD 13.1)**	**Male: 64.3%** **Mean age (years): 58** **Metastatic** **Unresectable, associated with disease progression with imatinib and sunitinib (100%)**	**Regorafenib**	**185**	**EQ-5D index score**	**Mean at baseline (day 1 of cycle 1) from the combined data set**	**0.769**
						Mean at First progression-free state (Progression-free state represented by baseline observation, QoL observations made on day 1 of cycle 1 before commencing blinded treatment)	**0.767**
						Mean at First post-progression State (The first post progression health state suggesting significantly impaired health-related quality of life after confirmed disease progression showed a decrease of -0.120)	0.647

Source: Bayer submission, Table 43, p.174

5.3.8 Resources and costs

5.3.8.1 Drug acquisition costs

Drug acquisition costs make up the vast majority of costs in the regorafenib arm. The drug prices used in the economic model are based on the list price and separately, a confidential Patient Access Scheme (PAS).

Regorafenib comes in 40mg tablets and all patients in the trial took multiples of 40 mg/day up to 160 mg/day. The unit costs and full per cycle costs assuming the 160mg dosage are shown below in Table 29.

Table 29. Drug costs

Drug	Unit cost	Drug cost per 28-day cycle	Source
Regorafenib 160mg per day (without PAS)	£44.57/40mg tablet	£3,744.00	Bayer UK
Regorafenib 160mg per day (with PAS)	£[redacted]/40mg tablet	£[redacted]	Bayer UK

5.3.8.2 Treatment duration

Bayer's method of modelling treatment duration of regorafenib changed substantially from the time of their original report submission to the time of our report submission.

Regorafenib treatment in the regorafenib arm of GRID was continued after disease progression. However, originally, Bayer modelled regorafenib treatment only up to progression, as they claimed this would be as in clinical practice in England & Wales, citing surveys of physicians. Originally, Bayer also assumed a dose intensity of 84.1% during this period.

As reported above, Bayer originally used OS corresponding to the 2015 cut off. In response to our question for clarification, they then provided OS data corresponding to the 2017 cut off. We then asked whether they also had updated treatment duration corresponding to the 2017 cut off. In response, they completely changed their method of modelling treatment duration. In particular, they now model treatment with regorafenib for the entire duration as experienced in GRID RCT, see Figure 32 below.

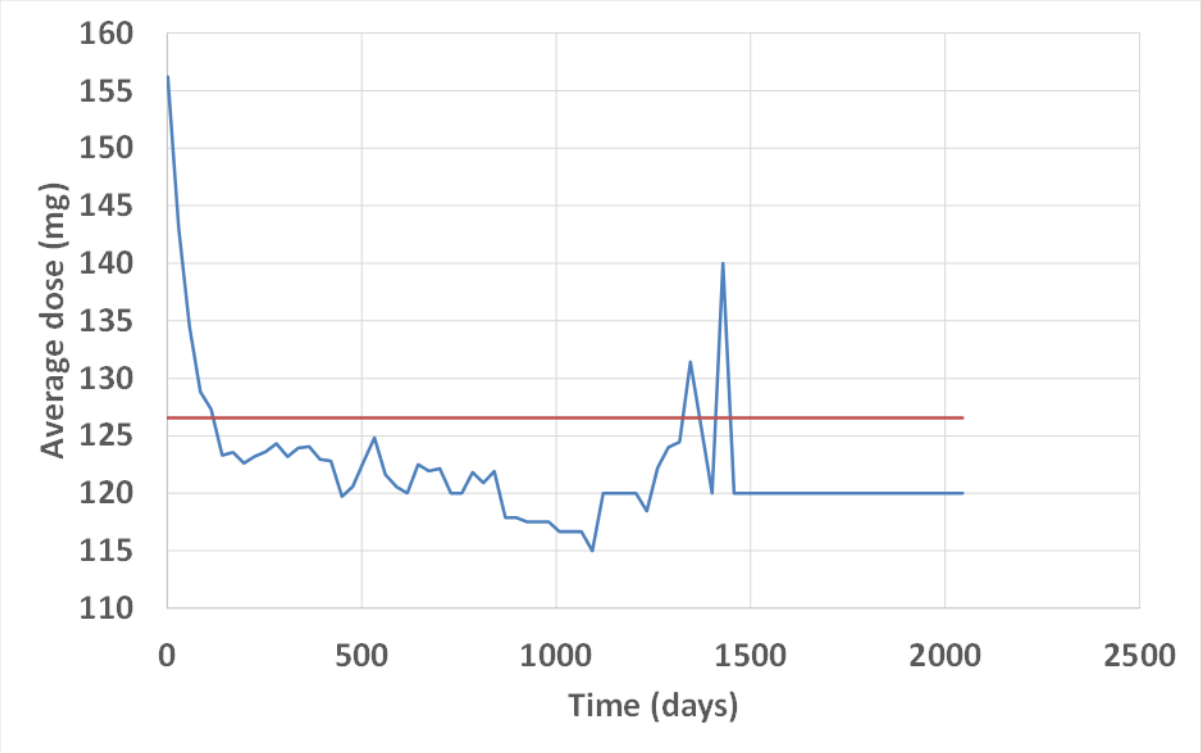
Figure 32. Time on regorafenib treatment in GRID RCT

We agree with this, their updated method of modelling treatment duration.

Bayer have also supplied different data for dose intensity of regorafenib (Figure 33). This is appropriate, because they now consider treatment with regorafenib over a different period, including post-progression. As explained above, previously, they assumed a dose intensity of 84%. By our calculations, the average dose, weighted for treatment duration is now 126.5mg, which gives a mean dose intensity of 79%. This dose intensity is implicit in Bayer's estimation of total cost of acquisition of regorafenib, and this is appropriate.

Notice that Bayer's updated method of modelling treatment duration acts to increase the ICER for regorafenib.

Figure 33. Average dose over time in GRID RCT



5.3.8.3 Health resources use and cost

Resource use information was gathered from a 2013 physician resource survey of 15 GIST medical oncologists with recent experience in managing GIST patients. Results were then updated and validated by two consultant oncologists in 2016 specialised in the management of metastatic or unresectable GIST. The physicians also showed that only 25.3% of patients receive TKIs post-progression, which informs Bayer's choice of only considering treatment costs during active progression.

The survey determined that the main tests for the patient population are CT scans, MRI scans, full blood counts and liver function tests. Table 30 below shows the proportion of patients taking each test *prior* to treatment. Fewer patients on BSC undergo tests compared to TKIs like regorafenib. The tests are continually performed, but CT and MRI scans are less common following disease progression, as shown in Table 31 and Table 32. Our clinical expert considers all estimates in Table 30, Table 31 and Table 32 reasonable, except that in the UK he estimates no MRI scans post-progression, and during PFS only for patients on a TKI, not BSC. Changing these values to those of our clinician increases the ICERs only marginally. Therefore, henceforth, we do not pursue this matter.

Table 30. Resource use prior to treatment

Test	Proportion of 3rd line patients receiving test prior to treatment with a TKI, Mean (SE)	Proportion of 3rd line patients receiving test prior to BSC, Mean (SE)
CT scan	0.85 (0.079)	0.24 (0.070)
MRI scan	0.12 (0.031)	0.01 (0.005)
Full blood count	0.92 (0.065)	0.56 (0.100)
Liver function test	0.92 (0.062)	0.49 (0.111)

Source: Bayer submission, Table 46, p.180

Table 31. Regular tests given to progression-free patients

Test	Patients on a TKI		Patients on BSC	
	Percentage of physicians responding that patients would be given the test regularly	**Average frequency (weeks between tests), Mean (SE)**	**Percentage of physicians responding that patients would be given the test regularly**	**Average frequency (weeks between tests), Mean (SE)**
CT scan	100%	12.1 (1.44)	60%	18.9 (3.26)
MRI scan	73%	19.9 (4.00)	27%	18.0 (2.58)
Full blood count	93%	6.4 (1.90)	67%	10.9 (2.36)
Liver function test	93%	6.4 (1.90)	60%	11.2 (2.61)

Source: Bayer submission, Table 47, p.180

Table 32. Regular tests given to patients in the post-progression state

Test	Percentage of physicians responding that patients would be given the test regularly, %	Average frequency (weeks between tests), Mean (SE)
CT scan	20%	14.5 (6.84)
MRI scan	7%	8.0 (-)
Full blood count	67%	8.8 (1.88)
Liver function test	60%	9.4 (2.03)

Source: Bayer submission, Table 48, p.180

All physicians consulted indicated that regular monitoring would be performed as an outpatient, as shown in Table 33. Our clinical expert estimates slightly different frequencies:

4 weeks between visits for patients on a TKI and 12 weeks for patients on BSC whilst in PFS and PD. Changing these values to those of our clinician increases the ICERs only marginally. Therefore, henceforth, we do not pursue this matter.

Table 33. Frequency of outpatient visits based on health state

Health state	Percentage of physicians responding that patients would be monitored on an outpatient basis	Average frequency (weeks between visits), Mean (SE)
Progression-free on a TKI	100%*	6.2 (0.86)
Progression-free on BSC	100%	7.9 (0.77)
Progressed disease on BSC	100%	6.9 (0.97)

Source: Bayer submission, Table 49, p.181

Average frequencies by health state for tests and monitoring were used to calculate per cycle (28) day probabilities.

Pain management medication is also common, and the physicians' responses to pain management usage are shown below in Table 34. Our clinical advisor considers these values reasonable. The physician survey also included the use of palliative surgical resection or palliative radiotherapy and indicated that this would not depend on whether a patient is on a TKI. These costs are shown below in Table 35. Our clinical advisor considers the data for radiotherapy reasonable, but consider the values for palliative surgical resection high. Instead, he advises proportions of 0.05 whilst in PFS (regardless of treatment) and 0.02 during progressed disease. Changing these values to those of our clinician increases the ICERs only marginally. Therefore, henceforth, we do not pursue this matter.

Table 34. Pain management resource use by health state

Treatment	Average proportion of patients treated with pain medication by health state and medicine	
	Progression-free Mean (SE)	**Progressed disease Mean (SE)**
Co-codamol, 2 tablets QDS (each containing 8mg codeine)	0.18 (0.039)	0.22 (0.043)
Tramadol, 100mg QDS	0.12 (0.028)	0.14 (0.036)
Paracetamol, 1g QDS	0.33 (0.074)	0.38 (0.085)
Morphine sulphate, 30mg immediate release every 4 hours	0.20 (0.057)	0.29 (0.065)
Dexamethasone, 4mg OD	0.11 (0.022)	0.19 (0.043)

Source: Bayer submission, Table 50, p.181

Table 35. Palliative care interventions by health state

Palliative intervention	Average proportion of patients who receive the palliative care intervention, Mean (SE)		
	Progression-free on a TKI	**Progression-free on BSC with no TKI**	**Progressed disease**
Palliative surgical resection	0.10 (0.024)	0.10 (0.031)	[redacted] (0.033)
Palliative radiotherapy	0.20 (0.053)	0.20 (0.061)	[redacted] (0.063)

Source: Bayer submission, Table 51, p.182

Full unit costs are given in Table 36 and full input costs per cycle associated with the intervention and the comparator in Table 37.

Table 36. Unit costs associated with health state resource use

Item	Cost (£)	Source	Assumption
Regular tests			
CT scan	40.23	NHS Reference costs 2015-16	Cost per scan (IMAG); code RD26Z - Computerised Tomography Scan of three areas, with contrast;
MRI scan	146.61	NHS Reference costs 2015-16	Cost per scan *(weighted average of all MRI – adult; codes: RD01A,*

			RD02A, RD03Z,RD04Z,RD05Z,RD06Z,RD07 Z)
Full blood count	3.10	NHS Reference costs 2015-16	Cost per test (DAPS); code DAPS05 - Haematology
Liver function test	1.18	NHS Reference costs 2015-16	Cost per test (DAPS); code DAPS04 - Clinical Biochemistry
Regular monitoring visit			
Outpatient visit (regular monitoring)	93.00	2016/17 National Tariff; OP	Cost of outpatient attendance Attendances - code 370 WF01A Follow Up Attendance - Single Professional
Pain management			
Co-codamol	0.89	MIMS, January 2017	Cost per 30-tab pack (non-proprietary), 8mg codeine phosphate per tab
Tramadol	2.87	MIMS, January 2017	Cost per 100-cap pack, 50mg per cap (non-proprietary)
Paracetamol	2.19	MIMS, January 2017	Cost per 100-tab pack, 500mg per tab (non-proprietary)
Morphine sulphate immediate release	5.31	MIMS, January 2017	Cost per 56-tab pack, 10mg per tab (Sevredol®)
Dexamethasone	42.85	MIMS, January 2017	Cost per 50-tab pack, 2mg per tab (non-proprietary)
Palliative care			
Palliative surgical resection	3,943.21	NHS Reference costs 2015-16	Single intervention for malignant GI Tract disorder *(weighted average; code: FZ92D, FZ92E, FZ92F)*
Palliative radiotherapy-	160.59	NHS Reference costs 2015-16	Cost per medical specialist palliative care attendance *(weighted average adult; code: SD01A, SD02A, SD03A, SD04A)*

Source: Bayer submission, Table 52, p.182

Table 37. Input costs per cycle in the economic model

Item	Regorafenib mean (CI)	Reference in submission	BSC mean (CI)	Reference in submission
Drug costs[§]	£3,271.09 (£2,616.87; £3,925.30)	Section 5.5.1	-	-
Management costs				
One-time costs for regorafenib	£55.72 (£44.58; £66.86)	Section 5.5.1	-	-
One-time costs post-progression	£[redacted] (£[redacted]; £[redacted])	Section 5.5.1	£[redacted] (£[redacted]; £[redacted])	Section 5.5.1
Regorafenib + BSC while progression-free	£124.21 (£99.37; £149.05)	Section 5.5.1	-	-
BSC while progression-free	-	-	£80.07 (£64.05; £96.08)	Section 5.5.1
BSC post-progression	£88.98 (£71.18; £106.78)	Section 5.5.1	£88.98 (£71.18; £106.78)	Section 5.5.1
End of life costs	£8,736.53 (£8,052.12; £9,422.00)	Section 5.5.8	£8,736.53 (£8,052.12; £9,422.00)	Section 5.5.8
Additional one-time costs for BSC	-	-	£13.82 (£11.05; £16.58)	Section 5.5.1
Adverse Events Costs				
Hand foot skin reaction	£0.00	Section 5.5.7	£0.00	Section 5.5.7
Diarrhoea	£7.02 (£5.62; £8.43)	Section 5.5.7	£7.02 (£5.62; £8.43)	Section 5.5.7

Hypertension	£11.86 (£9.48; £14.23)	Section 5.5.7	£11.86 (£9.48; £14.23)	Section 5.5.7

Source: Bayer submission, Table 53, p.185

5.3.8.4 Health state costs and resource use

Health state costs comprise one-time costs and per cycle costs, summarised in Table 38. The one-time costs consist of test costs prior to starting treatment, palliative surgical resection, and palliative radiotherapy. Palliative measures are only applied in the progressed disease state since resource use is zero for PFS regardless of treatment arm. One-off costs were estimated by unit cost of each test weighted by the proportion of patients undergoing each test/palliative measure, and then summed to get an expected one-off cost. Bayer made a minor error in modelling in their estimation of the number of new progressions in each cycle to which to apply the one-time costs. However, given that this error has a negligible effect on the ICERs per QALY, we pursue this no further.

Per-cycle costs consist of regular outpatient monitoring visits, regular tests and medication for pain management. Unit costs were weighted by the probabilities per cycle (see section 5.3.8.2). Standard errors are calculated assuming independence of variables – although this is unlikely, Bayer argue this results in larger standard errors and is a more conservative approach.

Table 38. Health state costs per cycle and one-off costs in the model

Cost component		Progression-free state on a TKI (£), Mean (SE)	Progression-free state on BSC with no TKI (£), Mean (SE)	Progressed disease (£), Mean (SE)
One-time costs	Tests	55.72 (5.53)	13.82 (2.93)	N/A
	Palliative resection	Not included	Not included	[redacted] (129.38)
	Palliative radiotherapy	Not included	Not included	[redacted] (10.11)
	Total one-time costs	55.72 (5.53)	13.82 (2.93)	[redacted] **(129.77)**
Regular per cycle costs	Regular tests	45.45 (5.46)	14.81 (4.08)	8.35 (36.00)
	Regular outpatient monitoring visits	60.49 (9.16)	46.91 (4.73)	53.68 (8.15)
	Pain management	18.27 (2.97)	18.35 (2.97)	26.95 (3.77)
	Total per cycle costs	124.21 (11.07)	80.07 (6.92)	**88.98 (37.11)**

Source: Bayer submission, Table 54, p.186

5.3.8.5 Adverse event costs and end-of-life criteria costs

Grade 3 and 4 adverse events were considered only if they were reported in at least 3% of patients and were: hand-foot skin reaction (HFRS), diarrhoea and hypertension. Bayer UK provides a free HFSR treatment kit and hence associate this AE with zero cost in the model.

Diarrhoea is treated with the drug loperamide. Hypertension is associated with an ACE inhibitor, and Bayer use the most common one according to their physician study, rampiril 10mg. Hypertension is also associated with 2 annual GP visits and two annual district nurse appointments. Treatments costs for both AEs and incidence rates are summarised in Table 39 to Table 42. We consider these values reasonable.

Table 39. AE incidence rates per cycle in the model

Adverse Event (Grade 3-4)	Estimated incidence rate per cycle (%)	
	Placebo	Regorafenib
Hypertension	1.35	5.16
Hand-foot skin reaction	0	4.25
Diarrhoea	0	1.07

Source: Bayer submission, Table 35, p. 162

Table 40. Diarrhoea drug treatment costs

Drug	Loperamide
Cost per pack	£2.15
No. tabs per pack	30.00
mg per tab	2.00
Cost per mg	£0.04
Average daily dose (mg)	7.00
Average weekly dose (mg)	49.00
Cost per cycle	**£7.02**

Source: Bayer submission, Table 55, p.187

Table 41. Hypertension drug treatment costs

Drug	Ramipril
Cost per pack*	£1.24
No. tabs per pack	28.00
mg per tab	10.00
Cost per mg	£0.004
Average daily dose (mg)	10.00
Average weekly dose (mg)	70.00
Cost per cycle	**£1.24**

Source: Bayer submission, Table 56, p.188

Table 42. Hypertension management costs

GP visit	£44	PSSRU Unit costs of Health & Social Care 2015, pg. 177 - Table 10.8b (62)
District nurse visit	£25	PSSRU Unit costs of Health & Social Care 2015, pg. 175 - Table 10.7 (62)

Source: Bayer submission, Table 57, p.188

End of life costs were taken from the study conducted by Abel et al [24], a UK hospice-based study. Costs were inflated to 2015/2016 level. The final EoL cost used is £8,736. Finally, Table 43 gives a complete summary of per-cycle variable costs and non-cost parameters.

Table 43. Summary of variables applied in the economic model (per cycle)

Variable	Value	Measurement of uncertainty and distribution: CI (distribution)	Reference to section in submission
Regorafenib cost	£[redacted]	(£[redacted]-£[redacted])	Table 45
One-time costs for regorafenib	£56	(£45-£67)	Table 54
Regorafenib + BSC while progression-free	£124	(£99-£149)	Table 54
BSC while progression-free	£80	(£64-£96)	Table 54
BSC post-progression	£89	(£71-£107)	Table 54
End of life costs	£8,736	(£8,052-£9,422)	Table 58
Diarrhoea costs	£7	(£6-£8)	Table 55
Hypertension costs	£12	(£9-£14)	Table 56
Progression-free state utility	0.767	(0.718-0.816)	Table 36
Post-progression state utility	0.647	(0.571-0.723)	Table 36
Discount rate (costs)	3.5%	(0-6%)	Table 29
Discount rate (benefits)	3.5%	(0-6%)	Table 29

Source: Bayer submission, Table 60, p.190

5.3.9 Cost-effectiveness results

Bayer's base case ICERs of regorafenib plus BSC compared to BSC alone are £[redacted]/QALY and £37,941/QALY without and with the PAS respectively. Table 44 and Table 45 below illustrate the base case results. Bayer present their base case as using the 2017 data cut off.

Table 44. Base case CE results. 2017 cut-off (no PAS)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (LYs)	ICER (£) incremental (QALYs)
Placebo + BSC	10,395	1.154	0.761					
Regorafenib	[redacted]	2.546	1.733					
				[redacted]	1.393	0.971	[redacted]	[redacted]

ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Source: Bayer response to clarification, Table 36, p. 47

Table 45. Base case CE results. 2017 cut-off (PAS)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (LYs)	ICER (£) incremental (QALYs)
Placebo + BSC	10,395	1.154	0.761					
Regorafenib	47,249	2.546	1.733					
				36,854	1.393	0.971	26,465	37,941

ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Source: Bayer response to clarification, Table 37, p. 47

Bayer argue that their model accurately reflects the trial data (Table 46).

Table 46. Summary of model results versus clinical data (2015 cut-off)

Outcome	Time horizon	Placebo + BSC		Regorafenib + BSC	
		Clinical trial result	Model result	Clinical trial result	Model result
Overall survival	1 year	0.38	0.42	0.65	0.66
	2 years	0.19	0.20	0.35	0.39
	3 years	0.15	0.12	0.26	0.26
Progression-free survival	168 days	n/a	n/a	0.43	0.44

5.3.9.1 Disaggregated base case cost-effectiveness results

Bayer provide disaggregated results for QALYs and predicted resource use (without and with the PAS) for the 2015 data cut-off, but not the 2017 cut-off. Given that the focus of our attention is now the 2017 data, we have recreated the disaggregated results from the Bayer model using the updated 2017 data, which are shown below.

Table 47. Summary of Bayer base case QALYs by health state, 2017 cut-off

Health state	QALY Regorafenib	QALY Placebo	Increment	Absolute increment	% absolute increment
Progression-free	0.566	0.095	0.471	0.471	40%
Post Progression	1.433	0.727	0.706	0.706	60%
Total	1.999	0.822	1.177	Total absolute increment	100%

QALY, quality-adjusted life year

*QALYs are undiscounted in line with the Bayer submission for 2015 results.

Table 48. Breakdown of Bayer base case costs, 2017 cut-off

	Component	Regorafenib + BSC	Placebo + BSC	Incremental
List price	Drug costs - progression-free	£[redacted]	£0	£[redacted]
	Drug costs - post-progression	£[redacted]	£0	£[redacted]
	Additional one-time cost post-progression	£[redacted]	£472	-£[redacted]
	Adverse event costs	£[redacted]	£3	£[redacted]
	Monitoring costs	£[redacted]	£1,418	£[redacted]
	End-of-life costs	£[redacted]	£8,503	-£[redacted]
	Total cost	**£[redacted]**	**£10,395**	**£[redacted]**
PAS price	Drug costs - progression-free	£[redacted]	£0	£[redacted]
	Drug costs - post-progression	£[redacted]	£0	£[redacted]
	Additional one-time cost post-progression	£[redacted]	£472	-£[redacted]
	Adverse event costs	£[redacted]	£3	£[redacted]
	Monitoring costs	£[redacted]	£1,418	£[redacted]
	End-of-life costs	£[redacted]	£8,503	-£[redacted]
	Total cost breakdown	**£47,249**	**£10,395**	**£36,854**

Source: Bayer response to clarification, Table 38, p.49

5.3.10 Sensitivity analyses

Bayer carried out both one-way sensitivity analyses (OWSA) and probabilistic sensitivity analyses to explore the effect of parameter uncertainty. Scenario analyses were also performed to explore the effects of assumptions in the model.

5.3.10.1 Probabilistic sensitivity analyses

Probabilistic sensitivity analysis (PSA) is a method of allowing all model parameters which are uncertain to vary simultaneously (for example, the exact HSUV for each state may be uncertain, but the list price of the drug is set by the company and is certain). Uncertain parameters were given suitable parametric distributions and repeatedly sampled 3,000 times and the ICERs recorded for each simulation. The probability of HFSR and diarrhoea were not varied in the PSA as there were 0 events in the GRID study making standard errors difficult to estimate. These probabilities were examined in the OWSA, but were found to have negligible effects on the ICERs per QALY. Table 49 shows the average of the simulated ICERs per QALY.

The base case PSA ICERs were £[redacted]/QALY without PAS and £38,494 with PAS. Results from the Monte Carlo simulations were also plotted in the (incremental cost QALY) space shown in Figure 34 and Figure 35 without and with PAS. The proportion of simulations which fall below the willingness-to-pay threshold (dotted line) gives the probability of the treatment being cost-effective. At a threshold of £50,000, regorafenib had a [redacted]% chance to be cost-effective without the PAS, and an 82% chance with the PAS.

Table 49. Average PSA ICER results. 2017 cut-off (with and without PAS)

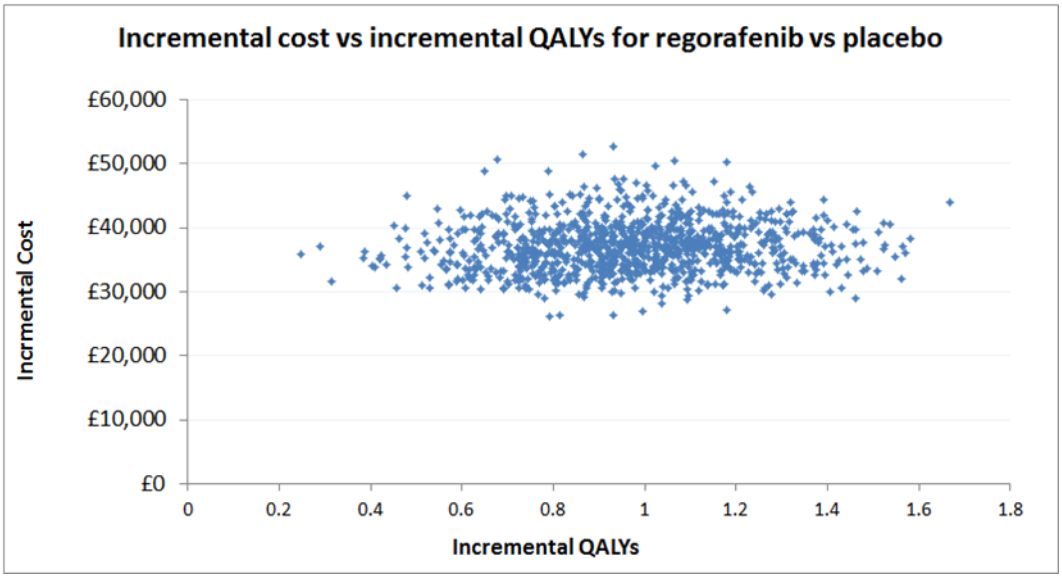
	Regorafenib + BSC			Placebo + BSC			Incremental			ICER (£/QALY)
	LYs	**QALYs**	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	**Costs**	
List Price	2.560	1.741	£[redacted]	1.178	0.776	£11,016	1.382	0.965	£[redacted]	£[redacted]
PAS price	2.533	1.745	£48,152	1.183	0.780	£11,021	1.380	0.965	£37,130	£38,494

Source: Bayer response to clarification, Table 39, p.47

Figure 34. PSA simulation results (no PAS)

Source: Bayer response to clarification, Figure 12, p. 52

Figure 35. PSA simulation results (with PAS)



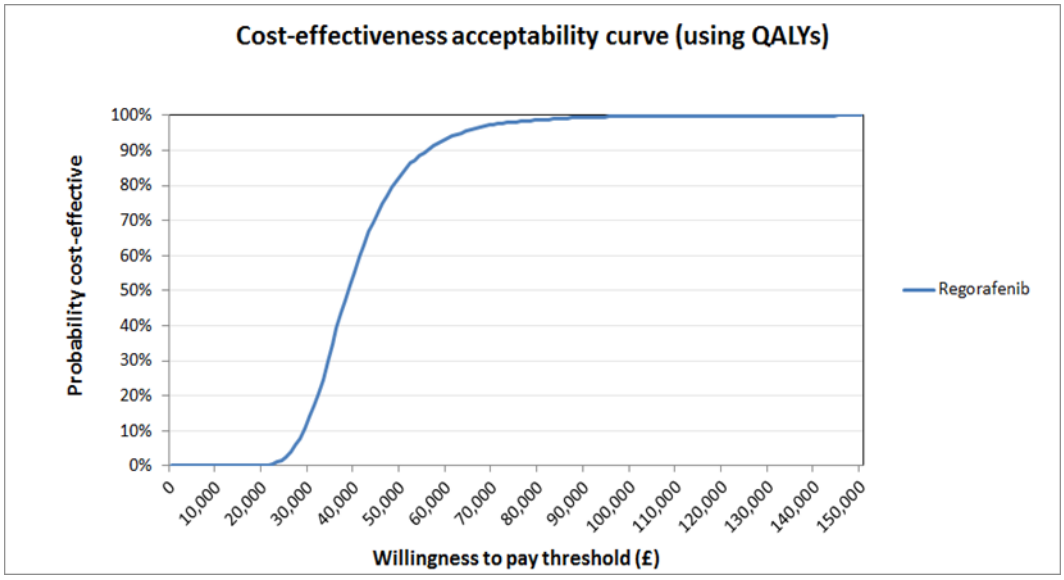
Source: Bayer response to clarification, Figure 14, p. 51

Cost-effectiveness acceptability curves (CEACs) show the probability of the treatment being cost-effective whilst varying the willingness to pay. CEACs without and with the PAS are shown below in Figure 36 and Figure 37.

Figure 36. Bayer base case CEAC (no PAS)

Source: Bayer addendum, Figure 13, p. 51

Figure 37. Bayer base case CEAC (with PAS)



Source: Bayer addendum, Figure 15, p. 52

5.3.10.2 Deterministic sensitivity analyses

The deterministic analyses carried out by Bayer are one-way sensitivity analyses. The input variables and their ranges are displayed below in Table 50, and the tornado diagrams for the top 15 drivers of the ICERs per QALY without and with the PAS are shown in Figure 38 and Figure 39. These variations resulted in ICERs per QALY varying between £[redacted]-£[redacted] at list price and £30,660-£45,222 with the PAS. See Tables 75 and 76 of Bayer's report for a full list of effects.

Table 50. Parameters varied in one-way sensitivity analyses

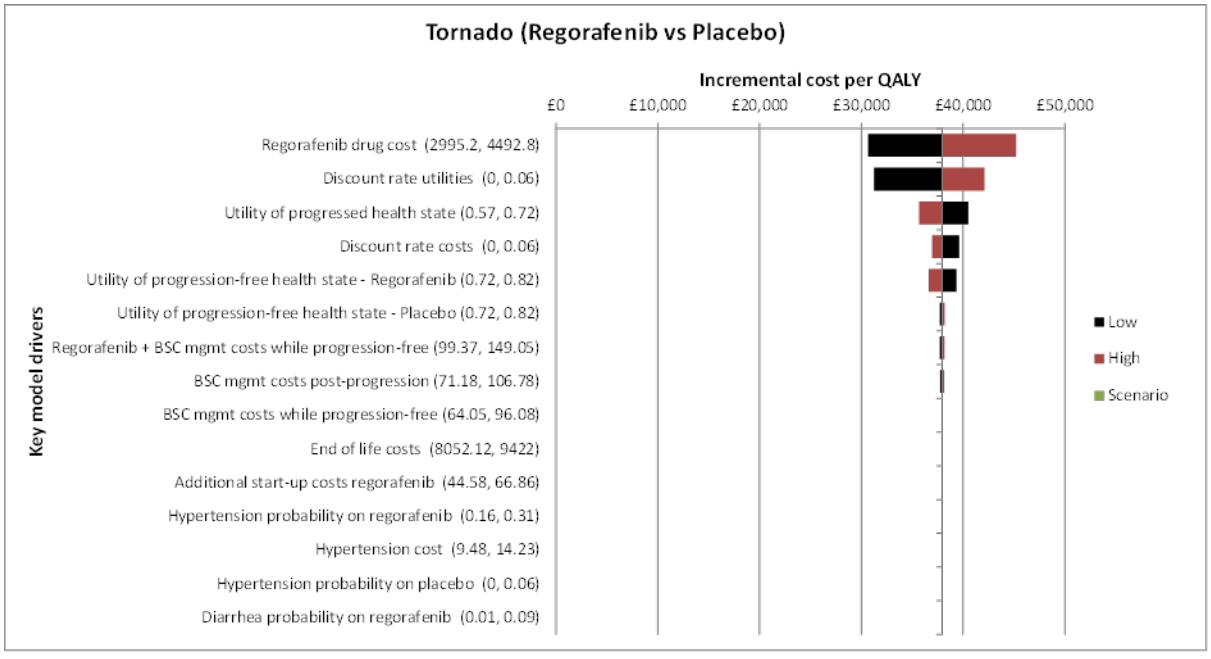
Variable	Input values used in OWSA		Source
	Lower input	**Upper input**	
Discount rate costs	0.00	0.06	Assumption
Discount rate utilities	0.00	0.06	Assumption
Additional one-time costs regorafenib	£44.58	£66.86	± 20% base case value
Regorafenib + BSC management costs while progression-free	£99.37	£149.05	± 20% base case value
BSC management costs while progression-free	£64.05	£96.08	± 20% base case value
BSC management costs post-progression	£71.18	£106.78	± 20% base case value
End of life costs	£8,052.12	£9,422	Abel et al (63)
Diarrhoea cost	£5.62	£8.43	± 20% base case value
Hypertension cost	£9.48	£14.23	± 20% base case value
HFSR probability on regorafenib	0.13	0.26	Base case ± 2 SE
Diarrhoea probability on regorafenib	0.01	0.09	Base case ± 2 SE
Hypertension probability on regorafenib	0.16	0.31	Base case ± 2 SE
Hypertension probability on placebo	0.00	0.06	Base case ± 2 SE
Utility of progression-free health state - Regorafenib	0.72	0.82	Base case ± 2 SE
Utility of progression-free health state - Placebo	0.72	0.82	Base case ± 2 SE
Utility of progressed health state	0.57	0.72	Base case ± 2 SE

Source: Bayer submission, Table 74, p. 208

Figure 38. Tornado diagram of top 15 model drivers, 2017 cut-off (no PAS)

Source: Bayer response to clarification, Figure 16, p. 57

Figure 39. Tornado diagram of top 15 model drivers, 2017 cut-off (with PAS)



Source: Bayer response to clarification, Figure 17, p. 57

5.3.10.3 Scenario analyses

Scenario analyses are designed to explore uncertainty around the structural assumptions of the model (see Table 18 in section 5.3.2 for a list of Bayer's assumptions). All scenario

analyses are reported here using the 2017 data cut-off. Bayer discuss 6 scenario analyses in their submission:

- OS extrapolation using Weibull and Gompertz curves;
- Using RPSFT crossover adjustment instead of IPE with loglogistic, Weibull and Gompertz curves;
- Changing resource use from the physician survey in line with clinical consultants' opinions;
- Costing for regorafenib post-progression in the regorafenib + BSC arm;
- Using repeated measures EQ-5D utility estimates over paired-samples;
- Using EORTC from GRID to derive utility estimates.

Bayer examined the effect of using Weibull and Gompertz parametric functions for OS data (for both placebo and regorafenib), although they argue that the statistical fit is worse than their base case (loglogistic). The Weibull curve caused both QALYs and costs to decrease, and the ICERs increased substantially: £[redacted]/£45,498 per QALY without and with PAS respectively.

The Gompertz model decreased both QALYs and incremental costs, and an increased ICERs of £[redacted]/£47,068 per QALY without and with the PAS respectively. Bayer also investigated the effect of using log-normal and exponential models, but argue that the effect was negligible and did not report results. We agree that the ICERs change only marginally when assuming log-normal distributions. However, using exponential distributions, we find that the ICERs increase substantially, to £[redacted] and £44,827 per QALY without and with the PAS respectively.

Bayer explore the effect of using the RPSFT method of crossover correction along with the loglogistic, Weibull and Gompertz parametric models for OS. The loglogistic case is still favoured using their AIC criterion. The resulting ICERs were:

- Loglogistic (no PAS/PAS): £[redacted]/£39,493 per QALY. These values are slightly higher than Bayer's base case, which use the IPE adjustment method: £[redacted]/ £37,941per QALY).
- Weibull(no PAS/PAS): £[redacted]/£46,996 per QALY
- Gompertz (no PAS/PAS): £[redacted]/£48,360 per QALY.

Bayer examine the effect of updating their resource use data from their physician survey in 2013 with suggestions from their clinical experts. The suggestions were:

- All patients should receive either a CT or a MRI scan prior to starting treatment.
- For progression-free patients on a TKI (i.e. regorafenib) a CT scan would be admitted about every 12 weeks.
- A lower frequency of outpatient visits (progression-free TKI patients from 6.2 weeks to 12 weeks, for BSC progression-free patients from 6.9 to 8-12 weeks).
- Reducing the proportion of progressed patients receiving either palliative resection or radiotherapy by 5%.

These changes resulted in the ICERs decreasing only very slightly: from £[redacted] / £37,941 to £[redacted]/£37,806 per QALY without and with the PAS respectively.

Bayer examine the use of utilities from repeated measures comparison (see Section 5.3.7, p95). Bayer maintain that this method is likely to be less reliable than paired-samples due to a heterogeneous progressed patient population. The ICERs decreased only very slightly, from £[redacted]/ £37,941 to £[redacted]/£36,765 per QALY without and with the PAS respectively.

Bayer also examine the use of the lower utility estimates from the GRID RCT for regorafenib in the PFS state to possibly account for disutility from AEs (see Section 5.3.7, p95). The resulting ICERs decrease only very slightly, to £[redacted]/£37,514 per QALY without and with the PAS respectively.

Finally, Bayer use utility values from the EORTC GRID data, using both repeated measures and paired-samples comparisons. The resulting ICERs were:

- Repeated measures (no PAS/PAS): £[redacted]/£34,281per QALY.
- Paired-samples (no PAS/PAS): £[redacted]/£33,964per QALY.

We agree with Bayer that these values are less relevant than those that underpin their base case, as the EQ-5D is the preferred instrument to measure health-related quality of life.

A summary of ICERs from the scenario analyses is presented below in Table 51.

Table 51. Summary of Bayer scenario analysis ICERs

Scenario analysis	ICER (list price)	ICER (PAS)
Bayer base case	[redacted]	£37,941
Weibull OS curve	[redacted]	£45,498
Gompertz OS curve	[redacted]	£47,068
RPSFT (loglogistic)	[redacted]	£39,493
RPSFT (Weibull)	[redacted]	£46,996
RPSFT (Gompertz)	[redacted]	£48,360
Resource use	[redacted]	£37,806
EQ-5D Repeated measures utility values	[redacted]	£36,765
EQ-5D repeated measures utility values by treatment arm (pre-progression)	[redacted]	£37,514
EORTC utility values (repeated measures)	[redacted]	£34,281
EORTC utility values (paired-samples)	[redacted]	£33,964

Bayer also initially submitted data from the 2015 cut-off. They do not present this as a scenario analysis in their updated report, but we consider it appropriate to present the cost-effectiveness results if only these data were available as a scenario analysis. Table 51 and Table 52 show the ICER per QALY without and with the PAS respectively. The ICERs per QALY were £[redacted]/£34,476 without and with the PAS respectively. This was also based on a different drug acquisition cost methodology; assuming that regorafenib was taken solely in PFS, which is superseded by their 2017 base case method of directly using the GRID data.

Table 52. Base case CE results. 2015 cut-off (no PAS)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (LYs)	ICER (£) incremental (QALYs)
Placebo + BSC	10,671	1.474	0.969					
Regorafenib	[redacted]	2.521	1.717					
				[redacted]	1.047	0.748	[redacted]	[redacted]

ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Source: Bayer submission, Table 62, p. 193

Table 53. Base case CE results 2015 cut-off (with PAS)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (LYs)	ICER (£) incremental (QALYs)
Placebo + BSC	10,671	1.474	0.969					
Regorafenib	36,457	2.521	1.717					
				25,786	1.047	0.748	24,623	34,476

ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Source: Bayer submission, Table 63, p. 193

5.3.11 Model validation and face validity check

Bayer described their validation checks as follows: (source: Bayer submission, p.233):

> *In the course of model development an independent health economic expert, familiar with oncology modelling was consulted. The health economic expert agreed that the modelling approach including the crossover adjustment methods was reasonable and proposed no major changes.*
>
> *A check of validity was performed by the model developers using a quality control process, and a model audit which was performed by an experienced health economist external to the team who built the model. This involved calculation spot checks, cross checks against source data and extreme value scenarios to check if the model behaved logically.*

We consider these checks appropriate.

Bayer describe their clinical validation as follows: (source: Bayer submission, p.233)

The two clinical experts were asked to validate the model inputs and model assumptions. The key points raised by the clinical experts were explored in the scenario analysis. The key points raised were:

- Gompertz and Weibull functions should be explored to reflect alternative long term OS predictions (explored in scenario analyses).

- Some of the resource use assumptions taken from the physician survey conducted in 2013 do not reflect current/best practice. More plausible resource use assumptions should be explored (explored in scenario analyses).

- For patients who progress from BSC to regorafenib the common treatment effect is clinically plausible given the quick progression of patients on the BSC arm (median PFS = 0.9 months).

6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

In this section we derive the PenTAG base case (Table 54 below). The impacts of the individual components of our base case on cost-effectiveness are shown, as well as selected combinations of components and finally the base case, which is composed of all relevant components applied simultaneously.

Total uncertainty in the cost-effectiveness of regorafenib versus BSC is high due to:

- Substantial uncertainty in the adjustment for widespread treatment switching on diseases progression, from BSC to regorafenib.
- Important uncertainty in the extrapolation of OS.

Table 54. Derivation of PenTAG base case ICERs Regorafenib vs. BSC (£ per QALY)

				Regorafenib price	
				PAS	List
	Bayer base case			**£38,000**	[redacted]
	PenTAG assumption	**Bayer assumption**			
1	OS from 2015 data-cut	OS from 2017 data-cut	(Section 5.3.6.2, p79)	£49,000	[redacted]
2	Include general mortality from UK population	Do not including general mortality from UK population	(Section 5.3.6.3, p87)	£41,000	[redacted]
3	OS average of Log-logistic / Weibull	OS average Log-logistic	(Section 5.3.6.3, p87)	£41,000	[redacted]
4	Utilities decrease with age	Utilities independent of age	(Section 5.3.7, p95)	£39,000	[redacted]
1 + 2				£52,000	[redacted]
1 + 3				£52,000	[redacted]
2 + 3				£43,000	[redacted]
1 + 2 + 3				£55,000	[redacted]
		ICER		**£56,000**	[redacted]
1+2+3+4	**PenTAG base case**	**Uncertainty**		**High, mostly due to switching adjustment, but also extrapolation.**	

Key: ICER = incremental cost-effectiveness ratio; OS = overall survival; PAS = patient access scheme; PFS= progression-free survival; QALYs = quality-adjusted life year(s);

Dark shading indicates ICER > £50,000 per QALY.

6.1 Key sensitivity analyses applied to PenTAG and Bayer base case

Here, we present key scenario analyses applied separately to the PenTAG and Bayer base cases. These scenarios were chosen either because they demonstrate key messages, e.g. the impact of adjustment for treatment switching (ITT analysis), or because they represent plausible alternatives (all other analyses).

Table 55. ICERs (£/QALY) for Regorafenib vs. BSC given important scenario analyses applied to Bayer base case

		PAS	List
Bayer base case		**£38,000**	[redacted]
ITT analysis	(Section 5.3.6.2, p79)	£149,000	[redacted]
Model costs and QALYs only up to progression	(Section 5.3.6.2, p79)	£52,000	[redacted]
OS from 2017 data cut	(Section 5.3.6.2, p79)	unchanged	unchanged
RPSFTM method (IPE method Bayer base case)	(Section 5.3.6.2, p79)	£39,000	[redacted]
Weibull distribution for OS (log-logistic base case)	(Section 5.3.6.3, p87)	£45,000	[redacted]
Gompertz distribution for OS (log-logistic base case)	(Section 5.3.6.3, p87)	£47,000	[redacted]

Key; PAS = patient access scheme; QALY, quality-adjusted life year
Dark shading indicates ICER > £50,000 per QALY.

Table 56. ICERs (£/QALY) for Regorafenib vs. BSC given important scenario analyses applied to PenTAG base case

		PAS	List
PenTAG base case		**£56,000**	[redacted]
ITT analysis	(Section 5.3.6.2, p79)	£235,000	[redacted]
Model costs and QALYs only up to progression	(Section 5.3.6.2, p79)	£51,000	[redacted]
OS from 2017 data cut	(Section 5.3.6.2, p79)	£44,000	[redacted]
RPSFTM method (IPE method Bayer base case)	(Section 5.3.6.2, p79)	£64,000	[redacted]
Weibull distribution for OS (log-logistic base case)	(Section 5.3.6.3, p87)	£59,000	[redacted]
Gompertz distribution for OS (log-logistic base case)	(Section 5.3.6.3, p87)	£64,000	[redacted]

Key; PAS = patient access scheme; QALY, quality-adjusted life year
Dark shading indicates ICER > £50,000 per QALY.

7 End of life

Bayer argues that their evidence supports inclusion into NICE's End of Life category; that the life expectancy for the patient population is under 24 months with the comparator and that there is sufficient evidence that regorafenib adds at least 3 months additional survival. They cite results from both their economic model and the GRID study. Bayer's model predicts a median OS for patients treated with BSC of about 7.5 months using the 2017 data cut, regardless of whether the crossover correction method is IPE or RPSFT. Using the 2015 data cut, median OS on BSC is between 11.1-11.9 months, depending on whether the crossover correction method is IPE or RPSFT respectively. Bayer also cite the crossover corrected median OS improvement from GRID for regorafenib to be at least 5.5 months, depending on whether IPE or RPSFT methods are used.

Under Bayer's base case, the mean OS for BSC, adjusted for treatment switching, is 1.25 years, substantially below the threshold of 2 years to quality for End of Life. Under our base case, mean OS on BSC is 1.37 years, again, clearly meeting the criterion.

Under Bayer's base case, the mean gain in OS for regorafenib over BSC, adjusted for treatment switching, is 20.5 months, substantially greater than the threshold of 3 months to quality for End of Life. Under our base case, mean OS benefit is 12.5 months, again, clearly meeting the criterion.

Considering all this, we agree with Bayer that regorafenib meets the End of Life criteria.

Based on the ITT data, i.e. without adjustment for treatment switching, Bayer estimate a mean survival benefit of regorafenib over BSC of just 1.4 months, clearly less than the 3 month threshold. So under the ITT analysis, regorafenib would not meet the End of Life criteria.

References

1. National Institute for Health and Care Excellence. Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours: Final scope. 2016.

2. Cancer Research UK. Gastrointestinal stromal tumour (GIST) 2015 [cited 2017 30/05/2017]. Available from: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/cancer-questions/what-is-the-treatment-for-gist-gastrointestinal-stromal-tumour.

3. Behazin NS. Gastrointestinal Stromal Tumors: Medscape; 2017 [31/05/2017]. Available from: http://emedicine.medscape.com/article/179669-overview.

4. Amelio JMS, Ruzafa JC, Desai K, Tzivelekis S, Muston D, Khalid JM, et al. Prevalence of gastrointestinal stromal tumour (GIST) in the United Kingdom at different therapeutic lines: an epidemiologic model. BMC cancer. 2014;14(1):364.

5. Sarcoma UK. Gastrointestinal stromal tumours (GIST) 2016 [cited 2017 12/04/2017]. Available from: https://sarcoma.org.uk/sarcoma-types/gastrointestinal-stromal-tumours-gist.

6. Cancer.Net. Gastrointestinal Stromal Tumor - GIST: Risk Factors 2015 [cited 2017 12/04/2017]. Available from: http://www.cancer.net/cancer-types/gastrointestinal-stromal-tumor-gist/risk-factors.

7. Lamba G, Gupta R, Lee B, Ambrale S, Liu D. Current management and prognostic features for gastrointestinal stromal tumor (GIST). Experimental hematology & oncology. 2012;1(1):14.

8. American Cancer Society. Treatment Choices for Gastrointestinal Stromal Tumor Based on Tumor Spread 2016 [cited 2017 15/05/2017]. Available from: https://www.cancer.org/cancer/gastrointestinal-stromal-tumor/treating/by-spread.html.

9. National Cancer Institute. Gastrointestinal Stromal Tumors Treatment (PDQ®)–Patient Version 2015 [cited 2017 16/05/2017]. Available from: https://www.cancer.gov/types/soft-tissue-sarcoma/patient/gist-treatment-pdq#section/_52.

10. Bond M, Hoyle M, Moxham T, Napier M, Anderson R. Sunitinib for the treatment of gastrointestinal stromal tumours: a critique of the submission from Pfizer 2009.

11. Demetri GD, Reichardt P, Kang Y-K, Blay J-Y, Rutkowski P, Gelderblom H, et al. Efficacy and safety of regorafenib for advanced gastrointestinal stromal tumours after failure of imatinib and sunitinib (GRID): an international, multicentre, randomised, placebo-controlled, phase 3 trial. The Lancet. 2013;381(9863):295-302.

12. Wilhelm SM, Dumas J, Adnane L, Lynch M, Carter CA, Schütz G, et al. Regorafenib (BAY 73-4506): A new oral multikinase inhibitor of angiogenic, stromal and oncogenic

receptor tyrosine kinases with potent preclinical antitumor activity. International Journal of Cancer. 2011;129(1):245-55.

13. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013.

14. Sanz-Granda Á, Hidalgo-Figueruela F, Granell M. Estimation of the treshold price of regorafenib in the treatment of unresectable and/or metastatic gastrointestinal stromal tumors after failure on imatinib and sunitinib in spain: cost-utility analysis. Value in health. 2015;18(7):a464.

15. Pitcher A, Grabbi E, Madin-Warburton, M. Vadgama S. Cost-effectiveness analysis of regorafenib in gastrointestinal stromal tumours in England using crossover adjustment methods. Value in Health. 2016;19(7):A741.

16. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. In: Excellence NIfHaC, editor. London2014.

17. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Med Res Methodol. 2011 Jan 11;11:4. PubMed PMID: 21223539. Pubmed Central PMCID: PMC3024998.

18. White I, Walker S, Babiker A. strbee: Randomization-based efficacy estimator. . The Stata Journal. 2002;2:140-50.

19. Yoon DH, Ryu MH, Ryoo BY, Beck M, Choi DR, Cho Y, et al. Sunitinib as a second-line therapy for advanced GISTs after failure of imatinib: relationship between efficacy and tumor genotype in Korean patients. Invest New Drugs. 2012 Apr;30(2):819-27. PubMed PMID: 21104107. English.

20. Reichardt P, Kang YK, Rutkowski P, Schuette J, Rosen LS, Seddon B, et al. Clinical outcomes of patients with advanced gastrointestinal stromal tumors: safety and efficacy in a worldwide treatment-use trial of sunitinib. Cancer. 2015 May 01;121(9):1405-13. PubMed PMID: 25641662. Pubmed Central PMCID: NIHMS688188. English.

21. Poole CD, Connolly MP, Chang J, Currie CJ. Health utility of patients with advanced gastrointestinal stromal tumors (GIST) after failure of imatinib and sunitinib: findings from GRID, a randomized, double-blind, placebo-controlled phase III study of regorafenib versus placebo. Gastric Cancer. 2015 Jul;18(3):627-34. PubMed PMID: 24957256. Pubmed Central PMCID: PMC4511071.

22. Craig R, Mindell J. Health Survey for England 2012. 2013.

23. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value Health. 2011 Jul-Aug;14(5):721-31. PubMed PMID: 21839411. Pubmed Central PMCID: PMC3811066.

24. Abel J, Pring A, Rich A, Malik T, Verne J. The impact of advance care planning of place of death, a hospice retrospective cohort study. BMJ Support Palliat Care. 2013 Jun;3(2):168-73. PubMed PMID: 23626905. Pubmed Central PMCID: PMC3632964.

Appendix 1. Baseline characteristic of trial participants

Table 57. Characteristics of participants in the studies across treatment groups (GRID study, ITT)

Characteristic	Regorafenib + BSC (n=133)	Placebo + BSC (n=66)
Median Age	60 (51-67)	61 (48-66)
Age group n (%)		
<65 years	90 (67.7)	46 (69.7)
≥65 years	43 (32.3)	20 (30.3)
Sex		
Men	85 (64%)	42 (64%)
Women	48 (36%)	24 (36%)
Ethnic Group		
White	90 (68%)	45 (68%)
Black or African American	0	1 (2%)
Asian	34 (26%)	16 (24%)
Not reported or missing	9 (7%)	4 (6%)
Geographic Region		
Asia	32 (24.1%)	15 (22.7%)
Rest of world	101 (75.9%)	51 (77.3%)
Geographic Region		
North America	22 (16.5%)	14 (21.2%)
USA	15 (11.3%)	11 (16.7%)
Canada	7 (5,3%)	3 (4.5%)
Non-North America	111 (83.5%)	52 (78.8%)
ECOG performance status		
0	73 (55%)	37 (56%)
1	60 (45%)	29 (44%)
Time since initial diagnosis to randomisation		
Mean (range), weeks	296.4 (32.3-774)	310.6 (47.0-657)
Median, weeks	256.0	272.2
Time since recent progression / relapse to randomisation		
Mean (range), weeks	13.29 (0.7-145)	16.7 (0.4-421)
Median, weeks	6.34	4.27
Extent of disease at baseline		
Metastatic	90 (67.7%)	38 (57.6%)
Unresectable	5 (3.8%)	10 (15.2%)
Metastatic and unresectable	35 (26.3%)	14 (21.2%)
Missing	3 (2.3%)	4 (6.1%)
Histology		
Missing	5 (3.8%)	4 (6.1%)
Spindle cells	66 (49.6%)	30 (45.5%)
Epithelioid	12 (9.0%)	4 (6.1%)
Mixed	18 (13.5%)	10 (15.2%)
Unknown	32 (24.1%)	18 (27.3%)
Number of tumour sites		
1	16 (12.0%)	9 (13.6%)
2	31 (23.3%)	20 (30.3%)
3	39 (29.3%)	13 (19.7%)
4	21 (15.8%)	9 (13.6%)
≥5	26 (19,5%)	15 (22.7%)
Previous systemic anti-cancer therapy		
2 lines	74 (56%)	39 (59%)
>2 lines	59 (44%)	27 (41%)
Duration of previous imatinib therapy		
≤ 6 months	18 (14%)	4 (6%)

Characteristic	Regorafenib + BSC (n=133)	Placebo + BSC (n=66)
6–18 months	26 (20%)	7 (11%)
> 18 months	89 (67%)	55 (83%)
Adapted from Pharmaceutical Benefits Advisory Committee (2008) Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Canberra: Pharmaceutical Benefits Advisory Committee		

Source: Bayer submission, section 4.5, Table 20, p86.

Appendix 2. Adverse events

Table 58. TEAEs (all grade) occurring in ≥10% regorafenib patients during GRID study (NCI CTCAE; SAF)

	Double-blind treatment (data cut-off 26 January 2012)		Data cut-off 08 June 2015	
	Regorafenib + BSC	Placebo + BSC	Regorafenib-treated at any time during study	Subgroup treated with regorafenib for >1 year
	N=132 n (%)	N=66 n (%)	N=190 n (%)	N=75 n (%)
Any TEAE	[redacted]	[redacted]	[redacted]	[redacted]
Blood and Lymphatics Anaemia	[redacted]	[redacted]	[redacted]	[redacted]
Cardiac	[redacted]	[redacted]	[redacted]	[redacted]
Ear and Labyrinth	[redacted]	[redacted]	[redacted]	[redacted]
Endocrine Hypothyroidism	[redacted]	[redacted]	[redacted]	[redacted]
Gastrointestinal Abdominal pain Constipation Diarrhoea Dyspepsia Mucositis oral Nausea Vomiting	[redacted]	[redacted]	[redacted]	[redacted]
General and Administrative Site Conditions Fatigue Fever Oedema limb Pain	[redacted]	[redacted]	[redacted]	[redacted]
Hepatobiliary disorders	[redacted]	[redacted]	[redacted]	[redacted]
Infection and Infestations Bronchial infection Rash pustular Upper respiratory infection	[redacted]	[redacted]	[redacted]	[redacted]
Injury, poisoning and procedural complications	[redacted]	[redacted]	[redacted]	[redacted]
Investigations Alanine aminotransferase increased (ALT) Aspartate aminotransferase increased (AST) Blood bilirubin increased Platelet count decreased Weight Loss	[redacted]	[redacted]	[redacted]	[redacted]
Metabolism and Nutrition Anorexia Hyperglycaemia Hypokalaemia	[redacted]	[redacted]	[redacted]	[redacted]
Musculoskeletal and Connective Tissue Arthralgia Back pain	[redacted]	[redacted]	[redacted]	[redacted]

	Double-blind treatment (data cut-off 26 January 2012)		Data cut-off 08 June 2015	
	Regorafenib + BSC	Placebo + BSC	Regorafenib-treated at any time during study	Subgroup treated with regorafenib for >1 year
	N=132 n (%)	N=66 n (%)	N=190 n (%)	N=75 n (%)
Myalgia Pain in extremity	[redacted]	[redacted]	[redacted]	[redacted]
Nervous System Dysgeusia Headache Paraesthesia	[redacted]	[redacted]	[redacted]	[redacted]
Psychiatric disorders Insomnia			[redacted]	[redacted]
Renal and urinary Proteinuria	[redacted]	[redacted]	[redacted]	[redacted]
Reproductive system and breast disorders			[redacted]	[redacted]
Respiratory, Thoracic and Mediastinal Cough Dyspnoea Hoarseness Voice alteration	[redacted]	[redacted]	[redacted]	[redacted]
Skin and subcutaneous tissue Alopecia Palmar-Plantar Erythrodysaesthesia Syndrome Pruritus Rash maculopapular	[redacted]	[redacted]	[redacted]	[redacted]
Vascular Hypertension	[redacted]	[redacted]	[redacted]	[redacted]

BSC=Best supportive care; TEAE=Treatment-emergent adverse event; NCI CTCAE=National Cancer Institute Common Terminology Criteria for Adverse Events version 4.0; A patient may have experienced more than one TEAE.

Source: Bayer submission, Section 4.12, Table 25, p122

Appendix 3. AICs/BICs for parametric OS extrapolation

Table 59. AICs for parametric OS extrapolation (2015 data cut off)

Parametric Model	**Regorafenib**	**Placebo**		**Sum AIC (placebo + regorafenib)**	
		RPSFT	*IPE*	*RPSFT*	*IPE*
Exponential	390.96	196.66	195.24	587.62	586.21
Loglogistic	**388.92**	**195.74**	**193.24**	**584.66**	**582.16**
Weibull	391.25	198.43	196.92	589.67	588.17
Lognormal	393.24	197.25	194.77	590.49	588.01
Gompertz	392.85	198.39	196.89	591.23	589.74

Source: Bayer submission, Table 32, p.160

Table 60. AICs for parametric OS extrapolation (2017 data cut off)

Parametric Model	**Regorafenib**	**Placebo**			**Sum AIC (placebo + regorafenib)**		
		Un-adjusted	**RPSFT**	**IPE**	**Un-adjusted**	**RPSFT**	**IPE**
Exponential	394.12	**201.84**	192.53	192.00	595.96	586.65	586.12
Loglogistic	**391.08**	204.83	**188.99**	**187.78**	**595.92**	**580.08**	**578.86**
Weibull	394.93	203.80	193.89	193.32	598.73	588.82	588.25
Lognormal	395.36	206.53	190.64	189.48	601.89	586.00	584.84
Gompertz	396.12	203.80	194.21	193.60	599.92	590.33	589.72

Source: Bayer addendum, Table 4, p. 13

Table 61. BICs for parametric OS extrapolation (2015 data cut off)

Parametric Model	Regorafenib	Placebo		Sum BIC (placebo + regorafenib)	
		RPSFT	***IPE***	***RPSFT***	***IPE***
Exponential	**393.85**	**198.85**	**197.43**	**592.7**	**591.28**
Loglogistic	394.7	200.12	197.62	593.97	591.47
Weibull	397.03	202.81	201.3	596.66	595.15
Lognormal	399.02	201.63	199.14	595.48	592.99
Gompertz	398.63	202.77	201.27	596.62	595.12

Source: Bayer submission, table 33, p.160

Table 62. BICs for parametric OS extrapolation (2017 data cut off)

Parametric Model	Regorafenib	Placebo			Sum BIC (placebo + regorafenib)		
		Un-adjusted	**RPSFT**	**IPE**	**Un-adjusted**	**RPSFT**	**IPE**
Exponential	397.01	**204.03**	194.72	194.19	**601.04**	591.73	591.20
Loglogistic	**396.87**	209.21	**193.37**	**192.16**	606.08	**590.24**	**589.02**
Weibull	400.71	208.18	198.27	197.70	608.89	598.98	598.41
Lognormal	401.14	210.91	195.02	193.86	612.05	596.16	595.00
Gompertz	401.90	208.18	198.59	197.98	610.08	600.49	599.88