Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours:

NICE STA

Addendum:

Between 1st and 2nd NICE Appraisal Committee meetings

11th August 2017

Confidential information that is commercial-in-confidence is highlighted and underlined.

Confidential information that is academic-in-confidence is highlighted and underlined.

Contents

Contents 2

1 Critique of Bayer responses to questions from NICE 3

1.1 Question 1 (Introduction) 3

1.2 Question 1a 4

1.3 Question 1b 8

1.4 Question 1c 9

1.5 Question 1d 10

1.6 Question 2 11

1.7 Question 3 11

1.8 Question 4 13

1.9 Question 5 13

2 PenTAG revised base case 14

3 References 17

1 Critique of Bayer responses to questions from NICE

The first NICE committee meeting for this STA was held on 28th June 2017. Afterwards, NICE reported that the appraisal committee require more information about Bayer's treatment switching adjustment for overall survival used in the 2017 data cut analysis. NICE asked Bayer to assume the following in all analyses:

- additional background mortality and
- age-related utility decrements

1.1 Question 1 (Introduction)

NICE asked for more justification for the assumptions underlying the RPSFTM and IPE methods of adjusting for treatment switching.

Bayer replied that the common treatment effect assumption states that the treatment effect received by switching patients must be equal to that received by patients initially randomised to the active treatment group, otherwise the crossover adjustment will produce biased results. We agree, and note that this assumption applies to both methods.

Bayer said that the ability to test the common treatment effect assumption is particularly limited in this case due to the small number of patients in the study. They continued that analysis of the counterfactual survival times (presented in 1c) indicated that the adjustment methods worked well, producing hazard ratios close to 1, providing evidence that the common treatment effect assumption holds.

We consider this response reasonable.

Bayer and we have previously agreed that the IPCW method is inappropriate due to the high proportion of placebo patients that switched treatment.

Therefore, we consider it reasonable to use the RPSFTM or IPE methods.

However, as discussed in our original report, we are not convinced by Bayer's rationale for choosing the IPE method over the RPSFT method in the base case. We consider both methods equally plausible. For example, Latimer et al (2016) preferred the RPSFT method

over the IPE method to adjust for treatment switching in a trail of metastatic melanoma. For this reason, in Section 2, p14, we now give equal credibility to these two methods.

1.2 Question 1a

NICE asked for an assessment of the impact of recensoring on the adjusted overall survival hazard ratios and cost effectiveness of regorafenib, specifically ICERs with and without recensoring.

In response, Bayer provided the hazard ratios below. The ITT and recensored values are the same as those previous reported by Bayer. The "no recensoring" values in the table below are new information.

Table 1. OS HRs (2017 data)

HRs	Recensored	No recensoring
Unadjusted		[redacted]
RPSFT	[redacted]	[redacted]
IPE	[redacted]	[redacted]

For completeness, we show the corresponding, higher, hazard ratios corresponding to the 2015 data.

Table 2. OS HRs (2015 data)

HRs	Recensored	No recensoring
Unadjusted		[redacted]
RPSFT	[redacted]	[redacted]
IPE	[redacted]	[redacted]

These values are summarised in Figure 1 below. This shows that in general hazard ratios are lower, and hence the estimated cost-effectiveness of regorafenib is better:

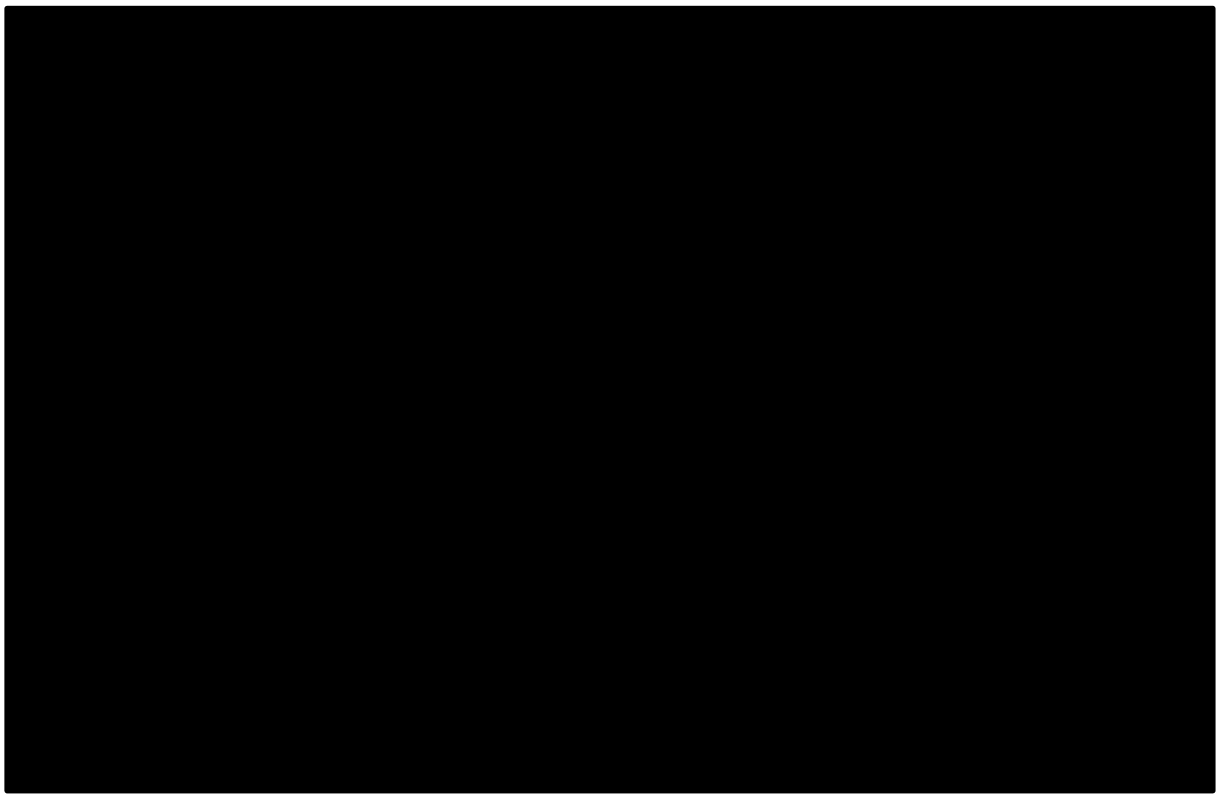
- For the IPE method compared to the RPSFT method,

- For the 2017 data compared to the 2015 data.
- Allowing for recensoring.

Recensoring reduces the hazard ratios more for the:

- 2017 data than the 2015 data.
- IPE method than the RPSFT method.

Figure 1 OS hazard ratio by data cut and whether data recensored



Next, Bayer present the impact of recensoring on cost-effectiveness (Table 3, Table 4) on the following basis:

- age-related utilities.
- additional background mortality.
- OS extrapolated as a 50%:50% average of the Weibull and log-logistic distributions (our original assumption).

- “updated dosing analysis”.
- 2017 data cut.

Although they apply the correction for age-related disutility, they claim this is unnecessary, as they believe this is already captured in the EQ-5D data from the GRID trial. We disagree. Age-related utility adjustment is standard practice in cost-utility analyses in general, and is certainly relevant in this case, given that some patients are predicted to survive far beyond the maximum follow up time of the trial.

Before the first NICE committee meeting, we also modelled OS as a 50%:50% average of the Weibull and log-logistic distributions. However, NICE stated that the committee preferred the Weibull distribution (Question 3). Therefore, this limits the relevance of the ICERs Bayer present in this section.

Originally, we did not understand the meaning of “updated dosing analysis”. On 10th August 2017, Bayer clarified as follows:

“*we refer to the revised dose intensity calculation of regorafenib including doses of 0 mg. Cost effectiveness analysis results based on the “updated dosing analysis” were already presented in our response to clarification question 5 from NICE received on June 15th, 2017. The inclusion of the revised dose intensity calculation in the ERG’s revised base case analysis was accepted by the Appraisal Committee on June 28th, 2017 (please see Cost Effectiveness slide 22). We agree with the ERG and the Appraisal Committee that cost effectiveness analyses should be based on the mean observed dose of regorafenib by cycle including 0 mg doses.*

When considering the actual doses from the GRID trial (including those of 0 mg), the mean observed dose of regorafenib by cycle is lower compared to when 0 mg doses are excluded from the calculation of the average. As shown during the first Appraisal Committee Meeting, the inclusion of the revised dose intensity calculation or “updated dosing analysis” had an impact on the ICER of approximately £2,000.”

As we have previously stated, we accept the logic of this argument. However, we caution that we do not have the underlying data to verify the change in the ICER. In other words, our version of Bayer’s economic model does not reflect the updated dosing analysis.

Table 3. Bayer ICERs with and without recensoring (with PAS, 2017 data)

	No recensoring	Recensoring
IPE	£51,629	£42,156
RPSFT	£49,573	£43,737

Table 4. Bayer ICERs with and without recensoring (without PAS, 2017 data)

	No recensoring	Recensoring
IPE	[redacted]	[redacted]
RPSFT	[redacted]	[redacted]

Bayer also present the analogous ICERs without the age-related utility adjustment. This reduces all ICERs by about £1,000 per QALY.

We attempted to recreate Bayer's ICERs in the tables above. We find the ICERs given in the tables below, on the same basis as used Bayer, but without the "updated dosing analysis". We are unable to calculate the ICERs in the absence of recensoring because we do not have the relevant OS data, although we make approximations in Section 2. Assuming the PAS, these ICERs are about £2,000 per QALY higher than those presented by Bayer, and without the PAS, about £3,000 per QALY higher. We agree with Bayer's estimates of total costs, life years and QALYs for BSC, and total life years and QALYs for regorafenib. However, we estimate slightly higher total costs of regorafenib. For example assuming the PAS, we estimate £46,997 versus Bayer's £45,459. We assume this difference is due to Bayer's "updated dosing analysis".

Table 5. PenTAG ICERs with and without recensoring (with PAS, 2017 data)

	No recensoring	Recensoring
IPE	unknown	£44,000
RPSFT	unknown	£45,652

Table 6. PenTAG ICERs with and without recensoring (without PAS, 2017 data)

	No recensoring	Recensoring
IPE	unknown	[redacted]
RPSFT	unknown	[redacted]

Our ICERs above are 4% higher than Bayer's ICER.

Relevance of recensoring

As stated in our original report, recensoring may lead to biased estimates of the average treatment effect when the proportional treatment effect assumptions do not hold, because longer term data on the effect of treatment may be lost. We understand that, whilst the relevant NICE Technical Support Document recommends recensoring, whether to perform recensoring remains a subject of academic debate. Indeed recent research recommends performing the adjustment both with and without recensoring (Latimer & Abrams 2017, Latimer et al 2016). This was confirmed in the email of 3rd August 2017 from Dr. Latimer to Bayer. The estimated treatment effect is generally greater when recensoring is performed compared to the analysis without recensoring (Latimer & Abrams 2017). Adjustment without recensoring was favoured in one recent dataset by Latimer et al 2016.

For these reasons, in our base case (Section 2, p14), we now consider analyses both with and without recensoring to be equally valid.

1.3 Question 1b

NICE asked for a comparison of results obtained with the IPE method under two distinct bases: (a) "on treatment" which assumes that the treatment effect applies only while a patient is on treatment and (b) "treatment group" which assumes that the treatment effect applies from the time of initiation of the drug until death.

In all analyses so far, Bayer have assumed the "treatment group" analysis. In response to NICE's request for further information, Bayer now present the results of the "on treatment" analysis. Bayer say they implemented this analysis based on an academic paper and

advice from Dr. Latimer. They estimated a much shorter tailed OS for placebo than using their base case "treatment group" analysis: median OS 69 days vs. [redacted] days. However they caution that the "on treatment" method "may not be reliable" due to the small sample size and large numbers of patients switching treatment. Using the RPSFT method, they found similar results with the same concerns about reliability.

Bayer then say they performed an exploratory analysis "adjusting only for discontinuation of regorafenib (crossover from regorafenib to BSC)". They claim this analysis suggests that the "on treatment" analysis will likely produce results more favourable for regorafenib than the "treatment group" analysis. In response, first, we do not understand this exploratory analysis. Second, Bayer already claim that the "on treatment" analysis yields a greater estimated treatment benefit for regorafenib than the "treatment group" analysis (e.g. median OS values quoted above). Therefore, we do not see the relevance of this exploratory analysis.

Bayer then say that due to time constraints, it was not possible to implement a further analysis suggested by Dr. Latimer. We assume Bayer refer here to Dr. Latimer's suggestion in his email of 3rd August 2017 to try the two-stage method of adjustment for treatment switching. We sympathise with Bayer's reason for not using this technique and we consider that they have considered a reasonable range of adjustment methods.

We believe that all this uncertainty further highlights the uncertainty in the results of switching adjustments in general.

1.4 Question 1c

NICE asked for a comparison of counterfactual survival times in the regorafenib and placebo arms (estimate of overall survival if no patients in either treatment arm had received regorafenib). NICE also requested a visual comparison of the counterfactual survival curves. They noted that a hazard ratio close to 1 would indicate that the estimation procedure had worked well.

In response, Bayer now present the counterfactual OS survival curves with recensoring applied to the 2017 data cut. They considered separately the IPE and RPSFT methods. In both cases, they found that the counterfactual OS survival curves were very similar, with OS hazard ratios close to 1.

We agree that this provides some evidence to support use of these methods. However, importantly, this does not necessarily mean that the assumptions associated with the method are justified, or that the data fit the model (Latimer et al 2016).

1.5 Question 1d

NICE asked for a detailed explanation of the cause of the 24% reduction in mean overall survival in the placebo arm after adjustment for treatment switching using the 2017 data compared to the 2015 data.

In response, Bayer again account for the reduction as a combination of (a) difference in events i.e. change in the Kaplan-Meier curves during the follow up period of the 2015 data cut and (b) increase in follow up using the 2017 data.

Concerning (a), we agree that the estimated benefit of regorafenib during the follow up period of the 2015 data cut has increased. However, this increase appears very small on inspection of the relevant Kaplan-Meier curves in Figure 5 of Bayer's response document.

Concerning (b), we agree that there is some further follow up for both treatment arms. But this is only small.

Overall, we are surprised that together these small effects can yield rather a substantial reduction of 24% in mean OS for the adjusted placebo arm. However, given that we have no conclusive evidence that Bayer have not performed the IPE method correctly, we accept Bayer's justification.

In our original base case, we preferred the 2015 data cut over the 2017 data cut, because of our concerns about the 24% reduction in OS.

In our revised base case (Section 2), we now prefer the 2017 data cut.

In their Tables 18 and 19, Bayer report total life years and QALYs for each treatment arm, separately for the 2015 and 2017 data cuts. We agree with the data they present in these tables.

1.6 Question 2

The appraisal committee noted that the p-values associated with the 2017 adjusted analyses for overall survival are incorrect. NICE requested the updated adjusted hazard ratios (stratified and unstratified analyses), 95% confidence intervals and associated p-values using both IPE and RPSFT methods.

Bayer have now provided the data requested in Table 20 of their response document. They provide hazard ratios separately for the unstratified and stratified analyses. We are unable to check the unstratified hazard ratios. For the stratified analysis, the mean hazard ratios for the ITT, RPSFT and IPE methods are appropriately the same as those given in Bayer's Clinical Study Report Addendum 2 (2017 data cut) at [redacted], [redacted] and [redacted] respectively. We expected the confidence interval for the ITT analysis quoted by Bayer to be the same as that given in Clinical Study Report Addendum 2. However, these differ: (0.676, 1.194) and (0.645, 1.250) respectively. Nonetheless, we do not dwell on this issue, as we believe this will not materially affect the committee's decisions.

1.7 Question 3

The appraisal committee heard an additional concern from us, the ERG that, whilst the Weibull distribution was assumed in the implementation of the IPE method, Bayer then extrapolated the adjusted OS data using a different distribution, the log-logistic. Related to this, the committee considered extrapolation of overall survival with the Weibull as more appropriate than the log-logistic, based on the estimated proportions of patients alive after several years. NICE asked Bayer to provide ICERs assuming a Weibull extrapolation for overall survival.

In response, Bayer estimate ICERs of £56,000 with the PAS and [redacted] without the PAS on the following basis, which is the same as that given in Section 1.2, p4, but assuming OS Weibull:

- age-related utilities.
- additional background mortality.
- OS extrapolated Weibull.
- updated dosing analysis.
- 2017 data cut.

When we try to create these ICERs, without the "updated dosing analysis", we estimate £47,000 with the PAS and [redacted] without the PAS. Applying the 4% reduction in ICERs corresponding to the "updating dosing analysis", these ICERs decrease to £45,000 with the PAS and [redacted] without the PAS. Our ICERs are substantially lower than those given by Bayer. Also, we estimate different total costs, life years and QALYs compared for both treatment arms to Bayer. We are unable to account for these differences. We believe that Bayer's ICERs are incorrect.

Based on Bayer's analysis, when we select the Weibull distribution over the 50% Weibull: 50% log-logistic, the ICERs increase substantially, from:

- £42,000 to £56,000 assuming the PAS and
- [redacted] to [redacted] without the PAS.

On the other hand, we estimate that the ICERs increase less, from:

- £42,000 to £45,000 with updated dosed, and £44,000 to £47,000 without updated dosing assuming the PAS and
- [redacted] to [redacted] with updated dosed and [redacted] to [redacted] without updated dosing without the PAS.

Next, Bayer cite advice from Dr. Latimer that the function used for the IPE method, namely the Weibull, does not necessarily need to be the same as the function used to extrapolate OS. We now have some sympathy for this argument. However, we note that Bayer chose the log-logistic distribution because it gave the best fit to the trial data. This would suggest that they should have used the log-logistic, rather than the Weibull, as part of the IPE method. Nonetheless, we do not dwell on this issue, as we have no evidence for the impact of using the log-logistic function in the IPE method.

The NICE appraisal committee favoured the Weibull, partly on advice from the clinical experts at the meeting. **Therefore, we now change our base case assumption for OS extrapolation from a 50%:50% average of the Weibull and log-logistic to 100% Weibull.**

1.8 Question 4

The appraisal committee noted that maximum follow up in the placebo adjusted arms were the same in the 2015 and 2017 analyses. NICE asked Bayer to complete a table summarising maximum follow up times.

In response, Bayer provide the required follow up times. The maximum follow up time for the placebo RPSFT-adjusted and IPE-adjusted data was [redacted] days for both the 2015 and 2017 data cuts. In our report, we noted that we expected the maximum follow up to be greater for the 2017 data cut compared to the 2015 cut, given that the 2017 data is more mature.

Bayer accounted for this as follows: "*Note that the maximum follow-up for placebo patients who do not crossover (N=8) is [redacted] days. As the counterfactual survival for all crossover placebo patients is estimated to be less than this, and patients who do not crossover are not affected by the crossover adjustment this results in the maximum follow-up for the adjusted analysis with both data cuts being equal*.".

It seems surprising to us that the counterfactual survival for all crossover placebo patients is estimated to be less than the maximum follow-up for placebo patients who do not crossover. Nonetheless, given no evidence to the contrary, we accept Bayer's explanation.

1.9 Question 5

NICE asked Bayer to provide all relevant log files for the treatment switching analysis for both 2015 and 2017 data for overall survival. This should be provided as a text file.

In response, Bayer have provided STATA log files for the 2015 and 2017 data separately for the ITT, IPE and RPSFT methods.

Due to time constraints, we have not had checked the STATA logs in detail. However, they do at least appear reasonable.

2 PenTAG revised base case

In our original report, we favoured:

- OS Weibull 50%, log-logistic 50%.
- Age-related utilities.
- IPE method.
- Analyses with recensoring.
- 2015 data cut.

We have now revised our preferred assumptions, in the light of (a) the committee discussion at the first NICE committee meeting and (b) Bayer responses above, to the following:

- OS Weibull (Section 1.7, p11).
- Age-related utilities (unchanged) (Section 1.2, p4).
- IPE and RPSFT methods equally plausible (Section 1.1, p3).
- Analyses with and without recensoring equally plausible (Section 1.2, p4).
- 2017 data cut (Section 1.5, p10).
- With or without Bayer's "updated dosing analysis" equally plausible (Section 1.2, p4).

Our corresponding ICERs are given in the Tables below. We consider all ICERs within each table equally valid. ICERs above NICE's £50,000 per QALY willingness to pay threshold for End of Life treatments are shown in grey shading.

The ICERs corresponding to no recensoring are approximations, because we do not have access to the relevant OS data. These are estimated by multiplying our relevant ICER in Table 7, Table 8, Table 9 or Table 10 corresponding to recensoring by the ratio of Bayer's relevant ICER from Table 3 or Table 4 on p7 without recensoring to their relevant ICER with recensoring. For example, our ICER of £55,230 (rounded to £55,000 in Table 9) = £45,096 (Table 9) x (£51,629 / £42,156).

We repeat from our original report that total uncertainty in the cost-effectiveness of regorafenib versus BSC is high due to:

- Substantial uncertainty in the adjustment for widespread treatment switching.
- Important uncertainty in the extrapolation of OS.

Table 7. PenTAG revised preferred ICERs without updated dosing, with PAS

	No recensoring	Recensoring
IPE	£57,000#	£47,000
RPSFT	£55,000#	£49,000

approximation, see text

Table 8. PenTAG revised preferred ICERs without updated dosing, without PAS

	No recensoring	Recensoring
IPE	[redacted]#	[redacted]
RPSFT	[redacted]#	[redacted]

approximation, see text

Applying the updated dosing, all ICERs are estimated to be 4% lower, as shown in the tables below. But we repeat our concern that we are unable to use Bayer's model to check these figures, because we have not been provided with the updated dosing data.

Table 9. PenTAG revised preferred ICERs with updated dosing, with PAS

	No recensoring	Recensoring
IPE	£55,000#	£45,000
RPSFT	£53,000#	£47,000

approximation, see text

Table 10. PenTAG revised preferred ICERs with updated dosing, without PAS

	No recensoring	Recensoring
IPE	[redacted]#	[redacted]

	No recensoring	Recensoring
RPSFT	[redacted]#	[redacted]

approximation, see text

3 References

Latimer NR & Abrams KR (2017). To re-censor, or not to re-censor, that is the question: critical considerations when applying statistical methods to adjust for treatment switching in clinical trials. Poster for ISPOR conference Boston. https://www.ispor.org/ScientificPresentationsDatabase/Presentation/73257?pdfid=50852

Latimer N, Bell, H, Abrams K, Amonkar M, Casey M. (2016) Adjusting for treatment switching in the METRIC study shows further improved overall survival with trametinib compared with chemotherapy. Cancer Medicine, 5(5):806–815.