[redacted]
National Institute for Health and Clinical Excellence
MidCity Place
81 High Holborn
London
WC1 6NA

BY EMAIL

3 September 2012

RE: Vemurafenib for the treatment of BRAF V600 mutation positive metastatic melanoma

Thank you for giving us the opportunity to comment on the ACD for the above appraisal. We welcome the Committee's consideration of the impact of crossover within the BRIM3 study and hope the information provided within sections 1.1 and 1.2 of this document will give further clarity on the crossover adjustment conducted.

As requested we have conducted the overall survival extrapolation sensitivity analysis specified in the ACD (detailed in section 2 of this document). This analysis resulted in clinically implausible results - a post-progression survival period 2.2 months shorter in the vemurafenib arm than in the dacarbazine arm. We do not believe this is reasonable and have therefore not presented cost-effectiveness estimates using this approach. Further detail on our rationale for this is provided below.

As stated in our previous ACD response we believe the most appropriate cost-effectiveness estimates for vemurafenib are those based up the RPSFT adjusted February 2012 data with long term extrapolation based upon the fair assumption that patients given vemurafenib die at the same rate as those given dacarbazine. The ICER for this analysis was stated as £52,327 in our previous ACD response – following correction of some minor errors identified by the ERG this figure has fallen to £51,757 (see Appendix 1).

Roche Products Limited 6 Falcon Way, Shire Park, Welwyn Garden City, Hertfordshire, AL7 1TW
Health Economics and Strategic Pricing
Tel. +44(0)1707 361032
Fax +44(0)1707 384123

We hope this information allows the Committee to make vemurafenib available to a group of patients with extremely poor prognosis and no alternative treatment options. Whilst the magnitude of long-term benefits provided by vemurafenib are clearly relevant to this appraisal it should be noted that this is not the key reason that clinicians wish to use vemurafenib – it is the clear, demonstrated and unprecedented benefits provided by vemurafenib in the BRIM3 study which are of key importance to clinicians and people with advanced melanoma.

If any further clarification of the content of this document is required, please do not hesitate to contact us.

Kind Regard

1.1 *A full explanation of the assumptions made and parameter values used for the rank preserving structural failure time (RPSFT) method to adjust survival estimates for patients who switched from dacarbazine to vemurafenib on disease progression*

The 'Rank Preserving Structural Failure Time' (RPSFT) (Robins and Tsiatis, 1991) approach to adjusting for crossover is a method that has been utilised in at least six NICE Technology Appraisals to date (NICE TA179, NICE TA215, NICE TA219, NICE TA257, NICE TA263 and NICE ID498 - this Appraisal)

The growing use of RPSFT by manufacturers is largely a result of the NICE appraisals of sunitinib for GIST (NICE TA179) and pazopanib for renal cell carcinoma (NICE TA215). In both these Appraisals the RPSFT method was used in order to adjust for crossover, in both Appraisals the RPSFT approach was accepted by the Appraisal Committee and resulted in positive guidance for both technologies under assessment.

Whilst other methods have been utilised in NICE Appraisals (for example the Inverse Probability Censoring Weight (IPCW) approach), in the cases where RPSFT has been presented alongside these alternative methods the Appraisal Committee have opted to use the RPSFT results as a basis for decision making (everolimus for renal cell carcinoma (TA219) and pazopanib for renal cell carcinoma (TA215)). As a result, the combination of these Appraisals has sent a clear signal to manufacturers that the RPSFT approach is acceptable for use in adjusting for crossover.

The use of RPSFT was further supported by a publication by members of NICE's Decision Support Unit (DSU) (Morden et al 2010). This publication reports the results of a simulation study which indicates that the RPSFT method and a parametric variant of the RPSFT method (the "Branson and Whitehead" approach) are the least biased of the methods assessed. In this report it was noted that the RPSFT methods gave treatment effect estimates *'close to the true treatment effect'.*

In light of the findings reported in this publication and the precedent set in the previous NICE Appraisals featuring RPSFT, the RPSFT method was utilised in order to adjust for crossover in the February 2012 data-cut of the BRIM3 study.

A brief introduction to RPSFT, the assumptions underlying it and their applicability in the case of BRIM3 is provided below.

The below was informed by consultation with an academic expert closely involved in developing, and assessing the validity of, methods to adjust for crossover and their use in decision analytic models.

An introduction to the RPSFT method

The RPSFT method attempts to simulate a counterfactual control arm - a control arm not confounded by crossover and representative of current clinical practice.

It does this by ‘speeding up’ or ‘accelerating’ the time patients who crossed over spent receiving crossover treatment. This ‘acceleration’ reduces the time patients who crossed over are alive in order to simulate what would have been observed had crossover not occurred.

The method operates by the simple application of an ‘acceleration factor’ to the time patients spent receiving the crossover treatment (as shown below).

RPSFT Simulated Event Time = Time Before Crossover + (Acceleration Factor * Time on Crossover)

An example of application of the method

Imagine we have data for a patient who is observed to have received their randomized treatment for 100 days, then experienced disease progression and crossed over to receive 200 days of the intervention arm prior to death. If we knew that the acceleration factor associated with that intervention was 0.5 we could apply RPSFT as follows:

Unadjusted Survival Time = 100 days + 200 days = 300 days

RPSFT Adjusted Survival Time = 100 days + (0.5 * 200 days)

100 days + 100 days = 200 days

As a result of the RPSFT adjustment this patient’s time to death would be reduced from 300 days to 200 days (a 100 day reduction). If the assumptions underlying RPSFT are met this new RPSFT adjusted survival time will portray what that patient’s survival time would have been if crossover had not occurred.

Deriving the acceleration factor

The acceleration factor used in RPSFT is derived iteratively by applying various potential acceleration factors to the time patients in **both** arms of the study spent receiving the intervention of interest.

This approach is designed to generate two identical survival curves that have ‘stripped out’ the impact of all of the intervention treatment (whether that was following randomisation or as a crossover treatment). When the difference between the two ‘stripped out’ curves is minimized (i.e.

the p-value testing for a difference between the curves via a Cox-model based, Log-rank or Wilcoxon test is equal to 1) the acceleration factor is assumed to be correct.

This acceleration factor is then applied to solely the time patients who crossed over spent receiving the crossover treatment in order to simulate a survival time representative of the decision problem.

If the patient was observed to die within the study then this acceleration would then reduce their time to death – whilst if the patient was censored at the time of study end this would then reduce their time to censoring.

Interpreting the acceleration factor

The acceleration factor defines the proportion of the crossover time observed that will remain following RPSFT adjustment. The smaller the acceleration factor estimated by RPSFT the more the application of that factor accelerates (reduces) a patient's time to death and moves the simulated survival curve to the left.

For example an acceleration factor of 0.1 would shrink the time spent alive following crossover to 10% of the original value (a 90% reduction in observed survival post-crossover) whilst an acceleration factor of 0.9 would shrink the time spent alive following crossover to 90% of the original value (a 10% reduction).

The acceleration factor estimated for vemurafenib in BRIM3

In the case of BRIM3 an acceleration factor of 0.34 was estimated using a Log-Rank model.

With the application of this acceleration factor every day spent receiving crossover vemurafenib in the dacarbazine arm would be reduced by 66% (i.e. 34% of that time will remain following adjustment).

An example of the impact the acceleration factor would have upon a theoretical patient in BRIM3 is demonstrated below.

Patient X is observed to receive 1.5 months of dacarbazine and then crosses over to receive vemurafenib for 6 months prior to dying. The unadjusted survival time of patient X would be 7.5 months (1.5 + 6 = 7.5). With the application of the RPSFT acceleration factor the 6 months this patient spent receiving vemurafenib would be reduced to approximately 2 months (6 * 0.34 = 2.04). As a result the RPSFT simulated survival time of patient X would now be 3.5 months (1.5 + 2.04 = 3.54).

In our previous ACD response this 0.34 acceleration factor was applied to the time each patient who crossed over in BRIM3 received crossover vemurafenib for in order to derive revised time to

event values. These revised time to event values were then combined with the unadjusted values (i.e. those for patients who did not crossover) in order to produce the RPSFT adjusted KM survival curves and associated statistics provided in our previous ACD response.

The assumptions underlying RPSFT and their validity in BRIM3

An RPSFT adjusted survival curve will only be unbiased if the assumptions underlying RPSFT are met.

Assumption 1: There is a single underlying acceleration factor associated with the intervention that is time invariant and valid for use in both the randomized and crossover arms

The RPSFT method defines an 'average' acceleration factor across all the patients who received the intervention (in both the control and intervention arm) and then applies this to solely those patients who crossed over. In doing this it is implicitly assumed that there is a single acceleration factor associated with the intervention adjusted for and that this does not vary with time or between the randomized and crossover population. If this assumption is violated then the RPSFT estimated survival curve will be biased.

Sub-assumption 1.1: The acceleration factor is time invariant

Patients who crossover are disproportionately likely to be in their early phases of treatment at the point of data cut-off, when compared to their randomised counterparts. As result if the effectiveness of the treatment is time variant then the application of an average acceleration factor to this crossover period will result in a biased RPSFT adjusted curve.

For example if a technology is more effective in the first few months of treatment than in the later phases of treatment the application of an average acceleration factor derived from the whole study to that crossover period is likely to under-accelerate (under-reduce) the control arm.

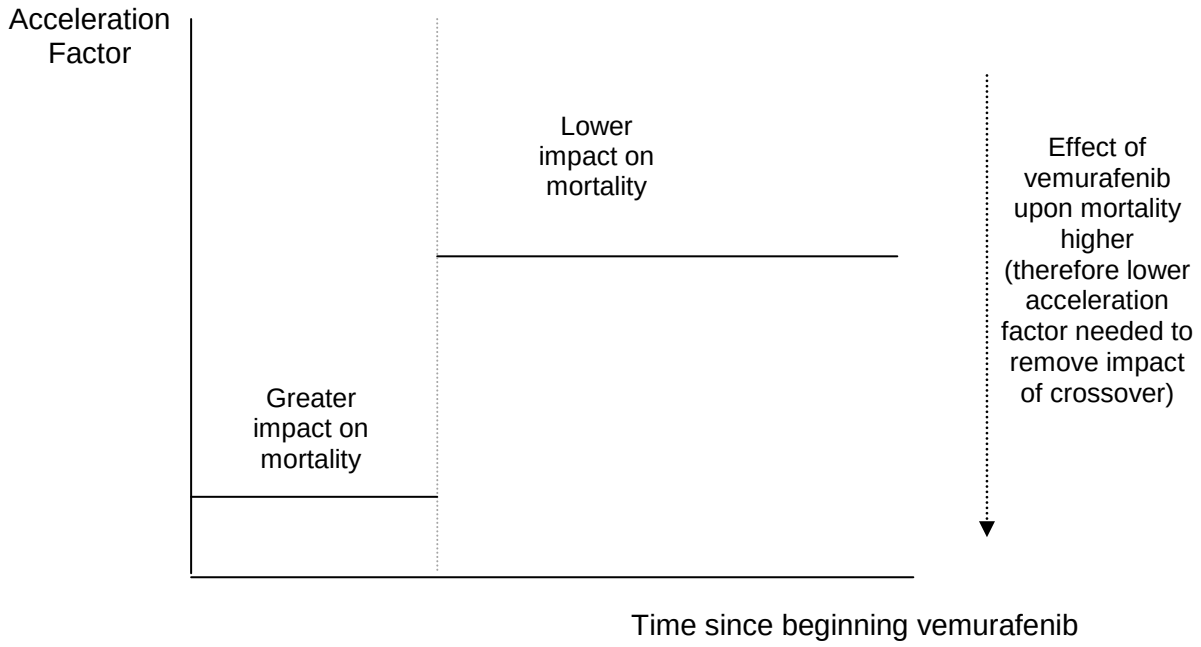
This assumption appears not to hold in the case of the BRIM3 study.

Both Roche and the ERG have developed overall survival modelling approaches that suggest the effect of vemurafenib upon mortality is highest for the first few months of treatment and then declines over time (albeit moving to a HR of 1 in the Roche model and a HR of greater than 1 in the ERG's model).

As a result it appears unlikely that a single 'acceleration factor' can be defined for vemurafenib. Both Roche and the ERGs models indicate that there should be a smaller acceleration factor in the first few months of treatment (treatment more effective therefore greater acceleration required) followed by a higher acceleration factor in the later months (treatment less effective therefore a smaller acceleration factor required).

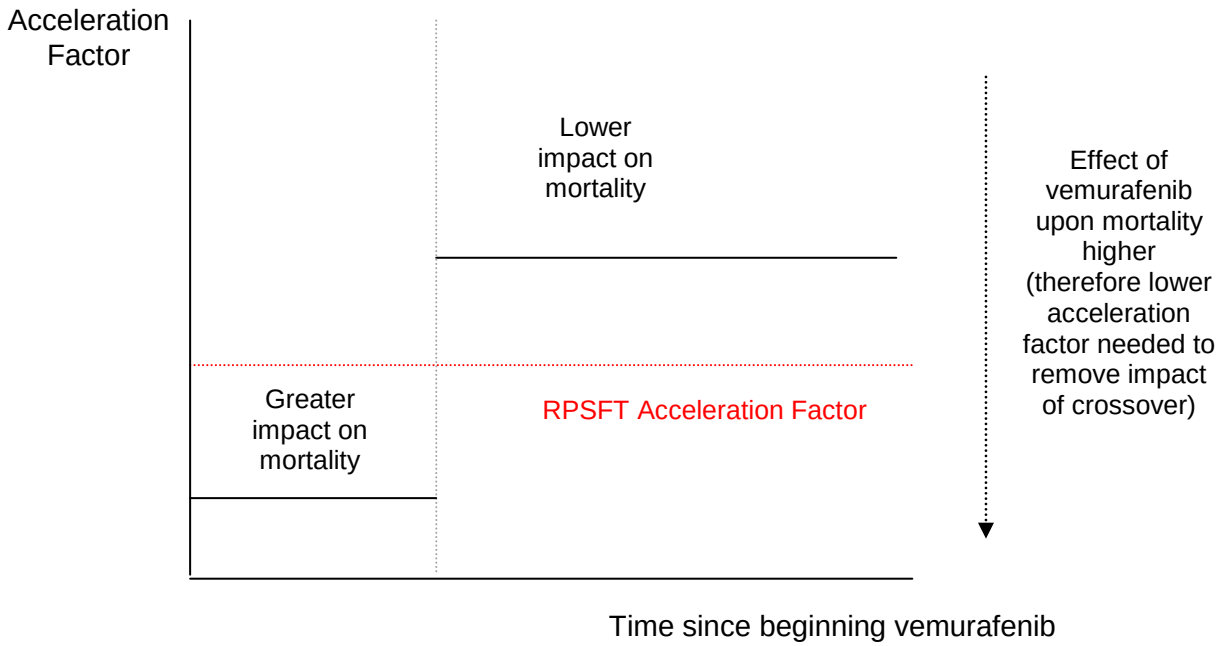
This is shown diagrammatically in Figure 1 below – note these figures are not to scale and are intended to be illustrative only.

Figure 1. Vemurafenib has a higher impact upon mortality in the earlier phases of treatment



The RPSFT method assumes that there is a single acceleration factor across time. It therefore generates a pooled acceleration factor across all periods of the BRIM3 study which will overestimate the acceleration factor in the early period of treatment and underestimate it in the later period of treatment.

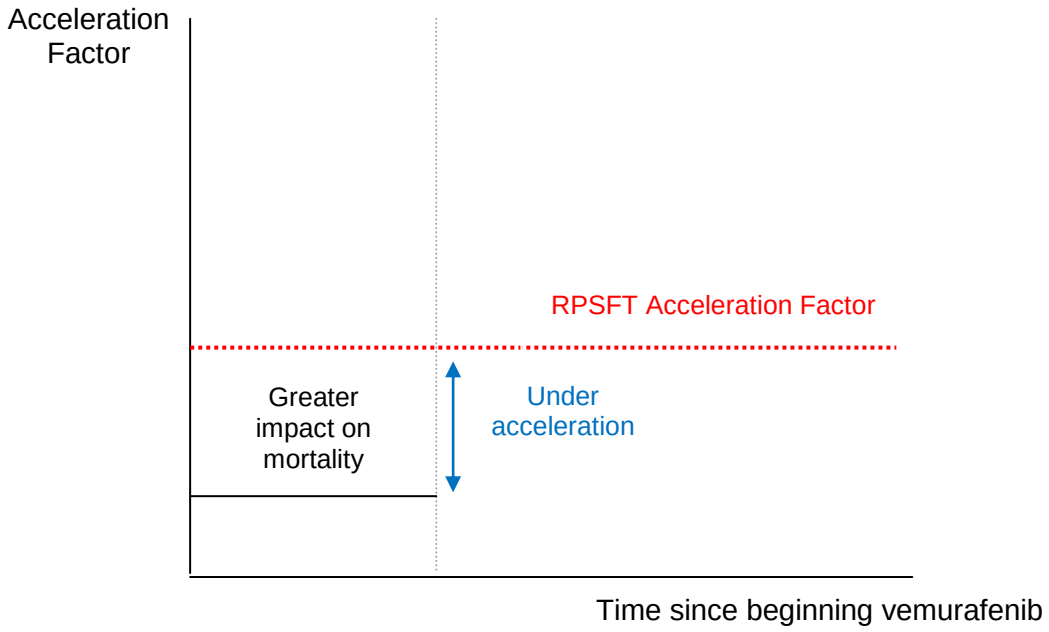
Figure 2. The RPSFT method derives an average treatment effect - as the treatment effect decreases over time this will result in the benefit of vemurafenib being underestimated in the short term and overestimated it in the longer term



In the BRIM3 study those patients who crossed over are still in their earlier phase of treatment at the point of data cut-off (with a mean crossover duration of around 4 months).

As a result the application of an average acceleration factor derived from the whole study to these patients crossover time is likely to under-accelerate the time to the event for these patients. This will result in a simulated dacarbazine survival curve that is biased to the right (i.e. the benefit of dacarbazine will be overestimated and the resultant benefit provided by vemurafenib underestimated).

Figure 3. As crossover patients are likely to be in the early, most effective, part of their vemurafenib treatment application of an average acceleration factor will under-accelerate the simulated comparator arm



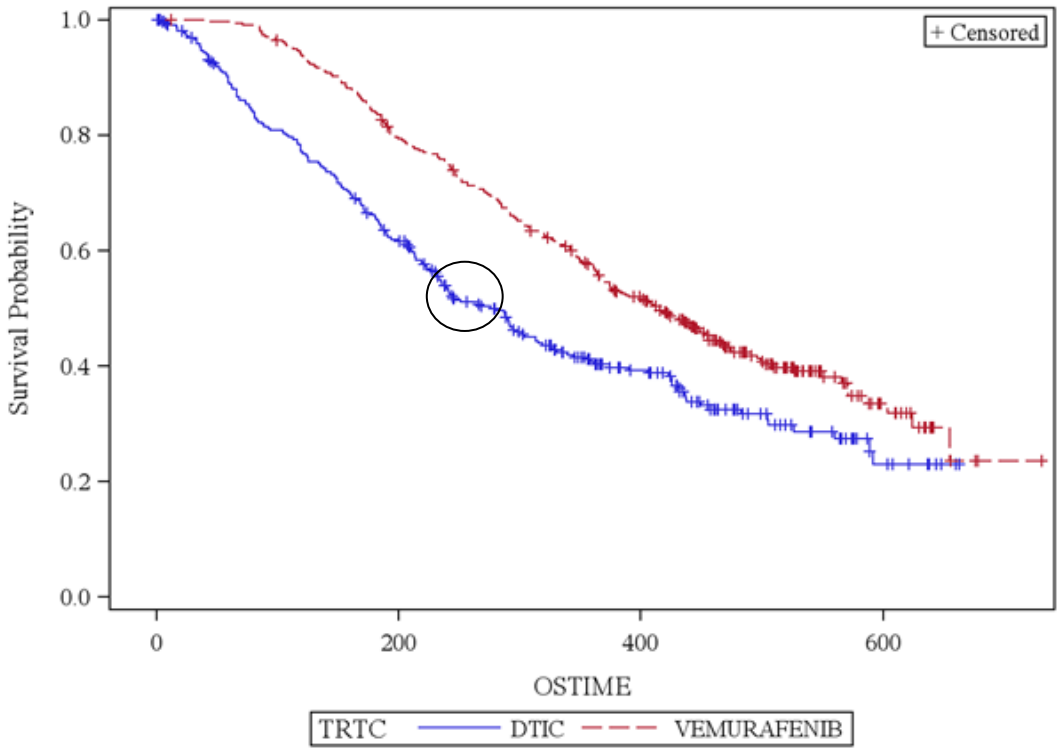
Our expert statistical advisor noted that this was an issue that would bias the RPSFT adjustment against vemurafenib and indicated that this problem has not yet been fully considered in academic research in this area (which to date has largely been focused on simulation studies in which proportional hazards was assumed in the data-sets simulated). Our expert suggested that it may be possible to conduct a stratified analysis in which the acceleration factor is derived in temporal phases (as informed by assessment of log cumulative hazard plots) but that this approach had not yet been fully explored by the academic community. Roche are currently considering how best to implement this time variant acceleration factor RPSFT approach. Given the complexities associated with further developing the RPSFT method we are unable to present this analysis in the time-frame available for consultation on the ACD.

It should be noted that in the first 3 months of the BRIM3 study the KM survival probability reduce only 3% (from 100% alive to 97% alive). In this initial period vemurafenib appears to completely stall a patient's disease with the result that the survival curve is virtually horizontal until after 3 months. Given this complete stalling of the disease it is clear that an acceleration factor of 0.34 is likely to under-accelerate the survival times of patients who crossed over.

This point is extremely relevant to the validity of the additional analysis requested by the Committee in the ACD. If the RPSFT adjustment is 'under-accelerating' the survival times of dacarbazine patients who crossed over then this could lead to the 'kink' observed the RPSFT adjusted survival curve at around 8 months (see Figure 4 below). It is notable that around month 8 the RPSFT adjusted dacarbazine arm becomes almost horizontal and appears to mirrors the trend

observed in the initial months of the vemurafenib arm – this is highly suggestive of this curve being under-accelerated in the initial period of crossover.

Figure 4. Rank Preserving Structural Failure Time (RPSFT) Adjusted Curve features 'kink' as a result of under-acceleration of initial, highest efficacy, crossover period due to application of average acceleration factor derived from whole study



As a result the atypical change in hazard observed in the RPSFT adjusted curve is unlikely to be representative of that expected in UK clinical practice (and as required by the decision problem). It is likely to be an artefact of under-acceleration due to application of an average, rather than time variant, acceleration factor and so should not be relied upon for modelling.

Sub-assumption 1.2: The acceleration factor derived for the intervention is valid for both the randomised and comparator arms

If there is a difference between the true treatment effect associated with the intervention in the crossover and randomised arms then the application of an ‘average acceleration factor’ from the whole study to the event times of those who crossed over will result in a biased RPSFT adjustment.

There are two key determinants of whether this would hold in a study; (1) the treatment is as effective as a first and second line treatment and (2) there are no differences between the randomised and crossover groups that would be expected to result in differential levels of treatment effect.

If these assumption do not hold then the RPSFT will be biased and result in under/over-acceleration of the crossover patients' time to events. The validity of these two assumptions are provided below:

(1) Vemurafenib is as effective as a first line treatment as it is as a second line treatment

If first line treatment with vemurafenib was more effective than as a second line treatment (perhaps due to some change in the biology of the disease caused by the comparator arm first line treatment) then the application of an average acceleration factor would over-accelerate (i.e. over-reduce) the control arm and visa-versa. As a result it is important to consider whether this assumption is valid when assessing whether the RPSFT analysis of the BRIM3 study is biased.

Upon consultation our clinical experts strongly supported the assumption that vemurafenib was as effective as a first and second line treatment. They noted that there was no biological rationale to suggest that prior treatment with dacarbazine would impact upon the efficacy of vemurafenib.

This assumption is strongly supported by the BRIM2 study. BRIM2 was a single arm phase 2 study conducted in previously treated advanced melanoma patients (n=132) which produced results comparable to those seen in BRIM3 (Sosman 2012).

Given this data and the views of our clinical experts it appears reasonable to assume that the acceleration factor associated with vemurafenib is not dependent upon whether it is used first or second line.

(2) There are no differences between the randomised and crossover groups that would be expected to result in differential levels of treatment effect.

If the efficacy of a treatment was improved by the presence of certain patient characteristics (for example biomarkers) and a higher proportion of patients in the crossover group expressed these characteristics the application of an average acceleration factor to this crossover arm would likely under-accelerate (under-reduce) the RPSFT adjusted control arm.

As a result it is important to consider whether there are any differences between the randomised and crossover populations that would be expected to be predictive of treatment effect.

In RCTs randomisation is typically relied upon to ensure predictive and prognostic patient characteristics and spread equally across both arms of a trial (both those characteristics that are observable and those that aren't). In the case of a comparison of the BRIM3 vemurafenib arm and those patients who crossed over to receive vemurafenib it is not possible to rely upon randomisation to assume these predictive factors are balanced as crossover to vemurafenib was not randomised.

Due to this lack of randomisation we will never be able to assess whether there was an imbalance in unobserved predictive factors of the efficacy of vemurafenib between the vemurafenib randomised and crossover patients. We can however assess the patient characteristics recorded for the two groups and assess whether or not these are predictive of the efficacy of vemurafenib.

Table 1 below presents a comparison of the characteristics of patients who received vemurafenib as a randomised and crossover treatment.

Table 1. Comparison of characteristics of patients who received vemurafenib as a randomised treatment and those that received vemurafenib as a crossover treatment

	Randomised to vemurafenib	Crossed over to vemurafenib
Age	55.23	52.48
Male	59.35%	57.83%
Female	40.65%	42.17%
Elevated LDH	42.14%	34.94%
Normal LDH	57.86%	65.06%
Stage M1a	10.09%	8.43%
Stage M1b	18.40%	27.71%
Stage M1c	65.58%	60.24%
Unresectable Stage 3c	5.93%	3.61%
ECOG PS 0	68.68%	67.95%
ECOG PS 1	31.32%	32.05%

The above table indicates that patients who crossed over to receive vemurafenib were similar to those who received it as a randomised treatment in terms of age (mean age 52.48 compared to 55.23), gender (57.83% male compared to 59.35%) and ECOG performance status (67.95% were PS 0 in the crossover arm compared to 68.68% in the randomised arm).

There does appear to be an imbalance between the proportion of patients in each group who had elevated LDH (a negative prognostic factor) - 34.94% of patients in the crossover arm had elevated LDH compared to 42.14% of patients in the randomised arm. This suggests that those patients who crossed over to receive vemurafenib are likely to be a marginally better prognosis group than

the randomised to vemurafenib population. This finding is supported by the staging information highlighted above which indicates a higher proportion of patients randomised to vemurafenib were stage M1c (worse prognosis – 60.24% in the crossover population compared to 65.58% in the randomised population) and a lower proportion of patients were stage M1b (better prognosis - 27.71% in the crossover population compared to 18.40% in the randomised population).

It therefore appears that patients who crossed over to receive vemurafenib were on average a slightly better prognosis group than those randomised to it. Upon consultation our clinical experts indicated that this may be due to the fact that the poorest prognosis patients randomised to dacarbazine would have died before crossover to vemurafenib was recommended by the independent data monitoring committee.

If these marginal prognostic imbalances are somehow predictive of the efficacy of vemurafenib then the assumption that a single acceleration factor can be applied to the randomised and crossover populations of the BRIM3 study will be violated. As the result of this violation the RPSFT adjusted survival curve will be biased.

However, given the relatively minor level of prognostic imbalance observed, and the lack of any evidence to suggest that these factors are predictive of the efficacy of vemurafenib, it appears reasonable to assume that a single acceleration factor can be applied to both the crossover and randomised populations.

Assumption 2: The assumptions underlying the hypothesis test utilised in order to derive the acceleration factor are valid

As the RPSFT method utilises an iterative, test based approach to define the acceleration factor, it is important to consider the assumptions underlying the tests available (i.e. a Log-Rank, Cox-model based or Wilcoxon test) when deciding which test to use.

Use of a Cox or Log-Rank test would require the assumption of proportional hazards (i.e. the relative efficacy of the intervention and comparator do not change over time). This assumption may not be valid in cases where one of the interventions is given for a specific period of time (for example as some chemotherapies are in lung cancer) whilst another is given up to the point of disease progression (as targeted therapies typically are). In this case one would expect the relative efficacy of these two treatment options to change over time (with a greater benefit of the targeted agents over the chemotherapy arm in the period after patients have completed their finite duration of chemotherapy) and so it may not be appropriate to assume proportional hazards model.

If the assumption of proportional hazards does not hold within a data-set then it may be more appropriate to utilise a test that does not rely upon this assumption (such as a Wilcoxon test). In

cases where event rates are imbalanced over time it may be more appropriate to utilise a test which weights events differentially with respect to time (i.e. a Wilcoxon test) rather than one which weights them equally (i.e. the Log-Rank test).

Following the receipt of the second ACD for vemurafenib Roche sought additional academic advice from an expert on the use of RPSFT. During the course of this advice the academic expert noted that we had utilised a Log-Rank model (reliant upon proportional hazards) in order to derive the RPSFT acceleration factor. The academic expert pointed out that this was inconsistent with our modelling of overall survival, and the data from BRIM3, which suggested that proportional hazards did not hold (with the hazard ratio associated with vemurafenib seeming to be lowest in the first few months of treatment).

In light of this the academic expert suggested the use of a Wilcoxon rather than a Log-Rank or Cox model-based test to derive the acceleration factor. Following this guidance the RPSFT analysis was repeated with use of a Wilcoxon test. However, this produced an acceleration factor equivalent to that estimated using the Log-Rank test (i.e. 0.34) suggesting minimal impact on the results presented in response to the first ACD.

Conclusion on the applicability of the RPSFT method for the BRIM3 study

The RPSFT method attempts to simulate the unobserved – a dacarbazine arm un-confounded by crossover representative of that expected in UK practice. As a result it is difficult to assess how well the method has achieved this in lieu of an un-confounded control arm of the BRIM3 study.

In theory if the assumptions underlying the approach are met then the RPSFT adjusted analysis will be unbiased and will represent what would have occurred if crossover had not happened.

What is clear from the above is that one of the key assumptions underlying the original RPSFT analysis conducted by Roche (i.e. the treatment effect is time invariant) does not hold in the case of the BRIM3 study. As a consequence the analysis presented previously based on the February 2012 using RPSFT is biased against vemurafenib in terms of both clinical and cost-effectiveness.

The ICER for this analysis was stated as £52,327 in our previous ACD response – following correction of some minor errors identified by the ERG this figure has fallen to £51,757 (see Appendix 1).

1.2 *A discussion of the plausibility of using alternative approaches to adjust from crossover, or the use of data from other trials to represent the clinical effectiveness of dacarbazine*

Whilst the RPSFT method has become the most commonly used form of crossover adjustment in NICE Technology Appraisals it is not the only method that has been, or could be, employed. There are a range of other approaches available for this purpose including:

- Censoring at the point of crossover
- The Branson and Whitehead method
- Inverse Probability Censoring Weighting (IPCW)
- Use of external data

These methods and plausibility of using each of them in the case of BRIM3 is discussed below.

Further detail on these methods can be found on the Medical Research Council 'Hub for Trials Methodology Research' website (link below).
http://www.methodologyhubs.mrc.ac.uk/workshop_summaries/treatment_switches.aspx

Censoring patients at the point of crossover

When faced with data containing crossover that does not fit the decision problem for an economic evaluation it appears logical to conduct a pseudo per-protocol/per-decision problem analysis in which patient are censored at the point of crossover. Theoretically this approach takes into account as much of the data as fits the decision problem faced whilst leaving the results uninfluenced by the confounded and unrepresentative data post-crossover.

However, there is a fundamental problem with this approach: crossover is not randomised and so is highly likely to be subject to selection bias. If this selection bias is related to the prognosis of patients (as one might expect) then censoring those patients who crossover will result in a comparison of two groups of patients with inherently different prognoses. As a result the difference in outcomes between the two groups cannot be assumed to be due to the intervention given.

This approach was one of the methods evaluated by members of NICE's DSU in their 2010 simulation study (Morden 2010). This publication reached the following conclusion on this methodology: *'Commonly adopted approaches of censoring patients at their switching time or ….. were found to be particularly inappropriate'.*

As a result of the violation of randomisation associated with simple censoring, the potential for selection bias and the findings of Morden et al (2010), the use of a 'per-decision problem' censored

analysis appears inferior to an RPSFT analysis. We have therefore not presented an economic analysis based upon such an analysis and would not advocate it.

It should be noted this same conclusion appears to have been reached in NICE TA179 and NICE TA215. In both these Appraisals the Appraisal Committees were presented with the possibility of making a decision on the basis of RPSFT or an analysis including censoring and in both cases the RPSFT based analysis was favoured.

The Branson and Whitehead Method

The Branson and Whitehead method is an 'Accelerated Failure Time' model similar to RPSFT. It is applied in the same manner as RPSFT (i.e. an acceleration factor is utilised to reduce the time to death for patients who crossed over) but differs in the way in which the acceleration factor is estimated.

Whilst the RPSFT acceleration factor is derived iteratively, and non-parametrically using a test-based approach the Branson and Whitehead method utilises a parametric likelihood based method to define the acceleration factor (Morden 2010).

As a result the Branson and Whitehead approach generally produces treatment effect estimates close to those estimated by RPSFT. The below paragraph is an excerpt from the Morden et al 2010 simulation paper referenced throughout this response which highlights this finding.

"Relationships between these (*the RPSFT treatment effect*) estimates and those from the Branson & Whitehead method are also strong, … . This is to be expected as the model used by Branson & Whitehead takes the same form as that presented by Robins & Tsiatis, differing only by the way in which the estimate of (*the acceleration factor*) is found."

Due to the similarity of the Branson and Whitehead and RPSFT approaches it is subject to many of the same assumptions associated with RPSFT (i.e. those around the time invariance of the acceleration factor) with the additional requirement for the assumption that the times for patients fit a certain parametric distribution. Whilst this approach was found to be marginally less biased than the RPSFT approach in the Morden et al simulation study it should be noted that the survival times used in the study were simulated using a parametric function (a limitation acknowledged by the authors themselves). As a result it would appear logical that this assumption that the assumption the survival times fit a parametric function would not be an issue in the study whereas with real world data this is likely to be questionable. Given the violation of proportional hazards in BRIM3 and the changing treatment effect of vemurafenib over time it appears unlikely that a parametric approach would be appropriate.

In light of this additional assumption, the expected similarity of the acceleration factor estimated, and in the absence of further detailed simulation studies which our academic expert states are required, the resultant survival curves and cost-effectiveness estimates defined by the Branson and Whitehead approach we have not applied this approach to the BRIM3 data.

Inverse Probability Censoring Weighting (IPCW)

The IPCW approach attempts to simulate a comparator arm unaffected by crossover by:

(1) censoring patients who crossed over at the point of crossover and

(2) creating a weighted analysis in which the event times of patients who did not crossover yet had the similar characteristics to those patients who did crossover are given a higher weighting.

The IPCW approach acknowledges that the event times for some patients have not been observed but recognises that it may be possible to draw inference about what their survival times would have been by utilising information on the event times of patients with similar characteristics – this is the logic underlying the method.

However whilst the RPSFT approach is largely objective the IPCW approach is highly subjective. When conducting the analysis all relevant patient characteristics must be pre-specified by the analyst (required in order to weight the final adjusted analysis to give a higher weighting to patients 'similar' to those censored). In addition it must be assumed that these covariates capture all variability in expected event times (i.e. there are no unmeasured confounders). This assumption is clearly contentious as clinical trials record only limited pre-specified prognostic factors and there could be a wealth of unobserved confounding factors that would influence survival times.

To date the IPCW approach has been utilised in two NICE Appraisals (NICE TA215, NICE TA219). However in each of these Appraisals both the ERGs assessing the manufacturer's submission and the Appraisal Committees have expressed a preference for the RPSFT method.

In light of the subjectivity of the IPCW approach, the challenge in defining all appropriate covariates and the view of the Appraisal Committee's in TA215 and TA219 we believe the RPSFT method is more appropriate than the IPCW approach. Due to this, and the intensive computational burden associated with IPCW, we have not conducted this analysis on the BRIM3 data.

Use of External Data

All the methods highlighted above have a common aim in attempting to simulate a control arm unaffected by crossover. Whilst the methods highlighted so far do this via use of statistical methods of varying complexity there is also a pragmatic method of simulating this – by using

external evidence on the efficacy of dacarbazine from alternative, similar, clinical trials. The control arms from other clinical trials conducted prior to the development of vemurafenib were not subject to crossover onto vemurafenib. It may therefore be reasonable to utilise these as a proxy for an un-confounded dacarbazine arm in the model. In order for this method to be unbiased it is important that those patients in the study used as a proxy are similar to those in the study impacted by crossover.

This 'external data' approach was utilised as a sensitivity analysis in our previous ACD response in which the data from the control arm of Bedikian 2011 was utilised in the model with the result that the ICER fell to £45,003 (note – this figure has been amended to £44,405 on the basis of comments made by the ERG following the first ACD - see Appendix 1 for further detail).

A comparison of the characteristics of patients in the dacarbazine arm of BRIM3 and that in the dacarbazine arms of the four studies digitised in our previous ACD response (Bedikian 2011, Robert 2011, Patel 2011 and Bedikian 2006) is provided below.

Table 2. Comparison of characteristics of patients randomized to dacarbazine in BRIM3 and Bedikian 2011

	BRIM3	Bedikian 2011	Robert 2011	Patel 2011	Bedikian 2006
Age (median)	52	62	56[2]	61%>65 years[3]	60
Male	54%	68%	59%	59%	66%
Female	46%	32%	41%	41%	34%
Elevated LDH	58%	NR	44%	34%	36%
Normal LDH	42%	NR	56%	66%	64%
Stage M1a	12%	7%	17%	NR	NR
Stage M1b	19%	24%	25%	NR	NR
Stage M1c	65%	69%	55%	NR	NR
Unresectable Stage 3c	4%	3.61%	3%	NR	NR
ECOG PS 0	68%	50%	71%	68%	91.7% PS0-1[4]
ECOG PS 1	32%	44%[1]	29%	32%	

1. 5.5% of patients were PS 2 and 0.5% of patients were PS 3, 2. Mean age – median not reported, 3. Median and mean age not reported, 4. 7.5% patients were PS 2 and 0.5% were PS

Given the information available on each of these studies it is difficult to evaluate the level of bias associated with use of these data sources as proxies for the BRIM3 control arm. There appears to be inconsistent reporting of known prognostic factors (i.e. proportion of patients with elevated LDH, disease staging) between studies which makes use of this data as a proxy problematic. However given the similarity of the BRIM3 data and these external data sources prior to the point of crossover it may be reasonable to utilise this data.

Upon consultation our academic expert noted that research looking at expanding this approach is currently underway. [redacted]

In light of this, the lack of data on known prognostic factors within the data identified and the lack of information available on unobserved patient characteristics in BRIM3/external data (which may confound the data observed), it may be wise to consider the naïve use of external data with caution and as *one* potential way of adjusting for crossover amongst many.

Conclusion on the methods available to adjust for crossover in BRIM3

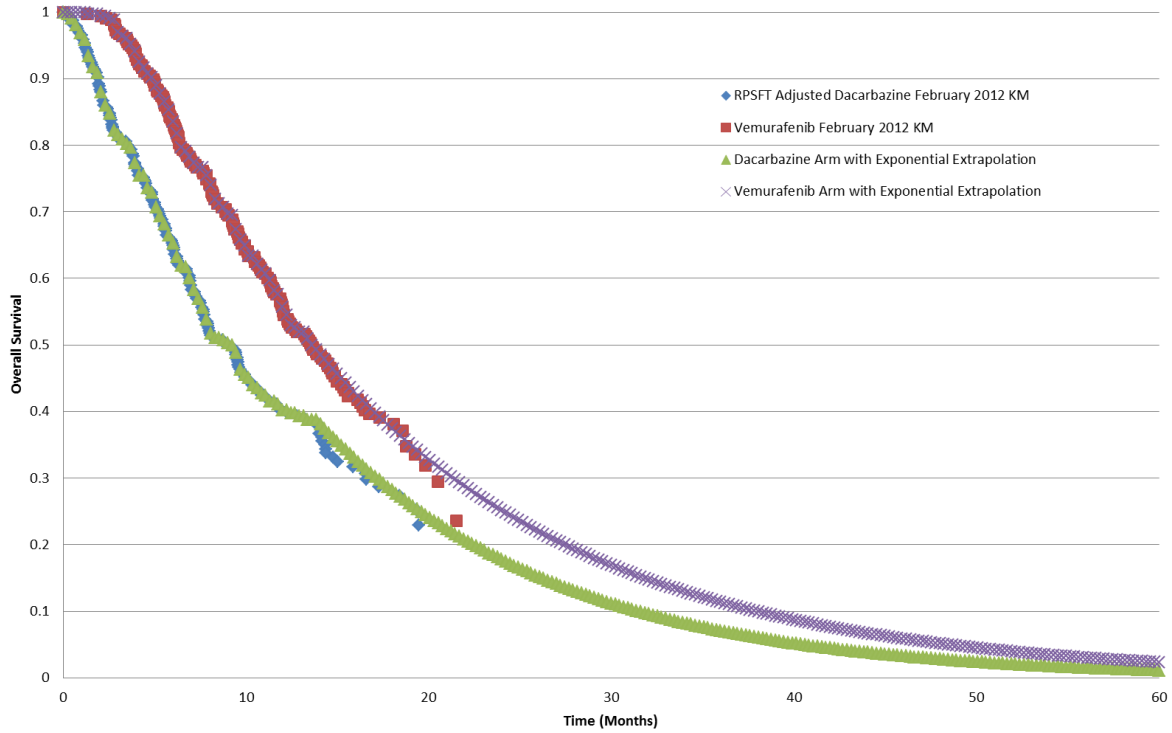
Adjusting for crossover will always be subject to uncertainty. As each method attempts to simulate the unobserved it is not possible to truly validate whether the output simulated is unbiased (given there is nothing to compare it to and if there were crossover adjustment would not be required). What can be assessed is the extent to which the assumptions underlying the methods are valid – whilst it is not possible to confirm that all these assumptions hold (as many rely upon assumptions about the unknown or unobserved) it is possible to utilise the data available in order to draw inference about the extent to which they are likely to hold.

Given the above and the information provided in the literature on this topic it appears the RPSFT or Branson and Whitehead accelerated failure time models are likely to produce the least biased adjustment of the BRIM3 data. Due to the fact that these two methods are highly related and are expected to produce similar acceleration factors we have not conducted the Branson and Whitehead approach and believe the revised RPSFT method provided within this document should provide a reasonable basis for modelling a dacarbazine arm in an economic model.

2. *An additional scenario analysis for vemurafenib compared with dacarbazine, estimated by separately applying exponential hazards to each arm of the BRIM3 study (using the February 2012 data cut-off) from 14 months*

Figure 5 below demonstrates the survival curves produced if the Committee's requested analysis is conducted.

Figure 5. Committee Requested Analysis – Exponential hazards fitted separately to each arm of the February 2012 RPSFT adjusted cut of BRIM3 and applied from month 14 onwards



As stated at the beginning of this response, we do not believe this extrapolation is appropriate for the following reasons:

1. When combined with PFS curves from BRIM3 this modelling approach suggests that post-progression survival following treatment with vemurafenib is 2.2 months shorter than post-progression survival following dacarbazine. This appears implausible.

2. Historical registry data (the SEER Registry data, the Balch AJCC staging data) and data from more mature advanced melanoma RCTs (i.e. Robins 2011) suggest that the probability of death associated with melanoma reduces over time – this issue was discussed in length in our initial evidence submission (pages 154-178). An exponential model assumes a constant probability of death over time. As a result the use of an exponential modelling approach has poor external validity.

3. We believe the ‘kink’ observed in the dacarbazine arm at around 8 months is a product of RPSFT under-accelerating the dacarbazine arm when crossover to vemurafenib was in its earliest stage (when vemurafenib is most effective). As the treatment effect of vemurafenib appears to be time dependent (with no events occurring in the vemurafenib arm in the first few months of treatment) and the RPSFT approach operates by deriving an *average* acceleration factor across the study and then applying it to solely this period it is highly likely that the dacarbazine arm is under accelerated. This has been described in detail on pages 5-9 above. As a result the hazard observed in the dacarbazine appears to change suddenly (i.e. the curve “kinks”). We therefore believe that after month 8 the dacarbazine arm should be treated with caution and not relied upon for modelling. We therefore believe that the Committee’s suggestion that the dacarbazine KM data be used up to month 14 is not an appropriate way of modelling a dacarbazine arm.

As a result of the above we do not believe this extrapolation is plausible. We have therefore not presented cost-effectiveness results associated with its use.

However, given the ICERs presented in our previous ACD response (notably £52,327/QALY gained if the RPSFT data is combined with the long term extrapolation utilised in the Roche submission) it is clear that use of this substantially more pessimistic extrapolation would produce cost-effectiveness estimates above the range typically considered acceptable.

In light of the above discussion we firmly believe the RPSFT model provided in our previous response provides the most appropriate basis for decision making.

References

Bedikian, Agop Y; Millward, Michael; Pehamberger, Hubert; Conry, Robert; Gore, Martin; et al. 'Bcl-2 antisense (oblimersen sodium) plus dacarbazine in patients with advanced melanoma: the Oblimersen Melanoma Study Group.' Journal of clinical oncology : official journal of the American Society of Clinical Oncology 24. 29: 4738-45. (Oct 10, 2006)

Bedikian, A Y; DeConti, R C; Conry, R; Agarwala, S; Papadopoulos, N; et al. 'Phase 3 study of docosahexaenoic acid-paclitaxel versus dacarbazine in patients with metastatic malignant melanoma'. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 22. 4: 787-93. (Apr 2011)

J Morden, P Lambert, N Latimer, K Abrams, A Wailoo. Assessing methods for dealing with treatment switching in randomised clinical trials: A simulation study *BMC Medical Research Methodology* 2011; Vol 11:4

NICE ID56: Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer http://guidance.nice.org.uk/TA/Wave19/56 (last accessed on 06/07/2012)

NICE TA179: Sunitinib for the treatment of gastrointestinal stromal tumours http://guidance.nice.org.uk/TA179 (last accessed on 06/07/2012)

NICE TA215: Pazopanib for the first line treatment of metastatic renal cell carcinoma http://guidance.nice.org.uk/TA215 (last accessed on 06/07/2012)

NICE TA219: Everolimus for the second-line treatment of advanced renal cell carcinoma http://guidance.nice.org.uk/TA219 (last accessed on 06/07/2012)

NICE TA257: Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2 http://guidance.nice.org.uk/TA257 (last accessed on 06/07/2012)

Patel, Poulam M; Suciu, Stefan; Mortier, Laurent; Kruit, Wim H; Robert, Caroline; et al. 'Extended schedule, escalated dose temozolomide versus dacarbazine in stage IV melanoma: final results of a randomised phase III study (EORTC 18032)'. European journal of cancer (Oxford, England : 1990) 47. 10: 1476-83. (Jul 2011)

Robert, Caroline; Thomas, Luc; Bondarenko, Igor; O'Day, Steven; M D, Jeffrey Weber; et al. 'Ipilimumab plus dacarbazine for previously untreated metastatic melanoma'. The New England journal of medicine 364. 26: 2517-26. (Jun 30, 2011)

Robins JM, Tsiatis AA: Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991, 20(8):2609-2631

Appendix - ERG Comments on ACD response model

Modelling problems and errors identified by the ERG in the model submitted by Roche in response to the ACD

Item	Issue	Detail	ERG commentary	Roche Response
1	Double discounting of terminal care costs	Terminal care costs (column U in 'Vemurafenib' sheet, column AA in 'Dacarbazine' sheet) are calculated from figures in column L in 'Vemurafenib', and column R in 'Dacarbazine', which have been discounted using the utility discount rate. However, the formula applied to derive terminal care costs applies a further discounting factor using the cost discount rate.	This error understates terminal care costs in each arm, and hence also understates incremental terminal care costs	The ERG are correct that this is an error within the model – revising the model to account for this caused the ICERs estimated to increase marginally (RPSFT adjusted base case increased by £213/QALY gained)
2	OS model options	In weeks 0-40 the 'Vemurafenib' sheet offers 6 OS options corresponding to the options detailed on the 'Model Inputs' sheet. However the corresponding cells on the 'Dacarbazine' sheet offer 7 OS options.	Option 6 (Bedikian 2011 comparator) is compared with ERG OS and not to the Kaplan-Meier February data which seems to be the logical choice consistent with a sensitivity analysis relative to the manufacturer's preferred option (#5). This suggests that the quoted ICER of £45,003/QALY may be inaccurate and/or misleading.	The ERG are correct – the Bedikian based extrapolation was intended to be compared to the February 2012 vemurafenib data rather than the ERG model - this has been amended accordingly. As a result the ICER associated with this sensitivity has reduced marginally (by approximately £500).
3	Use of RPSFT in model	RPSFT results have only been used up to week 40, reverting thereafter to the Robert/SEER based complex projection method.	The RPSFT method generates adjusted dacarbazine results up to week 80. The ERG notes that using the full RPSFT results is likely to give a less favourable ICER for vemurafenib.	This was intentional – the reasons for this are specified in our previous ACD response and in this document.
4	Overall Survival censoring	Figure 2 (p6) of the manufacturer's response to the ACD shows no censoring of patients in the dacarbazine from month 1 to month 15. By contrast, the RPSFT adjusted plot for	This anomaly needs to be resolved as it forms the basis of the manufacturer's justification for not using the RPSFT results beyond 8	This appears to be due to a lack of clarity in 'Figure 2' in our ACD response. There was censoring in the first few months of the ITT

Item	Issue	Detail	ERG commentary	Roche Response
		dacarbazine (Figure 3) shows multiple cases censored during this period.	months in Appendix 1.	data but this appears not to be shown clearly in the powerpoint figure provided. Below this table a small sample of ITT data from the dacarbazine arm is provided – the data in this table clearly shows censoring (the asterisks) in the early part of this data.
5	One-off progression costs	One-off costs arising following disease progression are estimated in column S of the 'Vemurafenib' sheet and column Y of the 'Dacarbazine' sheet. These costs are combined with the health state costs in PPS in a single formula. However, when disaggregated the one-off costs show unusual features: after the first few weeks negative estimates begin to appear and after the first year virtually all estimates are negative. The overall totals for the base case are £68.71 for the average vemurafenib patient, and £50.63 for the average dacarbazine patient, despite the standard one-off cost per patient being £648 and virtually all patients dying or progressing during the model.	The problem has arisen because the one-off cost has been calculated from the estimated post-progression survival (column H in 'Vemurafenib' and column N in 'Dacarbazine'), which has been incorrectly labelled as 'Progression'. Hidden columns in the 'Dacarbazine' sheet suggest that an earlier version of the model may have attempted to use a modelling approach based on following patients through different pathways from progression-free to early death, or post-progression leading to death. However, this seems to have been abandoned in favour of a simpler approach in which PPS is estimated as the difference between OS and PFS. This makes the estimation of one-off costs more difficult, especially as the one-off progression cost will only apply to those PFS failures which are not fatal (we estimate	This point appears valid – this error was not noted previously due to the aggregation of the two progression related costs. This has been corrected in the model by applying a one off undiscounted cost of progression to those patients who experienced progression in each arm (i.e. not implemented for those patients who died in PFS). This caused the ICER estimated to increase by £17.

Item	Issue	Detail	ERG commentary	Roche Response
			progression event fatalities at around 19% for vemurafenib patients and 24% for dacarbazine patients leading to greater costs for vemurafenib).	
6	BRAF testing cost double counted	In the 'Verumafenib' sheet, the cost of BRAF testing (cell O6) is counted twice in Q6 (in its own right and as part of P6).	This overstates vemurafenib costs.	This has been corrected with the result that the ICERs estimated are reduced marginally (a £240 reduction in the cost per QALY gained).
7	Missing OS data	When OS option 4 is selected, cells F1567:F1571 of the 'Verumafenib' sheet and J1567:J1568 of the 'Dacarbazine' sheet are zero, because the formulae refer to cells in the 'KM OS (with new OS)' sheet which are empty. As a result, the corresponding calculated PPS values are negative (albeit very small!).	This type of error (negative imputed values) is a common problem with the simple model structure used in the model, and can be readily detected by use of a check variable to alert the modeller to the presence of any infeasible values.	This has been corrected. As these extremely small negative values in the last 5 weeks of the model resulted in the survival associated with dacarbazine being overestimated by approximately 45 minutes the correction of this has no meaningful impact upon the ICER.
8	Half-cycle correction formulae	Half-cycle corrected values are calculated in columns G, H and I in the ‘Vemurafenib’ sheet and in columns M, N and O in the ‘Dacarbazine’ sheet. Incorrect time periods have been used in rows 1142-1571 in the ‘vemurafenib’ sheet and rows 1140-1568 in the ‘dacarbazine’ sheet.		Correcting this error reduced the ICER by £560.

The ICER associated with the RPSFT adjusted analysis including correction of errors 1, 5, 6 and 8 is £51,757 rather than the £52,327 stated with our previous ACD response.

Stratum 1: TRTC = DTIC						
Product-Limit Survival Estimates						
TTDIEDM		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0300	*				[redacted]	[redacted]
0.0700	*				[redacted]	[redacted]
0.0700	*				[redacted]	[redacted]
0.1000	*				[redacted]	[redacted]
0.1000	*				[redacted]	[redacted]
0.1000	*				[redacted]	[redacted]
0.1000	*				[redacted]	[redacted]
0.1300		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.1600	*				[redacted]	[redacted]

Stratum 1: TRTC = DTIC						
Product-Limit Survival Estimates						
TTDIEDM		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.2000	*				[redacted]	[redacted]
0.2000	*				[redacted]	[redacted]
0.2000	*				[redacted]	[redacted]
0.2600		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.2600	*				[redacted]	[redacted]
0.3000		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.3000	*				[redacted]	[redacted]
0.5300		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.5600					[redacted]	[redacted]
0.5600		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.6900	*				[redacted]	[redacted]
0.7200		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.7900		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.8200		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.8900		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0.9500	*				[redacted]	[redacted]