Author’s response to reviews

Title: Implementing the EffTox Dose-Finding Design in the Matchpoint Trial

Authors:

Kristian Brock (k.brock@bham.ac.uk; kristian.brock@gmail.com)
Lucinda Billingham (L.J.Billingham@bham.ac.uk)
Mhairi Copland (Mhairi.Copland@glasgow.ac.uk)
Shamyla Siddique (S.Siddique@bham.ac.uk)
Mirjana Sirovica (M.Sirovica@bham.ac.uk)
Christina Yap (C.Yap@bham.ac.uk)

Version: 1 Date: 01 Jun 2017

Author’s response to reviews:

Authors’ responses to referees’ comments on ‘BMC Medical Research manuscript BMRM-D-17-00049 “Implementing the EffTox Dose-Finding Design in the Matchpoint Trial” by Brock et al.

Dear Sir / Madam,

We extend our sincere gratitude to Emily Dressler, Peter Thall, and the editors of BMC Medical Research Methodology for devoting the time to review our manuscript and suggest ways in which it may be improved. We have annotated a revised manuscript. Additions are highlighted in yellow, and removals are shown struck-through.

In specific response to the two reviewers’ comments:
Responses to Points Raised by Emily Dressler (in order)

1. We apologise that the references were not shown in the manuscript. We appreciate that this must have been frustrating during review. It appears the TeX script and the BibTex bibliography file we uploaded did not build correctly in the BMC system. We will upload an article PDF built on our own machines to address this on re-submission.

2. We have added a sentence to Background clarifying that the optimal dose varies by trial design, and another sentence lower down to clarify that in EffTox, the optimal dose is the admissible dose with the greatest efficacy-toxicity trade-off score.

3. We have named in Background the TriCRM design.

4. We removed the clause in Background identifying EffTox as our choice because that specific section should just be about EffTox. Our motivations for choosing EffTox are detailed in the later section headed "EffTox in the Matchpoint trial".

5. We added a reference to the quote in Background.

6. We added a sentence to Methods to describe $L^p$ norms, and another to introduce the popular example of the $L^2$ norm, aka Euclidean norm, that determines the length of the hypotenuse in a right triangle.

7. These are not typos; the bars below and over $\pi_E$ and $\pi_T$ denote lowest & highest thresholds, respectively. We re-use the notation previously used by Thall et al.

8. We certainly agree that being able to appropriately observe efficacy and toxicity outcomes over a similar time horizon is a significant constraint to the use of joint efficacy-toxicity dose-finding designs. We describe our expectations on response with respect to time-horizon in the paragraph following your comment. We also added a reference to the Late Onset EffTox design, which ameliorates the issue.
9. The motivation to start at 30mg is an important point. Thus, we added sentences to the "EffTox in the Matchpoint trial" section to stress: the difference from a typical phase I scenario of starting low; the desire to avoid both under- and over-dosing; the acute nature of blast phase CML; the prior evidence of tolerability of 45mg ponatinib as monotherapy in CML patients.

10. We made the zero-utility contour bolder in Figure 1, made the other lines slightly lighter, and added blue triangles to show the neutral points.

11. Thanks for the compliment. We like the nomenclature too. It proved really useful, for e.g. it made monitoring reports more concise and intuitive.

12. We removed the paragraph in the Dose Ambivalence section that was perhaps too stringent in advocating supplementing the model with clinician view. Our intention was to say that dose ambivalence does manifest and when it does, supporting information can be considered.

13. The section on recalibrating pE to 0.03 was too big and the message was diluted. We have removed five paragraphs, and added a few sentences to stress the key point: trialists should analyse DTPs, especially early on, to monitor for undesirable behaviour. Furthermore, when doing so, they need to consider the interaction of contours and admissibility rules.

Responses to Points Raised by Peter Thall (with matching numbers)

1. Thank you for complimenting us on the extent of our preparation to use EffTox.

2. Thank you for bringing the 2014 paper and 2016 book to our attention. We heed the advice, and have cited each. We have addressed this point in multiple places. In the Methods section, we have added a paragraph to highlight the problem, the symptom, and
the solution to insufficiency steep contours. We have also stressed that the problem was not reported in earlier manuscripts, and that the 2004 paper contained a pathological example. We stressed that trialists should be mindful to select contours that are steep. We edited our existing prose to stress selecting points for steep contours rather than necessarily eliciting from clinicians, and referred to the advice contained in your review. We re-emphasise this topic in the Discussion section. Furthermore, we performed new simulations with the general neutral point set to \((\piE, \piT) = (0.6, 0.3)\). The outcomes of those simulations are discussed below.

3. We explained the above issues. Additionally: we described that EffTox has evolved; we pointed out that low ESS can lead to pathological design behaviour.

4. We agree that a greater sample size would be beneficial and that seamless phase I-II trials address the phase II objective too. We added a paragraph in the Operating Characteristics section to describe the performance gain from using \(n=60\); the afforded chance to optimise delivery of a new agent; and the abrogation of traditional phase II by seamless phase I-II designs.

5. We added sentences in two locations to point out that Thall et al. 2014 use \(pE = pT = 0.1\). We also presented in the Operating Characteristics section the benefit of using the standard thresholds, very slightly mitigated by an increased chance of stopping in some other scenarios where an optimal dose exists.

6. We strongly agree on the importance of starting at a dose believed a-priori to be therapeutic, and welcome the suggestion to elaborate. Thus, we added sentences to Methods to stress: the difference from a typical phase I scenario of starting low; the desire to avoid both under- and over-dosing; the acute nature of blast phase CML; the prior evidence of tolerability of 45mg ponatinib as monotherapy in CML patients.

7. Thank you for alerting us to the LO-ET design. We have cited the design and identified it as a formal approach to the general problem we described.

8. We added sentences to describe that uncertainty is present in all sequentially adaptive designs with small/moderate sample sizes and cited the Thall et al. 2017 example.
9. We removed the superfluous sentence from the caption of Figure 1.

10. We have added sentences to the DTPs section to describe the co-operative nature of admissibility criteria and contours; and noted that the recommendations can sometimes seem counter-intuitive.

Minors:

1. We added "on average".

2. We have changed the reference to Thall et al. 2014.

3. We removed the citation to EffTox software and added sentences with URL and version number at the start of the Methods section.

4. Thank you for clarifying this.

Further general edits

• We added a note to Background to clarify that the original authors use the term desirability where we use utility.

• We rewrote the paragraph in Methods on calculating the neutral utility points. Now it is more succinct and in-keeping with Thall’s advice on not necessarily eliciting these points. We also refer to the Discussion where we consider alternative contours.

• We removed a paragraph in the Dose Ambivalence section that was perhaps too stringent in advocating supplementing the model with clinician view. Our intention was to say that dose ambivalence does manifest and when it does, supporting information can be considered.
Further simulations undertaken

We ran further simulations:

A. with the general neutral desirability point \((\pi_E, \pi_T) = (0.6, 0.3)\), as suggested by Thall, but retaining our values of \(p_E = 0.03\) and \(p_T = 0.05\);

B. with the general neutral desirability point \((\pi_E, \pi_T) = (0.6, 0.3)\), as suggested by Thall, and the more typical values \(p_E = p_T = 0.1\);

C. using our design with \(p_E = p_T = 0.1\);

D. using our design with \(n=60\) to consider the benefit of a larger sample size.

Simulated operating characteristics (OCs) produced using v4.0.12 of the EffTox software are shown in Appendices A – E of this letter. Appendix E contains the OCs shown in the manuscript using our design with \(ESS = 1.3\), reproduced here for convenience of reference. Each simulation study was conducted using 10,000 iterations and a seed value of 10502.

With respect to the manuscript OCs in Appendix E:

- Appendix A shows that the steeper contours slightly improve appendix scenario 1, materially harm scenarios 2 & 5, marginally harm scenario 3, make no difference in 4, and redefine the best dose in scenario 6 (making comparison difficult);
• Appendix B shows that the steeper contours with traditional values for pE & pT improve scenario 1, materially harm scenarios 2 & 5, materially improve the stopping probability in scenario 3, and practically guarantee stopping in scenario 4, and once again redefine best dose in scenario 6;

• Appendix C shows that traditional values for pE & pT improve scenario 1, harm scenario 2, materially improve the stopping probability in scenario 3, and practically guarantee stopping in scenario 4, and leave scenarios 5 & 6 largely unchanged. In spite of this, our motivation for changing pE was to avoid premature trial stopping without exploring lower doses first. We describe this in the “Changing pE to avoid premature stopping” section;

• Appendix D shows that a sample size of n = 60 materially improve scenarios 1 & 3, and leaves the other four scenarios largely unchanged. We describe this in the Operating Characteristics section of the manuscript.

Please note that the order of the scenarios in our EffTox parameter files (and thus Appendices A – E) is not the same as the manuscript. If cross-referencing appendices and manuscript, please note the following equivalence of scenarios numbers:

Scenario number in manuscript Scenario number in this letter, Appendices A–E
1 6
2 1
3 2
4 5
5 3
6 4

However, cross-referencing should not be necessary because we have reproduced the manuscript OCs here (Appendix E) with the new OCs in the same congruent format.
We hope you agree that we have improved the manuscript by clarifying some of the key points whilst reducing the overall word count by pruning some superfluous sections. We did this thanks to the comments from our reviewers and the journal editor. Once again, we extend our sincere gratitude.

We include an annotated revised manuscript as requested. We are happy to provide a clean manuscript with all changes accepted if this is preferable to the referees or editors.

Finally, is it possible that Kristian Brock and Christina Yap (email: c.yap@bham.ac.uk) could both be denoted as corresponding authors? If it is a requirement that there be a single corresponding author, please leave it as KB.

Yours sincerely,

Kristian Brock

Cancer Research UK Clinical Trials Unit, University of Birmingham, UK

<simulation results appear here in the Word file of the cover letter, uploaded as a supplementary file>