Author’s response to reviews

Title: When to keep it simple – adaptive designs are not always useful

Authors:

James Wason (james.wason@mrc-bsu.cam.ac.uk)

Peter Brocklehurst (P.Brocklehurst@bham.ac.uk)

Christina Yap (C.Yap@bham.ac.uk)

Version: 2 Date: 06 Jun 2019

Author's response to reviews:

We would like to thank the editors and Drs Mills, Hey and Ivanova for their supportive and helpful comments. We have provided our point-by-point response to the comments below. All line/page numbers refer to the clean version of the manuscript.

Editorial comments

1.) Please add a list of abbreviations.

This has been added.

2.) Please structure your abstract. Suggested section headings are 'Background', 'Main text' and 'Conclusions'.

We have followed the suggested abstract structure. The abstract has been slightly edited to fit the structure suggested.
"In the section, "Long-term outcomes", the authors discuss the importance considering the delay between recruitment and assessment of outcomes, and how the outcome being used to make adaptations should be observed sufficiently quickly compared to the total recruitment length of the trial. They provide their first real-life example of two-arm, phase 2 immunotherapy trial (ISRCTN: 11889464). In this example, they discuss why a single-stage fixed sample design with a maximum sample size of 70 participants was chosen over another design that would have potentially allowed for early stopping for futility after 46 participants (23 per arm), based on projected recruitment rate of 2 participants per month.

1) The current figure is not very clear. I would recommend a different figure than Figure 1 provided to illustrate how at the time of 23 months (when 46 participants are observed for at least 12 months), all of participants would have already been recruited at projected rate of 2 participants per month."

We have replaced figure 1 with a new figure demonstrating the timeline of a trial (and its interim analysis) in the case of 2 participants recruited per month and 1 participant recruited per month.

"2) It would be important to highlight the potential savings if projected rate for recruitment was erroneous; for instance, if the actual recruitment rate of 1 participant per month is observed, there would have been some potential saving gained from a planned interim futility analysis. As well, there would have been some savings if an outcome of 6-month PFS rate that is correlated with 12-month PFS rate was used as an intermediate outcome for adaptation."

We agree this is useful additional detail and we have added it to page 6 lines 134-136.

"3) More detailed figure legend would be beneficial to explain Figure 2 for expected sample size (2a) and proportion of patients allocated to effective dose (2b)"

We have added more a more informative caption to figure 2.
"4) In lines 163-168, the authors mention that the actual time taken to conduct the interim analysis in their example of TAILoR trial. If the authors of this manuscript were the investigators of this trial, it would be nice to mention the time taken for the interim analysis, given that this trial has already been published."

None of the authors of this work were investigators on the TAILoR trial – we used this example as there was a lot of information in the public domain about the planned recruitment rate. Unfortunately, we do not know how long the interim analysis actually took to do.

"5) In section, "Limitations due to additional administrative and logistical complexity", the authors discuss logistical challenges associated with conducting data collection and cleaning in a timely manner to first conduct the interim analysis during the trial, as well as the challenges of having data monitoring committees (DMCs) and trial steering committees (TSCs). But it should be noted that ensuring high-quality data collection and having independent DMCs should be standards (as the investigators of the STAMPEDE trial were able to) that all clinical trial research should thrive for. These attributes will not only benefit future adaptive trial designs but also non-adaptive, single-stage fixed sample designs."

This is a good point. Intuitively it feels that an adaptive design that is making an irreversible decision at an interim analysis should have a higher bar for data quality (and inclusion of as much outcome data as possible) compared to what a non-adaptive trial would require for a DMC. However admittedly this is not based on any research that has been conducted so far. We (JW and CY) are involved in a UK working group that is planning to investigate what additional, if any, resources are required for an adaptive design to be run successfully. This may go some way to answer this question.

We have added the following the text on page 10 to soften what we previously said and ensure this point is clear:

It would be interesting to contrast the above requirements for adaptive designs to the needs for data monitoring in non-adaptive designs. It is our opinion that it is likely adaptive trials should require more resources, but this requires further research.

"6) The last section of "Weighting the pros and cons of adaptive designs" is really important. I recommend the authors to build up on this section. If they are constrained against by BMC Medicine's word limit, I would recommend taking the section out on the BATTLE-2 trial. In such a biomarker-guided trial, there are other issues of ensuring good quality genomic data that I feel is beyond the scope of the manuscript."
We agree this section is very important and have added additional text on that motivates further methodology research and guidelines. We are still within the suggested word count so have kept the BATTLE-2 section in, although will remove it if suggested by the editors.

"7) In this section, it would be nice to the authors to mention how no methods (adaptive or non-adaptive) should be a default. Careful considerations of recruitment, data quality, choices in outcomes, administrative and logistical complexity while following the KISS (keep it simple, stupid) principles using simulations and other ways should be a standard practice that future investigators should adopt for their trial planning."

Absolutely, we very much agree with this and have added this to the ‘Weighing the pros and cons of adaptive designs’ section.

Reviewer #2

"For the most part, this is a thoughtful and well-reasoned argument. However, there is an assumption throughout that adaptive designs are ethically advantageous because they may increase the likelihood that a participant will be allocated to the arm that is eventually found to be superior. Although this is not an uncommon assumption, it is in the very least controversial, and, in my view, it is simply incorrect. There are at least three reasons for this:

(1) Clinical trials cannot be ethically justified by direct benefits to the participants, because benefit is precisely what is being determined by the trial. Therefore, trials are justified by the *potential* benefits to *future* patients. And (part of) what makes a trial ethical is when these potential future benefits offset the actual risks and burdens to which the participants will be exposed. Thus, to claim that adaptive trials are more ethical because they increase the prospect of direct benefit is fundamentally at odds with the widely-held principles of research ethics."

We agree and have softened our reference to ethical properties of adaptive trials with new text on lines 60-63.
"(2) Notwithstanding the above, direct benefit to trial participants is not irrelevant to the ethical evaluation of trials. However, the way that patient benefit is often (although not always) invoked is through the ethical requirement of clinical equipoise, which demands that there is genuine uncertainty in the expert community about the relative therapeutic merits of all arms in a controlled trial. The consequences of clinical equipoise that are most relevant to the argument here (it seems to me) are (a) that patients should not be exposed to anything less than competent care and (b) patients should not be systematically disadvantaged by participating in the trial. In other words, clinical equipoise requires that patients will not be harmed by being allocated to one arm or another.

Adaptive trials interface with clinical equipoise is some interesting and complex ways—and it is probably beyond the scope and interests of this paper to really get into those matters. Nevertheless, I think there may be at least a few points that the authors should consider: On the one hand, adaptive trials which drop arms (like the MAMS design), seem to operationalize clinical equipoise in a really intuitive way: Because as soon as we know that an arm is ineffective, we drop it from the trial. This seems to straightforwardly satisfy equipoise requirements."

This is interesting as often MAMS or similar designs are set up to drop arms that are not showing sufficient promise. An example is the STAMPEDE trial that requires treatments to reach a certain level of activity on PFS. It would be more that the design drops arms that are insufficiently likely to demonstrate benefit rather than being definitively ineffective – presumably in this case there could still be equipoise about arms that are dropped.

"However, for adaptive trials that weight allocation in favor of "better performing" arms, the implications of equipoise are less intuitive. As long as there is (and should still be) uncertainty about which is truly the better intervention, then (consistent with point 1 above), there is no good reason to think that patients are benefitting more or less by ending up in one arm or the other. In other words: Because every arm must be consistent with clinical equipoise, then patients are not being advantaged or disadvantaged due to their allocation. In which case, it is mistake to say that there is an ethical advantage by allocating more patients to the superior arm. The (potential) ethical advantage of adaptive designs is rather from minimizing the total patient burden (in terms of sample size, time, number of visits, trial burdens, costs to society, etc.)."

We agree and have removed all references to ‘ethics’ or ‘ethical’ apart from the new text in the second paragraph of the background.
"(3) Finally, it may be important to acknowledge how adaptive trials can introduce a systematic disadvantage for patients who enroll earlier in the trial. In a MAMS trial, for example, since some arms may be dropped over the course of the study due to lack of efficacy or futility, it would be better for a patient to wait as long as possible to enroll, since this will maximize their chance of getting something that is effective. Assuming that equipoise holds throughout the trial, this temporal asymmetry does not make the trial less/more unethical from that perspective. But it may raise ethical concerns from justice insofar as adaptive trials redistribute some of the burdens of trial participation onto early-enrolling patients, who may be more vulnerable. For some conditions and patient populations, this may not be relevant. But in some kinds of adaptive trials in oncology, for example, I think this can be an important consideration that may intersect with or amplify some of the authors concerns about trial complexity. And this issue also interfaces with validity concerns, if the trial population is changing over time."

This an interesting point that is often thought about from a statistical point of view (i.e. does this introduce differences in patients during the trial that might require the analysis to account for) but we had not thought about this from an ethical point of view. It would certainly be important for an adaptive design to not intentionally disadvantage earlier patients in order to advantage later patients in the trial; we feel that all adaptive designs we know of would meet this requirement but clearly might be neglecting aspects of ethics that we are not experts in.

"Bottom line here: I don't think the authors need to get into all of this. But I do think that at least some of this ethical complexity needs to be surfaced, and that doing so would enrich their analysis and discussion."

Thank you we have attempted to do this in the new text. We think a dedicated paper that looked at this in general within adaptive trials would be very interesting and would encourage Dr Hey to do this!

"For more on many of these points, the authors may find the series of articles from 2015 in Clinical Trials helpful (starting with Hey and Kimmelman's "Are Outcome-Adaptive Randomization Trials Ethical?")), since many of these issues are discussed there as well."

Thank you, we have added this reference and a subsequent one that was relevant.
"1. Though the title reads well, I think the title "When to keep it simple - adaptive designs are not always useful" is too broad for what is in the paper. Perhaps it can be something like "When to keep it simple - adaptive designs are not useful when the outcome is delayed". If I saw a paper with the original title I would have expected to read about other reasons for adaptive designs not being useful. For example, about cases when adaptive is not that useful when the outcome is observed relatively quickly, and something like in Korn EL and Freidlin B (2011). Outcome-adaptive randomization: is it useful? Journal of Clinical Oncology 29, 771-776. This paper on a similar topic should be mentioned and referenced. Some of the similar points about outcome-adaptive randomization are made in Lai, Liao, Kim " Group sequential designs for developing and testing biomarker-guided personalized therapies in comparative effectiveness research" CCT 2013."

Regarding the title, we do feel that we have considered a broader range of issues than just delay in the outcome and just referring to that aspect might look too focused. We would therefore prefer to keep it as it is. Thank you for the suggestions for additional references, which we have added.

"2. The term "adaptive design" has been evolving. Some people mean outcome-adaptive randomization when they say adaptive design. I recommend you mention in the abstract and introduction that by adaptive designs you mean all kinds of adaptations such as group-sequential stopping, adaptive enrichment etc. You can mention that you focus on adaptations related to stopping early for futility of efficacy or dropping an arm for futility but your conclusions apply to other adaptations such as response adaptive randomization and adaptive enrichment."

Thank you, this is certainly an important point to clarify. We have attempted to be fairly general where issues apply more widely and have added the variety of adaptive designs to the abstract and first paragraph of the background to make this clear.

"3. Table 1. Replace "adaptive randomization" with "Outcome-adaptive randomization" as you don't seem to include minimization and other covariate adaptive randomization methods as the methods are you considering are outcome-adaptive."

Thank you, we have made this change throughout.
"Page 5 lines 15 "Methodological papers often do not consider the delay between recruitment and assessment of outcomes" I understand what you mean but may be better to say "consider the rate of enrollment versus the length of follow-up for outcomes". It might be good to point out to the reader that when you say "Delay in endpoint" you mean "length of follow-up" that is set before the trial (except for time to event endpoints) as opposed to something that is unplanned."

Thank you, we have edited this to make it clearer.

"Page 9, line 25 "This is impressive quickly given that the process involved an analysis being conducted" there seem to be a couple of typos in this sentence"

Apologies for missing these typos, the sentence has been rewritten.