Author’s response to reviews

Title: Value of information methods to design a clinical trial in a small population to optimise a health economic utility function

Authors:

Michael Pearce (m.a.l.pearce@warwick.ac.uk)
Siew Wan Hee (s.w.hee@warwick.ac.uk)
Jason Madan (j.j.madan@warwick.ac.uk)
Martin Posch (Martin.Posch@meduniwien.ac.at)
Simon Day (simon.day@ctct-ltd.co.uk)
Frank Miller (mail@frank-miller.eu)
Sarah Zohar (sarah.zohar@inserm.fr)
Nigel Stallard (N.Stallard@warwick.ac.uk)

Version: 1 Date: 15 Nov 2017

Author’s response to reviews:

Editor

The authors proposed a Bayesian decision-theoretic approach to the design of a clinical trial in a small population. The proposed framework is valuable and addresses a common problem encountered in the design of clinical trials where the target patient population is small (such as for patients with rare diseases), where recruiting sufficient patients to achieve conventional (yet arbitrary) error rates may be infeasible. The proposed method has to reply on multiple simplified assumptions, but the idea is sensible, and various extensions are possible (with part of which already being briefly sketched in the Discussion). I have a few specific comments below for potential improvement of the presentation of the paper:

1) After or in equation (1), it might be helpful to immediately point out, with the focus of the paper on selecting both the sample size of and the significance level of the test used in the analysis after a clinical trial, that the action $d(\bar{x}, n)$ is often also a function of the
significance level $\alpha$. I think this would help the reader follow the fundamental idea quicker and easier as to why the proposed framework allows 'optimisation' of both the sample size and significance level.

Response: Text has been added at the end of the paragraph as suggested.

2) The authors have already discussed some of the limitations and extensions related to the simplified assumption on the 'known' patient population size. Can the authors comment on what the default time period is for assessing the treatment benefit and cost among future patients? This is not the accrual time the authors refer to in the Discussion, right? These questions are relevant because the population size $N$ is approximately proportional to the time period being considered for treatment of future patients.

Response: We would take the time period to be the estimated lifetime of the drug, which might be the period of market exclusivity, less the accrual period. We have added some text to the discussion section illustrating how this value was obtained in the example section to hopefully make this clearer.

3) Can the authors comment on potential incorporation of information into the utility function on the likelihood (or proportion) of future patients receive the treatment, if it gets approved by the regulatory agency, as a function of the actual effect size of the treatment as observed in the clinical trial? For example, assuming after the treatment is approved, the efficacy of the treatment as estimated from the clinical trial would be easily accessible (or known) by the general population. Is the following scenario possible? That is, with an increased significance level, and correspondingly a potentially reduced observed efficacy of the treatment to meet approval requirement by the regulatory agency, future patients may have a reduced likelihood of choosing this treatment after its approval (based also on consideration of the monetary cost of the treatment), as compared to a scenario where the treatment is approved based on more stringent criteria (such as a conventional type I error rate).

Response: We have expanded the Discussion section to include consideration of this point.
4) Echoing one of Dr. Lin's comments, can the authors comment on the sensitivity of the proposed approach to the prior specification of the model parameters? It is conceivable that informative prior should be used as much as possible to yield a potentially more efficient study design. Is there any recommendation of a procedure to elicit the priors? Can the authors comment on the utility of using a vague prior? Regardless, it appears that some guideline or recommendation should be provided on what prior to use, as it is an integral part of the proposed method.

Response: We have conducted further calculations to evaluate the effect of varying the prior mean and variance, including additional figures in the Supplementary material. We have also added to the discussion section to include a key reference relating to prior elicitation and some discussion on the use of informative and vague priors.

5) Some language editing is needed before publication of the paper.

Response: We have read carefully through the paper and corrected a number of minor grammatical errors.

Reviewer 1 (Ruitao Lin)

The authors proposed a value of information method to design a clinical trial in a small population based on the Bayesian decision-theoretic approach. Specifically, the authors first introduced a health economic utility function that can reflect the cost as well as the benefit of a trial, and then utilized a Bayesian approach to find the optimal sample size. In general, the considered problem is interesting, the proposed method is sound, intuitive, and practical. I do have some reservations, mainly related to the presentation of the manuscript, which are detailed in the following comments.

1. There are some existing work on the Bayesian decision-theoretic clinical trials, for example, Lewis and Berry (1995, 'Group Sequential Clinical Trials: A Classical Evaluation of Bayesian
Decision-Theoretic Designs', JASA) and a recent review (Yin et al. 2017, 'Bayesian randomized clinical trials: From fixed to adaptive design', Contemporary Clinical Trials). Please compare the proposed method with the existing approaches and provide the novelty behind the new design.

Response: We have now cited these two papers. In contrast to the current work, both consider trials with a binary outcome rather than a continuous (normally distributed) outcome. As we now state clearly near the start of the discussion section, another difference between our method and those proposed in these papers is the explicity modelling in the utility function of the size of the population, with this decreasing with the size of the trial. This feature of our approach is driven by the focus on small population clinical trials.

2. The null hypothesis in the manuscript is incorrectly stated. In page 2 line 22, the authors wrote that 'the null hypothesis may state that there is no difference between the treatments'. However, the proposed method considers a one-sided hypothesis testing problem (page 4, line 7) and the null hypothesis should be that the treatment is superior to the standard. Please double check the hypothesis. It might be better to write out the formula of the null hypothesis as well as the alternative hypothesis.

Response: We agree that it is easier to discuss a one-sided null hypothesis, so have reworded the text here and in abstract and discussion accordingly. We considered defining the null hypothesis in a formula, but decided not to do this as the treatment effect theta is not defined until later in the paper.

3. page 3 line 7, the title 'A Bayesian Decision Theoretic Approach to Clinical Trial Design' may be too general, since the proposed method only considers sample size determination. It might be better to change the section title.

Response: We have changed the section title to 'A Bayesian Decision Theoretic Approach to Sample Size Determination'
4. page 3 line 55, it might be better to provide the formula for \( f(\bar{x}|n) \).

Response: The formula has been added as requested, with \( \sigma_x \) now defined here rather than in the next subsection.

5. page 5 line 38, I do not see there is a need to put the derivation for \( z_\alpha \) in the Supplementary Material. It would be better to move the formula in page 2 of the Supplementary Material to the main body.

Response: The derivation of \( z_\alpha \) has been moved from the Supplementary material into the main paper as suggested.

6. In Figures 2 and 3, all the powers decrease as the sample sizes increase for the proposed design (even if \( \theta_A > c_2 \)), which is very counterintuitive. Please provide more comments or discussions on this point.

Response: The power actually does increase for large \( N \) for smaller \( c_2 \), but this could not really be seen in the plots with \( N \) up to \( 10^6 \). The range of \( N \) for the plots has now been increased so that the gain in power for largest \( N \) and smallest \( c_2 \) is more apparent.

7. To show the benefits of the proposed design, the authors can consider to compare it with a standard design that is determined by restricting type I and II error rates. In particular, for the standard design, you can match the same \( \alpha^* \) level as the proposed design, and then you can calculate the sample size of a standard design given a specific power value. You should be able to show that the proposed design has a higher utility than the standard design.
Response: Since our optimisation is over $\alpha$ and $n$ (equivalent to choice of $\beta$), matching type I error and power in a standard design will yield exactly the same design, and hence the same utility. The optimal $\alpha$ and $\beta$ are, however, far from conventional values, so that the gain over a more traditional design can be considerable as we now illustrate.

8. Since the design is proposed under the Bayesian framework, it is better to provide a sensitivity analysis to investigate the performance with respect to different priors.

Response: We have now added some sensitivity analysis, with detailed figures given in the Supplementary material and a brief summary in the main paper.

9. Some suggestions for possible extensions of the work. For example, you can propose a full Bayesian approach where the decision is based on the posterior distribution of the treatment effect, instead of the frequentist hypothesis testing procedure. Such a full Bayesian approach may be more appealing for rare events. You can just mention such an extension in Discussion.

Response: The method proposed bases the final decision on the posterior expected value of the utility, so is thus fully Bayesian. For a single-stage trial, however, the decision will be to recommend the new treatment whenever the observed treatment difference exceeds some critical value, so that this is equivalent to a frequentist hypothesis test, albeit conducted at a level determined by this critical value. The first paragraph of the Discussion section now makes this clearer.

10. It would be better to combine 'Discussion' and 'Conclusions' sections.

Response: We have combined these sections as suggested, moving the 'Conclusions' paragraph to come at the beginning of the Discussion section
11. The paper can probably benefit from a serious polishing exercise. For example, page 2 line 32, 'that' should be 'than'; page 2 line 50, extra 'may' here; page 7 line 27, extra '.'; and so on.

Response: We have read carefully through the paper and corrected a number of minor grammatical errors including those identified.

Reviewer 2 (Donglin Yan)

1. Figure labels in section results, subsection Operating Characteristics are not correctly displayed.

Response: Apologies for this; these are now corrected.

2. A few minor editorial errors.

Response: We have read carefully through the paper and corrected minor grammatical errors.

3. As N increases, the proposed method doesn't seem to converge to standard 2-arm RCT. Why didn't the method converge or why should it not converge?

Response: The proposed design does not approach a standard design as $N$ becomes large as it is assumed that the final decision will be made so as to maximise the expected posterior utility rather than to control a specified type I error rate. For large N (and hence large n), the prior distribution and trial costs have decreasing influence and the decision will be to recommend the
new treatment if the observed treatment difference exceeds $c_2$. We have added to the Results section to make this clearer.

4. When $N$ is small, as discussed in the article, the recommended $n^*$ could be 0. Based on the proposed method, the optimal decision is to 'approve the experimental treatment based on prior belief alone' under certain circumstances. Under what range of $N$ is the proposed method appropriate?

Response: The range of $N$ for which it is optimal to make a decision without any trial depends on the parameters of the utility function and also on the specification of the prior distribution, and is more likely if the prior is informative; this makes sense as if there is sufficient a priori knowledge of the treatment effect it might be reasonable to conclude that further trial data is unnecessary. We have added a comment on this in the Discussion section.

5. How do the operating characteristics of the proposed method compare to that of commonly used methods?

Response: As shown in the Figures and now discussed more fully in the text, the proposed method can have larger or smaller type I error than a conventional approach, but mostly has lower power (except when the optimum is to recommend the new treatment with no trial at all, so that the power is equal to 1).