Author’s response to reviews

Title: A web tool for designing and conducting phase I trials using the continual reassessment method

Authors:

Nolan Wages (nwages@virginia.edu)
Gina Petroni (gpetroni@virginia.edu)

Version: 1 Date: 18 Sep 2017

Author’s response to reviews:

Reviewer 1:

A manuscript entitled "A web tool for designing and conducting phase I trials using the continual reassessment method" by Wages NA et al. has showed the web tools written in R to design and implement the phase I trial based on continual reassessment method. It could be the interesting knowledge for clinicians and statistician who involved to phase I trials. After some points are to be revised, the article has enough quality for publication for BMC Cancer.

Response: Thank you for your careful review and consideration for publication of our manuscript.

Major comment: Many part of "Methods" should be moved to the Introduction (e.g. P5L8-14). "Methods" should be simple and easy to understand.

Response: The revised version of the manuscript moves some portion of the Methods section to the Introduction.

In this article, "the web tools" have not validated. For clinical use, these tools have to be validated by using the data from real phase I trials. The authors should reveal concordance rate of MTD between the dose which estimated by the tool and the dose determined by the trials. In safety view, number of DLT in each dose should be compared between from web tools and real world data.
Response: To validate this tool with data from real phase I trials would be a substantial undertaking, and its results would be difficult to interpret. Phase I trials that did not use the CRM would perhaps result in a different MTD recommendation than the CRM, so we wouldn’t be able to discern whether or not MTD concordance was a good thing in this case. In Phase I trials that did implement the CRM, some design specifications (skeleton, prior, etc.) may be different so that the same data may not yield the same DLT probability estimates, or even the same recommended dose. Additionally, published phase I trials using a model-based design rarely report the final probability estimates at each dose level. We did find one that did report the final estimates, and employed the CRM using a one-parameter model (Doughtery et al., Anesthesiology, 2000; 92: 1010-1016). Using the data from this trial, our app recommended the same dose as the MTD, and yielded probability estimates similar to those observed in the trial, despite the trial using a different model and skeleton than our app. This is only one trial example, and in general it is difficult to “validate” our app with real trial data since each trial uses different designs to sequentially arrive at the estimated MTD.

Our app has been validated against other established CRM software. We have matched the probability estimates in Table 4 exactly with those output by the R package dfcrm, given the same data and design specifications. This was stated in the Discussion of the original version. Additionally, we closely matched simulation results with dfcrm and another R package (bcrm), with only minor differences resulting from unavoidable simulation variation. We have added a sentence about this in the section Simulation Results.

Reviewer 2:

1. Our review of Nolan et al’s manuscript about the description of how useful web app for the Bayesian CRM is to simulate and implement the maximum tolerated dose in phase 1 trials. They show the possibility that CRM might be more convenient than traditional or modified 3+3 models. The manuscript needs some minor improvements, but I do not believe they will affect their concept.

Response: Thank you for your careful review and consideration for publication of our manuscript.
Our compulsory revisions:

Major

1) I can see this program help the researchers use CRM more easily and correctly.

But we still use 3+3 designs instead of CRM in the situation of Phase 1 trials.

Could you describe the strong points of this app compared to 3+3 design models?

Response: In the Introduction, we have provided a few sentences on the strong points of the CRM compared to the 3+3. The operating characteristics of the CRM have been extensively evaluated, and compared to the popular 3+3 algorithmic design. There is substantial evidence that the CRM is the more accurate and efficient design [refs 3, 4, 7, among others] and poses no safety concerns [ref 5].

2) I like this app because there are less time and energy to get deal with CRM.

I want to know more information about how you validated this CRM app compared to another app.

Response: Our app has been validated against other established CRM software. We have matched the probability estimates in Table 4 exactly with those output by the R package dfcrm, given the same data and design specifications. This was stated in the Discussion of the original version. Additionally, we closely matched simulation results with dfcrm and another R package (bcrm), with only minor differences resulting from unavoidable simulation variation. We have added a sentence about this in the section Simulation Results.

Minor

1) Page6. Line 38 and 39. I can’t understand why the starting dose could be the prior MTD.

Response: We apologize for the confusion. Much of the work conducted on calibrating a CRM design (Lee and Cheung, 2009, 2011; among others) has been focused on recommending design specifications based on setting the prior MTD, for the purposes of specifying a skeleton, to be the
median dose. Based on this research, we have adjusted the app to automatically generate skeletons that place the prior MTD to be the median dose. This is a separate specification than the starting dose of the trial. Most trials begin at the lowest dose, or the next to lowest dose (with a -1 level). The user of the app has a choice to specify the starting dose level of the trial they will implement.

2) Page 7. Line 18. I am concerned about N(0, 2). How could you choose the number of “2”?

Response: In the revised version, we have incorporated the work of Lee and Cheung (2011), which provides algorithms for calibrating the prior variance of the normal distribution. Instead of fixing this value at 2, the app now relies on the algorithm to choose the least informative normal prior for use in the design. This prior distribution is vague in terms of which dose is the MTD.

3) Page 8. Line 34 and 35. I am a little bit confused of the sentence, “If the entered number is larger than the maximum sample size, each trial will accrue to the maximum sample size.” Could you explain this sentence more logically?

Response: The app allows for the user to specify the total number of patients accrued to a dose level that investigators would consider enough information to stop the trial and declare this dose the MTD. For instance, in a small study, if 10 patients have been accrued to one dose level before the maximum sample size is reached, it may be reasonable to conclude that this dose is the MTD, and stop accruing the remainder of the patients. However, a stopping rule is an option that does not have to be incorporated. If the user desires for the trial to accrue to the maximum sample size, he/she can specify the stopping rule criteria to be a value larger than the maximum sample size. Thus, this number of patients accrued to one dose level will not be possible, and the trial will accrue to the maximum size.

4) Page 9. Line 24. How can we describe the case of replacement?
Response: By replacement, we assume you are referring to the situation where a patient withdraws from the study without experiencing a DLT and without being followed for the minimum required time thus their contribution to the escalation decision is not known. If the a priori decision is to ‘replace’ these patients then the maximum sample size should refer to the maximum number of patients with complete information thus excluding replaced patients. The number of patients accrued in the implementation phase would not count replaced patients as well. This has been clarified in the revision.