Author's response to reviews

Title: Screened Selection Design for Randomised Phase II Oncology Trials: An Example in Chronic Lymphocytic Leukaemia

Authors:

Christina Yap (c.yap@bham.ac.uk)
Andrew Pettitt (a.r.pettitt@liverpool.ac.uk)
Lucinda Billingham (l.billingham@bham.ac.uk)

Version: 2 Date: 7 June 2013

Author's response to reviews: see over
Dear Editor-in-Chief,

Screened Selection Design for Randomised Phase II Oncology Trials: An Example in Chronic Lymphocytic Leukaemia (MS:4209306149495829)

On behalf of my co-authors I am pleased to submit a response to the editor’s and reviewers’ comments on the above manuscript together with the revised manuscript.

We are grateful for the thoughtful and detailed input from the three reviewers which have led to an improvement of this article. Our point-by-point responses to your specific comments are addressed in the following pages.

Please do not hesitate to contact me if you require any additional information on this submission. Thank you for considering our paper for publication in BMC Medical Research Methodology.

Yours sincerely,

Dr Christina Yap

Corresponding Author

Telephone 0121 414 6043
Email c.yap@bham.ac.uk
Response to comments on manuscript:

Associate Editor's Comment:

"Where the SSD is described at the bottom of p.6, the authors indicate that the SSD allows the user the flexibility to only select the treatment arm with the highest response rate if the activity rate is greater than the other arm by a specified value? this is the modified SSD, not the SSD. This needs to be clarified.

Thank you for pointing this out. This has been made clearer in the manuscript (Pg. 6) and the abstract (Pg.2).

“In addition, a variant of this design, the Modified Screened Selection Design also allows the user the flexibility to only select the treatment arm with the highest response rate if...”

p.15, paragraph 2 ? RR which are 15% better than historical control rates are atypical in phase II, at least in my experience. I suggest the authors intended to mean that the response rate hypotheses which would indicate clinical activity and further study of the agent. Please clarify.

The text has been amended to improve clarity.

“We illustrated our proposed design for use in trials where there are two potential treatment regimens that have response rates that are hypothesized to be at least 15% better the historical control rates, which is typically used in Phase II trial designs.”

p.16, line 3 ? this does raise the question if this design, ?? which design? I think the authors are trying to make the point that the SSD is more appropriate when it is hypothesized, before the trial, that both agents will be active. If one, or neither agent is believed active, then it is unlikely that the trial will move the selection comparison. This is a very good point and needs to be clarified.

Thank you for highlighting this point. The paragraph has been amended to improve clarity. Please see Pg 16-17.

The authors should discuss some of the drawbacks of the SSD. Specifically, as in the previous comment, there remains a good chance the SSD and modified SSD (though less so) will select an arm for further study, even though it has reduced activity. See scenario 2 of table 1, where both arms have response rates which are below the desirable response rate, yet the SSD still selected one arm over 90% of the time, and modified SSD >60% of the time. It should also be noted clearly that the overall trial error rate may still be quite high, and despite the use of randomization, results from any trial using a SSD or modified SSD should not be considered definitive. A large, phase III trial should be conducted following these trials, including the winning arm from a SSD trial.

Thank you for the helpful suggestion. We have added a paragraph in the manuscript to expand on our discussion of the limitations of the proposed design which it inherits from the SWE design (1st paragraph, Pg. 15).

We have also included your suggested statement that a trial using SSD or modified SSD is not definitive. (2nd paragraph, Pg. 17)
Table 2 ? the authors should clarify that the alpha and beta are specific to each arm in the first segment, not for the trial as a whole.

This table (referred to as Table 3 in the revised manuscript) has been updated to include information on alpha and beta as suggested.
Reviewer: ken cheung

Reviewer's report:

This is a nicely written article on a common and practical problem in early phase cancer trials. The proposed method (SSD) has the advantage of simplicity and is easy to implement. I have seen several trials that randomize subjects to two or more active treatment arms. This paper provides the necessary rigor to justify the selection stage after a Simon's two-stage design is implemented within each arm.

I have a few comments that I hope are helpful to improve the paper:

1. The modified SSD in the simulation study select the superior arm only if the difference in the activity rate is greater than certain margin. This rule seems counter-intuitive; at least in this particular simulation setting. As a result of this rule, the modified SSD selects no arm much more often than desired, especially when (pA,pB) = (0.2, 0.35). In this scenario, I would argue that selecting an inferior but active arm is better than selecting no arm. I think the real motivation of the modified rule is to have another outcome such as safety as a tiebreaker. Is it possible to incorporate that in the simulation? At least, the authors should clarify this point and the appropriate utility of the modified SSD.

    Thank you for raising this point. We have endeavoured to make the utility of the Modified SSD clearer in the manuscript.

    We have highlighted that the selection of the neither arm in the Modified SSD could be due to futility where both arms are considered inactive, as well as due to a lack of an observed superior difference of active arms. In the latter case, we have emphasized that no selection of either arm is made based on the primary endpoint but a secondary outcome measure would be used to choose the superior arm in the final selection process (last paragraph in Pg 6 and 2nd paragraph in Pg 15). It would be possible for the user to incorporate this secondary measure in the simulation for a specific trial, but this is outside the scope of this paper.

    Following the comments of the editor and reviewer, we have added the probability of selecting no arm due to a lack of superior difference in Table 1 (given in parenthesis under Modified SSD), which could point to a selection based on pre-specified secondary outcome measure.

    Also, in Page 11, we have added (in bold) to our original statement, “Modified SSD does substantially better by selecting neither arm based on the primary endpoint when both CR rates are desirable but are the same (Scenario 3 and 4) with over 33% increase in identifying arms that have less than 5% difference in response rates. Pre-specified secondary criteria could be used to select between the two active arms.”

2. An advantage of SSD is the provision for early stopping if one or both arms are inactive. Thus, we can conclude a (negative) trial with fewer patients. This is an important metric to compare methods. The authors should consider including this in Table 1.

    Many thanks for the helpful suggestion. Due to the limitation of space, we have incorporated this in another Table, referred to as Table 2 in the updated manuscript.
3. As a curiosity, how will the methods compare in case of more than 2 arms?

We would expect that the SSD and the Bayesian Selection Strategy would be comparable for more than 2 arms.

As the SSD is built on the SWE, it can be easily extended to more than 2 arms. We would first conduct independent evaluation of activity based on Simon’s 2 stage design, followed by SWE for k arms. As the sample size for SWE increases as number of arms increases (e.g. with the same pA and pB of 0.2 and 0.35 respectively, n=44 is required for each arm for a 3-arm trial using SWE), this implies that if we were to use the same sample size in SWE for SSD, we could allow for a tighter control of the error rates (alpha and beta) in the first segment with the larger sample size compared to a 2-arm trial.

Reviewer:
Christopher J Weir

Reviewer's report:

This study is well motivated, with a clearly stated objective to improve the design of drug screening trials in cancer. The appeal of the proposed method is its simplicity supported by the accompanying R code. The Methods are appropriate and generally well described. I have suggested some additions in the revision recommendations. The Discussion is well balanced and considers a range of possible design alternatives.

Major Compulsory Revisions

1. On p11, “sections 2 and 3” are referred to but it is unclear what this means. Initially it seemed to be referring to parts of the R code cited in the previous sentence but on inspection of the R file this did not seem to be the case. This section needs to be restructured to improve clarity.

   Thank you for highlighting this. This has been amended in the revised manuscript.

   “The focus so far has been based on a worked example of a trial with 2 experimental arms.....”

2. P14, paragraph 2 – can the sample size required by the Jung & George approach described be stated?

   We have expanded this paragraph to illustrate the difference between Jung & George’s approach and the Modified SSD using Example 4 in their paper.

   “In the approach suggested by Jung & George [5], they proposed carrying out Simon’s 2 stage in the initial part, before selecting the superior arm based on hypothesis testing. For a fixed probability of correctly selecting the superior treatment, this would again require a larger sample size than our proposed design of SSD based on selection theory. Alternatively, for a fixed sample size, the probability of correctly selecting the superior treatment is substantially higher with Modified SSD compared to Jung and George’s design. This can be illustrated by considering Example 4 from their paper [5] which for a fixed sample size of 45 patients designed to compare hypothesized response rates of 15% vs 35% the probability of correctly selecting the superior treatment when the true rates are 25% vs 45% is 0.927 with modified SSD compared to 0.639 with Jung and George.”
Minor Essential Revisions

3. p8 “...is a high probability...” change to “...should be a higher...”

_This has been amended accordingly._

4. It would be helpful if the notation in equation 1 were introduced directly rather than via a reference.

_The notation has been added to the revised manuscript._

5. 1 million replications were used in the simulations. How was this number determined?

_We chose a feasibly large number of replications to allow for an accurate level of precision within a reasonable time frame._

6. It would improve the clarity if the scenarios considered in the simulations were briefly introduced in the methods section.

_This has been added to the manuscript._

“The scenarios included settings whereby both arms have either the same or different response rates within a relevant range that is applicable for the CLL trial.”

7. The Discussion point about introducing more flexible stopping boundaries for futility in the SSD merits being expanded.

_This has been added to the manuscript._

“... incorporate more flexible stopping boundaries for futility in the first segment, for instance, using a three stage design which allows for two interim evaluation [14] if this is appropriate. Generally, a two stage design is more common to reduce disruption to the enrolment of the trial while an interim analysis is being carried out.”

Discretionary Revisions

8. It would be helpful for some indicative numerical results to be included in the results section of the abstract

_Thank you for your suggestion. We have added that the SSD is able to obtain similar probabilities of selecting the correct superior arm of at least 90% in the abstract._