Reviewer's report

Title: How do multi-stage, multi-arm trials compare to the traditional two-arm parallel group design - a reanalysis of 4 trials

Version: 1 Date: 7 November 2008

Reviewer: Susan Todd

Reviewer's report:

This manuscript presents a nice case-study of the use of an adaptive design approach to studies in Cancer. Four trials are presented and re-analysed using the methodology and the results from the re-analyses are compared with the actual outcomes of the trials.

Minor Essential Revisions

1. My main comment concerning the manuscript is that currently the methodology is presented rather as a 'black box', for example phrases like "given by stage2.ado program" and "using the n-stage program". The brief description of the method that is given on page 3 may be confusing to the medical reader. It would be useful to define the appropriate delta in terms of the target hazard ratio. How does alpha feed into the calculations? What does alpha mean here? How should a value for it be chosen? How are the required number of control arm events determined? It would perhaps be useful to give a brief 'long-hand' numerical example for one of the trials cited ahead of the summary results in Tables 3 and 4?

2. A second issue is how the multiple testing problem associated with the multiple treatment arms is dealt with in the FOCUS trial? This probably has an impact on several of the re-analyses of this trial presented in the paper.

3. In Section 2.2, a reference is given to the software [9] but this is not yet published. Are the authors able to make the program available to readers of this manuscript or is it commercially available?

4. In Section 2.2, more details could be given of where the figures for target hazard ratios were obtained. All values are around 0.7 - 0.75. Was this really the case for all the trials that ran, or have these values been chosen for illustrative purposes?

5. Last line, page 2, why target number of events in the control group, why not target number of events overall?

6. In the text, it would help to give details of how the 'mean time saved' was calculated.

7. Is it really possible to claim that recommendations for the timing of the first analysis can be made based upon a small number of re-analysed trials? Would it
be true to say that a more comprehensive simulation study is also needed?

8. A comment on the likely impact of including covariates in this framework would be interesting.

Discretionary Revisions

9. In "Background" the phrase 'suggested previously' could be clarified.

10. In "Methods" should 'fictive' be 'fictitious'?

11. Last line page 1, maybe consider adding the introductory reference to bootstrapping here?

12. In Table 2 it might be helpful to include lines for relevant overall alpha associated with the trials?

13. On page 3, possibly include numbers of simulation scenarios which had to be discarded?

14. In the text associated with Table 3 a note could be included which explains why and how the calculated hazard ratio varies with alpha

15. In the text and titles to Tables 5 and 6 it would help to clarify things if it were stated that PFS is the intermediate endpoint here.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests