Reviewer’s report

Title: Comparison of anticipated and actual control group outcome in randomised trials in paediatric oncology provides evidence that historically controlled studies are biased in favour of the novel treatment.

Version: 1

Date: 2 September 2014

Reviewer: Tim Morris

Reviewer’s report:

The paper is a very neat way to address concerns about using historically-controlled studies in paediatric research, and demonstrates that the concerns are well-founded. Overall it has a strong message. There is much to recommend this submission, particularly the discussion and recommendations.

My comments below focus on details relating to the methods and presentation of results, which could be improved.

MAJOR COMPULSARY REVISIONS

1. Methods -> Inclusion criteria. States that ‘Where publications reported more than one randomised question (RQ), each question was considered separately.’ This needs to be justified. A three-arm trial includes two randomised questions, but if you include both you are counting one control arm twice. I am also not convinced it is a good idea for factorial designs, because the control group rate for the two questions overlaps and is thus correlated. It would be reasonable where different randomised questions use different clinical outcomes, but if you do that why not look at the control group rates for secondary outcomes as well as primary?

MINOR ESSENTIAL REVISIONS

2. Abstract -> background -> final sentence. Makes sense, but is a bit convoluted. I would suggest something like 'The rationale was that the control group outcome used in an RCT is what would have been used if a historically-controlled study had been done instead.' (and the same with this sentence in the introduction).

3. Methods -> Analysis -> Absolute differences. Need to say which variables are correlations calculated for and why. The ‘why’ is not explained in the results or discussion either.

4. Methods -> Analysis -> Relative differences. ‘Rate ratio’ usually refers to count variables; ‘risk ratio’ is surely the appropriate term for binary variables (I grant that an outcome may be good or bad so ‘risk’ might sound strange).
5. Methods / Results. It might be nice to standardise the anticipated vs. observed outcomes in some way. For example, you could use: (Superiority) The ratio of observed-anticipated to the 'clinically important difference' used in sample size calculation; (NI/equivalence) The ratio of observed-anticipated to NI/equivalence margin (delta) used in sample size calculation. These ratios could be negative, but would have a nice interpretation for how wrong the corresponding historically-controlled study might have been.

DISCRETIONARY REVISIONS

6. Discussion. I would suggest the paragraph beginning 'HC studies compare outcome data...' should be at the start of the discussion, followed by a recap of your main findings.

7. Table 2. Is it relevant to include alpha and power? Why?

8. Figure 5. Caption says 'dashed vertical lines show the median HRs of RRs' - should say 'HRs or RRs'.

9. Appendix 2. I am not sure why the correlation between trial size and year of publication are interesting as the methods section is very vague about this (the same for figure 2). Why are the figures in appendix 2 separated from figure 2? Would a Bland-Altman plot not be more useful than any of these? Such a plot would be worth including in the main body of the paper.

10. There are several minor grammatical errors and I recommend a careful proofread by the authors.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I declare that I have no competing interests