Reviewer's report

Title: A tutorial on sensitivity analyses in clinical research: the what, why, when and how

Version: 3 Date: 18 June 2013

Reviewer: Cathie Spino

Reviewer's report:

The authors have addressed many of the concerns raised in my initial review; however, there a few areas where additional detail or clarity would improve the paper. I describe these below.

- Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)
  • There are a few grammatical errors that I'm sure the editorial staff will address.

- Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
  • Section 1.2: I am not aware that sensitivity analyses are generally powered and thus the last sentence of the 1st paragraph is problematic for me. In my clinical trials experience, generally sensitivity analyses are only provided for the primary endpoint and key secondary analyses --- not for all endpoints. I think it would be helpful for the readers to provide a bit more guidance about how to decide which endpoints or outcomes and which issues require sensitivity analyses. – the statement “… if they are sufficiently powered” still remains in the paper and I think is misleading.
  • The 2nd example in Section 2.1 seems unusual. Usually there is greater ability to detect an effect if subjects have severe disease with an effective treatment (i.e., no floor effect). In this example, did one of the groups have a larger number of patients with high values?
  • The first example in section 2.2 would benefit with more detail. How is the PP analysis set defined? How many subjects are excluded? Also, it may help the reader know that PP provides the best scenario for a treatment to be shown to be effective --- in the group of subjects who comply, while the ITT may provide a better estimate of effect in “real life” where there are subjects who comply and subjects who don’t.
  • Section 2.3 has several reasons for missingness that are no focused on RCTs (e.g., registries). The discussion of missing data mechanisms is much improved. However, could you provide an example of MAR, MCAR and non-ignorable missingness from a clinical trials perspective?
  • Please provide more details for the example in Section 2.4.

- Major Compulsory Revisions (which the author must respond to before a
decision on publication can be reached)

• Although you focus more on clinical trials, I believe that Section 1.3 is not informative because you report on only 3 RCTs. I think focusing solely on clinical trials for medical journals (over an expanded period of time) would help us have a more robust sense of the frequency of reporting on sensitivity analyses. It is hard to believe it’s only 20%. Most well-conducted and well-reported RCTs included an ITT and PP or AT analysis set --- a sensitivity analysis as defined in your manuscript.

• The references have not been edited to exclude meta-analyses and other observational studies.

• Table 1. Note that you provide the number of publications with statistical analyses, not the number of statistical analysis (right?) Your footnote (&) notes 3 + 6 studies which does not equal 13 provided in the table.