Author's response to reviews

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

Authors:

Lehana Thabane (ThabanL@mcmaster.ca)
Lawrence Mbuagbaw (mbuagblc@mcmaster.ca)
Shiyuan Zhang (zhangs27@mcmaster.ca)
Zainab Samaan (samaanz@mcmaster.ca)
Maura Marcucci (Maura.marcucci@gmail.com)
Chenglin Ye (yecl@mcmaster.ca)
Marroon Thabane (mthaban@mcmaster.ca)
Lora Giangregorio (lora.giangregorio@uwwaterloo.ca)
Brittany Denis (dennisbb@mcmaster.ca)
Daisy Kosa (Sarah.kosa@uhn.ca)
Victoria Borg Debono (borgdevi@mcmaster.ca)
Rejane Dillenburg (dillenburg@mcmaster.ca)
Vincent Fruci (vincent.fruci@medportal.ca)
Monica Bawor (baworm@mcmaster.ca)
Juneyoung Lee (jyleeuf@korea.ac.kr)
George Wells (gawells@ottawaheart.ca)
Charles H Goldsmith (cgoldsmi@sfu.ca)

Version: 3 Date: 31 May 2013

Author's response to reviews: see over
Dear Dr Taljaard

RE: MS: 7041318138686307 – A tutorial on sensitivity analyses in clinical research: The what, why, when and how

We thank you for your email with reviewers’ comments on our paper. The comments were very helpful. We have revised the paper in accordance with the reviewers’ suggestions. In particular, we have now restricted our focus on clinical trials and deleted all areas and examples that dealt with observational studies. We selected appropriate examples based on published trials to illustrate the issues. For each example, we have commented on the benefit that performing sensitivity analysis provides in interpreting the primary findings. Please find attached the revised version of the manuscript. We have highlighted all the changes in YELLOW.

We hope that we have adequately addressed all the issues. Below are the point-by-point responses to the reviewers’ comments

We look forward to hear from you.

Sincerely,

Lehana Thabane, PhD
Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University
**Response to reviewers**

**Reviewer 1:**

OVERALL COMMENTS:

The authors define sensitivity analyses, the types of issues that can be explored via sensitivity analyses, and provide examples and suggestions for reporting of the sensitivity analyses. This is an important topic in the context of clinical trials; however, the topic is also relevant to observational studies and health economics analyses. The manuscript is organized well and well written; however, the scope is not completely clear and the coverage is somewhat superficial. I think that providing a bit more depth on fewer topics and examples would be advantageous. There are also a few minor issues that would be helpful to address to strengthen the paper.

Overall, I thought this was an important addition to the field and would recommend it in my class that I teach to medical professionals on clinical trials.

**Response:** We thank the reviewer for her encouraging comments. The entire manuscript has been revised to reflect her recommendations. A point-by-point description of the modifications made is described below.

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

- I believe that the scope of the paper should be clearer: where the scope includes clinical trials, observational studies, health economics studies. In different parts of the paper, there is mention of these 3 types of studies. For example, the Abstract/Background section notes that the scope is “key methodological issues in design and analysis of clinical trials.” Section 1.3 includes medical/health research studies (not clear if clinical trials and observational studies) and studies published in health economics journals. More generally, are all of the types of sensitivity analyses described in Section 2 relevant to all 3 types of studies? Section 2.6 describes the impact of center or site in multi-center RCTs --- are the same issues relevant to multi-center observational or health economics studies?

**Response:** The manuscript has been revised to reflect a single focus - clinical trials. We have limited our scope to clinical trials. Health economics studies are mentioned
to illustrate that such studies use sensitivity analyses more often. We have also chosen examples that relate specifically to clinical trials.

• I fear that the non-statistical reader of Section 2.3 will not be able to follow the discussion. For example will the average non-statistical reader understand “missing completely at random” and “missing at random”? I think it would be helpful to provide a brief description of the missing data paradigm (MCAR, MAR, non-ignorable missingness) and multiple imputation. I think that a statement like “Multiple imputation (MI) technique is currently the best available method of dealing with missing data under the assumption that data are missing at random (MAR).” will not benefit the reader. I also feel it would be helpful to describe the approach that linear mixed models for repeated measures uses to deal with missingness and when it is valid.

Response: We thank the reviewer for highlighting these limitations. The missing data paradigm has been described in more detail. We have also made reference to the use of the linear mixed methods model. See section 2.3, page 8.

• I appreciate that examples are given for the many types of sensitivity analyses that can be performed; however, I believe that some of the examples are so superficial as to be not useful. For example, in Section 2.5, the first example states “Ma et al performed sensitivity analyses of different methods of analyzing cluster RCTs.” I think it would be better to have one example with more data than three examples that don’t describe enough detail to understand what types of sensitivity analyses were considered and their implication.

Response: We have revisited all the examples to ensure that they cover the concepts in detail. We have also provided the results from the examples in order to demonstrate the application of sensitivity analysis and their implication. Specifically for section 2.5, we have provided more detail for all the examples. See pages 11 and 12.

The third example in Section 2.4 would benefit with the actual definitions and numbers (results) for PSA failure and PSA progression results.

Response: This example has been changed to one involving a clinical trial. More information is provided. See section 2.4, page 10 and 11.

• In Section 2.9, please define confounding, stratification, propensity score adjustment and instrumental variables. The first example in this section is not very informative – it just notes that there is a paper on the use of propensity score
methods and doesn’t provide information on the impact of the approaches on the estimation of mortality from an in-hospital smoking cessation program.

**Response:** This section was related to observational studies, and has been removed from the paper.

- In Section 2.11, please describe some of the special analytic challenges that meta-analyses of trials with zero events pose.

**Response:** This section was related to meta-analysis, and has been removed from the paper.

- **Section 3, Question:** When should one perform sensitivity analysis? And “How many sensitivity analyses can one perform for a single primary analysis?” I agree with the authors’ position that sensitivity analyses should be incorporated into all clinical studies, but I believe that greater guidance can be given to the non-statistical reader about the extent of sensitivity analyses. For example, the primary endpoint in a clinical trial should have sensitivity analyses performed and perhaps key secondary endpoints, but not all secondary or exploratory endpoints need sensitivity analyses. There are practical limitations. Knowledge of key issues identified in blinded data reviews (e.g., if only 1% of data are missing on the primary endpoint, then it seems unlikely that the methods to address missingness will impact the study’s conclusion) or through knowledge of the substantive field (e.g., skewness of total bilirubin values in studies of jaundice) can help the analysts make decisions about the number of sensitivity analyses that should be performed.

**Response:** We thank the reviewer for bringing this up. We have provided more guidance on the extent of sensitivity analyses that should be performed. See answer to question: How many sensitivity analyses can one perform for a single primary analysis? Page 17.

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.

**Response:** We agree with the reviewer, and have clarified what we meant.
• Section 2.6. I believe that “clusters as units of analysis” would be clearer if described as “sites as units of analysis”

Response: "Cluster" has been replaced by “center”. See section 2.6, page 12. Section 2.5 (cluster RCTs) and section 2.6 (multicentre trials) have been merged since the sensitivity analyses for these kinds correlated data are similar.

• In the 2nd example in Section 2.6, can you please describe what “COMPLETE II” is? (Another example of insufficient information to help provide sufficient detail for the reader to understand the nature and impact of sensitivity analyses).

Response: The COMPETE II trial has been described in more detail. The impact of sensitivity analysis has been highlighted. See section 2.6, page 13.

• Section 2.7: please reword “if someone dies, they will not likely experience a subsequent event, or stroke or myocardial infarction”. I believe it is not only NOT likely but impossible!

Response: This statement has been reworded: “if someone dies, they cannot experience a subsequent event, or stroke or myocardial infarction” See section 2.7, page 13.

• Section 2.8, 2nd example: The example doesn’t seem to be relevant to the point that you’re trying to make. It seems more to describe what you should adjust for, not as an example of a sensitivity analysis for imbalances.

Response: A more appropriate example of sensitivity analysis by adjusting for baseline imbalances is now used. Section 2.8, page 14.

• In Section 3, 3rd question: It might be useful to note the sensitivity analyses that are defined a priori vs. those identified post hoc.

Response: We have made reference to post hoc sensitivity analysis. See section 3, third question, page 16.

• In section 3, question “What is the difference between secondary analyses and sensitivity analyses?”: Please consider that secondary analyses of secondary endpoints are often used to provide support that the treatment or other primary effects under study are consistent with the primary analysis and operate in a manner that is coherent with the biological or underlying science.

Response: We have described what we mean by secondary analysis: analysis of secondary outcomes. See section 3, page 17.
• Section 3, last question: Do you believe that most clinical trialists would consider subgroup analyses to be a type of sensitivity analysis? I haven’t seen this reported (and that doesn’t mean that it hasn’t occurred!). Do you have a reference or two to support this assertion?

**Response:** We have described the differences in subgroup analysis and subgroup sensitivity analysis and why there may be confusion. See section 3, last question, pages 17 and 18.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

• Background: the choice of term “consumers” seems a bit odd. I would delete the term in the first paragraph.

**Response:** The term has been deleted. See background, page 4.

• Section 1.3: how would the %s of articles that report sensitivity analyses change if you divide the medical journals into randomized clinical trials vs observational trials?

**Response:** We have reported the number of RCTs that report sensitivity analyses. The numbers are similar. See section 1.3, page 6.

• References 8 and 9 are equivalent.

**Response:** This error was corrected.

• Table 1. Can you please define “key assumption”

**Response:** We have described what we mean by key assumptions and provided examples. See Table 1.

**Reviewer 2:**

Major comments:

This manuscript could make an interesting overview of sensitivity analysis (SA) of randomized trials. It provides some guidance on what SA are and how to apply it. A lot of topics are touched upon, but not discussed in much detail. To be honest, I guess such general information is not really relevant. For example, when discussing missing data, the authors provide information that is redundant if one is familiar to
methods for handling missing data, or is a too short introduction to the topic if one is not familiar to the topic.

Another key issue is that in the abstract it is indicated that this manuscript focuses on trials. Why then discuss observational research (section 2.9)? In fact, I think a distinction between trials and observational research is necessary here.

Where the distinction between observational research and trials comes into play, is how to interpretation / what to do with results from SA. Suppose SA of a trial shows that the ITT effect is significant, whereas the per-protocol isn’t. What to conclude then? And if it would be vice versa, what then? I guess a bit of discussion on how to interpret results from SA should be added to the manuscript.

**Response:** We thank the reviewer for these comments. The focus of the manuscript has been changed to randomized controlled trials. All reference to observational research has been removed. We have provided more depth to our discussions, notably in the firsts section of the summary.

This overview of sensitivity analysis is not comprehensive. For example, misclassification is not addressed, and if observational research is part of this overview, then SA of unmeasured confounding is an important issue.

**Response:** Sensitivity analysis of observational research is no longer be covered in this paper.

Page 6. Regarding the review. I wonder whether the results of this review would change if secondary analyses are considered as a form of sensitivity analysis. In addition, sometimes the impact of certain assumptions or data points is so obvious that one would not even consider a formal sensitivity analysis, because the conclusion of that is known in advance. Is that taken into account here?

**Response:** We have re-examined the data. In our review we report only sensitivity analyses stated as such in the manuscripts.

Could the authors make a distinction in this review based on RCT and non-RCT analysis? For example, they consider an per-protocol analysis in addition to ITT analysis as a sensitivity analysis. I would expect that all trials published in these high-impact journal would report on a per-protocol analysis in a RCT.

**Response:** The paper is now focused on RCTs. We have re-examined the data and report findings relevant to RCTs

Minor comments:
I think the structure of the abstract is a bit odd. What is written under ‘Discussion’ I would consider ‘Methods’. Same holds for the main text. Why indicate the second part of the main text as ‘discussion’? There must be more informative headings available.

Response: We have adopted the recommended format for debate style articles in this journal: http://www.biomedcentral.com/bmcmedresmethodol/authors/instructions/debate

Page 6. “Therefore despite their importance, sensitivity analyses are under-used in practice and often misunderstood.” What is the basis for the second part of this claim?

Response: This statement has been rephrased. The second part has been deleted.

Page 7. Regarding the protocol violations in trials. What is written there focuses on trials with a single endpoint, but in case of repeated measures outcomes more alternative sensitivity analysis can be thought of.

Response: We have added some material on repeated measures outcomes in the subsequent section. See section 2.3, page 9.

Page 9. Recently a relevant paper on handling of missing data in trials (and particularly about sensitivity analysis for that) was published (Little et al. N Engl J Med. 2012 Oct 4:367(14):1355-60). I miss a reference to that paper. In addition, with the examples, I think it is a bit odd to consider a method that is known to be flawed (LOCF) as a sensitivity analysis. Better examples are available (e.g., de Ruyter et al N Engl J Med. 2012 Oct 11:367(15):1397-406).

Response: We thank the reviewer for these resources, which we have consulted and referenced. The limitations of LOCF have been highlighted. It is stated purely for academic purposes. The example using LOCF has been removed, and a more informative one added. See the revised section 2.3.

Page 11, examples: these examples do not add information to the paper; either make the examples more informative, or consider removing from the manuscript.

Response: We have chosen to provide more detail regarding the examples and make them more informative. See revised section 2.5.

Page 11, 2.6. To what extent is multicenter data different from clustered data (as discussed in 2.5). I think the main analytical approach (and thus also sensitivity analysis) would be the same.
Response: We agree with the reviewer. Sections 2.5 and 2.6 are now merged into section 2.5.

Page 13, second example. This example has nothing to do with baseline imbalance (as the authors also point out, this was a study with balanced covariates).

Response: We have chosen a more suitable example, describing a sensitivity analysis adjusting for baseline imbalances. See section 2.8, second example, page 14.

Page 14, example no. 4 (ref 65). This is a simulation study. I don’t think that is a nice example of how to apply SA in observational studies. Same holds for the example on page 15, ref 70

Response: We agree with the reviewer. Given our new focus on clinical trials, these examples have been removed.