Identifying relevant sensitivity analyses for clinical trials

Tim Morris
MRC Clinical Trials Unit at UCL, London
CSM seminar, LSHTM. 25 Jul 2014

Outline

Background story

Why we need to do sensitivity analysis

Criteria for deciding whether a candidate sensitivity analysis is relevant

Details, examples, discussion

Disclaimer

This work was done in August 2013.

I submitted my PhD thesis at the end of September 2013.

While doing this work I was probably in a 3-month bad mood. I apologise in advance for criticism of others' work. This does not represent the views of the MRC etc…

Some of the slides you see 'live' should probably be omitted from anything that goes online.

Thabane et al. BMC Medical Research Methodology 2013, 13:92
http://www.biomedcentral.com/1471-2288/13/92

BMC Medical Research Methodology

COMMENTARY Open Access

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

Lehana Thabane[1,2,3,4,5*], Lawrence Mbuagbaw[1,4], Shiyuan Zhang[1,4], Zainab Samaan[1,6,7], Maura Marcucci[1,4], Chenglin Ye[1,4], Marroon Thabane[1,8], Lora Giangregorio[9], Brittany Dennis[1,4], Daisy Kosa[1,4,10], Victoria Borg Debono[1,4], Rejane Dillenburg[11], Vincent Fruci[12], Monica Bawor[13], Juneyoung Lee[14], George Wells[15] and Charles H Goldsmith[1,4,16]

Tutorial on sensitivity analyses

The approach is undiscerning.

Note, in this laboured analogy:

Shoot = 'Do an analysis'

Questions = 'Do results agree?'

Why do sensitivity analysis?

In clinical trials, we should plan our analyses in precise detail before seeing the outcome data.

Should publish the plan and stick to it on receiving the data. We cannot check assumptions when planning analyses.

Choosing an analysis based on trial data has been shown to lead to questionable results in several settings.

Unless you have chosen the best possible analysis, **build sensitivity analysis in to your plan**!

Rationale for this work

Confusion about what sensitivity analysis is. Often used to refer to any secondary analysis of primary outcome.

'That would be a sensitivity analysis' – overused

Working definition

There are very general definitions of SA (quite computer science-y).

‘The study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs.’

– Saltelli et al. 2008. *Global sensitivity analysis*. Wiley.

Working definition

Assume we have chosen the principle analysis for our primary outcome.

A sensitivity analysis addresses the **same substantive question** in a **different way**.

What does *Tutorial* say?

‘Sensitivity Analysis (SA) is … a method to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, models, values of unmeasured variables, or assumptions with the aim of identifying results that are most dependent on questionable or unsupported assumptions.’

What does *Tutorial* say? Examples

Impact of outliers

Assess which observations are ‘outliers’ and perform a sensitivity analysis with and without outliers.

Give two examples where this makes a difference and conclude that the analysis including outliers was not robust.

What does *Tutorial* say? Examples

Impact of non-compliance or protocol deviations

Do an ITT analysis then run per-protocol and as-treated analyses.

Give two examples, in one the results differ for the two analyses.

What does *Tutorial* say? Examples

Missing data	– Analyse complete cases – Impute using single or multiple imputation
Definitions of outcomes	– Perform analysis using different cut-offs or definitions
Clustering	– Compare analysis that ignores clustering with various methods of accounting for it
Competing risks	– Perform analysis for each event separately – Use a proportional sub-distribution hazard model – Fit one model by taking into account all the competing risks together
Baseline imbalance	– Analyse with and without adjustment for baseline characteristics
Distributional assumptions	– Different distributions – Parametric vs. nonparametric methods – Classical vs. Bayesian methods

Three criteria for a relevant sensitivity analysis

1. The candidate sensitivity analysis addresses the **same question** as the primary analysis.

2. The proposed sensitivity analysis **can disagree** with the primary analysis.

3. If results disagree, there must be **genuine uncertainty** as to which analysis gives the more reliable result.

1. Addressing the same substantive question

If two analyses address different questions, we are talking about secondary analyses, not sensitivity analyses. These may be useful but should be framed correctly.

When did your train get in?

What time did you get here this morning?

Would you think, ‘Different answers – what does it mean?’ (Hint: the correct answer is ‘no’)

1. Addressing the same substantive question

Example: the Multicentre Aneurysm Screening Study (MASS) randomised 67,800 men to receive an invitation to an abdominal ultrasound or not.

20% of invited men did not accept their invites.

The primary analysis was by intention-to-treat.

Q – *What was the effect of being randomised to an invitation?*

A complier-average causal effect analysis was also performed.

Q – *What was the effect of abdominal ultrasound in patients who would have adhered to protocol however randomised?*

2. Analyses must be able to disagree

Sometimes two analyses that go by different names will *always* lead to the same result.

Such sensitivity analyses can be dangerous. It is like doing one analysis and being reassured that doing it again on the same data leads to the same results.

2. Analyses must be able to disagree

Example: Zheng et al. published a protocol for a study investigating the effect of Baduanjin exercise on health in college students. Some outcomes are anticipated to be missing, and the principal analysis involves a *t*-test in the complete records.

Multiple imputation as a sensitivity analysis?

Impute assuming outcome is normally distributed with different means and equal variances in the two treatment groups.

– Given sufficient imputations, the two results will agree!

2. Analyses must be able to disagree (a tip)

'What if the answer is *I don't know*?'

Your candidate sensitivity analysis should be motivated by concerns about certain features of the data. Try to construct datasets in which the sensitivity analysis disagrees with the primary analysis.

If you cannot, think about why you are considering this sensitivity analysis in the first place.

3. Genuine uncertainty about the best result

(Before seeing the data.)

For some candidate sensitivity analyses, it may be clear that you would always believe one over the other.

Example 1: Peters et al. found that for a cluster-randomised trial, analysis ignoring clustering and accounting for it led to different results.

No uncertainty as to the more reliable result. Futile to plan this as a sensitivity analysis.

What does *Tutorial* say? Another look.

Impact of outliers

Assess which observations are ‘outliers’ and perform a sensitivity analysis with and without outliers.

Give two examples where this makes a difference and conclude that the analysis including outliers was not robust.

What does *Tutorial* say? Another look.

Impact of non-compliance or protocol deviations

Do an ITT analysis then run per-protocol and as-treated analyses.

Give two examples, in one the results differ for the two analyses.

Another look at *Tutorial*'s examples

Missing data	– Analyse complete cases – Impute using single or multiple imputation
Definitions of outcomes	– Perform analysis using different cut-offs or definitions
Clustering	– Compare analysis that ignores clustering with various methods of accounting for it
Competing risks	– Perform analysis for each event separately – Use a proportional sub-distribution hazard model – Fit one model by taking into account all the competing risks together
Baseline imbalance	– Analyse with and without adjustment for baseline characteristics
Distributional assumptions	– Different distributions – Parametric vs. nonparametric methods – Classical vs. Bayesian methods

Thabane et al. BMC Medical Research Methodology 2013, 13:92
http://www.biomedcentral.com/1471-2288/13/92

BMC
Medical Research Methodology

Open Access

Morris et al. BMC Medical Research Methodology 2014, 14:11
http://www.biomedcentral.com/1471-2288/14/11

trials:

BMC
Medical Research Methodology

DEBATE

Open Access

Choosing sensitivity analyses for randomised trials: principles

Tim P Morris[1*], Brennan C Kahan[2] and Ian R White[3]

Article metrics

As of 24 Jul 2014,

Thabane et al: 9,059 accesses, 4 citations

Morris, Kahan & White: 1,386 accesses, 2 citations

Oh well.

Conclusions

Unless you think you have chosen the best possible analyses (regardless of what the data look like), it is advisable to plan sensitivity analyses in protocols and statistical analysis plans.

Ask yourself: **Is it asking the same question**; **can it disagree with the principle analysis**; **would I be uncertain about which to believe**?

Criteria ensure only relevant sensitivity analyses get planned. Can be applied to study designs beyond randomised trials.

You may be pleasantly surprised at how few you can come up with, and it may help to improve your principal analysis.

References

Thabane L, Mbuagbaw L, Zhang S et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Medical Research Methodology* 2013, 1(13):92.

Morris TP, Kahan BC, White IR: Choosing sensitivity analyses for randomised trials: principles. *BMC Medical Research Methodology* 2014, 14:11.

Ashton HA, Buxton MJ, Day NE et al. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. Lancet 2002, 360(9345):1531–1539.

Zheng G, Li M, Lan X et al. The effect of Baduanjin exercise for physical and psychological wellbeing of college students: study protocol for a randomized controlled trial. *Trials* 2013, 14(1):422.

Peters TJ, Richards SH, Bankhead CR, et al. Comparison of methods for analysing cluster randomized trials: an example involving a factorial design. *International Journal of Epidemiology* 2003, 32(5):840–846.

Discussion

Things are often less clear when we think through real examples. Particularly ‘addressing the same substantive question’, which can be fairly subjective. (E.g. one referee felt that ITT and PP answers his substantive question, ‘Does the treatment work?’)

Example: PanACEA trial uses ‘time to first negative culture conversion’ as the primary endpoint. This is defined as the first-of-two consecutive negative cultures. What is an analysis with first-of-one?