Author’s response to reviews

Title: Through the Looking Glass: Understanding Non-Inferiority

Authors:

Jennifer Schumi (jennifer@statcollab.com)
Janet T Wittes (janet@statcollab.com)

Version: 3 Date: 4 December 2010

Author’s response to reviews: see over
December 4, 2010

Doug Altman, Curt Furberg, Jeremy Grimshaw, Peter Rothwell
Editors, Trials

Dear Editors:

Many thanks for the thoughtful comments from the reviewers on our manuscript. Included in this letter please find a point by point response to the comments of the two reviewers. We have uploaded a revised manuscript with changes tracked, as well as the original figure files.

Please let us know if you have any problems or questions regarding the files we have uploaded.

Many thanks for the opportunity to submit our work to Trials.

Sincerely,

Jennifer Schumi & Janet Wittes

Enclosures: Responses to reviewers, revised manuscript, individual figure files

O:\ProDev\Papers\Non-Inferiority\Reviews\2010_12 Revised NI Paper Letter (JS,JW-Trials).doc\JS
Reviewer's report

**Title:** Through the Looking Glass: Understanding Non-Inferiority

**Version:** 2  **Date:** 26 October 2010

**Reviewer:** Anthony Marson

The manuscript is for a review addressing issues around non-inferiority trials, which is well written. The perspective taken is a regulatory one with reference to EMA and FDA regulations. I don’t know what brief the authors were given, but in my view it would be helpful to widen the scope to include issues that are relevant to publicly funded / phase 4 trials where the intention is to inform clinical practice and policy rather than regulators. Most of my comments relate to this and are thus discretionary.

We thank the reviewer for these thoughtful comments. While a rigorous discussion of post regulatory approval trials intended to shape clinical practice and policy is beyond the scope of this review, which focused on general concepts of non-inferiority and recent regulatory perspectives, we have expanded the discussion in a few places to incorporate the comments this reviewer raised.

1. It would have been helpful if the pages were numbered. Pages 2 and 3 highlight referencing problems. It appears that the software could not find the desired references or figures. This made it more of a challenge to read and review.

   Thank you for this comment. We have added page numbering and corrected the referencing.

2. Page 3 2nd para. Is it generally true that less rigour biases towards non-inferiority?
   Randomisation, blinding and ITT aim to prevent false positive results. I presume that ‘sloppy’ refers to ‘biocreep’ and problems with assay sensitivity. If so this needs better explaining.

   We thank the reviewer for these questions as they are important concepts to sort through. Less rigorously run clinical trials can be more likely to show no difference between treatment groups. For superiority trials, this tends toward the null hypothesis of no treatment difference, but for non-inferiority trials, it would actually bias towards a finding of non-inferiority. We have expanded our discussion below.
3. Page 7. In the sloppy hypothetical trial where treatments are mingled, would the fault here be with the randomisation? Surely the problem would be with implementation of the treatment or policy to which patients were allocated.

We have revised this section to clarify that we were referring to a hypothetically otherwise well run trial where there were in fact two distinct treatments (active control and new treatment), but a faulty randomization created two combined groups consisting of subjects from both treatments, rather than distinct groups. The treatments as implemented may be quite different in their effects, but when the groups are muddled together in this way, the different effects of the two distinct treatments could not be distinguished. This would not be a fault of the investigator or the pharmacy but a result of a flawed system for assigning subjects to treatment groups.

4. page 8. Choosing the margin. Should patients not participate in the process of deciding delta? Also, the authors state that this approach may be of limited success in a ‘scientific or regulatory setting’. What about trials that are attempting to inform clinical decisions and policy rather than regulators?

This is an interesting point. Individual patients receiving treatment for a disease with multiple options may have their own opinion of what non-inferiority would mean to them, but that opinion may also guided by the clinical expertise of their treating physicians. Regulators, drug developers, and clinical investigators take patients into account in considering clinically meaningful improvements in defining the margin, and sometimes may involve them directly in planning clinical trials or working with patient advocacy groups and performing market research in formulating a development program, if not formally in defining the non-inferiority margin. We have modified this paragraph accordingly.

The second point, regarding clinical decisions and policy, really speaks to the larger topic of comparative effectiveness research. We have added discussion on this to our last section.

5. There needs to be some discussion around the clinical relevance of 95-95 method and the synthesis method. While there approaches might allow trials that convince the regulators that a treatment has efficacy, large margins defined by this method might not be at all convincing to clinicians and patients that non-inferiority has been demonstrated.
6. The discussion in section 5 highlights a regulatory focus on being able to infer, from a non-inferiority trial, that a new treatment is better than placebo. Clinicians and patients are more interested to know whether a new treatment is ‘non-inferior’ to a standard treatment whereby the results exclude an important difference at the ‘non-inferior end’. This difference in emphasis results in regulatory trials that do little to inform clinical decision making.

These are interesting points. We had tried to address the first point about large margins leading to results that are not sufficiently convincing to patients and clinicians in our brief comments on vaccines, but we have now expanded the discussion to address these two comments.

A Marson
Professor of Neurology
University of Liverpool

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Reviewer’s report

Title: Through the Looking Glass: Understanding Non-Inferiority

Version: 2 Date: 21 October 2010

Reviewer: Mark Rothmann

I have the following comments.

On page 2 “larger positive effects are better than smaller positive effects” It is not clear whether you are talking about “positive effects” or outcomes. When would a smaller effect be more meaningful? The oncology example in reduction in tumor size does not make sense.

*We have clarified this sentence and the example to indicate we were referring to outcomes, rather than effect sizes.*

Page 2 “So the null hypothesis is backwards” Formally, the hypotheses are not reversed (compared to superiority), but shifted by the amount of the non-inferiority margin

*We have qualified this sentence slightly to indicate that the hypothesis is not literally backwards, but conceptually may appear backwards.*

Dual use of the capital greek letter delta is problematic as the assumed treatment effect (one use) is not analogous to the non-inferiority margin (another use)

*We have changed this to use γ for the treatment difference and capital delta for the margin, both non-inferiority and equivalence.*

Page 3 “Trials to show superiority penalize the sloppy investigator” I know that later you advocate (rightly) for ITT analyses. Sloppiness tends to introduce bias. In many cases, the existence of missing data that is ignored favors the more toxic, less efficacious treatment arm. This would increase the chance of a false finding in a superiority trial.
This is a good example that contradicts our broad, general comment. We have qualified our language in this section and added the reviewer's comment as an alternative example.

Page 4 should read “it is trying to mimic”

We have corrected this typographical error.

Page 4 how is “not unacceptably worse” the same as “not non-inferior” (I think you have an excessive “not”)

Thank you for this comment – we must have gotten tripped up with the double negatives! We have modified this to equate “non-inferior” with “not unacceptably worse” as pointed out.

Page 4 “margin of error” is related to the reliability of an estimate or measurement. Why is a non-inferiority margin being referred to as a “margin of error.”

We were using the phrase “margin of error” in a less formal sense than in the commonly used statistical sense dealing with surveys. We have modified these sentences with an alternative paraphrase to make this explicit.

Page 5 when there are multiple endpoints, testing for non-inferiority and superiority of the endpoints can be tricky (there are ways of doing it). However, if there is only one endpoint and the non-inferiority margin is pre-specified, the testing for superiority and non-inferiority can be simultaneous (e.g., based on a confidence interval). In places you seem to be discussing the lack of nesting for testing for a difference (superiority or inferiority) and non-inferiority, as opposed to the relationship between testing for superiority and non-inferiority.

We thank the reviewer for this thoughtful comment. While we discussed the testing sequentially for non-inferiority and superiority, we were referring to the same confidence interval, looked at for two different purposes. We have revised this section to add further clarity to this concept.

Page 6 “In some cases, …” I believe it should be “comparison” not “trial” in both places. I do not think that you are talking about two different trials, but within the same trial.

Agreed. We have modified this accordingly.
Page 9 “The synthesis method, …” Instead of a non-inferiority margin, a threshold is pre-specified for the fraction of the active control effect that is retained by the experimental regimen.

We have added clarifying language after this sentence as suggested by the reviewer.

Page 9 “This method assumes that the effect of the active control has remained relatively constant …” In [5 (7 in our renumbered bibliography)] and many papers by Snapinn, there is discussion on discounting the effect of the active control and in the synthesis method, when it is believed that the effect of the active control has diminished.

We thank the reviewer for this helpful suggestion, and have modified this sentence as well as incorporated additional references.

Page 11 “had a placebo been in the NI study” I understand that this is a quote. I do not think that it is literally true. In essence, the synthesis method can be applied (for zero percent retention) to test whether the experimental regimen is better than placebo, provided the historical estimation of the active control effect (and possibly with modification) unbiasedly estimates the effect of the active control relative to placebo in the setting of the non-inferiority trial.

See response to next comment.

Page 12 “the unobserved effect that the placebo group would have had in the non-inferiority trial” Placebos have no effect (relative to placebo). See comment above

Thank you for these comments. We have modified our interpretation of the initial quote from the guidance to incorporate the point raised above explaining how the synthesis method can be used to test against placebo, and have removed the sentence quoted above from Page 12 – we were being imprecise with when we were using “effect” to indicate effect of treatment (or placebo) and when we were using it as “effect relative to placebo”, which we agree was not clear.

Page 15 “with mean 0” This is only true if the common log odds ratio is zero (I do not think that this is being assumed)
We have modified this to clarify that this statement applies under the null hypothesis.

Page 16 should read “estimated log-hazard ratios or means, respectively”

Thank you; we have modified the language.

I particularly like your discussion on sample size determination.

Thank you! We have added further discussion in this section.

Page 20 “approved therapy exists for certain disease like cancer,” That may be true if the approved cancer agent provides a survival benefit. Otherwise, it likely a trade-off of anti-cancer effects with off-target effects, and it may be possible to do a study against best supportive care.

We have modified the sentence in question to make our point (which agrees with the reviewer’s comment) more clear.

Page 22 last sentence. A risk-benefit assessment would involve examining the safety profile. Additionally, one non-inferiority trial may not be sufficient to evaluate the efficacy of the experimental regimen.

We appreciate this point, and throughout the final section have emphasized our reference to the balance of risk and benefit with efficacy and safety.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

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