Reviewer's report

Title: What differences are detected by superiority trials or ruled out by noninferiority trials? A cross-sectional study on a random sample of two-hundred two-arms parallel group randomized clinical trials.

Version: 2  Date: 6 September 2010

Reviewer: Yuliya Lokhnygina

Reviewer's report:

This paper presents an interesting exploration of the variability in choices of differences to be detected in randomized superiority and noninferiority clinical trials.

- Major Compulsory Revisions

1) For the majority of the paper, the authors have chosen to treat standardized difference in means and standardized difference in proportions as the same standardized measure. However, there is no discussion about why this is a valid approach and whether standardized difference in proportions is an accepted measure of the effect size. For example, Cohen considers a difference in arcsine-transformed proportions for the effect size, while many trials with binary outcomes consider treatment differences expressed by odds ratios or by risk ratios. Were such trials included in this analysis (with effect size recalculated)? Why did the authors chose this specific measure of the effect size for the binary outcomes?

2) It is not clear why the p-values in Table 3 are presented separately by subgroup. It seems that an appropriate analysis approach would be to test both main effect of each factor and interaction between the factor and superiority/noninferiority type of the trial. If the interaction is not significant, a single p-value for the main effect of the factor should be presented.

3) Please clarify the last analysis of variance, results of which are described in the last paragraph of section "Trial characteristics associated with clinical differences used to estimate sample size". Was this a multivariable regression model with standardized difference as an outcome and predictors selected from factors listed in Table 3? In that case, what model selection method was used? What was the R-squared coefficient for the model? For each factor, does "adjusted" in "adjusted mean" refer to adjustment for other predictors in the model?

4) Have the authors considered the phase of the study (e.g. 2a, 2b, 3) as a potential predictor of the standardized difference? Typically, phase 3 trials aim to detect smaller treatment differences than phase 2 studies.
- Minor Essential Revisions
1) In Table 1, it appears that for the analyses of categorical variables sometimes chi-square test and sometimes Fisher's exact test were used. How was the statistical test selected in each case? Note that only chi-square test is mentioned in the "Statistical Analysis" section.

- Discretionary Revisions
1) It would be interesting to know how many (if any) of the trials considered in this study explained the choice of the treatment difference used to calculate the sample size of the trial.

**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.

**Declaration of competing interests:**

I declare that I have no competing interests.