Letter to Editor Rheumatology
www.lettertoeditor.org

An international open-access and peer-reviewed online journal

Letter

Lett Ed Rheumatol
doi:10.2399/ler.11.0002

In randomized trials, statistical tests are not helpful to study prognostic (im)balance at baseline

Maarten Boers, MSc, MD, PhD

Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, Netherlands

Abstract

The New England Journal of Medicine states that it complies with the CONSORT guidelines in reporting clinical trials. However, it does not follow these guidelines in reports of 7 of 13 clinical trials at the Journal website in September 2010. These reports include formal statistical tests to compare baseline characteristics of patients between different study groups. Reasons are cited why this practice is undesirable.

Dear Editor,

I am glad the New England Journal of Medicine has decided to endorse the CONSORT guidelines for the reporting of clinical trials.[1] However, this decision has apparently not been implemented in full. Reports of randomized trials in the Journal continue to include the results of statistical tests performed on baseline characteristics of the study groups. For example, more than half (7 of 13) of the clinical trials published this September on the Journal website did so. The P value equals the chance of getting the same or a more extreme result when the trial is repeated, given the null hypothesis. After randomization the null hypothesis is true by definition, so a statistical test makes no sense: it basically tells us how 'unlucky' we are to get this distribution of prognostic factors at baseline.

If we ignore this conceptual issue and press on, the fact remains that a statistical test is poorly suited to 'diagnose' prognostic imbalance of randomized groups as it is neither sensitive nor specific. The result of a statistical test is dependent on the absolute difference between the groups, the variability around the estimate of difference, and the sample size. In contrast the potential for bias is determined by the difference in the prognostic factor and the strength of the relationship between the factor and the outcome. If this relationship is strong, a small imbalance/difference may be important and need adjustment in the analysis. However, in studies where the variability around the imbalance is large or the sample size is small, the statistical test will not be significant (false negative, test insensitive). On the other hand, if the relationship between factor and outcome is weak, a large imbalance may be unimportant. If the variability around the imbalance turns out to be small or the sample size is large, the statistical test will be significant (false positive, test nonspecific).

I would greatly appreciate a statement from the editors that from now on they will endorse and adhere to CONSORT also in this respect, or explain why they continue to allow authors deviate from it.

Reference

1. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *J Clin Epidemiol* 2010;63:834-40.

***Correspondence:**
Maarten Boers, MSc, MD, PhD
Department of Epidemiology and Biostatistics, VU University Medical Center, PK 6Z 185 PO Box 7057
1007 MB Amsterdam, Netherlands
Tel: +3120 444 4475

Citation: Boers M. In randomized trials, statistical tests are not helpful to study prognostic (im)balance at baseline. Lett Ed Rheumatol 1:e110002. doi:10.2399/ler.11.0002

Received: December 9, 2010; **Accepted:** January 14, 2011; **Published:** February 16, 2011

Conflicts of Interest: The author has declared that no conflicts of interests exist.

Background and Editor's Note:

Hasan Yazici
Editor, LER

The above letter about the methodological issue of whether the baseline characteristics of the study arms in a controlled clinical trial could justifiably be compared by statistical tests of significance was rejected by the New Engl J Med. I comply with this journal's request about the privacy of the correspondence with Prof. Boers related to the rejection process. In brief, however, the statistical advisors to the New Engl J Med were of the opinion that such statistical comparisons could justifiably be made.

I now publish the original letter by Prof. M. Boers once intended for the New Engl J Med, with an added abstract according to LER's publication rules. The reader views regarding the important methodological issue discussed are very welcome.

