

Avoid significance tests for background variables in randomised controlled trials

MEDISIN OG TALL

STIAN LYDERSEN

E-mail: stian.lydersen@ntnu.no

Stian Lydersen, dr.ing. and professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU Central Norway) at the Department of Mental Health, Norwegian University of Science and Technology.

The author has completed the ICMJE form and declares no conflicts of interest.

In a randomised controlled trial there should be no systematic differences between the groups before treatment. Still, some researchers choose to significance test for possible differences in background variables. This is superfluous, and can be misleading.

An important strength of randomised controlled trials is the fact that background variables, for example age and sex, is randomly distributed between the treatment groups. According to the CONSORT guidelines, baseline demographic and clinical characteristics should be reported separately for each group (1). Table 1 shows an example which is an extract from a larger table with 23 variables, from a study comparing two treatment pathways for hip fractures. The authors write 'Baseline characteristics did not differ between the groups' ((2)page 1629). This is based on clinical judgement. The authors did not carry out significance tests for possible differences, and this is in line with the CONSORT guidelines (1).

Table 1

Extract from a table with background variables in a randomised controlled trial (2).
Reproduced with permission from Elsevier. Number (%) unless otherwise specified.

	Comprehensive geriatric care (N=198)	Orthopaedic care (N=199)
Age in years, mean (standard deviation)	83.4 (5.4)	83.2 (6.4)
Female	145 (73)	148 (74)
Living alone	115 (58)	124 (62)
Intracapsular fracture	119 (60)	127 (64)

There will always be some differences in background variables in a randomised controlled

trial. These differences are usually small. And following randomisation, these differences are known to be random. If significance tests were carried out for the background variables, one would expect to find statistically significant differences in about 5 % of the cases, that is, for about one variable out of twenty. In the study mentioned above, such a 'significant' finding could thus be expected in 1 of the 23 background variables.

However, such significance testing is still reported in some randomised controlled trials. What could the motivation be? We can think of two reasons: to test whether randomisation was performed properly, and to identify unbalanced background variables.

Was randomisation performed properly?

If there are reasons to suspect that randomisation was not performed properly, this can be tested. However, a significance level of considerably less than 5 % should be used. Fayers and King describe a trial with an overrepresentation of younger participants in one of the groups. The difference was highly significant, with $p < 0.0005$. A closer look revealed a breach of adherence to the randomisation protocol (3).

Unbalanced background variables?

A more common motivation is probably to identify possibly unbalanced background variables between the groups. Thereafter, one can adjust for these in the analyses. But statistical significance depends on both sample size and the degree of imbalance, so this approach is discouraged (4, 5). In a small study, a variable can be very unbalanced without causing the imbalance to be statistically significant. It could be more sensible to judge the degree of observed imbalance, and adjust for variables being unbalanced and judged as clinically important, than to use p-value driven selection. But this is also controversial, since this again would be a data-driven selection of variables for the analysis ((6) p. 417–8). And if carried out, these should be sensitivity analyses performed after the primary analysis.

No reason to significance test

There is no reason to significance test for differences in background variables in a randomised controlled study, unless there is reason to believe that randomisation was not performed properly. De Boer and colleagues call this an unhealthy research behaviour that is hard to eradicate (5). In some randomised controlled trials it may be sensible to adjust for certain background variables, but these must be predefined before data are studied. This will be discussed in the next article in this series.

REFERENCES:

1. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869. [[PubMed](#)][[CrossRef](#)]
2. Prestmo A, Hagen G, Sletvold O et al. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *Lancet* 2015; 385: 1623–33. [[PubMed](#)][[CrossRef](#)]
3. Fayers PM, King M. A highly significant difference in baseline characteristics: the play of chance or evidence of a more selective game? *Qual Life Res* 2008; 17: 1121–3. [[PubMed](#)][[CrossRef](#)]
4. Pocock SJ, Assmann SE, Enos LE et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21: 2917–30. [[PubMed](#)][[CrossRef](#)]
5. de Boer MR, Waterlander WE, Kuijper LD et al. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *Int J Behav Nutr Phys Act* 2015; 12: 4. [[PubMed](#)][[CrossRef](#)]
6. Vittinghoff E, Glidden DV, Shiboski SC et al. Regression methods in biostatistics linear, logistic, survival, and repeated measures models. 2nd utg. New York: Springer, 2012.

