Do not test for baseline imbalances unless they are known to be present?

Vance W. Berger

Accepted: 31 January 2009 / Published online: 25 February 2009
© US Government 2009

Editor

Fayers and King [1] are correct to “not suggest that one should *never* carry out significance tests on baseline characteristics” but they also state that “significance tests are pointless in a conventional (individual patient) randomized trial that has an effective randomization procedure [and] are usually [worth doing only] if potential violation of the randomization is suspected.” This advice is analogous to drivers wearing a seatbelt only when they expect to be in an accident. The flaw in the argument is the circularity in suggesting that one can know the results of an analysis without actually conducting this analysis. That is, “the basis for suspicion of flawed randomization can be the very tests of baseline balance that would not be performed, under this approach, without *prior* suspicion” [2, p. 125]. Moreover, it is not true that in individually randomized trials:

“at the time the patient enters into the trial neither the patient nor the clinical team know which treatment will be allocated; the randomized allocation is concealed until after the patient has been registered into the trial. This avoids all possibility of selection bias.”

The reality is that highly restricted randomization schemes, such as permuted blocks, allow for the prediction of upcoming allocations. In fact Chapter 3 of [2] enumerates 30 actual trials, almost all of them individually randomized, in which selection bias was at least suspected.

Hence, there is *always* a reason for suspicion, at least until this suspicion has been quelled by a combination of baseline tests and the more sensitive Berger-Exner test [2, p. 132].

Fayers and King [1] caution that “Significance tests of imbalance in baseline characteristics are [informative only] if the *P*-value is extremely small—we expect on average that one in 20 baseline characteristics will be ‘significant’ with $P < 0.05$, and *P*-values of 0.01 or smaller are not unusual, purely by chance.” One has to wonder if these authors are equally quick to dismiss significant efficacy *P*-values. Probability is a unitless quantity; one event with probability 0.05 is exactly as likely to occur as any other. Finally, note that selection bias need not condemn a trial to be “abandoned and deemed unworthy of publication,” as special methods [2, Sect. 7.3.1] can ensure a valid treatment comparison even in the presence of selection bias.

References

1. Fayers, P. M., & King, M. (2008). A highly significant difference in baseline characteristics: The play of chance or evidence of a more selective game? *Quality of Life Research*, 17, 1121–1123. doi:10.1007/s11136-008-9390-x.
2. Berger, V. W. (2005). *Selection bias and covariate imbalances in randomized clinical trials*. Chichester: John Wiley and Sons.

V. W. Berger (✉)

Biometry Research Group, National Cancer Institute, National Cancer Institute and University of Maryland Baltimore County, Executive Plaza North, Suite 31316130 Executive Boulevard, MSC 7354, Bethesda, MD 20892-7354, USA
e-mail: vb78c@nih.gov; vance917@gmail.com