Correct Baseline Comparisons in a Randomized Trial

Patrick Schober, MD, PhD, MMedStat,* and Thomas R. Vetter, MD, MPH†

KEY POINT: Standardized differences for baseline variables between study groups are more appropriate to report than hypothesis tests for the baseline variables.

In this issue of *Anesthesia & Analgesia*, Kim et al1 report a randomized controlled trial (RCT) comparing the incidence of epistaxis after nasotracheal intubation with a standard tube versus a thermosoftened tube. Participants’ baseline demographics and clinical characteristics are presented by treatment group in their above Table 1. This table typically includes central tendency and variability with means and standard deviations or medians and quartiles for continuous data and counts and percentages for categorical data.

The goal of an RCT is to compare study groups that differ only by the treatments received. Proper randomization ensures that treatment groups are balanced on average with repeated sampling; however, it does not prevent chance imbalances with 1 sampling. Therefore, Table 1 not only summarizes patient characteristics but also demonstrates actual study group balance.

Traditionally, baseline balance is assessed by a series of hypothesis tests comparing each baseline variable between the groups, and a nonsignificant result ($P > .05$) has commonly been considered to indicate baseline balance. This approach is flawed.2 First, all patients are randomly sampled from the same population, and any differences must be due to chance. It thus does not make sense to test hypotheses of group differences at baseline. Second, hypothesis tests are greatly affected by sample size. In a smaller trial, marked differences between the groups can be nonsignificant, whereas in a larger trial, negligible differences could be significant.

As done by Kim et al,1 an appropriate alternative is to report standardized differences between the groups, which are a measure of the difference in units of the pooled standard deviation.3 An absolute standardized difference of >0.1 is conventionally considered to indicate imbalance.3 Such imbalances should be noted by the authors. However, any such observed imbalances should be interpreted within the clinical context and the strength of the variable’s relationship with the outcome. In this example, the absolute study group difference in the mean body mass index of 0.5 is likely not clinically meaningful.

**REFERENCES**


From the *Department of Anesthesiology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; and †Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Austin, Texas.

Address correspondence to Thomas R. Vetter, MD, MPH, Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Health Discovery Bldg, Room 6.812, 1701 Trinity St, Austin, TX 78712. Address e-mail to thomas.vetter@austin.utexas.edu.