

### Correction Note

Correction for: Vanhove, Jan. 2015. Analyzing randomized controlled interventions: Three notes for applied linguists. *Studies in Second Language Learning and Teaching*, 5(1), 135-152.

#### Weighted t-test for cluster-randomised experiments

In Vanhove (2015), I wrote the following on the subject of analysing in which whole groups, rather than individual participants, are assigned to the experimental conditions (cluster-randomised experiments):

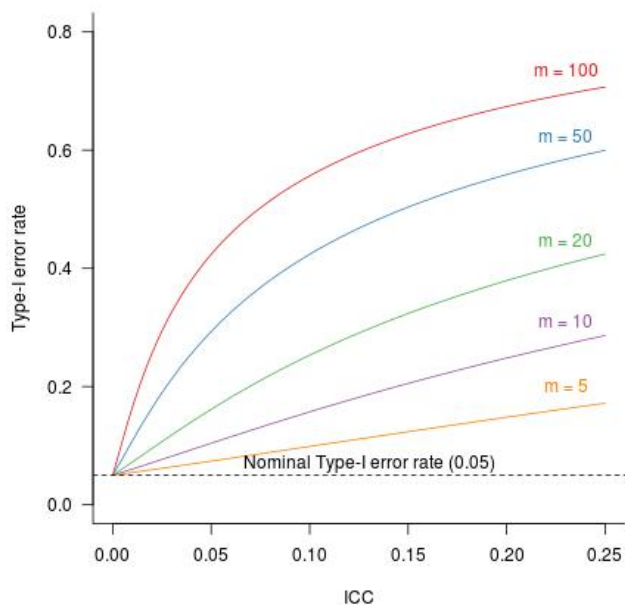
A conceptually straightforward approach [for taking clustering into account] is to calculate the mean (or another summary measure) of each cluster and run a *t* test on them rather than on the original observations. When the number of observations differs from cluster to cluster, a *t* test in which the cluster means are weighted for cluster size is recommended (see, e.g., Campbell, Donner, & Klar, 2007). This analysis is easy to compute and report, and it perfectly accounts for violations of the independence assumption: The Type-I error rate is at its nominal level (i.e., 5%). (p. 145)

I regret that the recommendation to weight cluster means for cluster size does not stem from M. J. Campbell et al. (2007) but from M. K. Campbell et al. (2000, pp. 193-194: "When the size of the clusters varies widely, it is preferable to carry out a weighted *t*-test, using cluster sizes as weights"), from where it can be traced back to Kerry and Bland (1998). More importantly, using cluster sizes as weights does *not* perfectly account for violations of the independence assumption, that is, it does not guarantee that the Type-I error rate will be at its nominal level (see <http://goo.gl/gGIZs7> for simulation results). As for appropriately weighting cluster means in the analysis of cluster-randomised experiments, Hayes and Moulton (2009) point out that, while "theoretically possible" (p. 178), it requires that the intraclass correlation (ICC) be known with great accuracy, which is not usually the case. In the absence of an accurate ICC estimate, Hayes and Moulton do not recommend use of the weighted *t* test.

In conclusion, weighting cluster means for cluster size is not generally recommended. Unweighted  $t$  tests on cluster means are still available as a straightforward alternative with a nominal Type-I error rate, whereas multilevel models offer greater flexibility as regards the inclusion of covariates, modelling more complex designs, and so on.

### Inflated Type-I error rate for cluster-randomised experiments

Figure 1 in the same article showed how the Type-I error rates for cluster-randomised experiments vary as a function of the ICC and the number of participants per cluster ( $m$ ) when the data are inappropriately analysed by means of a  $t$  test on the participants' scores. Due to an error in the underlying analytical derivation, this graph slightly exaggerates the Type-I error rate inflation. The graph below is based on the correct formula provided by Hedges (2007) and fixes this. The take-home message stays the same, however: Ignoring clustering drastically increases Type-I error rates, more so for larger clusters and larger ICC values.



### References

- Campbell, M. J., Donner, A., & Klar, N. (2007). Developments in cluster randomized trials and Statistics in Medicine. *Statistics in Medicine*, 26, 2-19.
- Campbell, M. K., Mollison, J., Steen, N., Grimshaw, J. M., & Eccles, M. (2000). Analysis of cluster randomized trials in primary care: A practical approach. *Family Practice*, 17, 192-196.

- Hayes, R. J., & Moulton, L. H. (2009). *Cluster randomized trials*. Boca Raton, FL: Chapman & Hall/CRC.
- Hedges, L. V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, 32(2), 151-179.
- Kerry, S. M., & Bland, J. M. (1998). Analysis of a trial randomised in clusters. *BMJ*, 316, 54.