Journal of the International Neuropsychological Society (2007), **13**, 903–904.

DOI: 10.1017/S1355617707071147

LETTER TO THE EDITOR

Use of covariates in randomized controlled trials

GERARD J.P. VAN BREUKELEN[1] AND KOENE R.A. VAN DIJK[2,3]

[1]Department of Methodology and Statistics, Maastricht University, The Netherlands
[2]Department of Neurocognition, Maastricht, The Netherlands
[3]Department of Psychiatry and Neuropsychology, Maastricht, The Netherlands

(RECEIVED March 9, 2007; FINAL REVISION April 5, 2007; ACCEPTED April 5, 2007)

A recent discussion in this journal illustrates some recurrent misunderstandings about the role of covariates in randomized controlled trials (RCTs). This letter aims at clarifying this role and at pointing out a pitfall in SPSS repeated measures ANOVA. We hope that our commentary will contribute to a further improvement in the use of advanced statistics in neuropsychology.

Anstey et al. (2006) reported the effects of cataract surgery on neuropsychological test performance in an RCT. They used repeated measures ANOVA to test for group by time interaction, adjusting for two covariates: age and baseline visual acuity. Gilmore (2007) disputed the correctness of their analyses, pointing out a significant group difference in baseline visual acuity and suggesting that a between-subject covariate cannot adjust within-subject effects or interactions involving within-subject effects. Anstey et al. (2007) replied that their model included a covariate by time interaction, thereby adjusting the group by time interaction of interest for baseline visual acuity. While the analysis by Anstey et al. (2007) is by and large correct, all three publications mentioned above show some statistical misconceptions that we briefly discuss in this commentary.

First of all, in an RCT baseline group differences are caused by chance fluctuation and will not be significant, except because of type I errors or if dropouts are excluded from the test on baseline differences. The only significant baseline difference in Anstey et al. (2006), that in visual acuity, had a p-value of .03, but this was one out of eight tests for baseline differences (see their Table 1). So the risk of a type I error caused by multiple testing was considerable. This baseline difference would not be significant after Bonferroni correction, that is, using $\alpha = 0.05/8$ (the alternative explanation of selective dropout is discussed at the end of this commentary). Relatedly, if treatment assignment is random and there is no selective dropout, then covariates are not needed to adjust for baseline differences; instead, they serve to increase the power of the treatment effect test by reducing unexplained outcome variance. So age and baseline visual acuity did not have to be included into the analysis by Anstey et al. (2007), but their inclusion may have increased the test power.

So let us first look at model (1) of Anstey et al. (2007), which does not contain covariates. This is a repeated measures ANOVA with group, time, and group by time effects. Their test of group by time interaction is equivalent to ANOVA of the group difference with respect to change from baseline (posttest minus pretest). The alternative method is ANCOVA of the group difference at posttest, with the pretest as covariate. The difference between both methods can be clarified with the following ANCOVA equation:

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 X_{ij} + e_{ij} \quad \text{(i)}$$

or equivalently,

$$(Y_{ij} - \beta_2 X_{ij}) = \beta_0 + \beta_1 G_{ij} + e_{ij} \quad \text{(ii)}$$

where Y_{ij} is the posttest score of person i in group j, G_{ij} indicates treatment (e.g. 0 for controls, 1 for treated), X_{ij} is the covariate, here the pretest score, and e_{ij} is a normally distributed residual. Equation (ii) shows that ANOVA of change is the special case of ANCOVA where $\beta_2 = 1$. Although both methods are valid for RCTs, ANCOVA is known to have more power than ANOVA of change, except if β_2 is close to one and the sample size is small. To prevent misunderstanding, we emphasize that for nonrandomized studies the choice between both methods is much more complicated (Van Breukelen, 2006).

Let us now turn to model (2) of Anstey et al. (2007), which adds the covariates baseline age and visual acuity. Running this model with SPSS repeated measures ANOVA comes down to two ANCOVAs following equation (i), but each with a different Y:

Correspondence and reprint requests to: Gerard J.P. Van Breukelen, Department of Methodology & Statistics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands. E-mail: gerard.vbreukelen@stat.unimaas.nl

a) Tests of Between-Subjects Effects: ANCOVA with the *average* of pretest and posttest as dependent variable *Y*, and group and covariates as independents.

b) Tests of Within-Subjects Effects: ANCOVA with the *difference or change* (post-minus pretest) as dependent variable *Y*, and group and covariates as independents.

Of these two ANCOVAs, the within-subject one is of interest here. The group effect β_1 on change is the group by time interaction. The covariate effect β_2 on change is the covariate by time interaction. The intercept β_0 is the main effect of time. Now, Gilmore (2007) suggests that a between-subject covariate cannot adjust within-subject effects. Anstey et al. (2007) correctly reply that it can by including covariate by time interaction, as they did. Anstey et al. (2007) are incorrect with respect to the consequences of this adjustment, however. They state that it adjusts the group by time interaction, but that is correct only if there is a correlation between group and covariate. In an RCT such a correlation can only arise by chance and so the adjustment is a chance adjustment. But there did not have to be any adjustment to start with, again because of the randomized assignment (unless there is selective dropout, see the end of this commentary). The real merit of covariates in an RCT is a gain of power by reducing the unexplained outcome variance.

Another mistake by Anstey et al. (2007), probably shared by the majority of SPSS users, is their belief that the covariate by time term does not adjust the within-subject effect Time. This within-subject Time effect is the intercept β_0 of the regression of change *Y* on group and covariates, and so it reflects the change for a person with value zero on all predictors. In ANOVA without covariates this is no problem because SPSS recodes a binary group factor from its original (0,1) coding into (−0.5,+0.5) coding, or a rescaled version like (−1,+1) or (−0.7,+0.7). As a result, β_0 is the outcome of a person halfway between both groups, that is, a "grand mean." But covariates like age and baseline visual acuity enter the ANOVA in their original form, and so β_0 is the expected change for a person with zero age and zero visual acuity. This is not what researchers believe the Time effect to be. The solution is simple: center covariates before the analysis, that is, subtract the sample mean of age from each individual age, and likewise for baseline visual acuity. Centered covariates have a mean of zero so that β_0 is a grand mean, here the grand mean of change, which is what researchers mean by the Time effect. To further confuse, whereas SPSS does not center covariates in its significance test tables, it does so in its marginal means and pairwise comparison tables. It can then happen that a within-subject effect is far from significant in the tests table but highly significant in the pairwise comparison table (or vice versa). The latter table should be chosen. This may be of importance to Anstey et al.'s (2006) statement that "Although there were no significant main effects for time, there did appear to be an improvement in means . . ." (p. 638). Perhaps the authors looked in the wrong table for testing the Time effect (but from their report we cannot tell).

Finally, 11 of all 56 patients dropped out, of which 8 in the control group, and only the 45 complete cases were included into the effect analysis. As Anstey et al. (2006) say, this ". . . may have resulted in higher functioning or more motivated participants being retained in the control group compared with the intervention group" (p. 638). This might perhaps explain (a) the group difference in baseline visual acuity and (b) the aversive treatment effect on face recognition. A good method for handling dropout is to include all available data of dropouts into the analysis. This is not possible with the GLM procedure in SPSS which uses listwise deletion. But it is possible with mixed regression in SPSS, which can be set up such that it either runs ANOVA of change or ANCOVA, always including dropouts. For details, see Little (1995) and Van Breukelen (2006).

REFERENCES

Anstey, K.J., Lord, S.R., Hennessy, M., Mitchell, P., Mill, K., & Von Sanden, C. (2006). The effect of cataract surgery on neuropsychological test performance: A randomized controlled trial. *Journal of the International Neuropsychological Society*, *12*, 632–639.

Anstey, K.J., Salim, A., Lord, S.R., Hennessy, M., Mitchell, P., Mill, K., & Von Sanden, C. (2007). Our correct use of ANCOVA yields acceptable results. *Journal of the International Neuropsychological Society*, *13*, 371.

Gilmore, G.C. (2007). Inappropriate use of covariate analysis renders meaningless results. *Journal of the International Neuropsychological Society*, *13*, 370.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121.

Van Breukelen, G.J.P. (2006). ANCOVA *versus* change from baseline: More power in randomized studies, more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*, 920–925.