Correction for baseline covariates in clinical trials and observational studies

Matilde Francisco[1], *Klaus Langohr*[2]

[1]fc57478@alunos.fc.ul.pt, Faculty of Sciences, University of Lisbon
[2]klaus.langohr@upc.edu, Department of Statistics and Operations Research, Polytechnic University of Catalonia

Longitudinal studies allow the repeated monitoring of health outcomes or risk factors, and the identification of differences in outcomes. Baseline measurements, demographic characteristics or measurements taken at the beginning of the study of the response variable or variables correlated with it, have several purposes, including the assessment of treatment effect based on the change from baseline. The question on whether to consider baseline as a covariate or a dependent variable is frequently asked.

Not accounting for baseline, can not only affect the magnitude of differences detected, but also the direction of these differences, which can result in different clinical conclusions. However, correcting for baseline effects can also introduce bias, which is problematic, especially in cases where sample sizes are small. The lack of consistency in the literature around this topic contributes to the difficulty to establish a standard statistical approach, so studies' specific characteristics influence the decision on what statistical approach should be used.

In clinical trials, randomization is a fundamental process, since it helps prevent bias associated with the selection of candidates receiving each treatment. It ensures that it is possible to compare the effect of the different treatments between groups, since they are similar in almost every other critical aspect. In observational studies, the lack of randomization can result in selection bias, since the baseline characteristics of individuals in exposure groups may differ. If these characteristics have a significant role in predicting the outcome, their imbalance between groups can cause bias.

When adjusting a model, it is necessary to adopt different strategies regarding the use of baseline values. The use of Analysis of covariance (ANCOVA) has been advocated when there is the need to correct for baseline. In this model, under the assumption of the existence of a correlation between baseline and post-baseline measurements, the baseline values are included in the model as a covariate, and the post-baseline values are the response variable. The constrained longitudinal data analysis (cLDA) is built under the assumption that the randomization of the subjects involved was efficient, so it is assumed that means at baseline are identical for the groups being compared. Due to this, both baseline and post-randomization values are dependent variables.

In this work, the models previously described are going to be applied to a data set from the Safety and efficacy of cognitive training plus epigallocatechin-3-gallate in young adults with Down's syndrome (TESDAD) clinical trial, in order to access if different approaches to baseline have impact on the findings.

Keywords: Baseline-value adjustment, linear mixed models, longitudinal studies.