Nonlinear interactions, confounding, and missing values: an interesting case study

Georg Heinze

Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria

Sophisticated methods exist to address challenges in the analysis of observational studies with survival outcomes, such as longitudinal continuous covariates, time-varying treatment, non-linear effects, confounding bias, or missing covariate values. However, often these issues occur in combination and the best analytical strategy is not obvious.

We report on our experience from the analysis of an observational cohort study of mortality in kidney transplant patients (Heinze et al, BMJ 2009;339:b4018), following the STROBE checklist. The main objective was to assess the interaction of erythropoietin use and achieved haemoglobin concentration using multiple Cox regression. This interaction was assumed non-linear and thus modeled using restricted cubic splines. Both variables were time-dependent, and modeling and interpreting their interaction required special care. Covariates to adjust for were selected by formalizing the Hosmer and Lemeshow guidelines for purposeful selection of potential confounders, within which a variable selection decision is made at each step of the modeling process. Missingness of covariate values was addressed by multiple imputation using chained equations, which significantly complicated the implementation of the methods outlined above. Model assumptions were assessed by sensitivity analyses.

We discuss advantages and disadvantages of alternative methods that could have been considered to deal with non-linear effects, longitudinal covariates, confounder selection and missing value imputation. We emphasize the importance of providing details of complex methods used and of sensitivity analyses in supplementary online materials.

Multiplicity of analysis challenges will often force statisticians, usually working on a tight schedule, to compromise. Aside from statistical optimality, three factors may influence the decisions made by the statistician during analysis: (i) the degree of confidence that the analyst has in the results, (ii) the expected effort in implementing specific methods, and (iii) the communicability of complex analysis strategies to non-statistical principal investigators and the target audience. These decisions may often have significant impact on the final results, and thus on the conclusions drawn by the principal investigators.