Modelling continuous variables with a spike at zero – on issues of a fractional polynomial based procedure

Patrick Royston[1], Willi Sauerbrei[2], Heiko Becher[3]

[1]MRC Clinical Trials Unit, London
[2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany
[3]Institute of Hygiene, University of Heidelberg, Germany

In clinical epidemiology, a frequently occurring problem is to model a dose/response function for a variable X which has value 0 for a proportion of individuals ("spike at zero"), and a quantitative value for the others, e.g. cigarette consumption or an occupational exposure. When the individuals with X = 0 are seen as a distinct sub-population, it may be necessary to model the outcome in the subpopulation explicitly with a dummy variable, and the rest of the distribution as a positive continuous variable using a dose-response relationship (Robertson et al 1994).

The concept of fractional polynomials (FP, Royston & Altman 1994) has been shown to be useful for estimating dose-response relationships for continuous variables. A multivariable procedure (MFP) is available to select variables and to determine the functional relationship in many types of regression models. A modification of the function selection component for variables with a spike at zero was proposed in chap 4 of Royston & Sauerbrei (2008). A binary variable indicating zero values of X is added to the model. The procedure considers in two stages whether X has any effect, whether individuals with X = 0 should be considered as a separate subgroup and whether an FP functional relationship for the positive values improves the model fit.

In three examples with substantial differences in the distributions of X, strength of the effects and correlations with other variables, we will discuss issues concerning the modelling of a continuous variable with a spike at zero. Adjustment for other important predictors is included in the models.

Analytical results can be derived in univariate models for certain special cases. This will be discussed in a related paper.

References

Royston P, Altman DG (1994): Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling (with discussion). Applied Statistics 43(3):429-467.

Robertson C, Boyle P, Hsieh CC, Macfarlane GJ, Maisonneuve P (1994): Some statistical considerations in the analysis of case-control studies when the exposure variables are continuous measurements. Epidemiology 5:164-70.

Royston P, Sauerbrei W (2008): MULTIVARIABLE REGRESSION MODELLING. A pragmatic approach based on fractional polynomials for modelling continuous variables. Wiley.