

Investigating interactions in multivariable model-building with continuous covariates

Willi Sauerbrei¹, Patrick Royston²

¹ Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany

² MRC Clinical Trials Unit, London, UK.

In multivariable model-building interactions between two continuous predictors are often ignored. If considered, the most popular approach is to assume linearity for both variables and test the multiplicative interaction term for significance. However, the model may fit poorly if one or both of the main effects is non-linear.

Fractional polynomials (Royston & Altman 1994, Sauerbrei & Royston 1999) have been proposed for investigating main effects of predictors for possible non-linearity. With the multivariable fractional polynomial procedure (MFP), selection of variables and simultaneous determination of functional relationships are possible. If the main effects of two continuous variables derived by MFP include non-linear functions, Royston & Sauerbrei (2008) proposed considering products of these FP functions, which may also be linear, as a candidate for the interaction between the two continuous covariates. This procedure called MFPIgen is a generalization of the MFPI procedure which investigates interactions between a continuous and a binary variable (Royston & Sauerbrei 2004).

Key issues of the method will be illustrated by assessing predictors for 10-year all-cause mortality in the Whitehall I data set, a large cohort study (N=17260, 1670 events) in British civil servants. Six continuous and one categorical predictor will be considered in a logistic regression model. Using MFPIgen we will illustrate that mismodelling of non-linear main effects as linear may introduce spurious interactions and that a systematic search for interactions will reveal further structure in the data.

Our new procedure allows the analyst to consider continuous by continuous interactions in a systematic way. It can also handle non-linear effects, an important issue too often ignored in practical data analysis. To reduce the chance of spurious interactions being detected, a stringent significance level such as 1% should be used.

References

- Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Applied Statistics* 1994; 43:429-467.
- Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *JRSS-A* 1999; 165:71-94.
- Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med* 2004; 23:2509-2525.
- Royston P, Sauerbrei W. *Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Wiley 2008