

---

## Seminar

25. Oktober 2016, 16:00  
Seminarraum 11/12, Gebäude W34, UKE

### On variable selection and shrinkage strategies to derive multivariable regression models

Prof. Dr. Willi Sauerbrei, Freiburg

---

In many areas of science where empirical data are analyzed, a task is often to identify important variables with influence on an outcome. Most often this is done by using a variable selection strategy in the context of a multivariable regression model. Many strategies have been proposed but each of them is criticized and none is generally accepted. We will discuss that the aim of a model is the most important criteria to decide whether a specific variable selection strategy may be appropriate to derive a suitable model. We will stress the difference between models for prediction and for explanation and consider model sparsity as an important criteria. In the context of low-dimensional data and excluding issues caused by very small sample size we will consider stepwise approaches with and without post-selection shrinkage and the LASSO procedure which selects variables and shrinks parameter estimates simultaneously. Based on results in real examples and a simulation study we will argue that predictions are often very similar, irrespective of the strategy used to derive the model, but that the stop criteria has an important influence on the model selected for explanation, including its sparsity, interpretation and transportability (Sauerbrei et al, 2015). We conclude that backward elimination followed by post-selection parameterwise shrinkage (PWSF) is a suitable approach when variable selection is a key part of data analysis, provided that the amount of information in the data is not too small (Sauerbrei, 1999; van Houwelingen and Sauerbrei, 2013). For continuous factors we will briefly discuss the importance to check whether the popular assumption of a linear effect is acceptable and introduce the multivariable fractional polynomial (MFP) approach as a suitable method for model building with continuous variables. For post-selection shrinkage of a selected MFP model the R package *shrink* has recently been provided (Dunkler et al, 2016). For two or more variables which are associated it extends the methodology by introducing 'joint shrinkage factors'.

#### References

- Dunkler D., Sauerbrei W. and Heinze G. (2016): Global, Parameterwise and Joint Post-Estimation Shrinkage. *J Stat Software*, 69:8.  
Royston P., Sauerbrei W. (2008): Multivariable Model-Building – A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. *Wiley*.  
Sauerbrei, W. (1999): The use of resampling methods to simplify regression models in medical statistics. *Appl Stat*, 48: 313-329.  
Sauerbrei W., Buchholz A., Boulesteix A.-L., Binder H. (2015): On stability issues in deriving multivariable regression models. *Biom J*, 57: 531-555.  
Van Houwelingen H.C., Sauerbrei W. (2013): Cross-validation, shrinkage and variable selection in linear regression revisited. *Open Journal of Statistics*, 3: 79-102.