# Using pseudo-values and simultaneous variable selection to extract a core set of variables for dynamic prediction models

Daniela Zöller[1], Irene Schmidtmann[1], Arndt Weinmann[2], Thomas Gerds[3], Harald Binder[1]

*1: University Medical Center Mainz, IMBEI, Mainz, Germany*

*2: University Medical Center Mainz, 1. Med, Mainz, Germany*

*3: University of Copenhagen, Department of Public Health, Copenhagen, Denmark*

In a competing risks setting for time-to-event data the cumulative incidence (CI) is the absolute risk that the event of interest occurs within a certain time. It is an easily interpretable quantity and patients can be compared with respect to their CI for different competing risks. By evaluating models for the CI at a series of landmark time points, dynamic prediction models can be obtained. Although the idea of dynamic prediction is that some covariates are only relevant for some of the landmarks, there might be a set of covariates that is relevant at all landmarks. One way to analyze effects on the CI in right censored data is by using a pseudo-value approach, which allows, among others, the application of direct regression models. Using a suitable link function, a straightforward interpretation of covariate effects is possible. We combine this approach with dynamic prediction to identify a potential set of covariates relevant for all landmark models and to deal with sparse information at early and late time points.

Specifically we propose a stagewise regression algorithm which performs variable selection simultaneously across time points but allows for separate estimates at each time point. The covariates relevant for all landmark models are identified with the help of inclusion frequencies. We apply the algorithm to clinical cancer registry data from hepatocellular carcinoma patients.

The combination of pseudo-values with dynamic prediction is seen to stabilize the variable selection and the result is a robust core set of variables which can be used for adjustment.