

**The analysis of clustered data - A comparison between hierarchical models
and the Generalized Estimating Equations approach**

P. Martus, P. Schlattmann

Charité University Medicine, Department of Biostatistics and Clinical Epidemiology,
Berlin, Germany

Clustered data appear in many biostatistical applications, e.g. longitudinal studies, clinical trials involving multiple measurement sites per subject, familial studies, or in cluster randomised interventional trials. In the last two decades several approaches have been developed to deal with such data. Two major approaches are the Generalized Estimating Equations and the Hierarchical Models. However, which of these methods are used often depends on the individual statistician and no criteria have been developed which could support the decision for a specific approach. In our study we compare these approaches in the framework of a large interventional study with a complex dependency structure in the data.

We use the data of a controlled, non randomised interventional study for primary prevention of smoking in schools obtained from 2246 pupils. The intervention consists of three parts: A presentation with basic information of health damages caused by smoking, an interview with a patient with lung cancer, and a live video from a bronchoscopy. All three parts are presented within a hospital. At each date two to five classes of the same or different schools are participating. Schools either belong to the intervention or to the control group. Thus the following levels of clustering have to be taken into account: (i) School, (ii) Presentation, and most important (iii) class. Demographical data, smoking behaviour, and an evaluation of the presentation (intervention group only) are documented. After one year follow up the smoking behaviour is reassessed. The Primary endpoint "smoking behaviour after one year" is measured on an ordinal scale. The statistical analysis is based on the proportional odds model with baseline adjustment and several correlational structures resp. hierarchies of varying complexity. We compare (1) the assumptions, (2) the results, (3) the interpretation, and (4) the numerical stability and of these approaches using our example. Discussion will be focussed on methodological aspects of this study.