A flexible modelling framework for overdispersed, Hierarchical data of a joint nature.

Geert Molenberghs *
Interuniversity Institute for Biostatistics and statistical Bioinformatics
(1) Katholieke Universiteit Leuven, Belgium
(2) Hasselt University, Diepenbeek, Belgium
– geert.molenberghs@uhasselt.be

Non-Gaussian outcomes are often modeled using members of the so-called exponential family. Notorious members are the Bernoulli model for binary data, leading to logistic regression, and the Poisson model for count data, leading to Poisson regression. Two of the main reasons for extending this family are (1) the occurrence of overdispersion, meaning that the variability in the data is not adequately described by the models, which often exhibit a prescribed mean-variance link, and (2) the accommodation of hierarchical structure in the data, stemming from clustering in the data which, in turn, may result from repeatedly measuring the outcome, for various members of the same family, etc. The first issue is dealt with through a variety of overdispersion models, such as, for example, the beta-binomial model for grouped binary data and the negative-binomial model for counts. Clustering is often accommodated through the inclusion of random subject-specific effects. Though not always, one conventionally assumes such random effects to be normally distributed. While both of these phenomena may occur simultaneously, models combining them are uncommon. This paper proposes a broad class of generalized linear models accommodating overdispersion and clustering through two separate sets of random effects. We place particular emphasis on so-called conjugate random effects at the level of the mean for the first aspect and normal random effects embedded within the linear predictor for the second aspect, even though our family is more general. The binary, count, and time-to-event cases are given particular emphasis. These can be modeled separately as well as jointly. Connections between joint modeling and a variety of areas are emphasized: group sequential trials; clusters with informative size; incomplete data.

Keywords: exponential family; repeated measures; overdispersion; longitudinal data.