A combined overdispersed and marginalized multilevel model

Peer-reviewed author version

A Combined Overdispersed and Marginalized Multilevel Model

Samuel Iddi[2] **and Geert Molenberghs**[1,2]

Interuniversity Institute for Biostatistics and statistical Bioinformatics

[1]*Universiteit Hasselt, Agoralaan 1, 3590 Diepenbeek, Belgium*

[2]*KU Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium*

Abstract

Overdispersion and correlation are two features often encountered when modeling non-Gaussian dependent data, usually as a function of known covariates. Methods that ignore the presence of these phenomena are often in jeopardy of leading to biased assessment of covariate effects. The beta-binomial and negative binomial models are well known in dealing with overdispersed data for binary and count data, respectively. Similarly, generalized estimating equations (GEE) and the generalized linear mixed models (GLMM) are popular choices when analyzing correlated data. A so-called combined model simultaneously acknowledges the presence of dependency and overdispersion by way of two separate sets of random effects. A marginally specified logistic-normal model for longitudinal binary data which combines the strength of the marginal and hierarchical models has been previously proposed. These two are brought together to produce a marginalized longitudinal model which brings together the comfort of marginally meaningful parameters and the ease of allowing for overdispersion and correlation. Apart from model formulation, estimation methods are discussed. The proposed model is applied to two clinical studies and compared to the existing approach. It turns out that by explicitly allowing for overdispersion random effect, the model significantly improves.

Keywords: Combined model; Correlation; Overdispersion; Partial marginalization

References

Heagerty, P.J. (1999) Marginally specified logistic-normal models for longitudinal binary data *Biometrics*, **55**, 688–698

Heagerty, P.J. and Zeger, S.L. (2000) Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science*, **15**, 1–26

Griswold, M.E. and Zeger, S.L. (2004) On marginalized multilevel models and their computation (November 2004). *Johns Hopkins University, Department of Biostatistics Working Paper #99*.

Hinde, J. and Demétrio, C.G.B. (1998a) Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G., Verbeke, G., and Demétrio, C. (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.

Molenberghs, G., Verbeke, G., Demétrio, C., and Vieira, A. (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Vangeneugden, T., Molenberghs, G., Verbeke, G., and Demétrio, C. (2011) Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Journal of Applied Statistics*, **38**, 215–232.