GENERALIZED LINEAR MIXED MODELS WITH SPARSE BINARY OUTCOME DATA: COMPARISON OF ESTIMATION METHODS

M.-E. Beauchamp[1], R.W. Platt[1], J.A. Hanley[1]

[1]*McGill University, Montréal, Canada*

Email: *marie-eve.beauchamp@mail.mcgill.ca*

A number of estimation methods for generalized linear mixed models (GLMMs) have been developed and implemented in commercial software packages. However, their limits in practical settings have not been studied completely and systematically. The penalized quasi-likelihood (PQL) is among the most commonly used methods in practice. However, PQL uses an approximate likelihood and may produce biased parameter estimates. For instance, the bias in the variance component estimates may be severe for binary outcome data where the number of occurrences is low for each covariate pattern within each cluster. In principle, numerical integration methods, such as the adaptive Gauss-Hermite quadrature, applied to the marginal likelihood can produce accurate results. In this project, the performance of a number of estimation methods accessible to a large number of end-users is evaluated and compared. We focus on datasets of small and large sample sizes with a low number of occurrences for a binary response variable. Simulated datasets are used for the evaluation. The bias in the parameter estimates is quantified for each fitting method in several settings and convergence problems are reported. Data and model characteristics influencing the bias in the estimators are also studied. Results obtained to date, in addition to a review of some existing knowledge, will be presented. The ultimate goal of this project is to provide guidelines for proper analysis of sparse binary outcome data with GLMMs.