Statistical Science
2012, Vol. 27, No. 3, 340–343
DOI: 10.1214/12-STS381A
Main article DOI: 10.1214/11-STS381

Discussion of "Multivariate Bayesian Logistic Regression for Analysis of Clinical Trial Safety Issues" by W. DuMouchel

Bradley W. McEvoy and Ram C. Tiwari

Key words and phrases: Meta-analysis, drug safety, hierarchical Bayesian model, data-mining, sparse data.

We would like to comment on this article by William DuMouchel, as it gives an interesting application of logistic regression to clinical safety data. Not to underscore the scope of the multivariate Bayesian logistic regression (MBLR) model, but the use of numerical integration is arguably its most important feature. Avoiding Markov chain Monte Carlo (MCMC) sampling techniques for other data-mining tools, such as the Multiple-item Gamma Poisson Shrinker (DuMouchel, 1999), has proven successful for Dr. DuMouchel in their acceptance among nonstatisticians. With MBLR this should not be an exception.

As most statisticians lack the clinical insight required to specify the appropriate MBLR model inputs, it makes MBLR an ideal tool for use by the clinicians. However, targeted users may not appreciate some subtleties of MBLR, which we present below. We also present findings from our empirical evaluation of the MBLR algorithm. This commentary provides some perspective that we have gained through multiple interactions with Dr. DuMouchel and from our reviews of different versions of MBLR formulation at FDA since 2009.

Bradley McEvoy is a reviewer in Division of Biometrics VII, Office of Biostatistics, Center for Drug Evaluation and Research, FDA, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993-0002, USA (e-mail: Bradley.McEvoy@fda.hhs.gov). Ram Tiwari is Associate Director in the Office of Biostatistics, Center for Drug Evaluation and Research, FDA, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993-0002, USA (e-mail: Ram.Tiwari@fda.hhs.gov).

The views expressed by authors are their own and do not necessarily reflect those of FDA.

1. MBLR AND META-ANALYSIS

In order to fully appreciate the MBLR methodology, one has to contrast it with a more traditional meta-analytical formulation when data from multiple trials are investigated. Dr. DuMouchel is correct in pointing out that the MBLR methodology is in the spirit of a full-data meta-analysis and does not consider it a meta-analytic model. The current MBLR model formulation does not render the flexibility of separating out patient- and trial-level variations in the model. Consequently, MBLR is very different from a multi-level/meta-analysis model that would consist of a patient-level model and a trial-level model, each with independent sources of variation. This makes MBLR effectively a patient-level model; the inclusion of trial-level variables (e.g., study identifiers) into equation (2) results in the variance components in equations (3)–(6) being influenced by both patient and trial heterogeneity.

This distinction between the MBLR and its meta-analytic formulation is critically important. The main advantage of a meta-analytic formulation is that it preserves the trial-specific randomized comparison between the treatment and control groups, thereby avoiding confounded estimates. With the MBLR formulation this is not necessarily the case, as Dr. DuMouchel aptly notes for the Pollakiuria example that the trial-specific estimates do not preserve the between-trial differences. Additionally, shrinkage estimates used to identify vulnerable patient subgroups depend on factors which are typically considered unrelated of patient characteristics.

The practical concern of applying a methodology that does not ensure the randomized comparison is preserved is that it may lead to a possible signal being missed or hidden. A recent high-profile example of