Computational Statistics and Data Analysis 83 (2015) 82–90

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Performance measures of the bivariate random effects model for meta-analyses of diagnostic accuracy

Mireya Diaz *

Vattikuti Urology Institute and Public Health Sciences, Henry Ford Hospital, 2799 W Grand Blvd., Detroit, MI 48202, USA

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form 29 August 2014
Accepted 29 September 2014
Available online 8 October 2014

Keywords:
Multivariate random effects
Meta-analysis
Diagnostic accuracy
Factorial design
NLMIXED

ABSTRACT

The bivariate random effects model has been advocated for the meta-analysis of diagnostic accuracy despite scarce information regarding its statistical performance for non-comparative categorical outcomes. Four staggered simulation experiments using a full-factorial design were conducted to assess such performance over a wide range of scenarios. The number of studies, the number of individuals per study, diagnostic accuracy values, heterogeneity, correlation, and disease prevalence were evaluated as factors. Univariate and bivariate random effects were estimated using NLMIXED with trust region optimization. Bias, accuracy, and coverage probability were evaluated as performance metrics among 1000 replicates in 272 different scenarios. Number of studies, individuals per study, and heterogeneity were the most influential meta-analytic factors affecting most metrics in all parameters for both random effects models. More studies improved all metrics while low heterogeneity benefited fixed and random effects but not the correlation. About twenty studies are required to obtain random effects estimates with good statistical properties in the presence of moderate heterogeneity, while only the univariate model should be used when few studies are summarized. In general, the bivariate model is advantageous for meta-analyses of diagnostic accuracy with complete data only when the correlation is of interest.

1. Introduction

The use of the bivariate random effects model has been advocated for meta-analyses of diagnostic accuracy studies (Chu et al., 2010; Harbord et al., 2008; Reitsma et al., 2005; Riley et al., 2007) despite the scarcity of information regarding the statistical performance of this model for non-comparative categorical outcomes. Very few studies have systematically assessed the statistical properties of bivariate random effects estimators. Although these studies have considered a range of conditions, they either restricted the evaluation to the fixed effects (Chu et al., 2010; Zhang et al., 2011) or if they assessed the variance components this was done under low heterogeneity (Paul et al., 2010; Riley et al., 2007; Kuss et al., 2014).

Paul et al. (2010) reported on an equivalent performance between the popular SAS PROC NLMIXED (SAS Institute Inc., 2009) and a new approach using Bayesian estimation through a deterministic Laplace approximation for bivariate meta-analysis in the setting of low heterogeneity. The estimates in this study were unbiased and accurate with coverage just slightly below nominal in certain scenarios. Riley et al. (2007) focused on the comparison between the bivariate and the univariate random effects models for a meta-analysis of diagnostic accuracy, again in the setting of low heterogeneity. They showed that both models were relatively unbiased in this setting, with a slight increase in the bias of the random effects

* Tel.: +1 313 916 4741.
E-mail address: diazemail@comcast.net.

http://dx.doi.org/10.1016/j.csda.2014.09.021
0167-9473/

in the bivariate model compared to the univariate model when correlation between random effects was high and only five studies were available. In a more recent study, Kuss et al. (2014) proposed a very flexible parameterization of the random effects via the beta-binomial model for the marginals linked via copulas. Among the advantages of this formulation are the existence of a closed form solution for the likelihood and the possibility to include any correlation structure depending on the copula chosen. The simulation study was extensive, but once again although a seemingly large variance for the random effect was used (0.75 in the logit scale) this is equivalent to an inconsistency $I^2 = 0.19$ (Higgins and Thompson, 2002) or still low heterogeneity. In a related simulation study, Austin (2010) compared the performance of different software implementations of univariate multilevel logistic regression for few clusters (between 5 and 20). The study showed that most packages estimated the random effects variance with a large bias unless that at least 15–20 clusters with a sufficient number of individuals per cluster were available. Thus, it is likely that this bias may worsen for more random effects particularly if few studies and/or large heterogeneity are present.

The current simulation study aims at filling the gaps in knowledge about the performance of random effects for meta-analyses of diagnostic accuracy in the setting of complete data with particular emphasis in greater heterogeneity. For this the metrics of estimators of fixed and random components for meta-analysis of diagnostic accuracy are evaluated within a comprehensive set of scenarios.

2. Methods

Four sets of staggered simulation experiments were performed using a factorial design in each. The factorial design approach was selected to screen a large number of parameters in an efficient manner. Findings from a given factorial design were used to guide the subsequent simulation study in order to explore further a particular finding. Values for the parameters were selected after running descriptive statistics of meta-analyses of diagnostic accuracy published in 2010 in order to cover the scenarios most frequently encountered in real practice.

The first experiment aimed at identifying the most influential meta-analytic parameters on the statistical properties of the fixed and random effects components for one of the most popular implementations (NLMIXED in SAS). The second experiment extended the range of diagnostic accuracy values examined. The third experiment assessed the effect of unequal proportions of diseased and non-diseased individuals. The fourth and last experiment extended the number of studies in a meta-analysis and increased variability in the sample size of the studies within a meta-analysis.

2.1. Characteristics of meta-analyses of diagnostic accuracy in the published literature

As a first step, a literature review of meta-analyses of diagnostic accuracy published in 2010 was conducted in order to select representative parameter values for the simulations (see Supplement 1, Appendix A). Sixty one studies with 112 arms were identified. The following information was retrieved from each meta-analysis: the number of studies, minimum, maximum, and average number of individuals per study, overall estimate of sensitivity, and specificity and heterogeneity. A subset of meta-analyses which provided information from each study was identified and used to characterize further these meta-analyses particularly in terms of the distribution of diseased and non-diseased individuals per study and the correlation between sensitivity and specificity.

2.2. Factorial designs as efficient tools for experiment planning

Factorial designs are experimental plans where the influence of variables or factors on a given outcome is assessed at two or more levels. The efficiency and elegance of these designs lie in their ability to examine simultaneously the independent and interactive effects of numerous factors by means of straightforward analysis of variance (ANOVA) methods. The most commonly used factorial design is the 2^k where each of k factors is examined at two levels each. The number of experiments increases rapidly with the number of factors and levels. Since the independent and interactive effect of several factors at multiple levels was of interest a sequence of full factorial designs was selected.

2.3. Random effects for categorical outcomes by the NLMIXED procedure

Evaluation of univariate random intercept logistic regression models with SAS-based and R-based frequentist estimation methods showed that SAS NLMIXED is the procedure that provides the most accurate parameter estimates under correct model assumptions (Zhang et al., 2011). That is why it was selected for the current simulation study. NLMIXED implements the integral approximation (SAS Institute Inc., 2009). This approach approximates the log-likelihood of the outcomes (i.e. a marginal likelihood obtained after integration over the random effects) and then maximizes this approximate function. NLMIXED implements the Adaptive Gaussian Quadrature (AGQ) which approximates the integral by a weighted sum over predefined points of the random effects. Key for the accuracy achieved with this method is the selection of the number of quadrature points, the larger these are the more accurate the results but also the more computationally demanding the process. Regarding maximization (optimization methods), NLMIXED offers three different classes depending on their use of derivatives and which derivatives. These are (1) methods which compute first derivatives and approximate second