Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 43: 156–171, 2016
doi: 10.1111/sjos.12170

Optimal Estimator for Logistic Model with Distribution-free Random Intercept

TANYA P. GARCIA
Department of Epidemiology and Biostatistics, Texas A&M University
YANYUAN MA
Department of Statistics, University of South Carolina

ABSTRACT. Logistic models with a random intercept are prevalent in medical and social research where clustered and longitudinal data are often collected. Traditionally, the random intercept in these models is assumed to follow some parametric distribution such as the normal distribution. However, such an assumption inevitably raises concerns about model misspecification and misleading inference conclusions, especially when there is dependence between the random intercept and model covariates. To protect against such issues, we use a semiparametric approach to develop a computationally simple and consistent estimator where the random intercept is distribution-free. The estimator is revealed to be optimal and achieve the efficiency bound without the need to postulate or estimate any latent variable distributions. We further characterize other general mixed models where such an optimal estimator exists.

Key words: exponential model, logistic regression, random intercept, robustness, semiparametric estimator, sufficiency and completeness

1. Introduction

Logistic models with a random intercept are frequently used in medical and social research when clustered or longitudinal data are analysed. Often the random intercept is assumed to follow a certain distribution (i.e., normal distribution) or be independent of model covariates. Such assumptions, however, inevitably raise concerns about model misspecification and misleading inference conclusions, especially when the random intercept depends on the covariates (Heagerty & Kurland, 2001). To circumvent these concerns, we consider a logistic random intercept model where no distributional assumption is made on the random intercept, and there may be dependence between the covariates and random intercept. We then use semiparametric methods (Tsiatis, 2006) to develop a computationally simple estimator for the model parameters with the following useful properties.

(i) The estimator is guaranteed to be consistent.
(ii) The estimator achieves optimal efficiency. Our estimator is most efficient among the class of consistent estimators that make no distributional assumptions about nor independence assumptions between the covariates and random intercept. We show this holds even though the random intercept distribution or quantities that rely on it are never estimated in our estimator.
(iii) Our estimator outperforms the simple penalized quasi-likelihood (PQL) estimator (Schall, 1991; Breslow & Clayton, 1993), which only assumes the random intercept to have zero mean and finite variance, but is biased for binary data (Breslow & Clayton, 1993; Breslow & Lin, 1995; Lin & Breslow, 1996).
(iv) Comparing our estimator with the traditional normal-based maximum likelihood estimator provides a practical way to assess whether the covariates and random intercept are independent (Section 2.3).

We consider data collected from n independent clusters where each cluster $i = 1, \ldots, n$ contains m_i subjects indexed by $j = 1, \ldots, m_i$. For each subject, we observe a binary random variable Y_{ij} and a p-dimensional covariate vector $X_{ij} = \left(X_{ij1}, \ldots, X_{ijp}\right)^{\mathrm{T}}$. We suppose individuals within a cluster share a common random intercept R_i so that a within-cluster correlation is induced. We further assume that conditional on R_i, observations from the ith cluster are independent.

A logistic random intercept model is then

$$\operatorname{logit}\left\{P\left(Y_{ij} = 1 | X_{ij} = x_{ij}, R_i = r_i\right)\right\} = x_{ij}^T \beta + r_i = \sum_{k=1}^{p} \beta_k x_{ijk} + r_i \tag{1}$$

with $\operatorname{logit}(p) = \log\{p/(1-p)\}$. Here, β is the p-dimensional parameter of interest where each β_k, $k = 1, \ldots, p$, measures the effect of increasing x_{ijk} by one unit on the log-odds ratio. For notational simplicity, we let $Y_i = \left(Y_{i1}, \ldots, Y_{im_i}\right)^T$ and $X_i = \left(X_{i1}, \ldots, X_{im_i}\right)$ denote a $p \times m_i$ matrix.

The model in (1) is frequently used in medical applications as in binary symptom data (Heagerty, 1999) and binary cardiac injury data (Tung *et al.*, 2004). For example, in a longitudinal study of people with schizophrenia (Heagerty, 1999), a key interest is in modelling the relationship between presence/absence of apathy (a negative symptom) in relation to schizophrenia onset and gender. Here, Y_{ij} represents the presence/absence of apathy for patient i in month j, X_{ij1} is age of onset, X_{ij2} is gender and R_i captures the within-person correlation induced from the repeated measures. A second example is a repeated measures study of cardiac injury (Tung *et al.*, 2004). Investigators at UC San Francisco were interested in studying which cardiac measures would affect whether or not troponin (an enzyme observed after heart damage) exceeds a certain threshold. One viable approach is to construct a mixed logistic model for subject i at time j where Y_{ij} represents whether or not the troponin level exceeds the threshold, X_{ij1} is a measure of heart performance via ejection fraction of the heart, X_{ij2} is systolic blood pressure, X_{ij3} is heart rate and R_i is a random intercept capturing the effect of the ith subject. In Section 4, we estimate the effects of these features and address which of them indeed affect troponin levels.

In these examples and others, estimating the covariate effects β begins with the likelihood based on model (1). Using f to denote various (conditional) densities described by the subindices, the likelihood for the ith cluster is

$$\begin{aligned} f_{Y,X}\left(y_i, x_i; \beta\right) &= \int f_{Y|X,R}\left(y_i \mid x_i, r_i; \beta\right) f_{X,R}\left(x_i, r_i\right) d\mu\left(r_i\right) \\ &= \int \prod_{j=1}^{m_i} \exp[y_{ij}\left(x_{ij}^{\mathrm{T}}\beta + r_i\right) - \log\left\{1 + \exp\left(x_{ij}^{\mathrm{T}}\beta + r_i\right)\right\}] f_{X,R}\left(x_i, r_i\right) d\mu\left(r_i\right). \end{aligned} \tag{2}$$

Here, $\mu(\cdot)$ denotes the dominating measure. Even when $f_{X,R}(x_i, r_i)$ is known, the integral in (2) has no closed form in general. Instead, numerical integration must be used. More seriously, incorrectly assuming that the random intercept (or random effects in general) has a specific parametric form or that X and R are uncorrelated can be problematic for inference. For example, the covariate effect can be strongly biased when the random effect is correlated with the model covariates and estimation is performed under independence. In the earlier schizophrenia example, for instance, Heagerty (1999) demonstrated that the variability of the random intercept depended on the covariate gender and that ignoring this feature resulted in completely misleading conclusions for the schizophrenia study. Misspecifying the random effect distribution can also seriously alter the variance estimate of the random effects distribution (Turner *et al.*, 2001). When the random effects distribution is truly heavy tailed (Pinheiro *et al.*, 2001) or skewed (Fernandez & Steel, 1998), variance estimates of the random effect can be largely inflated if the random effect distribution is falsely assumed to be normal. Empirical evidence also suggests that efficiency loss occurs when one incorrectly assumes a normal random effects distribution, although the true random effects distribution is a two-point discrete distribution (Agresti *et al.*, 2004) or the true distribution is a mixture of normals (Zhang & Davidian, 2001).

To avoid misspecifying the random effects distribution, more flexible methods have been developed. These include modelling the random effects using a mixture of normals (Lesaffre & Molenberghs, 2001), combinations of parametric distributions (Piepho & McCulloch, 2004), a seminonparametric approach (Zhang & Davidian, 2001; Chen *et al.*, 2002); and a nonparametric maximum likelihood method (Aitkin, 1999; Agresti *et al.*, 2004). These methods either assume a more flexible random effect distribution form or estimate the distribution nonparametrically. Therefore, the methods either do not resolve the misspecification problem completely or involve even more intense computation.

In this paper, we completely bypass the estimation of the random intercept distribution, either parametrically or nonparametrically. Through viewing the random intercept distribution as a nuisance parameter, we factor out its effect via semiparametric projection. Our approach yields a consistent estimator that also achieves the efficiency bound without estimating the random intercept distribution, nor specifying a working model for its form. In addition, we demonstrate how the flexibility of our method reveals other settings where a similar estimator exists.

The remainder is organized as follows. Section 2 describes our estimation procedure that yields a consistent and efficient estimator without relying on any assumption of the random intercept distribution. We describe the features and limitations of our approach and the general setting for which our estimator is valid. Simulation studies in Section 3 and an application to a cardiac injury study in Section 4 demonstrate the advantages of our estimator over traditional maximum likelihood estimators. Section 5 concludes the paper. All derivations are provided in the Supporting Information.

2. Main results

2.1. Semiparametric estimation

We assume throughout that $f_{X,R}(x_i, r_i)$ in (2) is an unknown density. We do not restrict the random intercept R_i to have mean zero, nor that $f_{X,R}(x_i, r_i) = f_X(x_i) f_R(r_i)$. Using semiparametric theory (Tsiatis, 2006), we demonstrate in Section S.1 (Supporting Information) that a consistent and efficient semiparametric estimator for β is the root of the estimating equation $\sum_{i=1}^{n} S_{\text{eff}}(Y_i, X_i; \beta) = \mathbf{0}$. Here,

$$S_{\text{eff}}(Y_i, X_i; \beta) = S_\beta(Y_i, X_i) - E\{h(X_i, R_i) \mid Y_i, X_i\},$$

with

$$\begin{aligned} S_\beta(Y_i, X_i; \beta) &= \partial \log f_{Y,X}(Y_i, X_i; \beta)/\partial\beta \\ &= E\{\partial \log f_{Y|X,R}(Y_i \mid X_i, R_i; \beta)/\partial\beta \mid Y_i, X_i\}, \end{aligned}$$

and h is an unknown, p-dimensional function that satisfies

$$E\{S_\beta(Y_i, X_i) \mid X_i, R_i\} = E[E\{h(X_i, R_i) \mid Y_i, X_i\} \mid X_i, R_i]. \tag{3}$$

In the terminology of semiparametric theory, S_{eff} is known as the efficient score vector so that estimators based on S_{eff} will have the smallest variance. Estimators based on S_{eff} are also consistent because $E\{S_{\text{eff}}(Y_i, X_i; \beta)\} = E[E\{S_{\text{eff}}(Y_i, X_i; \beta) \mid X_i, R_i\}] = \mathbf{0}$. The last equality holds from the definition of S_{eff} and because h is constructed to satisfy (3). Unfortunately, finding h to satisfy (3) is not a simple task. Directly solving (3) is computationally slow and numerically unstable. It requires solving an ill-posed problem similar to that in (Tsiatis & Ma, 2004), which can be cumbersome. Fortunately, however, these numerical issues can be avoided because S_{eff} actually has a closed-form solution as shown next.

2.2. *Explicit estimator*

To formulate an explicit estimator, our key idea is to construct an intermediate quantity that plays a similar role as the classical sufficient and complete statistic. To this end, we let 1_{m_i} and $\mathbf{0}_{m_i}$ be an m_i-dimensional vector of ones and zeros, respectively, and I_{m_i} be an $m_i \times m_i$ identity matrix. We also define $W_i = \mathbf{1}_{m_i}^{\mathrm{T}} \boldsymbol{Y}_i = \sum_{j=1}^{m_i} Y_{ij}$, and $V_i = \left(\mathbf{0}_{m_i-1}, I_{m_i-1}\right) Y_i = \left(Y_{i2}, \ldots, Y_{im_i}\right)^{\mathrm{T}}$. The variables W_i and $\boldsymbol{V}_i$ satisfy $\left(\boldsymbol{W}_i, \boldsymbol{V}_i^T\right)^T = \boldsymbol{A}_i \boldsymbol{Y}_i$, where

$$A_i = \begin{pmatrix} 1_{m_i}^{\boldsymbol{T}} \\ \mathbf{0}_{m_i-1} I_{m_i-1} \end{pmatrix}.$$

An immediately useful property of W_i and V_i is that they satisfy sufficiency and completeness conditions as stated in the proposition as follows.

Proposition 1. *In a logistic model with distribution-free random intercept,* $W_i = 1_{m_i}^{\boldsymbol{T}} Y_i$ *and* $\boldsymbol{V}_i = \left(\mathbf{0}_{m_i-1}, \boldsymbol{I}_{m_i-1}\right) Y_i$ *satisfy*

(i) Sufficiency: $f_{V|W,X,R}(v_i \mid w_i, x_i, r_i) = f_{V|W,X}(v_i \mid w_i, x_i)$ *and* $f_{R|W,X,V}(r_i \mid w_i, x_i, v_i) = f_{R|W,X}(r_i \mid w_i, x_i)$.

(ii) Completeness: $E\{g(W_i, X_i) \mid x_i, r_i\} = \mathbf{0}$ *implies* $g(W_i, X_i) = \mathbf{0}$ *for any p-dimensional function g.*

The proof of proposition 1 in Section S.2 (Supporting Information) uses calculations based on a change of variables. The result is of utmost importance as it allows us to show that S_{eff} actually has a closed-form solution as shown in proposition 2 (see Supporting Information, Section S.3 for a proof).

Proposition 2. *Because W_i and V_i satisfy the sufficiency and completeness properties in proposition 1, the efficient semiparametric estimator is*

$$S_{eff}(Y_i, X_i; \beta) = E\{S_\beta(Y_i, X_i, R_i) \mid W_i, V_i, X_i\} - E\{S_\beta(Y_i, X_i, R_i) \mid W_i, X_i\},$$

where $S_\beta(Y_i, X_i, R_i) = \partial \log f_{Y|X,R}(Y_i \mid X_i, R_i; \beta) / \partial\beta$ *with* $f_{Y|X,R}$ *given in (2).*

The result in proposition 2 thus provides a simple way to construct S_{eff}. First,

$$\partial \log f_{Y|X,R}(Y_i \mid X_i, R_i; \beta) / \partial\beta = \sum_{j=1}^{m_i} Y_{ij} X_{ij} - k(X_i, R_i) = X_i Y_i - k(X_i, R_i),$$

where $k(X_i, R_i) = \sum_{j=1}^{m_i} \exp\left(X_{ij}^{\boldsymbol{T}} \beta + R_i\right) X_{ij} / \left\{1 + \exp\left(X_{ij}^{\boldsymbol{T}} \beta + R_i\right)\right\}$.

Second, because of the sufficiency property in proposition 1, we have that $E\{k(X_i, R_i) \mid W_i, V_i, X_i\} = E\{k(X_i, R_i) \mid W_i, X_i\}$. This simplification along with $Y_i = A_i^{-1}\left(W_i, V_i^{\boldsymbol{T}}\right)^{\boldsymbol{T}}$ then implies that

$$\begin{aligned} S_{\text{eff}}(Y_i, X_i; \beta) &= E\{X_i Y_i - k(X_i, R_i) \mid W_i, V_i, X_i\} - E\{X_i Y_i - k(X_i, R_i) \mid W_i, X_i\} \\ &= X_i E\left\{A_i^{-1}\left(W_i, V_i^{\boldsymbol{T}}\right)^{\boldsymbol{T}} \mid W_i, V_i, X_i\right\} - X_i E\left\{A_i^{-1}\left(W_i, V_i^{\boldsymbol{T}}\right)^{\boldsymbol{T}} \mid W_i, X_i\right\} \\ &= X_i A_i^{-1}\left\{0, V_i^{\boldsymbol{T}} - E\left(V_i^{\boldsymbol{T}} \mid W_i, X_i; \beta\right)\right\}^{\boldsymbol{T}}. \end{aligned}$$

Therefore, a consistent and efficient estimator for β is the root of

$$\begin{aligned} \sum_{i=1}^{n} S_{\text{eff}}(Y_i, X_i; \beta) &= \sum_{i=1}^{n} X_i A_i^{-1}\left\{0, V_i^{\boldsymbol{T}} - E\left(V_i^{\boldsymbol{T}} \mid W_i, X_i; \beta\right)\right\}^{\boldsymbol{T}} \\ &= \sum_{i=1}^{n}\sum_{j=2}^{m_i}\left(X_{ij} - X_{i1}\right)\left\{V_{i,j-1} - E\left(V_{i,j-1} \mid W_i, X_i; \beta\right)\right\} = \mathbf{0}. \end{aligned} \tag{4}$$

2.3. Features of proposed estimator

The estimator $\hat{\beta}$ obtained from (4) has several key advantages including its simplicity, consistency and optimal efficiency, flexibility to non-standard distributions of the random intercept and possible dependence between the random intercept and covariates. We now discuss these advantages in more detail.

Theoretical advantages. The proposed estimator completely avoids all terms involving the random intercept. This contrasts from traditional maximum likelihood estimators (MLE) where computations rely on assuming a particular distribution for the random effect (e.g., normal). With our estimator, however, the sufficiency and completeness property of proposition 1 allows the component containing the random intercept, $k(X_i, R_i)$, to drop out from the efficient score. Dropping this term is advantageous because it means that our proposed estimator does not require positing a distribution $f_{X,R}$ nor does it require estimating this unknown distribution. That is, the estimator is implemented without making any reference to the true and unknown $f_{X,R}$. As such, our estimator is valid even when the variance of the random intercept is not finite, unlike other estimators that may assume the variance must be finite (Schall, 1991).

Eliminating the random effect term automatically excludes the possibility of estimating any quantity related to the random effect. This includes, for example, the random effect variance when it exists. A potential estimator to remedy this concern is the PQL estimator of Schall (1991) and Breslow & Clayton (1993). The PQL makes minimal modelling assumptions (i.e., the random effects are only assumed to have mean zero and finite variance) and obtains estimates of random effects parameters and their variances. Although this estimator performs well for a variety of mixed models, it unfortunately has bias for binary data from small clusters (Breslow & Clayton, 1993; Breslow & Lin, 1995; Lin & Breslow, 1996) and for models with heterogeneous random effects (i.e., the variability of the random effect depends on covariates) (Jang & Lim, 2009). Our own simulation study (Section 3) confirms this bias for the logistic random intercept model of interest.

Dropping the term $k(X_i, R_i)$ also leads to the consistency and efficiency of the proposed estimator being completely free of the unknown distribution $f_{X,R}$.

Theorem 1. *The estimator* $\hat{\beta}$ *solving* $\sum_{i=1}^{n} S_{eff}(Y_i, X_i; \beta) = \mathbf{0}$ *in (4) satisfies*

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \to Normal\left(\mathbf{0}, B^{-1}\right)$$

when $n \to \infty$. *Here* β_0 *is the true parameter value and* $B = var\{S_{eff}(Y_i, X_i; \beta_0)\}$. *Because* S_{eff} *is free of the unknown* $f_{X,R}(x_i, r_i)$, *the consistency and efficiency of* $\hat{\beta}$ *are completely independent of* $f_{X,R}$, *and the estimator* $\hat{\beta}$ *achieves the optimal estimation efficiency bound. That is, among the class of consistent estimators that do not make any distributional assumptions about the random intercept or its possible dependence with covariates, our proposed estimator has the smallest variability.*

A proof of Theorem 1 is in the Supporting Information, Section S.4. Theorem 1 indicates that $f_{X,R}(x_i, r_i)$ does not need to be estimated nor do any working models need to be posited in its place. This vastly differs from other semiparametric estimators where the estimator is constructed under posited working models for the unknown distributions, which can affect efficiency (Tsiatis & Ma, 2004; Ma & Carroll, 2006; Ma & Genton, 2010). For example, in most situations, when one is lucky enough to obtain a consistent estimator using a working model for the unknown distribution, the efficiency relies on the property of the working model. A classical scenario is when the working model happens to be the truth, then optimal efficiency is obtainable. Otherwise, only consistency is guaranteed. However, using our approach, the proposed estimator achieves consistency and optimal efficiency without the knowledge of the true random effect distribution or the attempt to estimate it. Therefore, the estimator is robust to misspecification of the unknown $f_{X,R}(x_i, r_i)$.

It is important to note that our estimator is not necessarily more efficient than estimators constructed under parametric $f_{X,R}(x_i, r_i)$. For example, the traditional normal-based MLE assumes that the random intercept is normally distributed and that the random intercept is uncorrelated with the covariates. In this case, the normal-based MLE may indeed have smaller variability because stronger model assumptions always reduce the variability of the parameter estimates. Despite this general phenomenon, we demonstrate in Section 3 that the reduction of variability is not necessarily a favourable aspect. Specifically, we show cases where the assumptions of the normal-based MLE do not hold, and the results are short confidence intervals that do not cover the truth.

Another advantage of our approach is that we do not require the covariates and random intercept to be independent, which is a common assumption with the traditional MLE implemented in standard software packages (e.g., glmer in the R package lme4 (Bates *et al.*, 2011)). Assuming X and R to be independent is not always valid, and testing for it may initially appear difficult because random intercepts are unobservable. Fortunately, however, we may formally test for independence along the lines of the Hausman chi-squared test (Hausman, 1978). It tests for a feature of interest (i.e., independence) by comparing two estimators as follows: under the null hypothesis, both estimators are consistent, and under the alternative, one estimator is inconsistent. Statistically significant differences between the estimators are then evidence against the feature of interest.

In our case, we will compare the difference between the estimate obtained assuming X and R are independent (i.e., the standard estimator available in software packages and denoted as $\hat{\beta}_{\text{IND}}$) and that obtained under no restrictions on the relationship between X and R (i.e., our proposed estimator denoted as $\hat{\beta}$). Under the null hypothesis that X and R are independent, both $\hat{\beta}_{\text{IND}}$ and $\hat{\beta}$ are consistent, whereas, under the alternative, $\hat{\beta}$ is consistent and $\hat{\beta}_{\text{IND}}$ is not. Therefore, a statistically significant difference between $\hat{\beta}$ and $\hat{\beta}_{\text{IND}}$ is evidence in favour of dependency between X and R. The Hausman chi-squared test statistic is

$\left(\hat{\beta} - \hat{\beta}_{IND}\right)^T \left\{var\left(\hat{\beta}\right) - var\left(\hat{\beta}_{IND}\right)\right\}^{-1} \left(\hat{\beta} - \hat{\beta}_{IND}\right)$, and it follows a chi-squared distribution with k degrees of freedom, where $k = \text{rank}\left\{var\left(\hat{\beta}\right) - var\left(\hat{\beta}_{IND}\right)\right\}$. See Hausman (1978) for the derivation of the test statistic.

Computational advantages. The proposed estimator is simpler and faster to compute than the traditional MLE, which requires numerically integrating out the random intercept in (2) and then maximizing. In comparison, the only integration in the proposed estimator is $E\left(V_i \mid W_i, X_i; \beta\right)$, which is straightforward because V_i is discrete and

$$f_{V|W,X}\left(v_i \mid w_i, x_i; \beta\right) = \frac{\exp\left\{\sum_{j=2}^{m_i}\left(x_{ij} - x_{i1}\right)^{\boldsymbol{T}} \beta v_{i,j-1}\right\}}{\sum_{R(v_i)} \exp\left\{\sum_{j=2}^{m_i}\left(x_{ij} - x_{i1}\right)^{\boldsymbol{T}} \beta v_{i,j-1}\right\}}.$$

Here, $\mathcal{R}\left(v_i\right)$ denotes the range of possible $v_i = \left(y_{i2}, \ldots, y_{i,m_i}\right)^T$ values such that $\sum_{j=1}^{m_i} y_{ij} = w_i$. See Section S.5 (Supporting Information) for a simple algorithm to construct $\mathcal{R}\left(v_i\right)$.

Comparing the proposed method to the traditional MLE (implemented in glmer in R) and the PQL estimator (implemented in glmmPQL in R) yielded the execution times in Table 1. Timings were carried out on an Intel Xeon 2.90 GHz processor. In general, our method was at least 12 times faster than either the MLE or PQL estimators, even as the cluster sizes increased. The fast performance of our method is attributed to the simple implementation of $E\left(V_i \mid W_i, X_i; \beta\right)$, which allows us to quickly establish the range $\mathcal{R}\left(v_i\right)$ (Supporting Information, Section S.5).

Limitations of the model. Relaxing the assumption on the random distribution does yield some limitations. First, we cannot estimate parameters associated with time-invariant covariates because time-invariant covariates cannot be distinguished from the random intercept, whose distribution is now left totally unspecified. In other words, time-invariant covariates form part of the random intercept and cannot be identified. Operationally, we can see the lack of identifiability from Equation (4). Assuming, for example, the first component of the covariate satisfies $X_{ij1} = X_{i11}$ for all j. Then the first component of the inner summand in (4) is identically zero, and the first component of the efficient score function degenerates to zero. This implies that the associated parameter β_1 is no longer estimable. Therefore, our method requires that $X_{ij} \neq X_{ij'}$ for $j \neq j'$. Methods with added assumptions to handle time-invariant covariates have been proposed in the economics literature. These include the Hausman and Taylor hybrid linear model (Wooldridge, 2002; Cameron & Trivedi, 2005) and quasi-differencing in some nonlinear models (Cameron & Trivedi, 2005).

Table 1. *Execution time (in seconds) for proposed estimator, traditional maximum likelihood estimator (implemented in glmer in R) and penalized quasi-likelihood estimator (PQL, implemented in glmmPQL in R). Results displayed for one simulation with cluster sizes* $\boldsymbol{m}$ *performed on an Intel Xeon 2.90 GHz processor*

m	Proposed estimator	Normal-based MLE	PQL
3	0.004	0.702	0.717
5	0.006	1.025	0.839
10	0.120	1.505	1.542

Second, clusters with a single observation will not contribute to the estimating equation in (4) because the estimating equation only involves observations from the second entry in the cluster onward. When a single observation is observed from a cluster, one cannot tell if this observation will change when a new observation from this cluster becomes available or not. Without any assumption on the random effect, this means we cannot distinguish if what we observe is a cluster property (random effect) or individual property (covariate effect). In comparison, the standard normal-based MLE and PQL estimators would keep those clusters with a single observation. For this reason, particular study design issues should be assessed early so that enough data within each cluster is collected for the clusters to be informative.

Note that the limitations are the price one must pay for relaxing the distributional assumptions on the random intercept.

3. Simulation study

3.1. Simulation design

We evaluated the performance of our estimator in comparison to two estimators: the traditional MLE that assumes the random effect is normally distributed and the PQL estimator. Both latter estimators also assume covariates are independent of the random intercept. All estimators were evaluated in terms of consistency, efficiency and 95 percent coverages in settings where the random intercept distribution is correctly specified and misspecified and when there is and is not dependency between covariates and the random intercept.

We generated 1000 data sets with sample size $n = 500$ and $m_i = 3$ for $i = 1, \dots, n$ for the logistic random intercept model in (1). We evaluated the robustness of the estimators to different distributional forms of R_i by considering the following: (i) R_i is $Normal(0, 1)$; (ii) R_i is a gamma distribution with shape parameter 1.5 and scale parameter 2; (iii) R_i is a mixture of normals with 90 percent of the data being $Normal(3, 1)$ and 10 percent of the data being $Normal(-3, 0.25)$; and (iv) R_i is a t-distribution with three degrees of freedom. These distributions allow the random intercept to be the standard bell shape, skewed, bimodal or heavy tailed. Finally, because the normal-based MLE and PQL estimators assume zero-mean random effects, the random intercepts generated were centred to have mean zero.

To evaluate the robustness of the estimators to varying levels of dependency between covariates and random intercept, we generated data according to three different dependency situations. Let $X_{ij} = \left(X_{ij1}, X_{ij2}, X_{ij3}\right)^T$. For $k = 1, 2, 3$, we considered the following: (i) X_{ijk} and R_i independent by generating X_{ijk} as $Normal(0, 1)$; (ii) X_{ijk} and R_i dependent and $\mathrm{corr}\left(X_{ijk}, R_i\right) = 0$ by generating X_{ijk} as $Normal\left(0, R_i^2\right)$; and (iii) X_{ijk} and R_i dependent and $\mathrm{corr}\left(X_{ijk}, R_i\right) \neq 0$ by generating X_{ijk} as $Normal\left(R_i, 1\right)$. In this last case, because $X_{ijk} \sim Normal\left(R_i, 1\right)$ and R_i is centred to have mean zero, we can easily show that $\mathrm{corr}\left(X_{ijk}, R_i\right) = \sqrt{E\left(R_i^2\right)} / \sqrt{1 + E\left(R_i^2\right)}$. The exact value of $\mathrm{corr}\left(X_{ijk}, R_i\right)$ thus depends on the distribution of R_i and is given in Table 4. Finally, the true $\beta_0 = (-0.25, -0.5, 0.25)^T$.

In summary, we considered all combinations between random intercept distribution settings (i)–(iv) and dependence settings (i)–(iii).

3.2. Simulation results

Results in Tables 2–4 show that regardless of the random intercept's true distribution or its possible dependence with the covariates, our proposed estimator had negligible bias, high efficiency and estimated coverage probabilities near the nominal 95 percent level. Such numerical results thus demonstrate the estimator's flexibility in handling a wide range of mixed model features:

a non-normal random intercept and a random intercept that may share dependency with the covariates. In addition, while the proposed estimator does have small variability, its variability is not smaller than that of the normal-based MLE or PQL. This is within our expectation, however, because stronger model assumptions always reduce the variability of the parameter estimates. Despite this general phenomenon, we argue next that the reduction of the variability is not necessarily a favourable aspect.

From Tables 2–4, we can see that the variance reduction from the proposed estimator to the normal-based MLE is smaller in Table 2 (where the components in X and R are independent) than in Tables 3 and 4 (where the components in X and R are dependent). Note also that the bias inflation is almost absent in Table 2 while it is evident in Tables 3 and 4. In these latter Tables, we see that estimates from the normal-based MLE are severely biased and result in coverages as low as 0 percent.

This is suggesting that the combined assumptions of normality and independence between covariates and the random intercept can have two effects simultaneously. The first effect is that the normality assumption tends to shrink the variability. The second effect is that the independence assumption tends to shift the estimator to the wrong place (unless the assumption happens to be correct).

When the two effects are not evident, the variance shrinkage and the estimator shift are both small, such as those in Table 2. In this case, assuming the additional normality and independence are advantageous as the normal-based MLE yields the smallest bias, smallest variance and the most unbiased variance estimate. When the two effects are evident, the variance shrinkage and the estimator shift are both large, such as in Tables 3 and 4. In this case, the seemingly advantageous small estimation variability can be very misleading, because the end result is a very precise but wrong estimator. This would lead to wrong inference results, for example, a

Table 2. *Simulation results when components of $\boldsymbol{X}$ are independent of $\boldsymbol{R}$. Bias, empirical variance* (var), average estimated variance* ($\widehat{\text{var}}$) and 95 percent coverage percentages (cov) for the proposed estimator, normal based maximum likelihood estimator (MLE), and penalized quasi-likelihood (PQL) estimator when the true random intercept distribution is as specified. $\beta_0 = (-0.25, -0.5, 0.25)^T$, $n = 500, m_i = 3$ and 1000 simulations*

	Proposed estimator				Normal-based MLE				PQL			
	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov
					$R \sim$ Normal (0,1)							
β_1	−0.0091	6.0	5.7	93.9	−0.0061	4.2	4.1	94.6	0.0198	3.4	3.0	91.8
β_2	−0.0041	5.5	6.2	95.8	−0.0033	4.1	4.5	96.6	0.0479	3.3	3.2	84.5
β_3	0.0042	5.8	5.7	94.2	0.0029	4.0	4.1	95.5	−0.0228	3.2	3.0	91.6
					$\sim$ Gamma$(1, 2)$ and centered							
β_1	−0.0036	7.4	7.0	94.6	−0.0037	5.6	5.5	94.6	0.0294	4.3	3.4	88.2
β_2	−0.0030	7.9	7.8	95.0	−0.0009	6.1	6.0	94.5	0.0642	4.8	3.5	77.0
β_3	0.0062	6.9	7.0	95.7	0.0036	5.1	5.5	96.0	−0.0292	4.0	3.4	89.9
					$R \sim 0.9(3, 1) + 0.1(-3, 0.25)$ and centered							
β_1	−0.0022	6.8	6.6	95.3	0.0002	4.8	5.1	95.9	0.0322	3.7	3.3	89.5
β_2	−0.0051	7.3	7.3	94.5	−0.0026	5.3	5.6	95.7	0.0619	4.1	3.4	78.6
β_3	0.0060	6.6	6.7	95.4	0.0036	5.2	5.1	95.7	−0.0288	4.0	3.3	89.3
					$R \sim t_3$							
β_1	−0.0028	6.0	6.2	96.1	0.0002	4.4	4.6	95.4	0.0303	3.4	3.2	90.7
β_2	−0.0049	7.1	6.8	94.4	−0.0024	4.9	5.1	95.4	0.0584	3.8	3.3	78.6
β_3	0.0027	5.9	6.2	95.4	0.0004	4.5	4.6	96.2	−0.0298	3.5	3.2	90.1

*Variances and estimated variances are multiplied by 1000.

Table 3. *Simulation results when components of $\boldsymbol{X}$ depend on $\boldsymbol{R}$, but are uncorrelated with $\boldsymbol{R}$. Bias, empirical variance* (var), average estimated variance* ($\widehat{\text{var}}$) and 95 percent coverage (COV) for the proposed estimator, normal based maximum likelihood estimator (MLE), and penalized quasi-likelihood (PQL) estimator when the true random intercept distribution is as specified.* $\boldsymbol{\beta_0 = (-0.25, -0.5, 0.25)^T}$, $\boldsymbol{n = 500}$, $\boldsymbol{m_i = 3}$ *and 1000 simulations*

	Proposed estimator				Normal-based MLE				PQL			
	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov
					$\boldsymbol{R} \sim$ Normal (0,1)							
β_1	−0.0105	9.1	8.7	94.9	0.0499	3.9	4.2	89.2	0.0687	3.2	3.3	76.1
β_2	−0.0100	9.1	10.0	97.2	0.1041	4.2	4.7	66.2	0.1413	3.4	3.5	34.2
β_3	0.0044	8.7	8.7	94.0	−0.0529	3.8	4.2	88.6	−0.0714	3.1	3.3	75.5
					$\boldsymbol{R} \sim$ Gamma$(1, 2)$ and centered							
β_1	−0.0061	5.4	5.5	95.1	0.1048	1.6	1.7	28.1	0.1208	1.2	1.1	7.7
β_2	−0.0112	7.1	6.8	93.9	0.2140	2.3	2.1	1.7	0.2453	1.8	1.3	0
β_3	0.0070	5.4	5.5	94.6	−0.1059	1.8	1.7	26.9	−0.1216	1.4	1.1	9.1
					$\boldsymbol{R} \sim$ 0.9 Normal (3,1)+0.1 Normal (-3,0.25) and centered							
β_1	−0.0042	6.4	7.2	96.7	0.1419	1.2	1.4	3.4	0.1536	0.9	1.0	0.7
β_2	−0.0134	8.1	9.1	96.9	0.2815	1.3	1.6	0	0.3053	1.1	1.1	0
β_3	0.0103	6.8	7.3	96.3	−0.1399	1.1	1.4	3.4	−0.1518	0.9	1.0	0.3
					$\boldsymbol{R} \sim t_3$							
β_1	−0.0086	8.0	7.8	95.6	0.1183	2.4	2.0	26.6	0.1327	1.9	1.6	11.6
β_2	−0.0159	12.7	9.5	95.4	0.2391	3.4	2.4	1.4	0.2676	2.8	1.8	0.3
β_3	0.0065	8.9	7.8	95.3	−0.1218	2.6	2.0	2.6	−0.1357	2.0	1.6	11.3

*Variances and estimated variances are multiplied by 1000.

very short confidence interval that does not cover the truth at all, which is even worse than a longer confidence interval that has a better chance to cover the truth.

These results suggest that there is no reason to make the additional normality or independence assumption when there is no clear evidence that the random intercept is indeed normal and independent of the covariates.

In comparison, the PQL estimator has more bias whether X and R are independent (Table 2) or not (Tables 3 and 4). The bias is larger when X and R are dependent, suggesting that the PQL estimator is most sensitive to violations of independent X and R. However, even when X and R are independent, the PQL estimates can have biases up to ten times more than the other two estimators and coverages as low as 77 percent. The sensitivity of the PQL estimator may just be due to the binary nature of the logistic model as suggested by earlier studies (Breslow & Clayton, 1993; Breslow & Lin, 1995; Lin & Breslow, 1996; Jang & Lim, 2009).

In addition to evaluating consistency and efficiency, we also evaluated the performance of the Hausman chi-squared test for independence (Section 2.3). We applied the test using the comparison between the proposed estimator and normal-based MLE, as well as between the proposed estimator and PQL estimator. Recall that the proposed estimator is consistent whether X and R are independent or not, whereas, both the normal-based MLE and PQL estimator are inconsistent when X and R are dependent. Therefore, when X and R are independent, the Hausman chi-squared test should fail to reject the null hypothesis of independence; conversely, when X and R are dependent, the test should reject the null.

The results in Table 5 show the percentage of times the Hausman test rejects the null across all combinations of random intercept distribution settings (i)–(iv) and dependence settings (i)–(iii). The test results from using the normal-based MLE and from using the PQL

Table 4. *Simulation results when components of $\boldsymbol{X}$ depend on $\boldsymbol{R}$ and are correlated with $\boldsymbol{R}$. Bias, empirical variance* (var), average estimated variance* ($\widehat{\text{var}}$) and 95 percent coverage (COV) for the proposed estimator, normal based maximum likelihood estimator (MLE) and penalized quasi-likelihood (PQL) estimator when the true random intercept distribution is as specified. $\beta_0 = (-0.25, -0.5, 0.25)^T$, $n = 500$, $m_i = 3$ and 1000 simulations*

	Proposed estimator				Normal-based MLE				PQL			
	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov
					$R \sim$ Normal (0,1), corr $(X, R) = 1/\sqrt{2}$							
β_1	−0.0085	5.3	5.0	93.8	0.2439	2.4	2.3	0	0.2436	2.3	2.2	0
β_2	−0.0054	5.1	5.5	95.7	0.2534	2.3	2.4	0	0.2573	2.2	2.2	0
β_3	0.0036	4.9	5.0	95.4	0.2335	2.5	2.6	0.02	0.2249	2.4	2.5	0.03
					$R \sim$ Gamma$(1, 2)$ and centred, corr $(X, R) = 0.926$							
β_1	−0.0025	5.6	5.5	95.5	0.2934	2.3	2.3	0	0.2925	2.2	2.2	0
β_2	−0.0020	5.8	6.1	95.7	0.3071	2.2	2.3	0	0.3088	2.1	2.2	0
β_3	0.0054	5.6	5.5	95.0	0.2780	2.6	2.7	0	0.2715	2.5	2.5	0
					$R \sim$ 0.9 Normal(3,1)+0.1 Normal (-3,0.25) and centred, corr $(X, R) = 0.535$							
β_1	−0.0029	5.4	5.5	95.5	0.2961	2.3	2.3	0	0.2952	2.2	2.2	0
β_2	−0.0050	6.0	6.1	96.2	0.3067	2.1	2.4	0	0.3086	2.1	2.2	0
β_3	0.0020	5.4	5.5	95.2	0.2766	2.7	2.8	0	0.2700	2.5	2.5	0.01
					$R \sim t_3$, corr $(X, R) = 0.896$							
β_1	−0.0036	5.2	5.2	95.6	0.2813	2.5	2.2	0	0.2804	2.4	2.1	0
β_2	−0.0026	5.6	5.8	95.1	0.2909	2.3	2.3	0	0.2933	2.2	2.2	0
β_3	0.0016	5.1	5.2	95.5	0.2606	2.8	2.7	0	0.2533	2.7	2.5	0

*Variances and estimated variances are multiplied by 1000.

Table 5. *Percentage of time the Hausman chi-squared test rejects that null hypothesis that covariates and the random intercept are independent. The test is applied using the comparison between proposed estimator and normal-based maximum likelihood estimator (MLE) and between proposed estimator and penalized quasi-likelihood (PQL) estimator. Results are based on different distributions for the random intercept and different dependency between covariates and random intercept as specified in the main text. $\beta_0 = (-0.25, -0.5, 0.25)^T$, $n = 500$, $m_i = 3$ and 1000 simulations*

Covariates		Proposed versus normal-based MLE				Proposed versus PQL			
		Distribution of R				Distribution of R			
Depend on R	Correlated with R	Normal	Gamma Normals	Mixed	t	Normal	Gamma	Mixed Normals	t
No	No	8.8	12.4	14.5	11.1	2.1	0.5	0.8	1.9
Yes	No	22.8	93.0	97.6	85.3	30.6	97.5	99.1	90.0
Yes	Yes	100	100	100	100	100	100	100	100

estimator are similar. When X and R are independent (Table 5, first row), the test correctly fails to reject the null; an incorrect decision is made at most 15 percent of the time. Conversely, when X and R are dependent, the test largely rejects the null. When the covariates depend on R, but the correlation is zero (Table 5, second row), the rejection rate is at least 85 percent when R is non-normal and is at least 22 percent when R is normal. The lower percentage when R is normal is perhaps because only one of the assumptions (i.e., independence of covariates and random intercept) is violated, and the correlation between the covariates and R is zero. When R is normal, but the covariates and R have non-zero correlation (Table 5, third row), the Hausman chi-squared test rejects the null 100 percent of the time. This steep increase from 22 percent to 100 percent suggests that the non-zero correlation has a strong impact. In fact, when the correlation between the covariates and R is non-zero (Table 5, third row), the null is rejected 100 percent of the time regardless of the distribution of R. Therefore, these results suggest that the Hausman chi-squared test works well in detecting dependence between the covariates and random intercept, especially when the correlation is non-zero.

Finally, it is worth noting that for smaller sample sizes (i.e., $n = 50, m_i = 2$), all estimators perform similarly as with larger sample sizes. See Table 6 for results from our proposed estimator and the normal-based MLE under the different dependency cases (i), (ii) and (iii), and R is normally distributed or a mixture of normals. We exclude results from the PQL estimator given that its bias has already been observed at larger sample sizes ($n = 500, m_i = 3$), and its bias only compounds further at smaller sample sizes.

In summary, our proposed estimator provides practical advantages in that it yields consistent and optimally efficient estimates without needing to make assumptions about the random effect distribution or its relationship to the model covariates.

4. Application to cardiology study

We applied our proposed method to a study of cardiac injury (Tung *et al.*, 2004) from UC San Francisco. The study involved 175 subjects who had undergone different cardiac measures

Table 6. *Simulation results for small sample setting for different distributions of* $\boldsymbol{R}$ *and dependence with* $\boldsymbol{X}$*. Bias, empirical variance (var), average estimated variance (*$\widehat{\text{var}}$*) and 95 percent coverage percentages (cov) for the proposed estimator and normal-based maximum likelihood estimator (MLE).* $\beta_0 = (0, 0, 0)^T$, $n = 50, m_i = 2$ *and 1000 simulations*

	Proposed estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	cov	bias	var	$\widehat{\text{var}}$	cov
				$R \sim$ Normal (0,1)				
				X and R independent				
β_1	−0.0118	0.2132	0.2455	98.7	−0.0058	0.0782	0.0717	95.2
β_2	−0.0057	0.2214	0.2464	98.6	0.0038	0.0788	0.0725	95.9
β_3	−0.0371	0.2398	0.2587	98.7	−0.0133	0.0757	0.0721	97.1
				X and R dependent, uncorrelated				
β_1	0.0023	0.5040	0.7761	99.8	−0.0065	0.0834	0.0828	97.2
β_2	−0.0155	0.4981	0.8560	99.8	0.0039	0.0842	0.0828	97.1
β_3	−0.0381	0.5134	0.8537	99.9	−0.0085	0.0807	0.0815	97.5
				X and R dependent, correlated				
β_1	−0.0118	0.2132	0.2455	98.7	0.2613	0.0540	0.0485	78.4
β_2	−0.0057	0.2214	0.2464	98.6	0.2632	0.0504	0.0484	79.3
β_3	−0.0371	0.2398	0.2587	98.7	0.2552	0.0475	0.0482	79.9
				$R \sim 0.9$ Normal(3,1)+0.1 Normal(-3,0.25) and centred				
				X and R independent				
β_1	0.0178	0.2984	0.3862	98.9	0.0030	0.1053	0.0975	95.8
β_2	−0.0269	0.2955	0.4334	99.3	−0.0117	0.1007	0.0968	96.8
β_3	−0.0107	0.2791	0.4327	99.7	−0.0033	0.0935	0.0971	97.3
				X and R dependent, uncorrelated				
β_1	0.0310	0.5526	1.2218	99.9	−0.0027	0.0363	0.0345	97.8
β_2	−0.0130	0.6049	1.3356	99.9	−0.0053	0.0313	0.0335	97.6
β_3	−0.0310	0.5517	1.2865	99.8	0.0081	0.0335	0.0347	97.7
				X and R dependent, correlated				
β_1	0.0178	0.2984	0.3862	98.9	0.3127	0.0585	0.0579	75.3
β_2	−0.0269	0.2955	0.4334	99.3	0.2923	0.0579	0.0568	77.5
β_3	−0.0107	0.2791	0.4327	99.7	0.2945	0.0608	0.0571	78.4

Table 7. *Results from the cardiology study based on the proposed estimator and normal-based maximum likelihood estimator (MLE). Parameter estimates (est), estimated variance ($\widehat{var}$) and 95% confidence interval (CI) for the effects of ejection fraction ($\beta_{ejecfrac}$), systolic blood pressure (β_{sbp}) and heart rate (β_{hrtrate})*

	Proposed estimator			Normal-based MLE		
	est	$\widehat{\mathrm{var}}$	95% CI	est	$\widehat{\mathrm{var}}$	95% CI
$\beta_{\mathrm{ejecfrac}}$	−2.7736	8.5965	(−8.5202, 2.9729)	−3.3445	1.4077	(−5.6700, −1.0191)
β_{sbp}	−0.0264	0.0002	(−0.0558, 0.0030)	−0.0157	4.25×10^{-5}	(−0.0285, −0.0030)
β_{hrtrate}	−0.0225	0.0004	(−0.0614, 0.0163)	0.0215	0.0001	(0.0071, 0.0360)

repeatedly over the study period (at most three times). The measures included ejection fraction that quantifies how much blood leaves the left ventricle with each contraction, systolic blood pressure and heart rate. Doctors also took a binary measure of cardiac injury of the subjects: whether or not troponin – an enzyme observed after heart damage – exceeded a specific threshold. One key interest is investigating the impact of ejection fraction, systolic blood pressure and heart rate on the binary measure of cardiac injury while accounting for the within-person correlation induced from the repeated measures.

To study this impact, we considered the logistic random intercept model in (1) where for subject i at time j, Y_{ij} is whether or not the troponin level exceeded the specified threshold, X_{ij1} is ejection fraction, X_{ij2} is systolic blood pressure and X_{ij3} is heart rate. Lastly, R_i is the random intercept to account for the subject effect.

The effects of these cardiac measures were estimated using both our proposed method and the normal-based MLE for comparison. The PQL estimator was not applied because of its observed bias from the simulation study. Results for both estimators are given in Table 7. The proposed estimator suggests that none of the cardiac measures is statistically significant as all 95% confidence intervals include the null effect. In stark contrast, the normal-based MLE suggests that all measures are statistically significant. To better compare the estimates, we applied the Hausman chi-squared test and obtained a p-value of 2.42×10^{-5}. This suggests evidence of dependence between the covariates and the random intercept and indicates that results from the proposed estimator are more reliable. Therefore, in terms of troponin impact, it appears that none of the cardiac measures has a significant effect. Such results agree with earlier analysis results (Vittinghoff *et al.*, 2005) that another feature, the severity of neurological injury, has an overwhelmingly dominating effect on troponin levels.

5. Discussion

For a logistic model with random intercept, our semiparametric approach yields a consistent and optimally efficient estimator that does not require estimating or modelling the random effect distribution. It does not even require proposing a working model in the process of estimation and inference. In our construction, the complete insensitivity of the estimator to the random effect is especially novel given that for most semiparametric estimators, efficiency loss is observed for misspecified working models (Tsiatis & Ma, 2004; Ma & Carroll, 2006). Our estimator also has practical advantages in detecting dependence between the covariates and random intercept via the Hausman chi-squared test. Our extensive simulation study also suggests that when there is doubt about the assumptions for traditional estimators, it is better to proceed with the proposed estimator that is consistent, efficient and simple to compute.

Our proposed estimator has similarity to the conditional likelihood estimator of Breslow & Day (1980) and Neuhaus & McCulloch (2006). However, the derivation leading to their estimator focuses on separating the between-cluster and within-cluster covariate effects and maximizing $f_{V|W,X}$ to achieve consistency. In comparison, we use a semiparametric approach that has the flexibility to reveal other potential settings where a consistent and optimally efficient estimator exists. For example, one may consider mixed models where the density for the ith cluster can be written as

$$
\begin{aligned}
f_{Y,X,Z}\left(y_i, x_i, z_i\right) &= \int f_{Y|X,Z,R}\left(by_i|x_i, z_i, r_i\right) f_{X,Z,R}\left(x_i, z_i, r_i\right) d\mu\left(r_i\right) \\
&= \int \prod_{j=1}^{m_i} \exp\left\{\frac{y_{ij}\left(x_{ij}^T\beta + z_{ij}^T r_i\right) - b\left(x_{ij}^T\beta + z_{ij}^T r_i\right)}{\phi_{ij}} + c\left(y_{ij}, \phi_{ij}\right)\right\} \\
&\quad \times f_{X,Z,R}\left(x_i, r_i\right) d\mu\left(r_i\right).
\end{aligned}
$$

Here, the conditional density $f_{Y|X,Z,R}\left(y_i|x_i, z_i, r_i\right)$ belongs to the exponential family where $b(\cdot), c(\cdot)$ are known functions, and the unknown parameters are now β and dispersion parameters ϕ_{ij}. We let R be a p_r-dimensional vector of random effects associated with covariates $z_{ij} = \left(1, z_{ij,1}, \ldots, z_{ij,p_r}\right)^T$. For example, in a simple case, $R = (R_0, R_1)^T$ with R_0, R_1 denoting the random intercept and slope, respectively. In this more general setting, we believe similar calculations to those carried out in Section 2.2 can be applied. However, given the added generality of an exponential model with p_r random effects, a careful investigation is needed and is beyond the scope of the current paper.

For binomial data, one may also be interested in considering other links such as probit, log–log and complementary log–log. While these links are useful in various applications, we unfortunately cannot use sufficiency and completeness properties to derive an optimally efficient estimator. The reason is that the terms $x_{ij}^T\beta + z_{ij}^T r_i$ in the aforementioned conditional density $f_{Y|X,Z,R}$ are now replaced with $q\left(x_{ij}^T\beta + z_{ij}^T r_i\right)$ for a nonlinear function $q(\cdot)$. For example, the probit link would have $q(t) = \text{logit}\{\Phi(t)\}$ with $\Phi(\cdot)$ being the standard normal distribution; the log-log link would have $q(t) = \text{logit}[\exp\{-\exp(-t)\}]$; and the complementary log–log link would have $q(t) = \text{logit}[1 - \exp\{-\exp(t)\}]$. For such a 'nonlinear-exponential' model, the results of propositions 1 and 2 do not hold because the nonlinear $q(\cdot)$ completely hinders the sufficiency property: No statistic exists that is independent of the random effect. Therefore, for these more complex link functions, other estimation procedures are needed. A semiparametric approach may still be applied, but it would most likely involve more intense computation. Details into this problem are of interest, but again are beyond the scope of this paper.

Lastly, an issue not considered is the potential colinearity among explanatory variables, which does occur frequently in medical and social research. This problem is not addressed by the traditional MLE and cannot be immediately addressed by our proposed method because the entire approach is conditional on the covariates. However, it is worth considering in the future.

Supporting information Additional information for this article is available online, including proofs and details of the proposed algorithm.

Acknowledgements

Garcia is supported by the Huntington's Disease Society of America Human Biology Project Fellowship. Ma is supported by grants from the National Science Foundation and National Institute of Neurological Disorders and Stroke. The authors thank Charles E. McCulloch and Jonathan Zaroff for providing the cardiology study data. The authors also thank the editor, associate editor and four anonymous referees whose comments substantially improved the quality and presentation of the work.

References

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Journal of Computational and Graphical Statistics* **47**, 639–653.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117–128.

Bates, D., Maechler, M. & Bolker, B. (2011). *lme4: linear mixed-effects models using S4 classes. R package version 0.999375-41*. http://CRAN.R-project.org/package=lme4.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Breslow, N. E. & Day, N. E. (1980). Statistical methods in cancer research. *International agency for research on cancer* **32**, 5–338.

Breslow, N. E. & Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

Cameron, A. C. & Trivedi, P. K. (2005). *Microeconomics: methods and applications*, Camridge University Press, New York.

Chen, J., Zhang, D. & Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear models with flexible random effects distribution. *Biostatistics* **3**, 347–360.

Fernandez, C. & Steel, M. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**, 359–371.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251–1271.

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

Heagerty, P. J. & Kurland, B. F. (2001). Misspecied maximum likelihood estimates and generalized linear mixed models. *Biometrika* **88**, 973–985.

Jang, W. & Lim, J. (2009). A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects. *Communications in Statistics–Simulation and Computation* **38**, 692–702.

Lesaffre, E. & Molenberghs, G. (2001). Multivariate probit analysis: a neglected procedure in medical statistics. *Statistics in Medicine* **10**, 1391–1403.

Lin, X. & Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**, 1007–1016.

Ma, Y. & Carroll, R. J. C. (2006). *Journal of the American Statistical Association* **101**, 1465–1474, Locally efficient estimators for semiparametric models with measurement error.

Ma, Y. & Genton, M. G. (2010). Explicit estimating equations for semiparametric generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B* **72**, 475–495.

Neuhaus, J. M. & McCulloch, C. E. (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B* **68**, 859–872.

Piepho, H. P. & McCulloch, C. E. (2004). Transformations in mixed models: application to risk analysis for a multienvironment trial. *Journal of Agricultural, Biological and Environmental Statistics* **9**, 123–137.

Pinheiro, J. C., Liu, C. & Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution. *Journal of Computational and Graphical Statistics* **10**, 249–276.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*, Springer, New York.

Tsiatis, A. A. & Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91**, 835–848.

Tung, P., Kopelnik, A., Banki, N., Ong, K., Ko, N., Lawton, M. T., Gress, D., Drew, B. J., Foster, E., Parmley, W. W. & Zaroff, J. G. (2004). Predictors of neurocardiogenic injury after subarachnoid hemorrhage. *Stroke* **35**, 548–551.

Turner, R. M., Omar, R. Z. & Thompson, S. G. (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* **20**, 453–472.

Vittinghoff, E., Glidden, D., Shiboski, S. & McCulloch, C. E. (2005). *Regression methods in biostatistics: linear, logistic, survival and repeated measures*, Springer, New York.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*, MIT Press, Cambridge.

Zhang, D. & Davidian, M. (2001). Linear mixed models with flexible distribution of random effects for longitudinal data. *Biometrics* **57**, 795–802.

Received December 2013, in final form June 2015

Tanya P. Garcia, Department of Epidemiology and Biostatistics, Texas A&M University, College Station, TX 77843-1266, USA.
E-mail: tpgarcia@srph.tamhsc.edu

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.