Prediction of random effects in mixed effects models under violations of the normality assumption for the random-effects and a graphical approach to detect violations

By

Zhiwen Wang

Submitted to the graduate degree program in Biostatistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Committee Chair: Francisco J. Diaz, Ph.D.

_____

Jo A. Wick, Ph.D.

_____

Jianghua He, Ph.D.

_____

John Keighley, Ph.D.

_____

Babalola Faseru, M.D., MPH

Date Defended: November 7, 2019

The dissertation committee for Zhiwen Wang certifies that this is
the approved version of the following thesis:

Prediction of random effects in mixed effects models under violations of the normality assumption for the random-effects and a graphical approach to detect violations

_____
Chair: Francisco J. Diaz, Ph.D.

_____
Date

_____
Graduate Director: Jo A. Wick, Ph.D.

_____
Date

Date Approved: November 11, 2019

Abstract

In longitudinal data analysis, the introduction of random effects provides statisticians with a convenient tool for modeling repeated measurements. Mixed effects linear models extend linear and generalized linear models for non-repeated measures to repeated measures or longitudinal data. One important assumption of these models is that the random effects are normally distributed. In this dissertation, we investigated via simulations the impact of violations of this assumption on the prediction of the random effects, by comparing the prediction accuracy and robustness of two methods: the empirical Bayes method and a semi-parametric method based on quadratic inference functions. Chapter 1 explores this impact for continuous responses modeled with the random effects linear model and Chapter 2 explore this impact for the random-effects logistic regression model. Finally, Chapter 3 proposes and examines a graphical method to examine this assumption in the context of two-dimensional time-dependent personalized medicine models with continuous responses that track the trajectories of patients' disease severities and individual treatment benefits when the patients are under medical or behavioral treatments. One important conclusion of these investigations is that the empirical Bayes approach is very robust to violations of the normality assumption. The EB approach has non-inferior but usually higher accuracy in random effects prediction, and is computational, numerical and algebraic simply. Thus, it is more recommendable for random-effects prediction than the method based on quadratic inference functions in statistical practice. Finally, our graphical approach successfully detected departures from the normality assumption and worked efficiently even with small and moderate sample sizes.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Francisco Diaz. I appreciate all his effort, enthusiasm, and patience in supervising me over the last five years. I remember the first course I took from him, Mathematical Statistics. I was immediately inspired by his knowledge and insight. He immersed me in the world of longitudinal data analysis and its applications to personalized medicine. My statistical knowledge grew significantly as a direct result of the opportunities he gave me, and the challenges he posed for me over these years. His insight and suggestions always quickly help me to solve any problems I encountered. Words cannot fully express how thankful I am to have such a wonderful advisor. I could not finish this difficult journey without his constant challenge and encouragement. He taught me how to be a researcher and he is my best role model for a scientist, mentor, and teacher. I have enjoyed my journey to becoming a PhD under his supervision.

We are grateful to Dr. Annie Qu from Department of Statistics at the University of Illinois, Urbana-Champaign, and Dr. Peng Wang from Department of Operations, Business Analytics, and Information Systems at the University of Cincinnati, for their generous and insightful comments for my first two Chapters.

My deepest gratitude goes to all my committee members for their constant support, kindness and counseling. I would like to thank Dr. Jo Wick for helping me to navigate the curriculum and coursework that ultimately put me on the right path in my PhD statistics program. I am also very grateful for all the encouragement from her in the face of frustration and confusion. I would like to thank Dr. Jianghua He for her generous counseling and her believing in my abilities. Her insights refined my work many times. Dr. John Keighley provided me with many practical

suggestions when I had problems with SAS/IML programming. His ingenuous advice helped me finished some of the most difficult parts of my work. I am very much thankful withy Dr. Babalola Faseru who has taught me to view and use statistics from an epidemiology perspective.

I also would like to thank the whole department for any and every role they have played in my education. I ultimately would like to thank Dr. Matthew Mayo for welcoming me to this program and giving me this opportunity. Without this opportunity, I would have never met so many talented professors and peers. I would also like to thank Rosemary Morrow and Mandy Rametta for their help and support in all these years.

Last but not the least, I would like to thank my mother and father, for encouraging me to chase my dreams, for supporting me in life, and showering me with endless love. I would also like to thank the University of Kansas Medical Center and associated departments and my university's friends. The time I spent at KU is balanced and colorful because of their accompanying.

The following is the acknowledgment for chapter II. The authors thank Drs. Annie Qu and Dr. Peng Wang for comments and suggestions. The depression trial was supported by the P. Carl Petersens Fond (Grants Nos. B.868 and B.975) and the Danish Medical Research Council (grants nos 512-2537, 512-3653, 512-4145, 512-5045, and 512-5719). None of the authors was involved in the planning or conduction of this trial and have never received any payment or salary from the trial's associated grant contracts or its investigators. Mr. Wang was supported by the Biostatistics and Informatics Shared Resource of the University of Kansas Medical Center, funded by the Cancer Center Support Grant P30 CA168524 of the National Cancer Institute, and the Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core which is partially supported by the National Institute of General Medical Science award P20GM103418, and by the

Department of Biostatistics of the University of Kansas Medical Center. Dr. Diaz was supported by an Institutional Clinical and Translational Science Award from the National Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366 (KL2/TL1), awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH.

The following is the acknowledgment for chapter III. Mr. Wang was supported by the Biostatistics and Informatics Shared Resource of the University of Kansas Cancer Center, funded by the Cancer Center Support Grant P30 CA168524 of the National Cancer Institute. Access to high performance computing was jointly supported by the Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core which is partially supported by the National Institute of General Medical Science award P20GM103418 and the Cancer Center Support Grant P30 CA168524 of the National Cancer Institute. Dr. Diaz was supported by an Institutional Clinical and Translational Science Award from the National Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366 (KL2/TL1), awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH. The authors thank Drs. Annie Qu and Dr. Peng Wang for comments and suggestions.

The following is the acknowledgement for chapter IV. The depression trial was supported by the P. Carl Petersens Fond (Grants Nos. B.868 and B.975) and the Danish Medical Research Council (grants Nos 512-2537, 512-3653, 512-4145, 512-5045, and 512-5719). None of the authors was involved in the planning or conduction of this trial and have never received any payment from the trial's associated grant contracts or its investigators. Mr. Wang was supported by the Biostatistics and Informatics Shared Resource of the University of Kansas Cancer Center, funded by the Cancer Center Support Grant P30 CA168524 of the National Cancer Institute. Dr. Diaz was supported by an Institutional Clinical and Translational Science Award from the National

Institutes of Health (NIH), NIH/NCATS Grant Number UL1TR002366 (KL2/TL1), awarded to the University of Kansas Medical Center. This manuscript may not reflect the opinions or views of the NIH.

Table of Contents

List of Figures

Chapter II

Appendix A

Chapter III

Chapter IV

List of Tables

Chapter II

Chapter III

Chapter IV

Appendix C

Chapter I: Introduction

Repeated measurements over time within individuals often occur in longitudinal studies in medicine, public health, social science, economy, education and agriculture. Two commonly used approaches to dealing with the repeated measurements are marginal models and mixed effects models. Marginal models are useful when the research is focused on average population effects and the mean responses depend on the fixed effects of the covariates of interest. In contrast, in regression models with mixed effects, the individual responses are modeled not only with the fixed effects but also with the random effects that constitute the unique individual regression parameters. The two approaches differ in how model parameters are estimated and, also important, how they are interpreted. In applications, these approaches allow statisticians to study the temporal changes of the responses within individuals and the relationships of the covariates with the responses, and to make inferences on the population and individual effects of the covariates and predictions of future responses (Diaz, 2016; Gillies et al., 2006; Have et al., 1998; Hedeker, 2003; Horrocks and van Den Heuvel, 2009; Kleinman et al., 2004; Lin and Breslow, 1996; D. Liu et al., 2008; Mann et al., 2018; Sashegyi et al., 2000; Skrondal and Rabe-Hesketh, 2003; Van Den Noortgate et al., 2003).

In mixed effects models, the random effects are usually assumed to follow a normal distribution with zero mean. In theory, any other distribution could be used in this assumption, but the normal distribution is widely used due to its ease in implementation and well-known and convenient theoretical properties. The estimation of the fixed effects is implemented by maximizing the marginal likelihood, which is the likelihood function of responses obtained by integrating over the distribution of the random effects. This process usually involves high-dimensional integration if approximation methods such as Gauss-Hermite quadrature is used. An alternative to maximum likelihood is the quasi-likelihood method. Marginal quasi-likelihood

(MQL) and penalized quasi-likelihood (PQL) are variants of this method (Breslow and Clayton, 1993).

In mixed effects models, the prediction of the random effects is frequently made via an empirical Bayes approach (Fitzmaurice et al., 2011, 2009; Frees, 2004; Hedeker and Gibbons, 2006). In linear mixed effects models the empirical Bayes predictor of random effects is an estimator of the best linear unbiased predictor, which, as suggested by previous research (McCulloch and Neuhaus, 2011a, 2011b) and confirmed in this dissertation, is relatively robust to violations of the normality assumption of the random effects. For the logistic model with random effects it is not possible to obtain a closed-form formula for the empirical Bayes predictor of the random effects. The robustness of this predictor to violations of the normality assumption of the random effects is also explored in this dissertation.

Using linear and logistic mixed effects models, Verbeke and Lesaffre (1996), McCulloch and Neuhaus (2011a, 2011b) found that histograms of the empirical Bayes predictors of the random effects may not reflect the true shape of the distribution of the random effects and thus they are not a convincing tool in evaluating the normality assumption. But prediction accuracy measured by mean square errors of prediction are less affected by violations of the normality assumption of the random effects. Agresti et al. (2004) found that in logistic models with random effects the estimates of the variance components are not severely affected by this violation except for very extreme cases. Marquart and Haynes (2019) showed that in the logistic model with only a random intercept, if the true random effects are from a three-components mixture of normal distributions, the bias of the estimator of the random intercept variance is large.

Some methods have been proposed to check the goodness-of-fit of mixed effects models (Abad et al., 2010; Alonso et al., 2008; Drikvandi et al., 2017; Efendi et al., 2017; Pan and Lin, 2005; Tchetgen and Coull, 2006; Verbeke and Molenberghs, 2013; Waagepetersen, 2006). In statistical practice, however, the normality assumption for the random effects is often hard to validate. Due to this limitation and the desire of incorporating population heterogeneity in a more realistic way, estimation methods robust to violations of the normality assumption have been proposed (Chen et al., 2002; Cho et al., 2017; Ghidey et al., 2004; Jiang, 1999; Shen and Louis, 1999; Ten Have et al., 1999; Ten Have and Localio, 1999; Wang et al., 2012; Zhang and Davidian, 2001; Zhu and Qu, 2016).

In addition to the empirical Bayesian approach, this dissertation examines the approach proposed by Wang et al. (2012), which is a semi-parametric method based on quadratic inference functions (QIFs), which in turn are an extension of generalized estimation equations (GEEs). An important concept of GEEs is the utilization a working correlation matrix. In the QIF approach, the inverse of this matrix is approximated by linear combinations of known basis matrices (Qu et al. 2000). The fixed and random effects are obtained by iteratively minimizing the quadratic inference functions based on the extended score equations of both fixed and random effects.

Little research has been done to compare random effects predictors and it is unclear to what extent the violations of normality for random effects affects these predations of random effects. This dissertation compares the prediction performance and the robustness to violations of the normality assumption of two approaches: the common empirical Bayes approach and the approach based on quadratic inference functions. Predictors in the context of linear mixed effects models and logistic mixed effects models are examined in Chapters II and III, respectively.

Random effects models have useful applications to the prediction of individual effects of medical treatments (Diaz, 2017, 2016), drug dose individualization (Diaz, Cogollo, et al., 2012; Diaz & de Leon, 2013; Diaz, Yeh, & de Leon, 2012; Zhu & Qu, 2016), and treatment individualization (Cho et al., 2017; Diaz, 2018). In Chapter IV, we propose a method to examine the normality assumption for the random effects when the goal of using a random effects model is to predict individual treatment benefits in severely ill patients. The essential idea of our graphical approach is to plot the quantiles of the empirical Bayes estimates of individual treatment benefits against the theoretical quantiles of the distribution of individual benefits which are derived under a normality assumption for the random effects. We used Monte Carlo simulations to study the performance of the proposed graphical approach assuming a variety of non-normal distributions for the random effects.

In Chapter V, the dissertation is concluded with a summary and a discussion of potential future research directions.

Chapter II: Prediction accuracy and robustness to non-normality of two methods of predicting random effects in linear mixed effects models for longitudinal data: empirical Bayes versus quadratic inference functions

Zhiwen Wang, John Keighley, Jianghua He, Jo Wick, Francisco J. Diaz*

ABSTRACT

Several methods for predicting random effects in linear mixed effects models have been proposed. The performances of these methods have not been thoroughly investigated when the normality assumption for the random effects is violated, except for the empirical Bayes (EB) approach, and comparisons of the methods have not been made. This simulation study compared the prediction accuracy of the EB approach with that of an approach based on quadratic inference functions (QIFs) under different distributional assumptions for the random effects, using a longitudinal linear model that included a random intercept and a random slope for time. The simulations revealed that the EB approach was generally superior to the QIF approach in predicting the random effects, even under non-normal distributions for the random effects, except in some scenarios with very large error variances. In addition, the EB approach is mathematically and computationally less complex. Thus, our study suggests that the EB approach is more recommendable as the first choice in statistical practice, even if non-normal random effects are suspected. An application to the prediction of individual benefits of an anti-depressant drug was considered.

KEYWORDS: Best linear unbiased predictors (BLUPs); Bivariate t distribution; Cross-validation; Distribution misspecification; Estimating equations; Mixtures of normal distributions.

1. Introduction

Linear mixed effects models are widely used in biostatistics applications (Cho et al., 2017; Diaz, 2018, 2017, 2016; Dimova et al., 2011; Fitzmaurice et al., 2011; Hooks et al., 2009; Laird and Ware, 1982; Verbeke and Molenberghs, 2000; Yau et al., 2003; Zhu and Qu, 2016). In particular, in the analysis of longitudinal data, they are used to model continuous responses measured over time from subjects under study. A basic assumption is the normality of both random effects and measurement error terms (Cho et al., 2017; Fitzmaurice, 2009; Frees, 2004; Laird and Ware, 1982; Verbeke and Molenberghs, 2009). Failure to include a subject-specific covariate, however, could result in a violation of the normality assumption for the random effects (Frees, 2001; Verbeke and Lesaffre, 1996). The assumption is also violated if the random effects follow heavy tailed distributions or mixtures of different distributions determined by unidentified different groups of subjects under study (Verbeke and Lesaffre, 1997, 1996). These assumption violations usually do not have a large impact on fixed effects estimation but may affect the estimation of variance components, which in turn may affect the prediction of individual random effects (Agresti et al., 2004; Alonso et al., 2008; Heagerty and Kurland, 2001; Mcculloch and Neuhaus, 2011a, 2011b; Verbeke and Lesaffre, 1997, 1996). Thus, inaccuracies in random effects prediction may be possible if the true random effects follow non-Gaussian distributions. This limitation of the normality assumption suggests the need for robust methods to estimate (or "predict") random effects in linear mixed effects models.

The goal of this paper is to compare two methods of predicting random effects: a method based on quadratic inference functions (Wang et al., 2012) versus the classic empirical Bayes approach that is based on best linear unbiased predictors (BLUPs) (Fitzmaurice et al., 2011; López

et al., 2007; Skrondal and Rabe-Hesketh, 2004; Verbeke and Molenberghs, 2000). Comparisons under Gaussian and non-Gaussian random-effect distributions were made.

There are situations in which an accurate prediction of random effects is necessary. For instance, in personalized medicine applications, a precise prediction of random effects is important for medical or behavioral treatment individualization, assessment of individual treatment benefits, and drug dosage individualization (Andrews and Cho, 2018; Cho et al., 2017; Diaz, 2017, 2016; Diaz et al., 2012a, 2012b, 2007; Diaz and de Leon, 2013; Zhu and Qu, 2016). In particular, a precise measurement of individual benefits requires an accurate prediction of both random slopes and random intercepts (Diaz, 2017, 2016). In these applications, one is not always convinced that the random intercepts and slopes that model patients' heterogeneity follow a Gaussian distribution. Thus, robust methods to estimate (or "predict") random effects in linear mixed effects models may be useful in personalized medicine.

The most common approach to predicting random effects is the empirical Bayesian (EB) method (Carlin and Louis, 2009; Efron and Morris, 1972; Fitzmaurice et al., 2011; Frees, 2004; Liu et al., 2008; Martínez et al., 2012; Morris, 1983; Robinson, 1991). The EB approach is implemented in popular statistical packages, such as the MIXED procedure in SAS (SAS Institute Inc, Cary, NC) or the meglm command in STATA (StataCorp LLC2, College Station, TX). An EB predictor can be justified as an estimator of the mean of the conditional distribution of the random effects given the data, assuming that the random effects follow a Gaussian distribution. Under this point of view, the distribution of the random effects is interpreted as an empirical prior distribution in the sense that it is objectively estimated with data (Robinson, 1991).

In the context of linear mixed effects models, however, EB predictors are also estimators of the Best Linear Unbiased Predictors (BLUPs) of the random effects (Frees, 2004; Rabe-Hesketh and Skrondal, 2012). This suggests that EB predictors must enjoy some robustness to non-normality because the BLUP property does not require any distributional assumption for the random effects, except for the existence of second moments (Frees, 2004). The estimation of the BLUPs, however, requires using estimates of variance components that are usually obtained under normality assumptions (Harville and Jeske, 1992). Thus, further research is needed to examine the accuracy and robustness of EB predictors and to compare them with alternative approaches.

Shen and Louis (1999) proposed a recursive smoothing-by-roughening approach to estimate random effects. Verbeke and Lesaffre (1996) proposed a heterogeneity model that can identify subgroups of random effects that come from a mixture of distributions. Zhang and Davidian (2001) proposed a semi-nonparametric representation to approximate random effects densities, which requires a user-chosen tuning parameter. Ghidey et al. (2004) proposed a penalized estimation of the marginal likelihood to estimate the random effects. These methods have been compared in cases in which the basic assumptions of linear mixed effects models are violated (Ghidey et al., 2010).

Qu et al. (2000) proposed a robust estimation method for marginal models. The method is based on quadratic inference functions (QIFs) with extended scores, which do not require estimating correlation parameters using the repeated measurements. This method represents the inverse of the conditional correlation matrix of responses given the random effects, using a linear combination of basis matrices. The improvement in estimation efficiency and robustness becomes most apparent when the working correlation is miss-specified. This common issue is usually handled with generalized estimation equations, which are widely used in longitudinal data

analysis. Building on this approach, Wang et al. (2012) proposed an efficient and robust estimation method applicable to generalized linear mixed models. The method is based on conditional extended scores that require only first and second conditional moments. Therefore, the method does not require a closed form likelihood function. It is robust in that neither normality distribution assumptions on the random effects nor independence assumptions on measurement errors are required. This method is implemented by iteratively minimizing fixed effect and random effect quadratic inference functions, utilizing an optimal tuning parameter. The asymptotic properties of the obtained fixed effects estimates have been extensively studied (Cho et al., 2017; Wang et al., 2012). The QIF approach has been proven useful in the prediction of personalized treatments by applying random forest algorithms on the estimated subject-specific treatment random effects (Cho et al., 2017). It has also been proposed as a tool for individualized dosage computations via a procedure that accounts for the patient-specific random effects that determine patients' heterogeneity (Zhu and Qu, 2016).

Since the QIF approach to the prediction of random effects does not assume normality, we can hypothesize that this approach is a robust alternative to the EB approach. To our knowledge, however, there are no published studies comparing the two approaches. The main purpose of this study was to elucidate via simulations the conditions under which the QIF approach predicts random effects more accurately in comparison with the traditional and more widely used EB approach.

This paper focuses on linear mixed effects models. Thus, we applied the QIF and EB approaches to continuous responses. The computational formulas of the QIF approach are usually presented under a very general modelling setting that encompasses these and other types of responses and model structures (Cho et al., 2017; Wang et al., 2012; Zhu and Qu, 2016). So, a

secondary aim of this paper was to translate this published work to the context of the linear mixed effects model and clarify the equations for computational purposes.

The article is organized as follows. Section 2 reviews the notation and assumptions of linear mixed effects models. Section 3 reviews the QIF and EB approaches, placing emphasis on computational aspects. Sections 4 describes the simulation scenarios which include Gaussian and non-Gaussian distributions. Simulation results are in Section 5. In Section 6, we apply the two approaches to the prediction of random effects and individual treatment benefits using data from a clinical trial of patients with depression (Diaz, 2017; Reisby et al., 1977). Discussion and conclusions are in Section 7.

2. The linear mixed effects model

For a subject i, $i = 1, \dots, N$, we consider the linear mixed effects model, $\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{Y}_i = \left(y_{i1}, \dots, y_{i,n_i}\right)^T$ is the subject's response vector, $\boldsymbol{\beta} = \left(\beta_1, \dots, \beta_p\right)^T$ is a fixed effects vector, $\boldsymbol{b}_i = \left(b_{i1}, b_{i2}, \dots, b_{iq}\right)^T$ is a random effects vector, $\boldsymbol{\varepsilon}_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{i,n_i}\right)^T$ is a measurement error vector assumed to be normally distributed with mean $\mathbf{0}$, and X_i and Z_i are design matrices. N is the number of subjects. We assume that $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, H_i)$ and $\boldsymbol{b}_i$ is independent of $\boldsymbol{\varepsilon}_i$. To estimate parameters through maximum likelihood or restricted maximum likelihood, it is usually assumed that $\boldsymbol{b}_i \sim N(\mathbf{0}, D)$ (Frees, 2004). Thus, the marginal distribution of the response is $\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i^T + H_i)$. Here, we will consider two cases: independent errors ($H_i = \sigma^2 I_{n_i}$ with I_{n_i} the identity matrix) and first order autocorrelated errors [the (j,k)-th and (k,j)-th elements of matrix H_i is $\sigma^2\rho^{|j-k|}$, where ρ is the autocorrelation coefficient].

3. Prediction methods for random effects

3.1. Empirical Bayesian approach

Under the assumption that $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ are Gaussian, the conditional mean of $\boldsymbol{b}_i$ given $\boldsymbol{Y}_i$ is

$$E[\boldsymbol{b}_i|\boldsymbol{Y}_i] = DZ_i^T V_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}) \qquad (1)$$

where $V_i = Z_i D Z_i^T + H_i$. If the distribution of $\boldsymbol{b}_i$ is viewed as an empirical prior distribution, $E[\boldsymbol{b}_i|\boldsymbol{Y}_i]$ is the mean of the posterior distribution of $\boldsymbol{b}_i$ given the subject's responses. From this perspective, $E[\boldsymbol{b}_i|\boldsymbol{Y}_i]$ is a Bayesian estimator of $\boldsymbol{b}_i$ (Frees, 2004). When $\boldsymbol{\beta}$ is replaced with its generalized least squares estimator in Equation (1), and D and H_i are known, $E[\boldsymbol{b}_i|\boldsymbol{Y}_i]$ can be shown to be the BLUP of $\boldsymbol{b}_i$ (Frees, 2004). If, in addition to this replacement, D and H_i are replaced by their maximum or restricted maximum likelihood estimators, $E[\boldsymbol{b}_i|\boldsymbol{Y}_i]$ is called the empirical Bayesian (EB) predictor of $\boldsymbol{b}_i$ (Fitzmaurice et al., 2009; Frees, 2004; Laird and Ware, 1982).

For the purposes of this study, it is important to keep in mind that the assumption of normality is not needed to demonstrate that Equation (1) gives the BLUP of $\boldsymbol{b}_i$ after replacing $\boldsymbol{\beta}$ with its estimator (Frees, 2004). This suggests that the EB predictor can be robust to violations of the normality assumption (McCulloch and Neuhaus, 2011b).

3.2. Approach based on quadratic inference functions

The random effects $\boldsymbol{b}_i$ may not be normally distributed. The QIF approach has been proposed as an estimation method robust to violations of this assumption or to misspecifications

of the structure of the variance covariance matrix D (Wang et al., 2012). Here, we review the computational aspects of the method in the context of linear mixed effects models. Let K be the total number of unique time points recorded in the dataset. Every subject was observed on at least one of these points. Thus, $1 \leq n_i \leq K$. We assume first that $n_i = K$ for all i, that is, balanced data with respect to time. Unbalanced data handling is explained in Section 3.2.1.

The estimating equations for the fixed and random effects are respectively

$$\sum_{i=1}^{N} X_i^T W_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \quad \text{and}$$

$$Z_i^T W_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \ , \ \ i = 1, \ldots, N$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}|\boldsymbol{b}_i) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i$ is the mean response for subject i, and $W_i^{-1} = A_i^{-1/2}R^{-1}A_i^{-1/2}$, with $A_i = \text{diag}(H_i)$. Here, R is the conditional correlation matrix of the responses given the random effects which is assumed to be the same for all subjects.

The QIF approach utilizes the fact that R^{-1} can usually be expressed as a linear combination of known basis matrices, that is, $R^{-1} = \sum_{j=1}^{m} a_j M_j$. The M_j's depend on the assumed structure for R and the a_j's are unknown constants that do not enter in the estimation process (Qu et al., 2000). For instance, for independent errors, $m = 1$ and M_1 is the identity matrix. For AR(1) errors, $m = 2$, M_1 is the identity matrix, and M_2 is a matrix whose entries directly above and directly below the main diagonal are all 1's and all other entries including those on the main diagonal are 0's (Qu et al., 2000).

Following Wang et al. (2012), we denote

$$g_i^f(\boldsymbol{\beta}) = \begin{pmatrix} X_i^T A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ X_i^T A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \end{pmatrix}_{mp \times 1} \quad \text{and} \quad g_i^r = Z_i^T A_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$$

and simultaneously estimate the fixed and random effects by iteratively minimizing two objective functions, namely the fixed-effects QIF (Wang et al., 2012)

$$L^f(\boldsymbol{\beta}|\boldsymbol{b}) = N\left(G_N^f\right)^T \left(C_N^f\right)^{-1} G_N^f$$

and the random-effects QIF

$$L^r(\boldsymbol{b}|\boldsymbol{\beta}) = \|G^r(\boldsymbol{b})\|^2$$

where $\boldsymbol{b} = (\boldsymbol{b}_1^T, \dots, \boldsymbol{b}_N^T)^T$,

$$G_N^f = \frac{1}{N} \sum_{i=1}^N g_i^f(\boldsymbol{\beta})$$

$$C_N^f = \frac{1}{N} \sum_{i=1}^N g_i^f(\boldsymbol{\beta}) \left(g_i^f(\boldsymbol{\beta})\right)^T$$

and

$$G^r(\boldsymbol{b}) = \left\{(g_1^r)^T, \dots, (g_N^r)^T, \lambda_1 \boldsymbol{b}^T, \lambda_2 \left(P_J \boldsymbol{b}\right)^T\right\}^T$$

The matrix P_J is computed as follows. Denote the design matrix $X = (X_1^T, \dots, X_N^T)^T$ with $X_i = \left(\boldsymbol{x}_{i1}^T, \dots, \boldsymbol{x}_{i,n_i}^T\right)^T$, and the block diagonal matrix $Z = \text{diag}(Z_1, \dots, Z_N)$ with $Z_i = \left(\boldsymbol{z}_{i1}^T, \dots, \boldsymbol{z}_{i,n_i}^T\right)^T$. Compute the Q-R decomposition of $A = \left((I - P_X)Z\right)^T$ such that $A = Q \times R$, where $P_X = X(X^T X)^{-1} X^T$, and obtain $r = \text{rank}(A)$. Let J be the matrix whose columns are the

columns of Q beyond the rth column (Boyd and Vandenberghe, 2004). Thus, J has $Nq - r$ columns. Then, $P_J = J(J^T J)^- J^T$.

The iterative process starts with an initial value for $\boldsymbol{\beta}$, denoted $\widehat{\boldsymbol{\beta}}^{(0)}$, obtained through linear regression. Then $L^r\left(\boldsymbol{b}|\widehat{\boldsymbol{\beta}}^{(0)}\right)$ is minimized with respect to $\boldsymbol{b}$ to obtain $\widehat{\boldsymbol{b}}^{(1)}$. Next $L^f\left(\boldsymbol{\beta}|\widehat{\boldsymbol{b}}^{(1)}\right)$ is minimized with respect to $\boldsymbol{\beta}$ to obtain $\widehat{\boldsymbol{\beta}}^{(1)}$, and so on. The process stops when $\left|\widehat{\boldsymbol{\beta}}^{(s+1)} - \widehat{\boldsymbol{\beta}}^{s}\right| + \left|\widehat{\boldsymbol{b}}^{(s+1)} - \widehat{\boldsymbol{b}}^{s}\right| < 10^{-5}$ with $\widehat{\boldsymbol{b}}^{(0)} = \mathbf{0}$.

For homoscedastic errors, $A_i = \sigma^2 I_{n_i}$ where I_{n_i} is an identity matrix. Therefore, $L^f(\boldsymbol{\beta}|\boldsymbol{b})$ does not depend on σ^2 because it cancels out. Also, since λ_1 and λ_2 can be written as proportional to σ^2, $L^r(\boldsymbol{b}|\boldsymbol{\beta})$ is proportional to σ^2. Thus, the value of σ^2 does not need to be known for the minimization of $L^r(\boldsymbol{b}|\boldsymbol{\beta})$. Therefore, in practice, we use $A_i = I_{n_i}$ for the minimization of these objective functions.

Finally, note that the QIF approach does not make distributional assumptions about the random effects and, therefore, about D. Instead, the approach works directly with the conditional variance-covariance matrix W_i. Also note that, in contrast to the default estimation methods implemented in the most common statistical packages, H_i may differ among subjects. This provides additional generality and gives flexibility for handling unbalanced data.

3.2.1. Handling of unbalanced data

The assumption that the subjects have the same number of responses is needed to ensure that the dimension of $g_i^f(\boldsymbol{\beta})$ does not change with i (the M_js are the same for all subjects). Thus, a transformation of the response vectors is necessary for unbalanced data. For subject i with

unbalanced responses such that $n_i < K$, let Λ_i be a transformation matrix of dimension $K \times n_i$ obtained by removing the columns of the $K \times K$ identity matrix corresponding to the time points with missing observations. In the iterative algorithm, we use $\boldsymbol{Y}_i^* = \Lambda_i \boldsymbol{Y}_i$, $X_i^* = \Lambda_i X_i$, $Z_i^* = \Lambda_i Z_i$, $A_i^* = \Lambda_i A_i \Lambda_i^T$, and $(A_i^*)^{-1} = \Lambda_i A_i^{-1} \Lambda_i^T$, respectively in place of $\boldsymbol{Y}_i$, X_i, Z_i, A_i and A_i^{-1} (Wang et al., 2012). This approach provides matrices of equal sizes by assigning zeros to the missing responses, inserting rows with zeros to the corresponding rows of X_i and Z_i, and inserting zeros to the corresponding elements of the diagonals of A_i and A_i^{-1}. In this way, missing responses are both replaced with zeros and predicted as zeros without affecting parameter estimation and random effects prediction.

3.2.2. Penalization parameters

The penalization parameter λ_2 is usually fixed to $\log(N)$ (Cho et al., 2017; Wang et al., 2012; Zhu and Qu, 2016). Since λ_1 controls the variability of random-effect predictors, we chose λ_1 using cross-validation as in Cho et al. (2017). Given a value of λ_1, let $\widehat{\boldsymbol{\beta}}_{\lambda_1}^{-k}$ and $\widehat{\boldsymbol{b}}_{\lambda_1}^{-k,i}$ be respectively the fixed-effects estimate and the random effects predictor for subject i, computed after excluding all observations at the k^{th} time point. The cross-validation error is

$$CV_{\lambda_1} = K^{-1} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(y_{ik} - \left(\boldsymbol{x}_{ik}^T \widehat{\boldsymbol{\beta}}_{\lambda_1}^{-k} + \boldsymbol{z}_{ik}^T \widehat{\boldsymbol{b}}_{\lambda_1}^{-k,i} \right) \right)^2$$

We chose the value of λ_1 that minimized CV_{λ_1}.

4. Simulation scenarios for comparing the QIF and EB approaches

The simulation study is based on the model:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{i1} + b_{i2} t_{ij} + \varepsilon_{ij}, \qquad i = 1, \ldots, N;\ j = 1, \ldots, n$$

where Y_{ij} is the continuous response of the ith subject at time t_{ij}. This model and parameter values were motivated by an analysis of the Framingham Study data (Zhang and Davidian, 2001).

For a given number of subjects N, we examined 3 different values of n, making sure the range of values for the time variable is the same. Thus, for $n = 3$, $t_{ij} = 0$, 5 or 10. For $n = 6$, $t_{ij} = 0$, 2, 4, 6, 8 or 10. And for $n = 11$, $t_{ij} = 0$, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. Thus, for instance, for $n = 3$, $X_i = Z_i = \begin{bmatrix} 1 & 0 \\ 1 & 5 \\ 1 & 10 \end{bmatrix}$.

Table 1 describes the simulation scenarios. For non-Gaussian random effects, we used $\beta_1 = 2.35$ and $\beta_2 = 0.28$ and, as a variance-covariance matrix for $\boldsymbol{b}_i = (b_{i1}, b_{i2})$, we used $D^* = \begin{bmatrix} 0.15 & 0.02 \\ 0.02 & 0.04 \end{bmatrix}$. This gives a correlation coefficient between the random intercept and random slope of 0.258. Variations about these values were used for Gaussian random effects. All random errors were simulated as Gaussian either independently or correlated with AR(1) structure, with $\sigma^2 = 1$, 10, or 30.

Three general cases for the random effects $\{\boldsymbol{b}_i, i = 1, \ldots, N\}$ were used:

1. Bivariate Gaussian with correlated or uncorrelated random effects.

2. A mixture of two bivariate Gaussian distributions:

$$\boldsymbol{b_i} \sim \frac{1}{2} N\left\{\begin{bmatrix}-0.35\\-0.1\end{bmatrix}, \begin{bmatrix}0.028 & -0.015\\-0.015 & 0.03\end{bmatrix}\right\} + \frac{1}{2} N\left\{\begin{bmatrix}0.35\\0.1\end{bmatrix}, \begin{bmatrix}0.028 & -0.015\\-0.015 & 0.03\end{bmatrix}\right\}, i = 1, \ldots, N,$$

so that the mean is **0** and the variance-covariance matrix is D^*

3. A bivariate t-distribution (d.f.=3) with mean **0** and variance-covariance matrix D^*. This distribution allows examining prediction performance in the presence of heavy tails.

Since QIFs computations are time consuming, only thirty datasets were simulated for each scenario. Mean square prediction errors (MSPEs) were used to compare the prediction performances of the two approaches. The ratio of the average MSPE for the QIF approach to the average MSPE for the EB approach was calculated for each simulation scenario. For the l-th dataset, the MSPE for the random intercept was

$$MSPE_1^{(l)} = \frac{\sum_{i=1}^{N}\left(\hat{b}_{i1,l} - b_{i1,l}\right)^2}{N}$$

where $\hat{b}_{i1,l}$ is a predictor of the random intercept and $b_{i1,l}$ is the true (simulated) value. The MSPE for the random slope is defined analogously, replacing subscript 1 with 2. Ratios comparing QIF with EB were calculated as $MSPE_{QIF,j}/MSPE_{EB,j}$, where $MSPE_{QIF,j} = \frac{\sum_{l=1}^{30} MSPE_j^{(l)}}{30}$ when the random effects were predicted using the QIF approach, and analogously for $MSPE_{EB,j}, j = 1, 2$. Thus, a ratio smaller than 1 indicates that the QIF approach is more accurate in predicting the random effect. Boxplots of ratios were also used to compare the two approaches (Figure 1). Simulations were programmed in SAS IML and EB predictors were computed with SAS PROC MIXED (SAS Institute Inc, Cary, NC). The computer code is available as Supplementary Material.

5. Simulation results

5.1. Random effects with a bivariate Gaussian distribution and Gaussian independent errors

Results for true random effects simulated from a bivariate Gaussian distribution, independent errors, large error variance, and independent random intercept and slope are shown in Tables 2A (random intercept) and 2B (random slope). QIF predictions on the random intercept were substantially more accurate than EB predictions (Table 2A). In contrast, EB predictions on the random slope tended to be more accurate than QIF predictions, although the two approaches performed similarly for large N and n (Table 2B).

When the error variance was small, the average MSPEs obtained with the QIF approach were always larger than those for the EB approach for both the random intercept (Table 3A) and slope (Table 3B), suggesting a better accuracy for the EB approach. When predicting intercepts with relatively large numbers of subjects and repeated measures, MSPE ratios were close to 1 (Table 3A); thus, the accuracy of the QIF tended to be similar to that of the EB approach. In contrast, when predicting slopes, larger values of N and n improved the accuracy of the EB approach faster, which is suggested by the considerably larger ratios for $n = 11$ in Table 3B. The EB approach thus seems to be more responsive to increases in the number of observations than the QIF approach. Interestingly, the accuracies of the QIF and EB approaches did not depend substantially on the coefficient of variation C_v of the last measure.

5.2. Bivariate Gaussian distributions for the random effects and AR (1) Gaussian errors

Tables 4-5 show simulation results when the random effects followed a bivariate Gaussian distribution and the Gaussian measurement errors were correlated with an AR(1) structure. The random intercept and slope were correlated in Table 4 and uncorrelated in Table 5. For the random intercept (Tables 4A and 5A), MSPE ratios were smaller than 1 only for $n = 3$. This suggests that the QIF approach is more accurate than the EB approach only when there is a relatively small number of repeated measures. For larger numbers of repeated measures, the EB approach performed substantially better in intercept predictions.

Tables 4B and 5B, which compare slope predictors, reveal that, in the presence of autocorrelated errors, the EB approach produces more accurate predictions on the random slope, except perhaps under small sample sizes and small numbers of repeated measures.

5.3. Mixtures of bivariate Gaussian distributions for the random effects

Table 6A shows ratios of average MSPEs for the random intercepts comparing the QIF to the EB approach, when the random intercept and slope were simulated from a mixture of bivariate Gaussian distributions. Except for relatively small error variances σ^2, the ratios were < 1. Thus, when predicting intercepts under relatively large error variances, the prediction accuracy of the QIF approach was less deteriorated than that of the EB approach by the violation of the normality assumption, regardless of sample size and number of repeated measures.

In contrast, Table 6B shows that ratios for the random slopes tended to be close or higher than 1 regardless of error variance, sample size and number of repeated measures. This suggests

that the accuracy of the EB approach was better than that of the QIF approach when predicting slopes under the violation of the normality assumption for the random effects. However, for large numbers of subjects and repeated measures, the QIF and EB approaches had similar accuracies, except for small error variances in which case the EB approach was superior. For fixed N and n, an increased error variance σ^2 tended to make the accuracy of the two approaches more similar.

5.4. Bivariate $t(3)$ distribution for the random effects

When the random effects were simulated from a bivariate t distribution, the results (Table 7) were very similar to those corresponding to bivariate Gaussian distributions in Section 5.1. If error variances are relatively large, the QIF approach tended to be more accurate than the EB approach to predict intercepts, but less accurate to predict slopes. For small error variances, the EB approach was always better. The important point that the EB approach tended to be more accurate and robust in the presence of small error variances emerges from these simulations.

6. Application: prediction of random effects in patients with depression.

As an illustration, we use clinical trial data consisting of a sample of 66 patients with two types of depression diagnosis: endogenous (N=37) and nonendogenous (N=29) (Reisby et al., 1977). The data is available in Hedeker and Gibbons (2006). Here, the response variable is the Hamilton Rating Scale (HRS) for depression. Data collection started two weeks before initiation of imipramine treatment and continued for four weeks during treatment. We fitted the same polynomial model as in Diaz (2017), which, as covariates, included diagnosis, time t (weeks on

treatment) and t^2. The model had random effects for the intercept and time covariates. The SAS procedure MIXED, which assumes Gaussian random effects, was used to obtain maximum likelihood estimates (MLEs) of the fixed effects and EB predictors of the random effects (SAS Institute Inc. Cary, NC). An unstructured covariance matrix for the random effects was assumed in this case.

The predictions on the intercept by the EB and QIF approaches correlated very strongly (Table 8), and similarly for the predictions on the slope of time. Predictions for the slope of t^2 were only moderately correlated. Figure 2 shows scatterplots comparing the predictions of the random effects by the QIF and EB approaches under homoscedastic independent errors. Scatterplots for exchangeable and AR(1) errors are in the Supplementary Material (Figures S1, S2). Overall, the two prediction methods agreed strongly, except perhaps for the predictions on the random intercept. Specifically, scatterplots for the intercept were slightly tilted with respect to the y=x line, suggesting that QIF intercept predictions tended to be larger in absolute value than EB predictions (Figure 2A). This deviation was more pronounced in the AR(1) structure (Figure S2C in the Supplementary Material). The histograms for the random intercepts from the two approaches exhibited similar shapes, and similarly for the random slopes (Figure S3 in Supplementary Material).

As additional illustration, we use the approach in Diaz (2017) to explore the individual benefits of imipramine treatment for the 66 subjects after 4 weeks of treatment, using an HRS score ≤ 7 as the therapeutic target. Since the model is a 2-dimensional personalized medicine model, predictions of both random intercepts and slopes are needed for these computations (Diaz, 2017, 2016). Diaz (2017) calculated individual benefits for these subjects by predicting the random effects using the EB approach. Here, we additionally use the QIF approach. Percentiles of

individual benefits by diagnosis are shown in Table 9. Both approaches suggest that, after 4 weeks of treatment, nonendogenous patients tended to achieve greater imipramine benefits than endogenous patients. However, on average, the EB approach predicted greater benefits than the QIF approach for both endogenous and nonendogenous patients.

7. Discussion and conclusions

Two methods of predicting random effects in the linear mixed effects model were compared under various assumptions for the random effects and measurement errors: the classical empirical Bayes approach and an alternative approach based on QIFs (Wang et al., 2012). An important conclusion of this study is that, under the investigated non-Gaussian random effects and small to moderate error variances, the EB approach was more accurate than the QIF approach for predicting both slopes and intercepts. For larger error variances, QIF was a better predictor for the random intercept but not for the random slope. This latter observation is consistent with the findings presented in Verbeke and Lesaffre (1996) who investigated mixtures of bivariate Gaussian distributions and found that the EB approach does not capture accurately the shape of the random effects distribution when the error variance is large. Our results are also consistent with the findings of other studies suggesting that prediction accuracy of EB predictors is not significantly affected by mild to moderate violations of the normality assumption (Mcculloch and Neuhaus, 2011a, 2011b). This robustness is probably due to the fact that EB predictors inherit the optimality properties of the BLUPs, which do not depend on the normality assumption (see Frees, 2004).

In general, in the presence of Gaussian random effects, the EB approach outperformed the QIF approach for predicting random slopes; and the QIF approach was superior to the EB approach only under relatively large error variances and, in some scenarios, under a relatively small number of repeated measures. The box-plots in Figures 1A and 1B illustrate some of the situations in which very large error variances were associated with a higher prediction accuracy for the QIF approach when the normality assumption was violated. In contrast, Figures 1C and 1D illustrate situations in which the EB approach performed relatively better than the QIF approach under the Gaussian assumption for the random effects.

For relatively large sample sizes, QIF computations are time consuming. For instance, using a high-performance computation cluster, for $N = 150$, $n = 11$ and Gaussian random effects, the average computation time was about 240 hours, regardless of the magnitude of the error variance. In light of this limitation and considering the fact that the EB approach is mathematically and computationally less complex, our simulations suggest the EB approach is more recommendable in most practical situations, even if non-Gaussian random effects are suspected.

The QIF approach ignores the serial correlation of repeated measures in the prediction of random effects, although it takes it into account in the estimation of the fixed effects (Wang et al., 2012). It is possible to circumvent this limitation by incorporating an empirical correlation matrix into the random-effects QIF. This approach was followed in applications to personalized medicine (Cho et al., 2017; Zhu and Qu, 2016). Further research is needed to compare these modifications of the QIF with the EB approach.

References

Agresti, A., Caffo, B., Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. Comput. Stat. Data Anal. 47, 639–653.

Alonso, A., Litière, S., Molenberghs, G., 2008. A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. Comput. Stat. Data Anal. 52, 4474–4486.

Andrews, N., Cho, H., 2018. Validating effectiveness of subgroup identification for longitudinal data. Stat. Med. 37, 98–106.

Boyd, S.P., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press, Cambridge, UK ; New York.

Carlin, B.P., Louis, T.A., 2009. Bayesian methods for data analysis, 3rd ed, Chapman & Hall/CRC texts in statistical science series. CRC Press, Boca Raton.

Cho, H., Wang, P., Qu, A., 2017. Personalize Treatment for Longitudinal Data Using Unspecified Random-Effects Model. Stat. Sin. 27, 187–205.

Diaz, F.J., 2018. Construction of the Design Matrix for Generalized Linear Mixed-Effects Models in the Context of Clinical Trials of Treatment Sequences. Rev. Colombiana Estadíst. 41, 191–233.

Diaz, F.J., 2017. Estimating individual benefits of medical or behavioral treatments in severely ill patients. Stat Methods Med Res 962280217739033.

Diaz, F.J., 2016. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat. Med. 35, 4077–4092.

Diaz, F.J., Cogollo, M.R., Spina, E., Santoro, V., Rendon, D.M., de Leon, J., 2012a. Drug dosage individualization based on a random-effects linear model. J Biopharm Stat 22, 463–484.

Diaz, F.J., de Leon, J., 2013. The Mathematics of Drug Dose Individualization Should Be Built With Random-Effects Linear Models. Ther. Drug Monit. 35, 276–277.

Diaz, F.J., Rivera, T.E., Josiassen, R.C., de Leon, J., 2007. Individualizing drug dosage by using a random intercept linear model. Stat. Med. 26, 2052–2073.

Diaz, F.J., Yeh, H.W., de Leon, J., 2012b. Role of Statistical Random-Effects Linear Models in Personalized Medicine. Curr Pharmacogenomics Pers. Med 10, 22–32.

Dimova, R.B., Markatou, M., Talal, A.H., 2011. Information methods for model selection in linear mixed effects models with application to HCV data. Comput. Stat. Data Anal. 55, 2677–2697.

Efron, B., Morris, C., 1972. Limiting the Risk of Bayes and Empirical Bayes Estimators--Part II: The Empirical Bayes Case. J. Am. Stat. Assoc. 67, 130–139.

Fitzmaurice, G.M., 2009. Longitudinal data analysis, Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Boca Raton.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2011. Applied longitudinal analysis, 2nd ed, Wiley series in probability and statistics. Wiley, Hoboken, N.J.

Frees, E.W., 2004. Longitudinal and panel data : analysis and applications in the social sciences.

Cambridge University Press, Cambridge, UK ; New York.

Frees, E.W., 2001. Omitted variables in longitudinal data models. Can. J. Stat. Can. Stat. 29, 573–595.

Ghidey, W., Lesaffre, E., Eilers, P., 2004. Smooth random effects distribution in a linear mixed model. Biometrics 60, 945–953.

Ghidey, W., Lesaffre, E., Verbeke, G., 2010. A comparison of methods for estimating the random effects distribution of a linear mixed model A comparison of methods for estimating the random effects distribution of a linear mixed model. Stat. Methods Med. Res. 19, 575–600.

Harville, D.A., Jeske, D.R., 1992. Mean Squared Error of Estimation or Prediction under a General Linear-Model. J. Am. Stat. Assoc. 87, 724–731.

Heagerty, P.J., Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88, 973–985.

Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. Wiley-Interscience, Hoboken, N.J.

Hooks, T., Marx, D., Kachman, S., Pedersen, J., 2009. Optimality Criteria for Models with Random Effects TT - Criterios de optimalidad para los modelos con efectos aleatorios. Rev. Colombiana Estadíst. 32, 17–31.

Laird, N.M., Ware, J.H., 1982. Random-Effects Models for Longitudinal Data. Biometrics 38, 963–974.

Liu, X.Q., Rong, J.Y., Liu, X.Y., 2008. Best linear unbiased prediction for linear combinations in general mixed linear models. J. Multivar. Anal. 99, 1503–1517.

López, L.A., Franco, D.C., Barreto, S.P., 2007. Sobre la construcción del mejor predictor lineal insesgado (BLUP) y restricciones asociadas TT - About the Best Linear Unbiased Predictor (BLUP) and Associated Restrictions. Rev. Colombiana Estadíst. 30, 13–36.

Martínez, C.A., Elzo, M., Manrique, C., Grajales, L.F., Jiménez, A., 2012. Random Regression Models for Estimation of Covariance Functions, Genetic Parameters and Prediction of Breeding Values for Rib Eye Area in a Colombian Bos indicus-Bos taurus Multibreed Cattle Population, Revista Colombiana Estadíst. 35, 309-330.

Mcculloch, C.E., Neuhaus, J.M., 2011a. Prediction of Random Effects in Linear and Generalized Linear Models under Model Misspecification. Biometrics 67, 270–279.

McCulloch, C.E., Neuhaus, J.M., 2011b. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Stat. Sci. 26, 388–402.

Morris, C.N., 1983. Parametric Empirical Bayes Inference - Theory and Applications. J. Am. Stat. Assoc. 78, 47–55.

Qu, A., Lindsay, B.G., Li, B., 2000. Improving generalised estimating equations using quadratic inference functions. Biometrika 87, 823–836.

Rabe-Hesketh, S., Skrondal, A., 2012. Multilevel and longitudinal modeling using Stata, 3rd ed. Stata Press Publication, College Station, Tex.

Reisby, N., Gram, L.F., Bech, P., Nagy, A., Petersen, G.O., Ortmann, J., Ibsen, I., Dencker, S.J., Jacobsen, O., Krautwald, O., Søndergaard, I., Christiansen, J., 1977. Imipramine: Clinical

effects and pharmacokinetic variability. Psychopharmacology (Berl). 54, 263–272.

Robinson, G.K., 1991. That BLUP is a Good Thing: The Estimation of Random Effects. Stat. Sci. 6, 15–32.

Shen, W., Louis, T.A., 1999. Empirical Bayes estimation via the smoothing by roughening approach. J. Comput. Graph. Stat. 8, 800–823.

Skrondal, A., Rabe-Hesketh, S., 2004. Generalized latent variable modeling : multilevel, longitudinal, and structural equation models, Chapman & Hall/CRC interdisciplinary statistics series. Chapman & Hall/CRC, Boca Raton.

Verbeke, G., Lesaffre, E., 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Comput. Stat. Data Anal. 23, 541–556.

Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. J. Am. Stat. Assoc. 91, 217–221.

Verbeke, G., Molenberghs, G., 2000. Linear mixed models for longitudinal data, Springer series in statistics. Springer, New York.

Wang, P., Tsai, G.F., Qu, A., 2012. Conditional Inference Functions for Mixed-Effects Models With Unspecified Random-Effects Distribution. J. Am. Stat. Assoc. 107, 725–736.

Yau, K.K.W., Lee, A.H., Ng, A.S.K., 2003. Finite mixture regression model with random effects: Application to neonatal hospital length of stay. Comput. Stat. Data Anal. 41, 359–366.

Zhang, D., Davidian, M., 2001. Linear mixed models with flexible distributions of random

effects for longitudinal data. Biometrics 57, 795–802.

Zhu, X.L., Qu, A., 2016. Individualizing drug dosage with longitudinal data. Stat. Med. 35, 4474–4488.

Table 1. Summary of simulation scenarios for comparing QIF with EB approaches

Case #	Tables #	Fixed effects		Random effects $\boldsymbol{b}_i \sim \varphi\left(\begin{bmatrix}0\\0\end{bmatrix}, D\right)$	Correlation structure of errors	Error variance σ^2	Coefficient of variation of last response[a] C_v	Correlation coefficient of measurement errors ρ
		β_1	β_2					
1	2	2	1.83	$\boldsymbol{b}_i \sim$ Bivariate Gaussian $\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0\\0 & 0.04\end{bmatrix}\right)$	Independent	30	10%	0
		2.35	0.274				40%	
		1.5	0.105				80%	
	3	2	1.83	$\boldsymbol{b}_i \sim$ Bivariate Gaussian $\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0\\0 & 0.04\end{bmatrix}\right)$	Independent	1	10%	0
		2.35	0.274				40%	
		1.5	0.105				80%	
	4	2.35	0.28	$\boldsymbol{b}_i \sim$ Bivariate Gaussian $\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0.039\\0.039 & 0.04\end{bmatrix}\right)$	$AR(1)$	1	43%	0.1, 0.4, 0.7
	5	2.35	0.28	$\boldsymbol{b}_i \sim$ Bivariate Gaussian $\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0\\0 & 0.04\end{bmatrix}\right)$	$AR(1)$	1	40%	0.1, 0.4, 0.7
2	6	2.35	0.28	$\boldsymbol{b}_i \sim$ Mixture of bivariate Gaussians $\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0.02\\0.02 & 0.04\end{bmatrix}\right)$	Independent	1,10,30	41%	0
3	7	2.35	0.28	$\boldsymbol{b}_i \sim$ Bivariate $t_3\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.15 & 0.02\\0.02 & 0.04\end{bmatrix}\right)$	Independent	1,10,30	41%	0

[a]The coefficient of variation C_v at time point $t_{i,10}$ was defined by $C_v = \frac{(\mathrm{Var}(\beta_{i1}^* + \beta_{i2}^* t_{i,10}))^{\frac{1}{2}}}{E[\beta_{i1}^* + \beta_{i2}^* t_{i,10}]}$, with $\beta_{i1}^* = \beta_1 + b_{i1}$, $\beta_{i2}^* = \beta_2 + b_{i2} t_{i,10}$ where b_{i1} and b_{i2} are simulated random effects.

Table 2. Ratios of the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *bivariate Gaussian distribution*, and the independent errors had a large error variance. Averages summarized 30 simulated datasets.

(A) Random Intercept

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 30$		Sample Size (N)			
n	C_v	60	90	120	150
3	10%	0.4164	0.4588	0.1674	0.2800
	40%	0.3343	0.3708	0.2049	0.2539
	80%	0.3190	0.5253	0.3091	0.2118
6	10%	0.7229	0.8787	0.6589	0.3998
	40%	0.5716	0.7112	0.7512	0.4471
	80%	0.4854	0.8231	0.9512	0.4758
11	10%	0.4143	0.7864	0.7556	0.7267
	40%	0.3622	0.6298	0.8955	0.7650
	80%	0.4773	0.8345	0.8135	0.8474

(B) Random Slope

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 30$		Sample Size N			
n	C_v	60	90	120	150
3	10%	1.403	1.373	1.118	1.097
	40%	1.342	1.980	0.9664	1.074
	80%	1.778	1.237	1.177	1.095
6	10%	1.049	1.094	1.031	0.9508
	40%	1.085	1.059	0.9989	0.9648
	80%	1.069	1.080	1.057	0.9518
11	10%	0.8021	0.9818	1.024	0.9849
	40%	0.8790	0.9732	1.014	0.9979
	80%	0.9247	1.000	0.9746	0.9990

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.
C_v: Coefficient of variation at the last measure.

Table 3. Ratios of the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *bivariate Gaussian distribution*, small error variance and independent random effects and errors. Averages summarized 30 simulated datasets.

(A) Random Intercept

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 1$		Sample Size N			
n	C_v	60	90	120	150
3	10%	1.035	1.082	1.103	1.087
	40%	1.005	1.042	1.118	1.069
	80%	1.032	1.078	1.059	1.119
6	10%	1.263	1.270	1.228	1.270
	40%	1.168	1.180	1.211	1.219
	80%	1.178	1.164	1.213	1.233
11	10%	1.517	1.494	1.090	1.067
	40%	1.496	1.305	1.058	1.066
	80%	1.290	1.489	1.048	1.059

(B) Random Slope

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 1$		Sample Size N			
n	C_v	60	90	120	150
3	10%	1.171	1.166	1.159	1.139
	40%	1.162	1.118	1.157	1.141
	80%	1.197	1.157	1.175	1.164
6	10%	1.169	1.170	1.133	1.156
	40%	1.134	1.098	1.106	1.113
	80%	1.118	1.085	1.109	1.141
11	10%	441.4	404.3	1.138	12.73
	40%	381.1	255.1	1.140	14.12
	80%	282.0	384.2	1.128	13.01

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.
C_v: Coefficient of variation at the last measure.

Table 4. Ratios from the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *bivariate Gaussian distribution* and the error terms had an AR(1) correlation matrix. The random intercept and slope were correlated. Averages summarized 30 simulated datasets.

(A) Random Intercept

$\text{corr}(b_{i1}, b_{i2}) = 0.5$ $\sigma^2 = 1$		Sample Size N			
n	ρ	60	90	120	150
3	0.1	0.6523	0.6933	0.4813	0.5760
	0.4	0.4766	0.5679	0.5387	0.7356
	0.7	0.3647	0.3604	0.3225	0.5666
6	0.1	1.193	1.441	1.335	1.472
	0.4	3.007	4.021	4.226	4.290
	0.7	3.495	3.960	4.072	3.661
11	0.1	1.508	1.316	1.386	1.612
	0.4	2.923	2.861	3.084	3.167
	0.7	4.349	1.035	1.023	1.033

(B) Random Slope

$\text{corr}(b_{i1}, b_{i2}) = 0.5$ $\sigma^2 = 1$		Sample Size N			
n	ρ	60	90	120	150
3	0.1	1.084	1.130	0.1883	1.108
	0.4	1.267	1.332	1.388	1.505
	0.7	0.2766	1.951	0.2798	1.967
6	0.1	1.129	1.299	1.235	1.274
	0.4	1.750	2.060	2.094	2.030
	0.7	1.724	1.713	1.689	1.672
11	0.1	1.413	1.305	1.350	1.461
	0.4	1.941	1.783	1.900	1.960
	0.7	2.266	1.085	1.074	1.089

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.

Table 5. Ratios from the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *bivariate Gaussian distribution* and the error terms had an AR(1) correlation matrix. The random intercept and slope were uncorrelated. Averages summarized 30 simulated datasets.

(A) Random Intercept

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 1$		Sample Size N			
n	ρ	60	90	120	150
3	0.1	0.6189	0.5694	0.5820	0.8422
	0.4	0.4796	0.5274	0.3958	0.6191
	0.7	0.4075	0.6029	0.3991	0.7015
6	0.1	1.189	1.238	1.124	1.194
	0.4	2.477	2.843	2.969	3.516
	0.7	2.726	3.215	3.326	3.371
11	0.1	1.049	1.123	1.052	1.107
	0.4	1.903	2.325	2.236	2.309
	0.7	3.774	1.003	1.003	1.000

(B) Random Slope

$\text{corr}(b_{i1}, b_{i2}) = 0$ $\sigma^2 = 1$		Sample Size N			
n	ρ	60	90	120	150
3	0.1	0.3935	0.5339	1.001	1.129
	0.4	0.9899	1.275	0.2371	1.290
	0.7	0.2181	1.787	1.931	1.954
6	0.1	1.101	1.163	1.130	1.141
	0.4	1.547	1.656	1.726	1.745
	0.7	1.441	1.555	1.554	1.496
11	0.1	1.106	1.112	1.091	1.109
	0.4	1.408	1.573	1.569	1.612
	0.7	1.769	1.045	1.043	1.000

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.
ρ: Correlation coefficient of errors within a subject.

Table 6. Ratios of the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *mixture of bivariate Gaussian distributions*. Averages summarized 30 simulated datasets.

(A) Random Intercept

		Sample Size N			
n	σ^2	60	90	120	150
3	30	0.1174	0.2715	0.2732	0.3989
	10	0.4624	0.4042	0.4235	0.5421
	1	0.9435	1.037	1.018	1.138
6	30	0.2969	0.2724	0.3962	0.4800
	10	0.5453	0.6740	0.7236	0.6142
	1	1.252	1.218	1.312	1.291
11	30	0.3734	0.5247	0.4441	0.6250
	10	0.6729	0.8064	0.8960	0.9452
	1	1.327	1.339	1.131	1.151

(B) Random Slope

		Sample Size N			
n	σ^2	60	90	120	150
3	30	1.090	1.558	1.057	1.044
	10	1.414	1.342	1.348	1.222
	1	1.584	1.682	1.664	1.712
6	30	0.9573	0.8363	0.9980	0.9323
	10	0.9444	1.028	1.058	0.9905
	1	1.223	1.185	1.278	1.292
11	30	0.9637	0.9720	0.9128	0.9511
	10	0.8943	1.119	0.9996	1.077
	1	199.9	274.5	1.194	11.56

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.

Table 7. Ratios of the average MSPE from the QIF approach to the average MSPE from the EB approach, when the true random effects were simulated from a *bivariate t(3) distribution*. Averages summarized 30 simulated datasets.

(A) Random Intercept

		Sample Size N			
n	σ^2	60	90	120	150
3	30	0.4433	0.5014	0.5791	0.4581
	10	0.4631	0.6332	0.7834	0.9097
	1	1.484	1.424	1.819	1.525
6	30	0.5919	0.6715	0.5876	0.6047
	10	0.8593	0.9498	0.9586	1.048
	1	1.242	1.385	1.286	1.270
11	30	0.6210	0.8303	0.8661	0.9197
	10	1.040	1.052	1.075	1.172
	1	1.483	1.230	1.148	1.151

(B) Random Slope

		Sample Size N			
n	σ^2	60	90	120	150
3	30	1.207	1.314	1.242	1.187
	10	1.269	1.260	1.288	1.358
	1	1.325	1.296	1.398	1.319
6	30	1.109	1.034	1.007	0.9689
	10	0.9942	1.015	1.057	1.120
	1	1.164	1.215	1.203	1.181
11	30	0.8940	1.321	1.020	1.005
	10	4.624	2.682	1.036	1.191
	1	302.7	169.7	1.166	12.60

σ^2: Variance of the measurement error.
N: Number of simulated subjects.
n: Number of repeated measures for each subject.

Table 8. Correlations between QIF and EB predictors by error covariance structure for the depression data.

Predicted Random Effect	Error covariance structure		
	Independent	**Exchangeable**	**AR(1)**
Intercept	0.9900	0.9902	0.9969
Time	0.8951	0.8951	0.8914
Time square	0.6258	0.6259	0.6587

EB: Empirical Bayes approach.
QIF: Quadratic inference function approach.

Table 9. Percentiles of individual imipramine benefits after 4 weeks of treatment. The therapeutic target was an HRS score ≤ 7. Individual benefits were multiplied by 100. An independent error structure was used.

	QIF					**EB**				
	Percentiles					**Percentiles**				
	10%	25%	50%	75%	90%	10%	25%	50%	75%	90%
Endogenous	0.00	0.00	3.67	52.31	98.29	0.00	0.05	5.77	38.59	63.49
Nonendogenous	0.00	0.09	6.32	58.89	98.80	0.00	2.03	10.15	47.70	88.19

EB: Empirical Bayes approach.
QIF: Quadratic inference function approach.

Figure 1. Boxplots of ratios comparing MSPEs of QIF versus EB. Each boxplot corresponds to 30 simulated datasets, $N = 120, n = 11$. Red boxplots are for random intercepts, and blue ones are for random slopes. (A) The true random effects were simulated from mixtures of bivariate Gaussian distributions (Table 6). (B) The true random effects were simulated from a bivariate $t(3)$ distribution (Table 7). (C) The independent true random effects were simulated from bivariate Gaussian distributions with independent errors (Table 3). (D) The independent true random effects were simulated from bivariate Gaussian distributions with AR(1) error structure (Table 5).

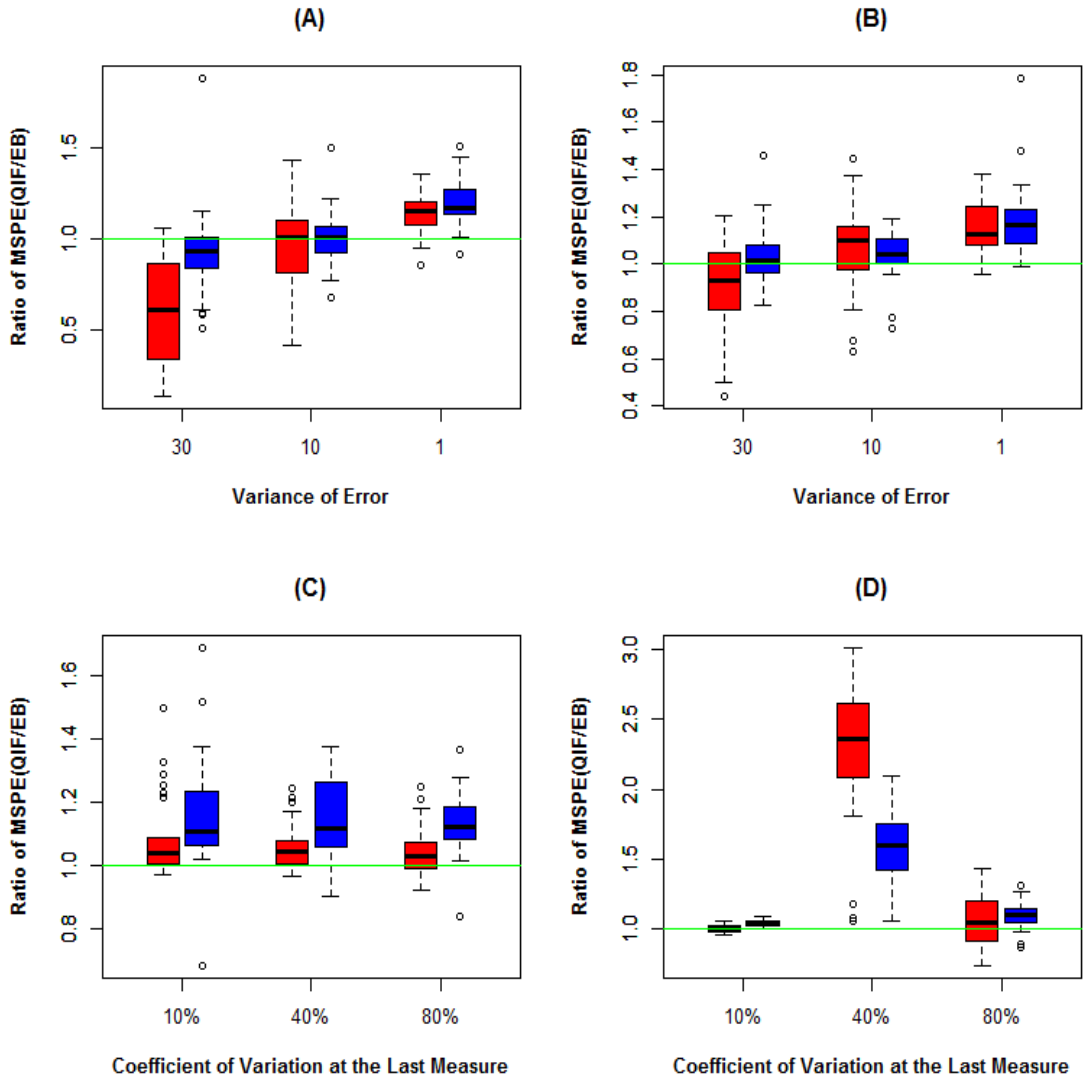
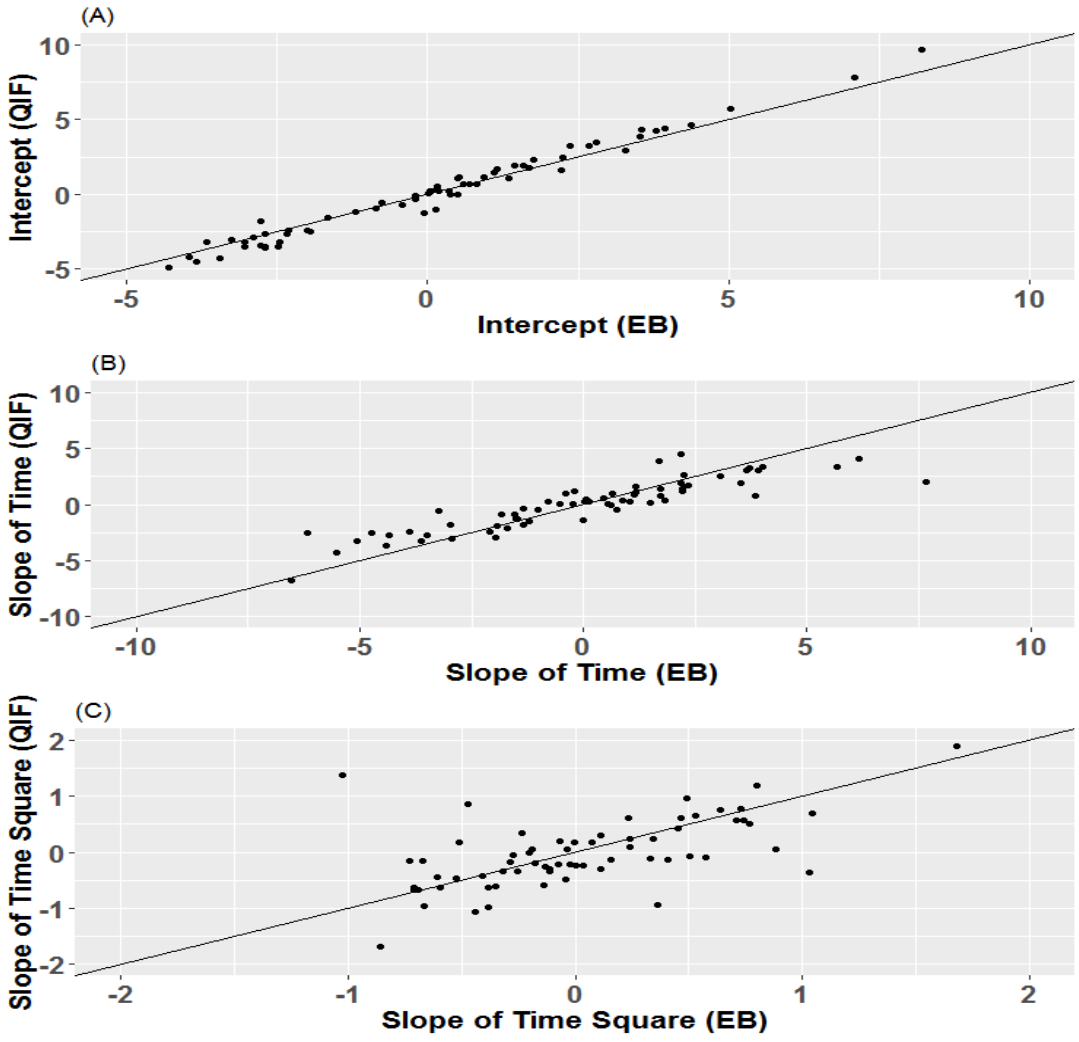


Figure 2. Scatterplots of random effects for depression data, predicted by QIF versus EB approaches under homoscedastic independent measurement errors. (A) Random intercept predictors. (B) Predictors of random coefficient of time. (C) Predictors of random coefficient of time square. The solid line is the y=x line.



Chapter III: The impact of violations of the normality assumption for the random effects in the logistic mixed effects model for longitudinal data on two methods of random-effect prediction: empirical Bayes versus quadratic inference functions

Zhiwen Wang and Francisco J. Diaz*

ABSTRACT

For the logistic mixed effects model in longitudinal data analysis, research on the accuracy and robustness to violations of the normality assumption of predictors of random effects is scarce. In this paper we use the mean square prediction error to assess the performance of predictors of random effects and compared the random effects prediction accuracy and robustness of two approaches: the empirical Bayes (EB) and a robust approach based on quadratic inference functions (QIFs). Our simulation study showed that in logistic models with random intercepts only, more accurate predictions were generated using the EB approach when the true random intercept variance was large. For logistic models with both random intercept and slope, the performances of the EB and QIF approaches were comparable, but the EB approach outperformed the QIF approach when the random effects were independent, and the sample size was relatively large. Our conclusion holds regardless of whether the true distribution of random effects is normal or non-normal. A consequence for statistical practice is that in logistic mixed effects models the EB approach from random effects prediction is relatively robust to violations of the normality assumption. Considering that the EB approach is easier to implement computationally, this approach is more recommendable for predicting random effects than the QIF approach. An application to a schizophrenia study using both approaches to predict random effects was illustrated.

KEY WORDS: Logistic model, prediction, random effects, BIC, distribution misspecification

1. Introduction

Mixed effects models are widely used when modeling repeatedly measured outcomes. In this family, linear mixed effects models, generalized linear mixed effects models and nonlinear mixed effects models are the three most commonly used models for continuous or discrete types of responses. Among generalized linear mixed effects models, the logistic mixed effects models are one of the most useful tools in modeling binary repeated outcomes for cluster data via logit link function. Logistic mixed effects models are widely used in biomedical research (Diaz, 2016; Have et al., 1998; Hedeker, 2003; Horrocks & van Den Heuvel, 2009; Lin & Breslow, 1996; Liu et al., 2008), epidemiology (Kleinman et al., 2004; Sashegyi et al., 2000; Skrondal & Rabe-Hesketh, 2003), social science (Mann et al., 2018), psychometrics (Van Den Noortgate et al., 2003) and animal ecology (Gillies et al., 2006). Statistical procedures, such as PROC GLIMMIX, PROC NLMIXED in SAS, or packages such as lme4 or nlme in R, are widely used for statistical practice.

One of the basic assumptions for estimation and inference in mixed effects models is that the random effects are assumed to follow normal distribution for simplicity and ease of computation. This assumption may be too restricted to follow in practice. For instance, when a categorical covariate is omitted, a multimodal distribution could be the best description for the distribution of random effects.

Research is heavily focused on estimation accuracy of the fixed effects parameters and their robustness to the violation of normality assumption of the random effects. There are some conflicting conclusions about whether the fixed effects estimates are affected by the distribution misspecification of the random effects. It has been shown that for logistic random intercept models, the estimates for fixed effects parameters are relatively robust with respect to the

distribution misspecification due to small biases occurring in the estimates (Neuhaus et al., 1992). This conclusion is also validated by presenting the conditions for obtaining consistent fixed effects parameter estimates under misspecification of the random effect distribution (Neuhaus et al., 1994). Agresti et al. (2004) showed that generally the fixed effects and variance components have small biases when the true random effects distribution was mis-specified, except in cases where the true random effects were simulated from an extreme two-point distribution with relatively large variance. On the contrary, Heagerty and Kurland (2001) illustrated that significantly biased estimates result for random intercept in logistic models, when the variance of the random effects depends on a between-cluster covariate. Heagerty and Zeger (2000) showed that in modeling longitudinal binary responses, for conditional logistic model with random intercept only, the estimated coefficients for between-cluster covariates could be severely biased; while small biases result for marginal logistic models. Using logistic models with random intercept and slope, and assuming a wide variety of true distributions for the random effects, Litière et al. (2008) illustrated that the misspecification of the random effects distribution has a severe effect on the fixed effects and variance components parameters estimates. Furthermore, the bias is more severe when there is more than one random effect included in the model. In Litière et al. (2007) they also showed that the type I error rate and statistical power of Wald tests on the mean structure, which are the fixed effects, are affected.

As important as the fixed effects and variance components, random effects play an essential role for inference, especially in personalized medicine, since the individual disease severity and treatment benefits are calculated based on individual random effects (Diaz, 2017, 2016). By comparing the mean square prediction errors under different distributions of the random effects, it is indicated that the prediction accuracy was little to mildly affected (Mcculloch & Neuhaus,

2011a). In McCulloch and Neuhaus (2011b) the authors validated their conclusion and stated that although the prediction accuracy was mildly affected, the exact shape of the distribution of random effects predictors could be severely biased. Agresti et al. (2004) showed that when the true random effects were distributed from an extreme 2-point mixture distribution with a large variance, severe prediction bias occurred on the random effects. A recent study showed that by using random intercept logistic models, and simulating the true random effects from a three-components mixture of normal distributions, the random intercept variance was biasedly estimated and this bias was relatively large when the true random intercept variance was large (Marquart and Haynes, 2019).

Thus, some authors have proposed robust methods for parameter estimation or prediction in generalized linear mixed effects models. For random effects in logistic models, Ten Have and Localio (1999) proposed two extensions of the empirical Bayes approach; one is based on Kass and Steffey (1989), and the other is based on Breslow and Clayton (1993). The performance of the two approaches were shown in terms of different cluster sizes, number of clusters and covariance of random effects. Shen and Louis (1999) proposed an empirical Bayes method based on smoothing by roughening estimate of the prior. The simulation studies in which the true prior was distributed from a mixture of Gaussian distribution and Poisson distribution showed its robustness to the violation and efficiency in estimation of prior distribution. Their methods could be applied in mixed effects models as a robust method for predicting the random effects. Requiring only that the random effects distribution be smooth, Chen et al. (2002) proposed a method which was based on a Monte Carlo EM algorithm with rejection sampling to estimate the fixed effects, variance components and to predict random effects. Wang et al. (2012) proposed an inference method by iteratively minimizing the extended score equations of both fixed and random effects. And their approach based on quadratic inference functions (QIF) was widely applied in treatment

personalization (Cho et al., 2017) and dosage individualization (Zhu and Qu, 2016) for longitudinal data via mixed effects models.

Since in logistic mixed effects model, little research has been done for evaluating the prediction performance of random effects, especially comparing the empirical Bayes approach with the modern robust approach, this paper focuses on comparing the prediction accuracy to the violation of the normality assumption of random effects, using the common empirical Bayes approach and the robust approach based on quadratic inference functions (Wang et al., 2012).

The paper is organized as follows. Section 2 presents the formulas and notations of logistic mixed effects models. Sections 3 reviews the empirical Bayes method of predicting random effects in mixed effects models and derives the formula of QIFs in the context of logistic models. Section 4 describes the simulation scenarios used in this paper and section 5 presents the simulation results and discussions. A clinical trial dataset was applied to both approaches for predicting random effects, and the results are shown in section 6. Conclusion and discussion are given in section 7.

2. Logistic mixed effects model

We present a logistic mixed effects model in the following. The total number of subjects is N. For subject i at time point j there are n_i number of repeated binary measurements y_{ij}. We assume the probabilities $p_{ij} = P\left(y_{ij} = 1 \middle| \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}\right)$, and build the model via logit link function such that

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i, \qquad i = 1, \dots, N; j = 1, \dots, n_i$$

where $\boldsymbol{Y}_i = \left(y_{i1}, \dots, y_{i,n_i}\right)^T$ is $n_i \times 1$ vector of responses such that $Y = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N)^T$ is $(\sum_{i=1}^{N} n_i) \times 1$ vector of responses of all subjects; $\boldsymbol{\beta} = \left(\beta_1, \dots, \beta_p\right)^T$ is $p \times 1$ vector of fixed effects; $\boldsymbol{x}_{ij} = \left(x_{ij1}, \dots, x_{ijp}\right)^T$ is $p \times 1$ vector of covariates for subject i at time j, $X_i = \left(\boldsymbol{x}_{i1}^T, \dots, \boldsymbol{x}_{in_i}^T\right)^T$ is $n_i \times p$ matrix of fixed effects covariates for subject i, such that $X = (X_1^T, \dots, X_N^T)^T$ is $(\sum_{i=1}^{N} n_i) \times p$ matrix for all subjects; $\boldsymbol{b}_i = \left(b_1, \dots, b_q\right)^T$ is $q \times 1$ vector of random effects for subject i and $\boldsymbol{b} = (\boldsymbol{b}_1^T, \dots, \boldsymbol{b}_N^T)^T$. Similarly, $\boldsymbol{z}_{ij} = \left(z_{ij1}, \dots, z_{ijq}\right)^T$ is $q \times 1$ vector of covariates for subject i at time j, $Z_i = \left(\boldsymbol{z}_{i1}^T, \dots, \boldsymbol{z}_{in_i}^T\right)^T$ is $n_i \times q$ matrix of random effects covariates for subject i, such that $Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & Z_N \end{bmatrix}$ is $(\sum_{i=1}^{N} n_i) \times (N \times q)$ block-diagonal matrix for all subjects. The random effects $\boldsymbol{b}_i$ follow a distribution φ with mean of zero and variance covariance matrix G; most commonly it is normal distribution, but violation is possible. In this paper we assume that conditional on the random effects, the binary responses are independent.

3. Prediction methods for random effects

3.1. Empirical Bayes (EB)

Empirical Bayes method of predicting random effects are widely used in biostatistics and medical research (Albert, 2012; Candel, 2009; Diaz, 2016; Feng et al., 2006; Mikulich-Gilbertson et al., 2019; Parzen et al., 2011; Sammel et al., 1997). The empirical Bayes predictor of the random effects is the posterior expectation of the distribution of random effects while treating the fixed

effects and variance components as fixed and replaced by their maximum likelihood estimates (MLE). The random effects predictor can be obtained by

$$\widehat{\boldsymbol{b}_i} = \frac{\int \boldsymbol{b}_i \, \varphi\left(\boldsymbol{b}_i, \hat{G}\right) \prod_{j=1}^{n_i} f\left(y_{ij} \middle| \boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \widehat{\boldsymbol{\beta}}\right) d\boldsymbol{b}_i}{\int \varphi\left(\boldsymbol{b}_i, \hat{G}\right) \prod_{j=1}^{n_i} f\left(y_{ij} \middle| \boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \widehat{\boldsymbol{\beta}}\right) d\boldsymbol{b}_i}$$

where $f\left(y_{ij} \middle| \boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \widehat{\boldsymbol{\beta}}\right)$ is the probability distribution function of the responses given the MLE of fixed effects $\widehat{\boldsymbol{\beta}}$, and $\varphi\left(\boldsymbol{b}_i, \hat{G}\right)$ is the assumed probability distribution function of random effects given the MLE of variance components $\hat{G}$ (Fitzmaurice et al., 2009). In logistic mixed effects models, there is no closed form expression for it. This predictor is also called the best predicted values (best predictor), since it is the one that has the minimum overall mean square prediction errors (McCulloch et al., 2008), which is the conditional mean of the random effects given the distribution of random effects. From Bayesian perspective, the empirical Bayes predictor is the Bayes rule that minimizes the expected posterior loss using quadratic loss function (Carlin and Louis, 2009; Skrondal and Rabe-Hesketh, 2004). It is shown in Carlin and Louis (2009) that using a Gaussian/Gaussian model, empirical Bayes point estimator has smaller empirical Bayes risk, compared with frequentist (MLE), pure Bayesian and hierarchical Bayesian estimators in most cases. And this superior performance naturally extends to the empirical Bayes coverage of the nominal 95% intervals of the parameter estimate.

3.2. An approach based on quadratic inference function (QIF)

Proposed by Wang et al. (2012) the approach based on quadratic inference functions is robust to the distributional assumption on the random effects when it is mis-specified from the default normal distribution. The fixed and random effects are obtained by iteratively minimizing

the fixed and random effects QIFs, and normality assumption is not required for the random effects. Denoting K as the total number of unique time points observed for all subjects, we assume that for each subject there is at least one measurement on at least one of the time points, such that, $1 \leq n_i \leq K$. Section 3.2.1 illustrates an implementation strategy for unbalanced data. We show the derivation of the QIFs for logistic mixed effects models in the following.

For the previously defined logistic mixed effects model, for response y_{ij}, we have the conditional mean

$$\mu_{ij}^{\boldsymbol{b}_i} = E\left(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i\right) = \frac{\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right)}{1+\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right)} = \frac{1}{1+\exp\left(-\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right)\right)}$$

and the conditional variance

$$v_{ij} = Var\left(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i\right) = \mu_{ij}^{\boldsymbol{b}_i}\left(1-\mu_{ij}^{\boldsymbol{b}_i}\right) = \frac{\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right)}{\left(1+\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right)\right)^2}.$$

For subject i its mean vector is $\boldsymbol{\mu}_i^{\boldsymbol{b}_i} = \left(\mu_{i1}^{\boldsymbol{b}_i}, \dots, \mu_{in_i}^{\boldsymbol{b}_i}\right)^T$ and variance vector is $\boldsymbol{v}_i = \left(v_{i1}, \dots, v_{in_i}\right)^T$.

Thus, we have

$$A_i = diag\left(Var(\boldsymbol{Y}_i|X_i, Z_i, \boldsymbol{b}_i)\right) = diag(\boldsymbol{v}_i)$$

$$= \begin{bmatrix} \mu_{i1}^{\boldsymbol{b}_i}\left(1-\mu_{i1}^{\boldsymbol{b}_i}\right) & 0 & \dots & 0 \\ 0 & \mu_{i2}^{\boldsymbol{b}_i}\left(1-\mu_{i2}^{\boldsymbol{b}_i}\right) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \mu_{in_i}^{\boldsymbol{b}_i}\left(1-\mu_{in_i}^{\boldsymbol{b}_i}\right) \end{bmatrix}_{n_i \times n_i}$$

Then,

$$\frac{d\boldsymbol{\mu}_i^{\boldsymbol{b}_i}}{d\boldsymbol{\beta}} = \begin{bmatrix} \frac{d\mu_{i1}^{\boldsymbol{b}_i}}{d\beta_1} & \frac{d\mu_{i1}^{\boldsymbol{b}_i}}{d\beta_2} & \cdots & \frac{d\mu_{i1}^{\boldsymbol{b}_i}}{d\beta_p} \\ \frac{d\mu_{i2}^{\boldsymbol{b}_i}}{d\beta_1} & \frac{d\mu_{i2}^{\boldsymbol{b}_i}}{d\beta_2} & \cdots & \frac{d\mu_{i2}^{\boldsymbol{b}_i}}{d\beta_p} \\ \vdots & \vdots & & \vdots \\ \frac{d\mu_{in_i}^{\boldsymbol{b}_i}}{d\beta_1} & \frac{d\mu_{in_i}^{\boldsymbol{b}_i}}{d\beta_2} & \cdots & \frac{d\mu_{in_i}^{\boldsymbol{b}_i}}{d\beta_p} \end{bmatrix}_{n_i \times p}$$

$$= \begin{bmatrix} \frac{\exp(\boldsymbol{x}_{i1}^T\boldsymbol{\beta} + \boldsymbol{z}_{i1}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{i1}^T\boldsymbol{\beta} + \boldsymbol{z}_{i1}^T\boldsymbol{b}_i))^2} x_{i11} & \cdots & \frac{\exp(\boldsymbol{x}_{i1}^T\boldsymbol{\beta} + \boldsymbol{z}_{i1}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{i1}^T\boldsymbol{\beta} + \boldsymbol{z}_{i1}^T\boldsymbol{b}_i))^2} x_{i1p} \\ \frac{\exp(\boldsymbol{x}_{i2}^T\boldsymbol{\beta} + \boldsymbol{z}_{i2}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{i2}^T\boldsymbol{\beta} + \boldsymbol{z}_{i2}^T\boldsymbol{b}_i))^2} x_{i21} & \cdots & \frac{\exp(\boldsymbol{x}_{i2}^T\boldsymbol{\beta} + \boldsymbol{z}_{i2}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{i2}^T\boldsymbol{\beta} + \boldsymbol{z}_{i2}^T\boldsymbol{b}_i))^2} x_{i2p} \\ \vdots & \vdots & \vdots \\ \frac{\exp(\boldsymbol{x}_{in_i}^T\boldsymbol{\beta} + \boldsymbol{z}_{in_i}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{in_i}^T\boldsymbol{\beta} + \boldsymbol{z}_{in_i}^T\boldsymbol{b}_i))^2} x_{in_i1} & \cdots & \frac{\exp(\boldsymbol{x}_{in_i}^T\boldsymbol{\beta} + \boldsymbol{z}_{in_i}^T\boldsymbol{b}_i)}{(1+\exp(\boldsymbol{x}_{in_i}^T\boldsymbol{\beta} + \boldsymbol{z}_{in_i}^T\boldsymbol{b}_i))^2} x_{in_ip} \end{bmatrix}_{n_i \times p}$$

$$= \begin{bmatrix} \mu_{i1}^{b_i}(1-\mu_{i1}^{b_i}) & 0 & \cdots & 0 \\ 0 & \mu_{i2}^{b_i}(1-\mu_{i2}^{b_i}) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \mu_{in_i}^{b_i}\left(1-\mu_{in_i}^{b_i}\right) \end{bmatrix}_{n_i \times n_i} \begin{bmatrix} x_{i11} & x_{i12} & \cdots & x_{i1p} \\ x_{i21} & x_{i22} & \cdots & x_{i2p} \\ \vdots & \vdots & & \vdots \\ x_{in_i1} & x_{in_i2} & \cdots & x_{in_ip} \end{bmatrix}_{n_i \times p}$$

$$= A_i X_i$$

Similarly, we have that $\frac{d\boldsymbol{\mu}_i^{\boldsymbol{b}_i}}{d\boldsymbol{b}_i} = A_i Z_i$.

Thus, the fixed effects QIF is

$$L^f(\boldsymbol{\beta}|\boldsymbol{b}) = N\left(G_N^f\right)^T \left(C_N^f\right)^{-1} G_N^f$$

where $C_N^f = \frac{1}{N}\sum_{i=1}^N g_i^f \left(g_i^f\right)^T$ and $g_i^f = \begin{pmatrix} X_i^T\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i^{\boldsymbol{b}_i}\right) \\ X_i^T A_i A_i^{-\frac{1}{2}} M_2 A_i^{-\frac{1}{2}}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i^{\boldsymbol{b}_i}\right) \end{pmatrix}_{2p \times 1}$.

And the random effects QIF is

$$L^r(\boldsymbol{b}|\boldsymbol{\beta}) = G_N^r\,(G_N^r)^T$$

where the extended score for random effects $\boldsymbol{b}$ is $g_i^r = \left(\frac{d\boldsymbol{\mu}_i^{\boldsymbol{b}_i}}{d\boldsymbol{b}_i}\right)^T A_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i^{\boldsymbol{b}_i}\right) = Z_i^T\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i^{\boldsymbol{b}_i}\right)$ and $G_N^r(\boldsymbol{b}) = \left\{(g_1^r)^T, \ldots, (g_N^r)^T,\ \lambda_1\boldsymbol{b}^T,\ \lambda_2\left(P_J\boldsymbol{b}\right)^T\right\}^T$.

Tuning parameter λ_2 is chosen to be $\log(N)$ which works fine from numerical studies (Cho et al., 2017; Wang et al., 2012; Zhu and Qu, 2016). It is known that the magnitude of the variance of random effects predictors is controlled by parameter λ_1 and thus should be selected by the value that gives the smallest BIC, in which

$$BIC = N\left(G_N^f\right)^T\left(C_N^f\right)^{-1}G_N^f + (\log(N))\left(P_J\boldsymbol{b}\right)^T\Sigma^{-1}\left(P_J\boldsymbol{b}\right), \text{ where } \Sigma = \text{cov}\left(P_J\boldsymbol{b}\right).$$

We compute the projection matrix P_J as follows. Apply Q-R decomposition to $A = \left((I - P_X)Z\right)^T$ such that $A = Q \times R$, where $P_X = X(X^TX)^{-1}X^T$, and obtain $r = \text{rank}(A)$. We will obtain a matrix J whose columns are the columns of Q beyond the rth column (Boyd and Vandenberghe, 2004). Thus, J has $Nq - r$ columns. The projection matrix is $P_J = J(J^TJ)^{-}J^T$.

To start the iteration process, logistic regression was used to obtain an initial estimate for $\boldsymbol{\beta}$, denoted $\widehat{\boldsymbol{\beta}}^{(0)}$. After the initial estimate, $L^r\left(\boldsymbol{b}|\widehat{\boldsymbol{\beta}}^{(0)}\right)$ is minimized with respect to $\boldsymbol{b}$ to obtain $\widehat{\boldsymbol{b}}^{(1)}$. Then, $L^f\left(\boldsymbol{\beta}|\widehat{\boldsymbol{b}}^{(1)}\right)$ is minimized with respect to $\boldsymbol{\beta}$ to obtain $\widehat{\boldsymbol{\beta}}^{(1)}$. And we repeat this iterative process until $\left|\widehat{\boldsymbol{\beta}}^{(s+1)} - \widehat{\boldsymbol{\beta}}^{s}\right| + \left|\widehat{\boldsymbol{b}}^{(s+1)} - \widehat{\boldsymbol{b}}^{s}\right| < 10^{-5}$ with $\widehat{\boldsymbol{b}}^{(0)} = \boldsymbol{0}$.

3.2.1. Strategy for unbalanced data

The approach of handling unbalanced data was the same with Cho et al. (2017) and Wang et al. (2012). Assuming for subject i with unbalanced responses such that $n_i < K$, a transformation matrix Λ_i of dimension $K \times n_i$ was defined by removing the columns of the $K \times K$ identity matrix corresponding to the time points with missing observations. In the iterative algorithm, we use $\boldsymbol{Y}_i^* = \Lambda_i \boldsymbol{Y}_i$, $\boldsymbol{\mu}_i^* = \Lambda_i \boldsymbol{\mu}_i$, $\dot{\boldsymbol{\mu}}_i^* = \Lambda_i \dot{\boldsymbol{\mu}}_i$, where $\dot{\boldsymbol{\mu}}_i = \frac{d\boldsymbol{\mu}_i}{d\boldsymbol{\beta}}$, $A_i^* = \Lambda_i A_i \Lambda_i^T$, and $(A_i^*)^{-1} = \Lambda_i A_i^{-1} \Lambda_i^T$, respectively in place of $\boldsymbol{Y}_i$, $\boldsymbol{\mu}_i$, $\dot{\boldsymbol{\mu}}_i^*$, A_i and A_i^{-1} (Wang et al., 2012). Both the missing responses and their predicted values are replaced with zeros.

4. Simulation scenarios for comparison of EB and QIF

Simulations were carried out to examine the prediction performance of both approaches under different distributions of random effects. We used two models that were motivated by NIMH schizophrenia collaborative study (Hedeker and Gibbons, 2006). The normal and non-normal distributions shared the same mean and variance-covariance in each model. For simplicity we assumed $n_i = n$ for all i, for all scenario in both models, so that all subjects have the same number of repeated measures, thus were of the same size.

Model 1: logistic random intercept model:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 drug_i + \beta_2\sqrt{week_{ij}} + \beta_3\left(drug_i \times \sqrt{week_{ij}}\right) + b_i, i = 1, \dots, N; j = 1, \dots, n$$

Sample size $N = \{50, 100, 200\}$, and the total number of repeated measurements for time covariate "week" was chosen to be either 4 or 7. For a subject i, when $n = 4$, $week_{ij} \in \{0, 1, 3, 6\}$;

when $n = 7$, $week_{ij} \in \{0, 1, 2, 3, 4, 5, 6\}$. The binary "drug" covariate indicated whether a subject i was under treatment (drug=1) or placebo (drug=0) and was simulated from a Bernoulli distribution such that $\text{drug}_\text{i} \sim \text{Bernoulli}(0.5), i = 1, \dots, N$.

The true fixed effects vector was $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (5.5, -0.025, -1.5, -1)^T$.

Random effects were simulated from 5 different normal and non-normal distributions with the same mean 0 and variance $\sigma^2 \in \{2, 4, 8, 10\}$. Necessary transformations were used when the distributions were chi-square and exponential to make sure that they all share the same mean and variance. We summarize the scenarios in the following: for subject i, $i = 1, \dots, N$,

1. Normal distribution: $b_i \sim \text{Normal}\ (0, \sigma^2)$.
2. t distribution: $b_i \sim t$ with degree of freedom $df = \frac{2\sigma^2}{\sigma^2 - 1}$.
3. Exponential distribution: $b_i \sim \text{exponential}\left(\sqrt{\sigma^2}\right)$.
4. Chi-square distribution: $b_i \sim \chi_2\left(\frac{\sigma^2}{2}\right)$.
5. Symmetric mixture of normal distributions: $b_i \sim \frac{1}{2}\text{Normal}(-1, \sigma^2 - 1) + \frac{1}{2}\text{Normal}(1, \sigma^2 - 1)$.

Similarly, we used the following bivariate random effects logistic model:

Model 2: logistic random intercept and slope model:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_1 drug_i + \beta_2\sqrt{week_{ij}} + \beta_3\left(drug_i \times \sqrt{week_{ij}}\right) + b_{0i} + b_{1i}\sqrt{week_{ij}},$$

$$i = 1, \dots, N; j = 1, \dots, n$$

The binary "drug" covariate for subject i was simulated from a Bernoulli distribution such that $\text{drug}_\text{i} \sim \text{Bernoulli}(0.75)$.

The true fixed effects vector was $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (6, 0.3, -1.5, -1.6)^T$.

We assume the random effects were distributed from a distribution φ such that, for subject i, $i = 1, \dots, N$,

$$\boldsymbol{b}_i = (b_{1i}, b_{2i})^T \sim \varphi\left(\begin{bmatrix}0\\0\end{bmatrix}, V = \begin{bmatrix}d_1 & d_{12}\\ d_{12} & d_2\end{bmatrix}\right)$$

In detail, we used the following distributions for φ:

1. Normal distribution:

$$\boldsymbol{b}_i \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, V_k\right), k = 1, 2, 3$$

where $V_1 = \begin{bmatrix}2.7 & 0\\ 0 & 1.6\end{bmatrix}$ such that the correlation between random effects is zero, i.e. $corr(b_{1i}, b_{2i}) = 0$; $V_2 = \begin{bmatrix}2.7 & -0.8\\ -0.8 & 1.6\end{bmatrix}$ such that $corr(b_{1i}, b_{2i}) = 0.4$; $V_3 = \begin{bmatrix}2.7 & -1.65\\ -1.65 & 1.6\end{bmatrix}$ such that $corr(b_{1i}, b_{2i}) = 0.8$.

2. Symmetric mixture of normal distributions:

$\boldsymbol{b}_i \sim \frac{1}{2} N(\boldsymbol{\mu}, D_k) + \frac{1}{2} N(-\boldsymbol{\mu}, D_k), k = 1, 2, 3$ where $\boldsymbol{\mu} = \begin{bmatrix}0.45\\0.45\end{bmatrix}$ and $D_k = V_k - \mu\mu^T$.

3. t distribution with 3 degrees of freedom: $\text{df} = 3$:

$\boldsymbol{b}_i \sim t_{\text{df}=3}\left(\begin{bmatrix}0\\0\end{bmatrix}, \Sigma_\text{k}\right), k = 1, 2, 3$ where $\Sigma_\text{k} = \frac{\text{df}-2}{\text{df}} V_k$.

For each scenario, 100 datasets were simulated and both EB and QIF approaches were used to predict the individual random effects assuming a normal distribution for the random effects.

Mean square prediction error (MSPE) were used to compare the prediction performance of the two approaches. For the $l-th$ dataset, the MSPE for the random intercept was

$$MSPE_l = \frac{\sum_{i=1}^{N}\left(\hat{b}_{1i,l} - b_{1i,l}\right)^2}{N}$$

where $\hat{b}_{1i,l}$ was a predictor of the random intercept and $b_{1i,l}$ was the true simulated one. Similarly, MSPE was defined for the random slope by replacing 1 with 2. The ratio R of average MSPE for the QIF approach to the average MSPE for the EB approach was calculated for random intercept and random slope for each scenario. $R = MSPE_{QIF}/MSPE_{EB}$, where $MSPE_{EB} = \frac{\sum_{l=1}^{\Delta_l} MSPE_{EB,l}}{\Delta}$ and $MSPE_{QIF} = \frac{\sum_{l=1}^{\Delta_l} MSPE_{QIF,l}}{\Delta}$ and $\Delta \leq 100$. Δ is the number of converged datasets $\Delta_l, l = 1, \dots, \Delta$. The SAS procedure GLIMMIX was used to obtain the empirical Bayes predictor of random effects. Adaptive quadrature was used for maximization of the likelihood functions; hence, not all datasets were converged. Thus, the comparisons of MSPEs were based on the converged datasets from the EB approach. If the ratio was smaller than 1, the QIF approach outperformed the EB approach with respect to the prediction accuracy, and vice versa.

5. Simulation results

5.1. Model 1: logistic random intercept model

The MSPEs of the two approaches and the ratios comparing the two approaches are shown in Table 1 and Table 2. Generally, regardless of the true distribution of the random intercept, the MSPEs obtained using the EB approach were always smaller than those obtained using the QIF approach. Therefore, the ratios of MSPEs were all larger than 1. This indicated that for the

prediction accuracy of the random intercept, the EB approach was better compared with the QIF approach, no matter whether the true distribution of random intercept was normal or non-normal.

For a fixed sample size, as the random intercept variance increased, the MSPEs and the ratios of the MSPEs increased as well, maybe except for the t scenarios, in which the ratios decreased slightly. This illustrated that although the performance of both approaches worsened due to the larger MSPEs in scenarios with large random intercept variance, the EB approach was still superior to the QIF approach. Although for the t scenarios, the performance of two approaches were becoming similar with each other as the variance of random intercepts increased. The MSPEs obtained when $n = 7$ are generally smaller compared with those from the same scenario when $n = 4$, and the ratios were generally larger when $n = 7$. The more information we have, either by increasing the number of subjects or increasing the number of repeated measurements, the smaller the MSPE values were, and the larger the ratios were. These conclusions were true not only when the random intercept was simulated from normal distributions, but also the non-normal distributions. This observation proved that the EB method for random effects was sufficiently accurate when compared with the relatively robust QIF approach. The larger the sample size, the larger the number of repeated measurements, the more accurate the prediction of both approaches. As the variance of random effects increased, the EB approach outperformed the QIF approach greatly.

5.2. Model 2: logistic random intercept and slope model

The MSPEs of the two approaches and the ratios comparing the two approaches are shown in Table 3 and 4 for $n = 7$ and Table 5 and 6 for $n = 4$. Generally, for a fixed correlation

coefficient between random intercept and random slope, the larger the sample size, the smaller the MSPEs obtained for the EB approach, not for the QIF approach, and the larger the ratios R. For a fixed sample size, the larger the correlation coefficient between random intercept and random slope, the larger the MSPEs obtained for both approaches, and the smaller the ratios R. For a scenario in which the true random effects distribution was known and the sample size was fixed, the larger the number of repeated measurements, the smaller the MSPEs. Generally, when there exists strong correlation between random effects, the QIF approach is more accurate in predicting random effects, and this was especially true when true random effects were from bivariate t distributions with 3 degrees of freedom. But when the sample size was sufficiently large, more accurate predictors were obtained by the EB approach, and this is especially true when the true random effects were from bivariate normal distributions and a symmetric mixture of bivariate normal distributions.

In conclusion, when the sample size was sufficiently large, and a weak correlation existed between random intercept and random slope, generally, the EB approach outperformed the QIF approach. But when the correlations between random effects were relatively strong, the QIF approach outperformed the EB approach even when the sample size was not relatively large. This was especially true for random intercept prediction.

6. Application to a schizophrenia study

As an illustration, to check agreement on random effects prediction two approaches, we applied both prediction methods to a NIMH schizophrenia collaborative study (Hedeker and Gibbons, 2006), in which the binary scores of "Item 79 of Inpatient Multidimensional Psychiatric

Scale(IMPS)" were measured for 4 treatment periods, for 108 patients from placebo arm and 329 patients from treatment arm. We denote $y_{ij} = 1$ if a patient i is mildly ill based on this score at treatment duration j, and $y_{ij} = 0$ if the patient is moderately ill. The treatment arm is a combination of treatments chlorpromazine, fluphenazine and thioridazine. We fit the same model as in Hedeker and Gibbons (2006), where the random effects were intercept and square root of treatment duration, in which $drug$ is an binary indicator of patient i being in either treatment or placebo group.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 drug_i + \beta_2\sqrt{week_{ij}} + \beta_3\left(drug_i \times \sqrt{week_{ij}}\right) + b_{0i} + b_{1i}\sqrt{week_{ij}},$$

$$i = 1, \dots, N; j = 1, \dots, n_i$$

The maximum number of repeated measurements $n = 4$, such that for subject i, $week_{ij} \in \{0, 1, 3, 6\}$. Missing responses were observed in all time points, thus missing handling method was used in the QIF approach. The SAS procedure GLIMMIX, which assumes random effects follow normal distribution, was used to obtain maximum likelihood estimates of the fixed effects and EB predictors of the random effects (SAS Institute Inc. Cary, NC). An unstructured covariance matrix for the random effects was assumed in this application.

The predictions on the random slope by the EB and QIF approaches correlated relatively strongly, with correlation of 0.78, while the correlation for random intercept is 0.71. Figure 1 shows scatterplots comparing the predictions of the random effects by the QIF and EB approaches under homoscedastic independent errors. Overall, the two prediction methods agreed moderately, except for the prediction on the random slope. Specifically, predictors for the intercept were scattered around mean zero, suggesting that EB intercept predictions tended to be larger in absolute

value than QIF predictions. The random slope predictors were scattered more evenly around the diagonal line.

7. Discussion and conclusion

In this paper we compared the random effects prediction performance of two approaches, i.e. the EB approach and the QIF approach. The EB approach requires that the random effects follow a parametric distribution, usually normal distribution. The QIF approach does not require the random effects to follow any specific parametric distribution, thus it is a robust method.

In the simulation study, the true random effects were distributed from both normal and non-normal distributions, such as t distribution, mixture of normal distributions, chi-square distribution, etc. To obtain the empirical Bayes predictors, SAS procedure GLIMMIX was used assuming independent working correlation assumption for responses. The comparison of the MSPEs was based on the converged datasets only. Out of the 100 simulations, for model 1, the proportion of converged simulated datasets was no smaller than 91% when $n = 7$, and no less than 71% when $n = 4$. And the smallest proportion was observed when the true random effects were from an exponential distribution. The proportions were relatively smaller for the bivariate logistic models. When $n = 7$, it was as small as 30% when the true random effects were from normal distribution and a mixture of normal distributions. When $n = 4$, this proportion was only 25% when the true random effects were from a t distribution with 3 degrees of freedom. Our results were consistent with the findings reported in the simulation study in Wang et al. (2012).

From the simulation study we conclude that in logistic models with random intercept only, the EB approach generates more accurate random intercept predictors compared with the QIF

approach. This conclusion is unrelated with the true distribution of random effects, number of subjects, and number of repeated measurements. Generally, the larger the sample size, and the larger the true variance of random intercept, the performance of the EB approach is superior compared with the QIF approach. The exception occurs when the true distribution of the random effects is t distribution: the smaller the variance, the greater accuracy of the EB approach. For logistic models with more than one random effect, we observe that in most cases, the performance of the two approaches are comparable. Generally, the EB approach performs slightly better than the robust QIF approach, when the random effects are independent, and the sample size is relatively large. And this observation is especially true when the true random effects are from normal or a mixture of normal distributions. When the random effects were from bivariate t distribution, and the random effects were strongly correlated, the QIF approach has greater accuracy in random effects prediction.

Generally, in logistic models with random intercept only, the MSPEs from exponential distributions are relatively larger than those from other distributions. This is because the exponential distribution has limited support, while the assumed true normal distribution does not. This is observed regardless of the number of repeated measures, and our findings are consistent with the ones reported in Mcculloch and Neuhaus (2011a). An important message stated in McCulloch and Neuhaus (2011a) is that the prediction performance of the EB approach was less affected by the misspecification of the distribution of random effects. Their conclusion was also validated in Mcculloch and Neuhaus (2011a) in which they derived the formulas for the MSPE of the random intercept for both linear and logistic models. The authors concluded that although the shape of the distribution of the EB predictors may not match the true underlining distribution, the performance measured by MSPE was not affected significantly by the misspecification. The

excellent prediction performance of the EB approach is not a surprise to us. As shown in Searle et al. (1992) for linear models, the EB predictor is the best predictor which minimized the mean square error of prediction given the sampling distribution of the response and all parameters are known. The well-known optimal property of BLUP for linear models further illustrated the robustness of EB predictor of random effects. Although there is no closed form of EB predictor for logistic model, the optimality from the linear case shed some light on understanding its robustness in the binary case. And we also proved this property through simulations in the paper.

Thus, for statistical practice, when focus is on prediction of random effects, even when one has doubts about the normal assumption of the random effects distribution, the empirical Bayes approach is still recommended, due to its noninferior performance in predicting random effects compared with a more robust method, i.e. the QIF approach, its robustness to violations of the normality assumption of predictors of random effects and its easy implementation with existing statistical packages.

References

Agresti, A., Caffo, B., Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. Comput. Stat. Data Anal. 47, 639–653.

Albert, P.S., 2012. A linear mixed model for predicting a binary event from longitudinal data under random effects misspecification. Stat. Med. 31, 143–154.

Boyd, S.P., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press, Cambridge, UK ; New York.

Breslow, N.E., Clayton, D.G., 1993. Approximate Inference in Generalized Linear Mixed Models. J. Am. Stat. Assoc. 88, 9–25.

Candel, M.J.J.M., 2009. Optimal designs for empirical bayes estimators of individual linear and quadratic growth curves in linear mixed models. Stat. Methods Med. Res. 18, 397–419.

Carlin, B.P., Louis, T.A., 2009. Bayesian methods for data analysis, 3rd ed, Chapman & Hall/CRC texts in statistical science series. CRC Press, Boca Raton.

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

Cho, H., Wang, P., Qu, A., 2017. Personalize Treatment for Longitudinal Data Using Unspecified Random-Effects Model. Stat. Sin. 27, 187–205.

Diaz, F.J., 2017. Estimating individual benefits of medical or behavioral treatments in severely ill patients. Stat. Methods Med. Res. 962280217739033.

Diaz, F.J., 2016. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat. Med. 35, 4077–4092.

Feng, S., Wolfinger, R.D., Chu, T.M., Gibson, G.C., McGraw, L.A., 2006. Empirical Bayes analysis of variance component models for microarray data. J. Agric. Biol. Environ. Stat. 11, 197–209.

Fitzmaurice, G.M., Davidian, M., Verbeke, G., Molenberghs, G., 2009. Longitudinal data analysis, Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Boca Raton.

Gillies, C.S., Hebblewhite, M., Nielsen, S.E., Krawchuk, M.A., Aldridge, C.L., Frair, J.L., Saher, D.J., Stevens, C.E., Jerde, C.L., 2006. Application of random effects to the study of resource selection by animals. J. Anim. Ecol.

Have, T.R. Ten, Kunselman, A.R., Pulkstenis, E.P., Landis, J.R., 1998. Mixed Effects Logistic Regression Models for Longitudinal Binary Response Data with Informative Drop-Out. Biometrics 54, 367–383.

Heagerty, P.J., Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88, 973–985.

Heagerty, P.J., Zeger, S.L., 2000. Marginalized multilevel models and likelihood inference. Stat. Sci. 15, 1–19.

Hedeker, D., 2003. A mixed-effects multinomial logistic regression model. Stat. Med. 22, 1433–1446.

Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. Wiley-Interscience, Hoboken,

N.J.

Horrocks, J., van Den Heuvel, M.J., 2009. Prediction of pregnancy: A joint model for longitudinal and binary data. Bayesian Anal. 4, 523–538.

Kass, R.E., Steffey, D., 1989. Approximate bayesian inference in conditionally independent hierarchical models (Parametric empirical bayes models). J. Am. Stat. Assoc. 84, 717–726.

Kleinman, K., Lazarus, R., Platt, R., 2004. A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism. Am. J. Epidemiol. 159, 217–224.

Lin, X., Breslow, N.E., 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. J. Am. Stat. Assoc. 91, 1007–1016.

Litière, S., Alonso, A., Molenberghs, G., 2008. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. Stat. Med. 27, 3125–3144.

Litière, S., Alonso, A., Molenberghs, G., 2007. Type I and type II error unde random-effects misspecification in generalized linear mixed models. Biometrics 63, 1038–1044.

Liu, D., Ghosh, D., Lin, X., 2008. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. BMC Bioinformatics.

Mann, R.P., Spaiser, V., Hedman, L., Sumpter, D.J.T., 2018. Choice modelling with Gaussian processes in the social sciences: A case study of neighbourhood choice in stockholm. PLoS One.

Marquart, L., Haynes, M., 2019. Misspecification of multimodal random-effect distributions in logistic mixed models for panel survey data. J. R. Stat. Soc. Ser. A Stat. Soc. 182, 305–321.

McCulloch, C., Searle, S., Neuhaus, J., 2008. Generalized, Linear, and Mixed Models, 2nd ed. Wiley Series in Probability and Statistics, Hoboken, N.J.

Mcculloch, C.E., Neuhaus, J.M., 2011. Prediction of Random Effects in Linear and Generalized Linear Models under Model Misspecification. Biometrics 67, 270–279.

McCulloch, C.E., Neuhaus, J.M., 2011. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Stat. Sci. 26, 388–402.

Mikulich-Gilbertson, S.K., Wagner, B.D., Grunwald, G.K., Riggs, P.D., Zerbe, G.O., 2019. Using empirical Bayes predictors from generalized linear mixed models to test and visualize associations among longitudinal outcomes. Stat. Methods Med. Res. 28, 1399–1411.

Neuhaus, J.M., Hauck, W.W., Kalbfleisch, J.D., 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. Biometrika 79, 755–762.

Neuhaus, J.M., Kalbfleisch, J.D., Hauck, W.W., 1994. Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. Can. J. Stat. 22, 139–148.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G.M., Mallick, B.K., Ibrahim, J.G., 2011. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. Ann. Appl. Stat. 5, 449–467.

Sammel, M.D., Ryan, L.M., Legler, J.M., 1997. Latent variable models for mixed discrete and continuous outcomes. J. R. Stat. Soc. Ser. B Stat. Methodol. 59, 667–678.

Sashegyi, A.I., Brown, K.S., Farrell, P.J., 2000. Application of a generalized random effects regression model for cluster-correlated longitudinal data to a school-based smoking prevention trial. Am. J. Epidemiol. 152, 1192–1200.

Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components, 1st ed, Book. Wiley-Interscience.

Shen, W., Louis, T.A., 1999. Empirical Bayes estimation via the smoothing by roughening approach. J. Comput. Graph. Stat. 8, 800–823.

Skrondal, A., Rabe-Hesketh, S., 2004. Generalized latent variable modeling : multilevel, longitudinal, and structural equation models, Chapman & Hall/CRC interdisciplinary statistics series. Chapman & Hall/CRC, Boca Raton.

Skrondal, A., Rabe-Hesketh, S., 2003. Some Applications of Generalized Linear Latent and Mixed Models in Epidemiology. Nor. Epidemiol. 13, 265–278.

Ten Have, T.R., Localio, A.R., 1999. Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. Biometrics 55, 1022–1029.

Van Den Noortgate, W., De Boeck, P., Meulders, M., 2003. Cross-Classification Multilevel Logistic Models in Psychometrics. J. Educ. Behav. Stat. 28, 369–386.

Wang, P., Tsai, G.F., Qu, A.N., 2012. Conditional Inference Functions for Mixed-Effects Models With Unspecified Random-Effects Distribution. J. Am. Stat. Assoc. 107, 725–736.

Zhu, X.L., Qu, A.N., 2016. Individualizing drug dosage with longitudinal data. Stat. Med. 35, 4474–4488.

Table 1. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPEs comparing QIF approach to EB approach ($R = MSPE_{QIF}/MSPE_{EB}$) for random intercept prediction in Model 1, when the true random effects were simulated from normal, symmetric mixture of normal, t, exponential and chi-square distributions with difference variances, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 7$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.02	0.974	0.936	1.45	1.45	1.54	1.42	1.49	1.64
4	1.61	1.56	1.52	2.72	2.80	2.54	1.69	1.79	1.67
8	2.71	2.59	2.45	5.40	5.24	5.14	1.99	2.02	2.09
16	5.34	4.97	4.51	10.9	11.5	11.6	2.05	2.32	2.58
	Symmetric mixture of normal distributions								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.02	0.975	0.962	1.45	1.54	1.51	1.42	1.58	1.57
4	1.67	1.53	1.47	2.88	2.57	2.41	1.73	1.68	1.64
8	2.76	2.51	2.39	5.55	4.93	4.95	2.01	1.96	2.07
16	5.21	4.64	4.73	10.6	11.1	11.6	2.04	2.38	2.45
	t distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.14	1.09	1.09	1.59	1.49	1.52	1.40	1.36	1.40
4	2.20	2.07	2.13	2.86	2.81	2.76	1.30	1.36	1.30
8	6.90	3.38	4.11	7.98	4.23	5.02	1.16	1.25	1.22
16	7.47	7.20	7.57	8.81	8.23	8.72	1.18	1.14	1.15
	Shifted exponential distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.47	1.42	1.46	1.77	1.77	1.83	1.21	1.24	1.25
4	2.65	2.50	2.56	3.32	3.34	3.20	1.25	1.33	1.25
8	4.49	4.80	4.50	6.24	6.52	5.92	1.39	1.36	1.32
16	9.67	8.38	8.54	14.2	13.5	14.4	1.47	1.61	1.69
	Shifted Chi-square distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.57	1.64	1.53	1.93	1.93	1.87	1.23	1.18	1.22
4	2.70	2.55	2.58	3.39	3.26	3.35	1.26	1.28	1.30
8	3.93	3.75	3.75	6.22	5.77	5.37	1.58	1.54	1.43
16	6.63	6.17	6.10	12.0	12.3	12.8	1.81	1.99	2.09

Table 2. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPEs comparing QIF approach to EB approach ($R = MSPE_{QIF}/MSPE_{EB}$) for random intercept prediction in Model 1, when the true random effects were simulated from normal, symmetric mixture of normal, t, exponential and chi-square distributions with difference variances, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 4$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.50	1.37	1.32	1.80	1.71	1.69	1.20	1.25	1.28
4	2.50	2.17	2.05	3.20	3.34	3.05	1.28	1.53	1.48
8	3.84	3.38	3.36	6.48	6.26	6.10	1.69	1.85	1.82
16	6.85	6.06	5.61	12.8	12.9	13.0	1.88	2.12	2.31
	Symmetric mixture of normal distributions								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.51	1.36	1.32	1.74	1.68	1.69	1.15	1.24	1.28
4	2.36	2.12	2.02	3.27	3.12	3.13	1.39	1.47	1.55
8	3.85	3.49	3.29	6.31	6.55	6.05	1.64	1.88	1.84
16	6.76	5.86	5.72	13.3	12.5	13.1	1.96	2.13	2.29
	t distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.56	1.42	1.38	1.73	1.63	1.68	1.11	1.15	1.22
4	2.75	2.62	2.43	3.36	3.20	2.95	1.22	1.22	1.22
8	5.91	5.88	5.96	7.15	7.01	6.71	1.21	1.19	1.13
16	12.1	7.36	6.65	13.3	8.59	7.63	1.09	1.17	1.15
	Shifted exponential distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.82	1.68	1.64	1.99	1.90	1.86	1.10	1.13	1.14
4	3.13	3.14	2.96	3.54	3.46	3.22	1.13	1.10	1.09
8	5.06	5.62	5.47	6.73	6.58	6.19	1.33	1.17	1.13
16	10.4	10.1	9.74	13.7	13.3	12.7	1.33	1.32	1.31
	Shifted Chi-square distribution								
Variance	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
2	1.77	1.81	1.76	1.95	1.99	1.97	1.10	1.10	1.12
4	3.16	3.08	3.11	3.51	3.36	3.46	1.11	1.09	1.11
8	5.06	4.93	4.92	6.73	6.53	6.20	1.33	1.32	1.26
16	8.16	7.77	7.56	13.2	13.5	13.3	1.61	1.73	1.76

Table 3. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPEs comparing QIF approach to EB approach ($R = MSPE_{QIF}/MSPE_{EB}$) for random intercept prediction in Model 2, when the true random effects were simulated from bivariate normal, symmetric mixture of normal and t distributions with difference correlations, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 7$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	3.79	2.28	2.17	2.36	2.40	2.44	0.623	1.05	1.12
0.4	2.65	2.67	2.53	2.53	2.62	2.58	0.955	0.98	1.02
0.8	3.36	2.97	2.87	2.83	2.75	2.74	0.841	0.926	0.955
	Symmetric mixture of normal distributions								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	2.60	2.21	2.15	2.48	2.39	2.49	0.956	1.08	1.16
0.4	3.43	2.53	2.51	2.55	2.52	2.59	0.745	0.993	1.03
0.8	3.06	2.83	2.77	2.78	2.67	2.71	0.908	0.946	0.979
	t distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	2.54	2.09	2.55	2.48	2.25	2.75	0.978	1.07	1.08
0.4	4.73	2.55	2.51	4.27	2.47	2.46	0.903	0.971	0.981
0.8	3.00	3.62	2.85	2.90	3.43	2.83	0.968	0.948	0.993

Table 4. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPEs comparing QIF approach to EB approach ($R = MSPE_{QIF}/MSPE_{EB}$) for random slope prediction in Model 2, when the true random effects were simulated from bivariate normal, symmetric mixture of normal and t distributions with difference correlations, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 7$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.96	0.938	0.862	1.01	1.03	1.07	0.512	1.10	1.24
0.4	1.02	1.04	0.969	1.06	1.14	1.18	1.04	1.10	1.22
0.8	1.25	1.16	1.15	1.18	1.17	1.28	0.941	1.01	1.11
	Symmetric mixture of normal distributions								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.02	0.904	0.865	0.988	1.05	1.09	0.971	1.16	1.26
0.4	1.82	1.02	0.989	1.08	1.12	1.18	0.597	1.09	1.19
0.8	1.66	1.18	1.11	1.13	1.21	1.25	0.680	1.03	1.13
	t distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.41	1.41	1.56	1.24	1.36	1.59	0.878	0.965	1.02
0.4	1.48	1.30	1.33	1.21	1.25	1.32	0.819	0.961	0.991
0.8	1.61	1.60	1.59	1.23	1.28	1.30	0.765	0.803	0.819

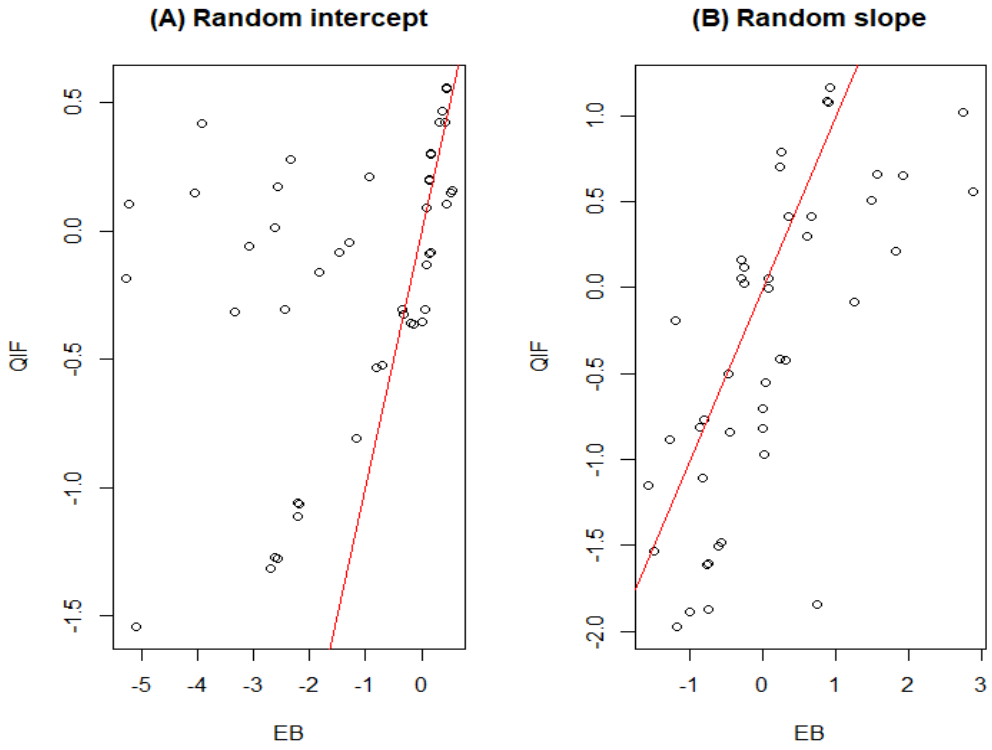
Table 5. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPE comparing EB approach to QIF approach ($R = MSPE_{EB}/MSPE_{QIF}$) for random intercept prediction in Model 2, when the true random effects were simulated from bivariate normal, symmetric mixture of normal and t distributions with difference correlations, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 4$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	3.59	2.41	2.32	2.52	2.50	2.52	0.700	1.04	1.09
0.4	3.69	2.68	2.68	2.60	2.58	2.66	0.703	0.963	0.993
0.8	3.35	3.14	3.00	2.70	2.81	2.79	0.806	0.895	0.928
	Symmetric mixture of normal distributions								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	2.82	2.72	2.30	2.45	2.48	2.57	0.871	0.912	1.12
0.4	3.84	2.83	2.58	2.71	2.71	2.61	0.706	0.959	1.01
0.8	2.96	3.08	2.89	2.69	2.91	2.81	0.908	0.944	0.972
	t distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	3.73	2.50	3.40	3.13	2.25	2.35	0.839	0.901	0.691
0.4	4.45	3.51	2.53	2.65	2.73	2.45	0.596	0.777	0.969
0.8	3.33	2.65	3.58	2.38	2.52	3.52	0.716	0.954	0.981

Table 6. Average MSPE ($MSPE_{EB}, MSPE_{QIF}$), and ratio of the average MSPE comparing EB approach to QIF approach ($R = MSPE_{EB}/MSPE_{QIF}$) for random slope prediction in Model 2, when the true random effects were simulated from bivariate normal, symmetric mixture of normal and t distributions with difference correlations, and the independent working correlation assumption. Averages summarized 100 simulated datasets. $n = 4$.

	$MSPE_{EB}$			$MSPE_{QIF}$			R		
	Normal distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.95	1.02	1.02	1.18	1.19	1.26	0.606	1.16	1.23
0.4	1.50	1.20	1.15	1.26	1.32	1.34	0.843	1.10	1.16
0.8	1.41	1.28	1.31	1.33	1.27	1.28	0.943	0.985	0.976
	Symmetric mixture of normal distributions								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.18	1.27	0.998	1.20	1.22	1.27	1.02	0.963	1.27
0.4	1.86	1.19	1.14	1.24	1.26	1.30	0.666	1.06	1.14
0.8	1.61	1.41	1.30	1.27	1.37	1.30	0.786	0.969	0.999
	t distribution								
Correlation	N=50	N=100	N=200	N=50	N=100	N=200	N=50	N=100	N=200
0	1.65	1.36	1.28	1.33	1.19	1.19	0.803	0.871	0.934
0.4	1.71	1.81	1.31	1.34	1.44	1.29	0.785	0.794	0.985
0.8	1.42	1.62	1.62	1.18	1.39	1.62	0.831	0.855	0.999

Figure 1. Scatterplots of random effects for schizophrenia study, random effects were predicted by QIF versus EB approaches under independent working correlation assumption. (A) Random intercept predictors. (B) Predictors of random coefficient of time. The solid line is the y=x line.



Chapter IV: A Graphical Approach to Assess the Goodness-of-Fit of Random Effects Linear Models When the Goal is to Measure Individual Benefits of Medical Treatments in Severely Ill Patients

Zhiwen Wang and Francisco J. Diaz*

ABSTRACT

Two-dimensional personalized medicine (2-PM) models are tools for measuring patients' individual benefits of medical treatments for chronic diseases that have potential applications in personalized medicine. These models assume normality for the distribution of random effects. It is necessary to examine the appropriateness of these assumption. We propose a graphical approach to assessing the goodness-of-fit of 2-PM models. Our approach plots the empirical quantiles of individual benefits predicted through an empirical Bayes approach versus the quantiles of the theoretical distribution of individual benefits derived under the assumption of normality for the random effects. We examine the performance of the proposed approach by computing Cramer-von Mises discrepancies between the empirical and theoretical distributions of individual benefits. Simulations showed that the proposed approach is sensitive to violations of the normality assumption and suggest that the approach is a useful tool to examine the goodness of the fit of 2-PM models.

KEY WORDS: Empirical Bayes, goodness-of-fit, personalized medicine models, quantiles, severity, treatment benefits

1. Introduction

Two-dimensional personalized medicine (2-PM) models are tools for measuring the severity of a patient's chronic disease and the individual benefits of medical or behavioral treatments (Diaz, 2019, 2016). The patient's disease severity at a specific time point is defined as the probability of missing the therapeutic target, and the individual benefit is therefore measured as the reduction in disease severity produced by the treatment. The severity and individual benefits are functions of known and unknown patient's characteristics. In practice, 2-PM models are built using linear models with random coefficients and severities and benefits are calculated with the random coefficients (Diaz, 2019, 2016).

Mixed effects models are efficient tools to build 2-PM models for understanding the individual time trajectory of treatment effects (Diaz et al., 2007, 2012a, 2012b; Diaz and de Leon, 2013; Diaz, 2016, 2019). In mixed effects models the distributional assumption of the unobserved random effects is important for estimation and inference, since the marginal likelihood function, which can be obtained by integrating out the random effects, depends on the assumed distribution of the random effects. Classically empirical Bayes (EB) approach is used to predict the random effects. In linear mixed effects models the EB predictors of the random effects are estimates of the best linear unbiased predictor (BLUP), which do not require the normality assumption for the random effects. Thus, a BLUP is robust to violations of the normality assumption. The prediction accuracy of BLUP for random effects was little affected by the distribution misspecification, as shown from both theoretical and numerical perspectives (Mcculloch and Neuhaus, 2011a). Traditionally, in mixed models, the unobserved random effects are assumed to follow normal distributions, for ease of computation. Violation of this normality assumption is possible. For example, a between-individual binary characteristic that is associated with the individual mean

profiles could be omitted. And a bimodal distribution may describe the induced variance. Although violation of this normality assumption has little to mild effect on maximum likelihood estimates (MLE) of the fixed effects, it may affect the random effects prediction by increasing bias of the variance components estimates, especially in generalized linear mixed effects models (Verbeke and Lesaffre, 1996, 1997; Verbeke and Molenberghs, 2000; Agresti et al., 2004; Litière et al., 2007, 2008). Therefore, for a particular dataset, it is crucial to assess the goodness-of-fit of the 2-PM models with respect to the distribution misspecification of the random effects.

Several graphical methods have been proposed for examining the goodness-of-fit of mixed models for longitudinal data. The most well-known graphical approach is based on conditional residuals which are computed with estimates of the BLUPs of the random effects (Gregoire et al., 1995). Verbeke and Molenberghs (2013) proposed a graphical diagnostic tool using gradient functions which checks the appropriateness of the random effects distribution assumption. Pan and Lin (2005) proposed graphical and numerical techniques to check the link function and functional forms of covariates through cumulative sums of residuals. Grady and Helms (1995) assess the fit of the assumed covariance structure, by plotting lagged covariances or correlations. Diaz et al. (2008) assessed goodness-of-fit of a random intercept model by plotting random-effect-adjusted observations based on EB predictors of random intercepts versus expected observations. Formal statistical tests have also been proposed. To check the normality assumption for the random effects, Efendi et al. (2017) used a bootstrap test based on gradient functions. Drikvandi et al. (2017) proposed a diagnostic test based on Cramer-von Mises discrepancies. Alonso et al. (2008) proposed test that use the eigenvalues of the variance-covariance matrices of fixed effects estimates obtained from robust inference methods. Similarly, for generalized linear mixed effects

models, two versions of diagnostic tests using information matrices were proposed in Abad et al. (2010).

This article proposes a graphical approach to assessing the goodness-of-fit of 2-PM models for continuous responses of severely ill patients. The approach compares the quantiles of the empirical Bayes estimates of individual treatment benefits of the patient sample against the theoretical quantiles of the distribution of individual benefits derived under the normality assumption for the random effects. We conducted a Monte Carlo simulation study that showed that the approach is sensitive to deviations from the normality assumption for the random effects. Specifically, the graphical approach captures the discrepancy between multivariate non-normal distributions for the random effects and normal distributions with the same mean and variance-covariance matrix.

This paper is organized as follows. The introduction of the conceptual and theoretical idea of the 2-PM models presents the functions of individual treatment benefits for continuous responses of severely ill patients. Then we present a motivation for the graphical approach and describe it in detail. Next, the approach is illustrated using data from a clinical trial of the antidepressant imipramine. An extensity simulation study was conducted with the presence of multivariate normal and non-normal distributions to evaluate the performance of the proposed graphical approach. Then we describe how Cramer-von Mises discrepancies are used to quantify deviations from the normality. The paper ends with a discussion and conclusions.

2. Methods

2.1. Individual severity and treatment benefits using time dependent 2-PM models

Time dependent 2-PM models allow understanding the evolution of individual treatment benefits over time (Diaz, 2019). Let Y be a continuous measure reflecting the patient's disease. Before a treatment Q is initiated, the responses for patient ω are measured $k_{0,\omega}$ times and modeled by

$$Y_{0,\omega,j} = \Lambda_\omega + \varepsilon_{0,\omega,j}, \qquad j = 1 \dots k_{0,\omega},$$

and after the treatment is initiated, the responses are measured $k_{1,\omega}$ times and modeled by

$$Y_{Q,\omega,j} = \Lambda_\omega + \beta_{Q,\omega,j} + \varepsilon'_{\omega,j}, \qquad j = 1 \dots k_{1,\omega},$$

where $\Lambda_\omega = \alpha_\omega + \boldsymbol{\lambda}^T \boldsymbol{X}_\omega$ and $\beta_{Q,\omega,j} = \theta_{1,\omega} t_{\omega,j} + \theta_{2,\omega} t_{\omega,j}^2 + \cdots + \theta_{d,\omega} t_{\omega,j}^d$. Here, $\boldsymbol{X}_\omega$ is a vector of subject characteristics that do not change during the trial. For patient ω, Λ_ω is a constant number that reflects the patient's disease state before treatment initiation and $\beta_{Q,\omega,j}$ is the individual (time-dependent) treatment effect after $t_{\omega,j}$ time units of treatment. We write $\beta_{Q,\omega}(t)$ in place of $\beta_{Q,\omega,j}$ to express the treatment effect at a generic time point t. We view Λ_ω and $\beta_{Q,\omega}(t)$ as individual realizations of population-level random variables Λ^* and $\beta_Q^*(t)$, respectively (Diaz, 2019). Also, $\varepsilon_{0,\omega,j}$ and $\varepsilon'_{\omega,j}$ represent measurement errors or within-subject variability due to patient's internal or external factors, assumed to be $N(0, \sigma_\varepsilon^2)$ and $N(0, \sigma_\varepsilon'^2)$, respectively.

Here, we assume that the therapeutic target is to achieve $Y \leq y$, where y is a prespecified value. Thus, a patient ω has baseline disease severity (Diaz, 2019, 2016)

$$s_{0,\omega} = 1 - \Phi\left(\frac{y - \Lambda_\omega}{\sigma_\varepsilon}\right)$$

where Φ is the cumulative distribution function of a standard normal distribution, and the patient's severity after a treatment duration t is

$$s_{2,\omega}(t) = 1 - \Phi\left(\frac{y - \Lambda_\omega - \beta_{Q,\omega}(t)}{\sigma_\varepsilon'}\right).$$

The patient's individual benefit from treatment Q at time point t is the reduction in disease severity

$$b_\omega(t) = s_{0,\omega} - s_{2,\omega}(t).$$

Here, we assume that all patients are severely ill, that is $s_{0,\omega} \approx 1$ for all ω. Thus, it is shown in Diaz (2019) that under the reasonable assumption that $\sigma_\varepsilon' \geq \sigma_\varepsilon$, the patient's benefit can be computed as

$$b_\omega(t) \approx \Phi\left(\frac{y - \Lambda_\omega - \beta_{Q,\omega}(t)}{\sigma_\varepsilon'}\right) \qquad (1)$$

In the following, we assume $\sigma_\varepsilon' = \sigma_\varepsilon$ which is usually clinically reasonable. Here, α_ω and $\theta_{1,\omega}, \ldots, \theta_{d,\omega}$ are characteristic constants of patient ω that are viewed as realizations of random coefficients α^* and $\theta_1^*, \ldots, \theta_d^*$ that do not necessarily have mean 0. In the terminology of mixed effects models, $E(\alpha^*)$, $\boldsymbol{\lambda}$ and $E(\theta_1^*), \ldots, E(\theta_d^*)$ are the fixed effects, and $\alpha^* - E(\alpha^*)$ and $\theta_i^* - E(\theta_i^*)$, $i = 1, \ldots, d$, are the random effects which are usually assumed to be normally distributed. Here, we propose a graphical method to examine the assumption of normality.

2.2. Quantiles of individual benefits under the normality assumption

Under the assumption of normality for the random effects, since the patients are severely ill, the cumulative distribution function of individual benefits for patients with covariate value $\boldsymbol{X} = \boldsymbol{x}$ at time t is (Diaz, 2019)

$$F(z) = F(z; \boldsymbol{x}, t) = \Phi\left(\frac{\Phi^{-1}(z) - \mu}{\gamma}\right), \quad 0 < z < 1, \qquad (2)$$

where $\mu = \mu(\boldsymbol{x}, t) = \frac{y - E\left(\Lambda^* + \beta_Q^*(t)\right)}{\sigma_\varepsilon'}$ and $\gamma^2 = \gamma^2(t) = \frac{\mathrm{Var}\left(\Lambda^* + \beta_Q^*(t)\right)}{\sigma_\varepsilon'^2}$. Further, the p-th quantile of the probability distribution function of individual treatment benefits is (Diaz, 2019)

$$B(p, \boldsymbol{x}, t) = \Phi(\gamma \Phi^{-1}(p) + \mu), \qquad 0 < p < 1. \qquad (3)$$

The quantities in (2) and (3) are functions of treatment duration t, since μ and γ are. They are also functions of the fixed effects and the variance components (i.e., the variances and covariances of the random effects and the error variance).

2.3. A motivation for the proposed graphical approach

Here, we estimate (predict) the individual treatment benefits using the EB approach described in Diaz (2019, 2016). The EB predictors of individual treatment benefits are obtained by replacing the fixed effects, error variance and individual random effects in Equation (1) with their estimates or predictors. The fixed effects and variance components are usually estimated through maximum or restricted maximum likelihood (Verbeke and Molenberghs, 2000; Hedeker and Gibbons, 2006; McCulloch et al., 2008; Fitzmaurice et al., 2011). We predict the random effects following an EB approach (Diaz, 2019, 2016; Diaz et al., 2012b; Frees, 2004; Verbeke and Lesaffre, 1996; Verbeke and Molenberghs, 2000). Importantly, the EB predictors of random

effects are estimates of the best linear unbiased predictors (BLUPs) which do not assume normality for the random effects (Frees, 2004). Moreover, the EB predictors of random effects are relatively robust to violations of the normality assumption (McCulloch and Neuhaus, 2011a; 2011b). Because of this, the sample quantiles of the estimated individual benefits can be viewed as robust estimates of the quantiles of the probability distribution of individual benefits. Alternatively, the quantiles can be directly estimated by replacing the fixed effects and variance components in Equation (3) with their corresponding estimates. Therefore, if the normality assumption is violated, we expect the quantiles estimated with Equation (3) to be substantially different from the sample quantiles based on the BLUPs because Equation (3) was derived under the assumption of normality. Thus, we propose to compare the sample quantiles based on the BLUPs with the quantiles calculated with Equation (3) in order to evaluate the assumption of normality for the random effects.

2.4. Goodness-of-fit plot

Suppose the sample of patients can be divided into G subgroups. This is possible, for instance, when the subject characteristics are categorical or, if a characteristic is continuous when it is split into categories based on published cut-off values or percentiles. Therefore, we assume that $\boldsymbol{X}_\omega$ includes only binary (dummy) covariates and that $\boldsymbol{X}_\omega$ has G distinct possible values $\boldsymbol{x}_1, \dots, \boldsymbol{x}_G$. Let N_g be the number of patients in the subpopulation of patients for whom $\boldsymbol{X}_\omega = \boldsymbol{x}_g$, and let $N = \sum_{g=1}^{G} N_g$ be the total number of patients. Let $\hat{b}_{g,1}, \dots, \hat{b}_{g,N_g}$ be the EB predicted individual benefits for the N_g patients in group g, and $\hat{b}_{g,(1)} < \hat{b}_{g,(2)} < \cdots < \hat{b}_{g,(N_g)}$ be the

corresponding order statistics. For a particular time t, a benefit quantile-quantile (BQQ) plot consists of plotting in an x-y plane the N points

$$\left(\hat{b}_{g,(\omega)}, \hat{B}\left(\frac{\omega - 0.5}{N_g}, \boldsymbol{x}_g, t\right)\right), \omega = 1, \ldots, N_g, g = 1, \ldots, G.$$

where $\hat{B}$ is obtained by replacing fixed effects and variance components in Equation (3) with their maximum likelihood or restricted maximum likelihood estimators (RMLEs). Thus, a BQQ plot compares the sample quantiles of individual benefits predicted with the EB approach versus estimates of the theoretical quantiles derived under the normality assumption for the random effects. In practice, we use the maximum time point available in the dataset as a value for t.

If the points on the BQQ plot do not deviate asymmetrically much about the $y = x$ line, then we conclude that the normality assumption for the random effects of the 2-PM model is appropriate. If, in addition, the usual conditional residuals for the model suggest normality for the errors and do not show apparent outliers, we can have reasonable confidence in the EB predictors of the individual benefits computed with Equation (1).

3. Application to depression study

As an illustration, clinical trial data of 66 patients under imipramine treatment with two types of depression diagnosis were used (Reisby et al., 1977). The diagnosis was endogenous (N=37) and non-endogenous (N=29). The data is available in Hedeker and Gibbons (2006) and was also analyzed by Diaz (2019). The response variable, the Hamilton Rating Scale (HRS) for depression, was recorded at the beginning and end of the week before imipramine treatment

initiation and at the end of each of the next four weeks during treatment. Diaz (2019) fitted a random effects linear model of the HRS scores in order to predict individual treatment benefits but did not provide evidence for the model's goodness-of-fit. As covariates, the model included a polynomial time trend and diagnosis (1=endogenous, 0=nonendogenous). The intercept and the linear and quadratic terms had random effects in addition to the fixed effects. The SAS procedure MIXED, which assumes normally distributed random effects, was used to obtain MLE of the fixed effects and EB predictors of the random effects for all patients (SAS Institute Inc. Cary, NC). Similar to Diaz (2019), an unstructured covariance matrix for the random effects and homoscedastic independent errors were assumed. Parameter estimates are shown in Table 1 in Diaz (2019). The Figure S1 in the Supplementary Material indicates that for the depression data, the normality assumption on the random errors of the linear mixed effects models are satisfied.

Histograms and kernel densities of EB predictors of the random effects are shown in Supplementary Figure S2. The shapes of the histograms seem to suggest approximate normality for the distribution of the random effects. However, Verbeke and Lesaffre (1997) and Mcculloch and Neuhaus (2011b) have found that the shape of the histograms for EB predictors may be misleading and may not reflect the true distribution of the random effects. Equation (3) was used to calculate some $p \times 100\%$ percentiles of the individual benefits of imipramine as functions of treatment duration. Figure 1 shows estimated $p \times 100\%$ theoretical percentiles computed from Equation (3) for the individual benefits of nonendogenous patients (A) and the corresponding sample percentiles of the EB predicted benefits (B) for some values of p, as functions of treatment duration. Analogous figures for endogenous patients are Figure S3 in the Supplementary Material. In general, the sample percentiles seem to match the estimated theoretical percentiles. However, a better comparison of the theoretical and sample percentile functions is provided by the BQQ plot,

which is shown in Figure 2. The points in the plot are scattered around close to the $y = x$ line, without exhibiting any asymmetric deviations. Thus, the sample percentiles of individual benefits match closely the theoretical percentiles derived under the normality assumption, suggesting the adequacy of this assumption and the goodness-of-fit of the 2-PM model for the imipramine data.

We also applied the diagnostic tests proposed by Alonso et al. (2008) on this dataset. The null hypothesis is that the normality assumption for the random effects is reasonable. The two determinant tests and the determinant-trace test yielded the test statistics $\delta_{d1} = 2.9$, $\delta_{d2} = 1.19$ and $\delta_{d3} = 1.28$, with corresponding p-values of 0.085, 0.276 and 0.258. All three p-values were larger than the chosen 0.05 significant level. These tests suggest that the normality assumption for the random effects was appropriate and further validate the conclusions from our proposed graphical approach.

4. Simulation study

We conducted a simulation study to assess the performance of BQQ plots under violations of the normality assumption for the random effects. Motivated by the application study in Diaz (2019), data from the following two models were simulated:

Model 1: (Random intercept and random slope for time).

$$Y'_{\omega,j} = \psi_0 + \psi_1 x_\omega + \psi_2 t_{\omega,j} + \psi_3 t^2_{\omega,j} + \tau_{0,\omega} + \tau_{2,\omega} t_{\omega,j} + \varepsilon_{\omega,j},$$
$$\omega = 1, \dots, N, \;\; j = 1, \dots, n, \qquad (4)$$

such that $\Lambda_\omega = \psi_0 + \tau_{0,\omega} + \psi_1 x_\omega$ and $\beta_{Q,\omega}(t) = (\psi_2 + \tau_{2,\omega})t + \psi_3 t^2$, N is the number of patients and n is the number of observations per patient. Here, $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)^T$ are the fixed

effects and $\boldsymbol{\tau}_\omega = \left(\tau_{0,\omega}, \tau_{2,\omega}\right)^T$ are the random effects with mean 0. Moreover, $x_\omega \sim \text{Bernoulli}(0.6)$ represents a patient's time-independent characteristic (for instance, gender, smoking, etc.) and the $\varepsilon_{\omega,j}$'s are mutually independent with $\varepsilon_{\omega,j} \sim N(0, \sigma_\varepsilon^2 = 10)$.

Model 2: (Random intercept and random slopes for time and time square):

$$Y'_{\omega,j} = \psi_0 + \psi_1 x_\omega + \psi_2 t_{\omega,j} + \psi_3 t_{\omega,j}^2 + \tau_{0,\omega} + \tau_{2,\omega} t_{\omega,j} + \tau_{3,\omega} t_{\omega,j}^2 + \varepsilon_{\omega,j}, \quad (5)$$

$$\omega = 1, \ldots, N, j = 1, \ldots, n,$$

such that $\Lambda_\omega = \psi_0 + \tau_{0,\omega} + \psi_1 x_\omega$ and $\beta_{Q,\omega}(t) = \left(\psi_2 + \tau_{2,\omega}\right)t + \left(\psi_3 + \tau_{3,\omega}\right)t^2$. In this case, $\boldsymbol{\tau}_\omega = \left(\tau_{0,\omega}, \tau_{2,\omega}, \tau_{3,\omega}\right)^T$ are the random effects with mean 0. No missing responses were assumed. An unstructured variance-covariance matrix for the random effects was assumed for both models (Fitzmaurice et al., 2011, 2009).

Varying values for N were used and $n = 4$ or 6. For either model, we simulated 2 baseline measurements and 2 or 4 measurements under medical treatment. Thus, for all models, $k_{0,\omega} = 2$, and $t_{\omega,1} = t_{\omega,2} = 0$. When $n = 4$, $k_{1,\omega} = 2$, $t_{\omega,3} = 1$ and $t_{\omega,4} = 4$; and when $n = 6$, $k_{1,\omega} = 4$, $t_{\omega,3} = 1$, $t_{\omega,4} = 2$, $t_{\omega,5} = 3$ and $t_{\omega,6} = 4$. For all models, $Y_{0,\omega,j} = Y'_{\omega,j}$ for $j = 1, 2$, and $Y_{Q,\omega,j} = Y'_{\omega,j+2}$ for $j = 1, \ldots, k_{1,\omega}$.

The therapeutic target was to achieve $Y \leq y$ with $y = 7$. The MLEs of $\boldsymbol{\psi}$ and σ_ε^2 are denoted by $\widehat{\boldsymbol{\psi}} = \left(\hat{\psi}_0, \hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3\right)^T$ and $\hat{\sigma}_\varepsilon^2$; and the EB predictor of $\boldsymbol{\tau}_\omega$ by $\hat{\boldsymbol{\tau}}_\omega = \left(\hat{\tau}_{0,\omega}, \hat{\tau}_{2,\omega}\right)^T$ for Model 1 or $\hat{\boldsymbol{\tau}}_\omega = \left(\hat{\tau}_{0,\omega}, \hat{\tau}_{2,\omega}, \hat{\tau}_{3,\omega}\right)^T$ for Model 2. Here, we will investigate BQQ plots computed at the last time point, namely $t = 4$. We used Equation (1) to predict the individual benefits after

replacing σ_ε, Λ_ω and $\beta_{Q,\omega}(t)$ with their estimates $\hat{\sigma}_\varepsilon$, $\widehat{\Lambda}_\omega = \hat{\psi}_0 + \hat{\tau}_{0,\omega} + \hat{\psi}_1 x_\omega$ and $\hat{\beta}_{Q,\omega}(t) = (\hat{\psi}_2 + \hat{\tau}_{2,\omega})t + \hat{\psi}_3 t^2$ for Model 1 or $\hat{\beta}_{Q,\omega}(t) = (\hat{\psi}_2 + \hat{\tau}_{2,\omega})t + (\hat{\psi}_3 + \hat{\tau}_{3,\omega})t^2$ for Model 2.

Let $\Sigma_{i,j}$ be the (i,j)-th entry of the variance covariance matrix of $\boldsymbol{\tau}_\omega$ and $\hat{\Sigma}_{i,j}$ be its maximum likelihood estimator. Thus, $\Sigma_{i,j}$ is of dimension 2×2 for Model 1, and 3×3 for Model 2. The μ in Equation (3) is estimated with $\hat{\mu} = \left(y - (\hat{\psi}_0 + \hat{\psi}_1 x_\omega + \hat{\psi}_2 t + \hat{\psi}_3 t^2)\right)/\hat{\sigma}_\varepsilon$ for both models, whereas γ is estimated with $\hat{\gamma}^2 = (\hat{\Sigma}_{1,1} + t^2\hat{\Sigma}_{2,2} + 2t\hat{\Sigma}_{1,2})/\hat{\sigma}_\varepsilon^2$ for Model 1, and $\hat{\gamma}^2 = (\hat{\Sigma}_{1,1} + t^2\hat{\Sigma}_{2,2} + t^4\hat{\Sigma}_{3,3} + 2t\hat{\Sigma}_{1,2} + 2t^2\hat{\Sigma}_{1,3} + 2t^3\hat{\Sigma}_{2,3})/\hat{\sigma}_\varepsilon^2$ for Model 2.

Table 1 shows the "true" fixed effects employed in simulations. These were chosen so that the majority of the patients were severely ill under all examined non-normal and normal distributions for the random effects, i.e., $P(s_{0,\omega} > 0.9) \geq 0.95$.

4.1. Simulation of random effects

We implemented four simulation scenarios to represent situations in which the normality assumption for the random effects is violated (Table 1). For comparison purposes, in each scenario, $\boldsymbol{\tau}_\omega$ was simulated from both a non-normal distribution and a reference normal distribution with the same mean and variance-covariance matrix.

4.1.1. Scenario 1: Model 1 with a symmetric mixture of two bivariate normal distributions

Here, we explore the effect on the BQQ plot of the distance between the means of the two components of a mixture of normal distributions for $N \in \{50, 100, 150, 200, 300, 500\}$. The true $\boldsymbol{\tau}_\omega$ was distributed as

$$\boldsymbol{\tau}_{\omega} \sim \frac{1}{2} N(\boldsymbol{m}_1^* = w * \boldsymbol{m}_1, V) + \frac{1}{2} N(\boldsymbol{m}_2^* = w * \boldsymbol{m}_2, V)$$

where $\boldsymbol{m}_1 = (0, -1)^{\mathrm{T}}$, $\boldsymbol{m}_2 = (0, 1)^{\mathrm{T}}$ and $V = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. The distance between the mean vectors is $w\sqrt{(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T \times (\boldsymbol{m}_1 - \boldsymbol{m}_2)}$. We examined $w \in \{1, 2, 3, 4, 5\}$. The reference normal distribution with the same mean and covariance matrix was $N(\boldsymbol{m}, D^*)$, where $\boldsymbol{m} = (0, 0)^{\mathrm{T}}$ and $D^* = \frac{1}{2}\boldsymbol{m}_1^* \boldsymbol{m}_1^{*T} + \frac{1}{2}\boldsymbol{m}_2^* \boldsymbol{m}_2^{*T} + V$. Here, a greater distance between the means of the two component distributions represents a greater deviation from normality. Thus, we expect the BQQ plot to show greater departures from the diagonal line (Figures 3, 7; Table S1).

4.1.2. Scenario 2: Model 1 with an asymmetric mixture of two bivariate normal distributions for the random effects

Here, we explore how the variance of the components of a mixture of normal distributions affects the BQQ plot, for sample sizes $N \in \{60, 100, 160, 200, 300, 500\}$. The true random effects vector $\boldsymbol{\tau}_{\omega}$ was distributed as

$$\boldsymbol{\tau}_{\omega} \sim \frac{3}{4} N(\boldsymbol{m}_1, V) + \frac{1}{4} N(\boldsymbol{m}_2, V)$$

where $\boldsymbol{m}_1 = (0, -1)^{\mathrm{T}}$, $\boldsymbol{m}_2 = (0, 3)^{\mathrm{T}}$ and $V = \begin{bmatrix} \sigma_1^2 & 0.9 \\ 0.9 & \sigma_2^2 \end{bmatrix}$. We examined $\sigma_1^2 = \sigma_2^2 \in \{1, 2, 3, 4, 5\}$. In this case, the overall mean and variance are $\boldsymbol{m} = (0, 0)^{\mathrm{T}}$ and $D^* = \frac{3}{4}\boldsymbol{m}_1 \boldsymbol{m}_1^T + \frac{1}{4}\boldsymbol{m}_2 \boldsymbol{m}_2^T + V$. Thus, for comparison purposes, $\boldsymbol{\tau}_{\omega}$ was also simulated from the reference bivariate $N(\boldsymbol{m}, D^*)$. Here, since the mean vectors are fixed, a greater variance for the components of the mixture imply a "less bimodal" distribution. Therefore, we expect BQQ plots for the non-normal

cases to be more like their corresponding reference normal cases when the variance of the components is large (Figures 4, 8; Table S2).

4.1.3. Scenario 3: Model 2 with a trivariate t distribution for the random effects

Here, the true random effects were simulated from a trivariate t distribution with degrees of freedom $v \in \{3, 5, 7, 9, 11, 13\}$, location parameters $\boldsymbol{m} = (0, 0, 0)^{\mathrm{T}}$, and shape parameter Γ given in Table 1. The purpose was to study how BQQ plots are affected by the heaviness of the tails of the t distribution, using $N \in \{50, 100, 150, 200, 300, 500\}$. The reference normal distribution with the same mean and variance-covariance matrix was $N\left(\boldsymbol{m}, D^* = \frac{v}{v-2} \times \Gamma\right)$ (Anderson, 2003). In this scenario, smaller degrees of freedom are associated with heavier tails for the distribution of random effects. Thus, we expect BQQ plots for non-normal cases to resemble more the reference normal plots when v is large (Figures 5, 9; Table S3).

4.1.4. Scenario 4: Model 2 with a symmetric mixture of two trivariate normal distributions

This scenario is analogous to Scenario 1, except that we used trivariate normal distributions for the components of the mixture. The goal was also to examine the effect of the distance between the means of the two normal components on BQQ plots. Since a greater distance represents a greater deviation from normality, we expect the BQQ plots to show greater departures from the diagonal line (Figures 6, 10; Table S4).

4.2. Cramer von Mises discrepancy statistic

We used Cramer-von Mises discrepancy statistic (CVM) to quantify the deviation of the BQQ plot from the y=x line under violations of the normality assumption (Anderson, 1962; Darling, 1957). Let $F_{N_g}(x)$ be the empirical distribution function of $\hat{b}_{g,1}, \ldots, \hat{b}_{g,N_g}$ and denote $U_{g,1} = F\left(\hat{b}_{g,1}\right), \ldots, U_{g,N_g} = F\left(\hat{b}_{g,N_g}\right)$. The CVM discrepancy between $F_{N_g}(z)$ and $F(z; \boldsymbol{x}_g)$ was computed as

$$\Omega_g = \int_{-\infty}^{+\infty} \left\{F_{N_g}(z) - F\left(z; \boldsymbol{x}_g\right)\right\}^2 dF(z; \boldsymbol{x}_g) = \frac{1}{12N_g^2} + \frac{1}{N_g}\sum_{k=1}^{N_g}\left(U_{g,k} - \frac{2k-1}{2N_g}\right)^2$$

(Anderson, 1962; CSöRgő and Faraway, 1996; Darling, 1957). The overall discrepancy was computed as the weighted average

$$\overline{\Omega} = \frac{\sum_{g=1}^{G} N_g\, \Omega_g}{N}.$$

Larger values of $\overline{\Omega}$ reflect more severe violations of the normality assumption for the random effects.

Five hundred datasets were simulated for each combination of values of N, n and random-effects distribution parameters. For illustration purposes, selected BQQ plots are presented for $N = 100$ and $n = 6$ (Figures 3-6). These plots corresponded to the datasets producing the closest $\overline{\Omega}$ to $\overline{\overline{\Omega}}$, where $\overline{\overline{\Omega}}$ is the average of the 500 values of $\overline{\Omega}$.

To examine the sensitivity of BQQ plots to detect deviations from normality, each simulated non-normal case was compared with its corresponding reference normal distribution by using the ratio $R = \frac{\overline{\overline{\Omega}}_{\text{non-normal}}}{\overline{\overline{\Omega}}_{\text{normal}}}$ (Figures 7-10). On average, we expect the $\overline{\Omega}$ obtained from a non-normal case to be larger than that from its reference normal case and, therefore, $R > 1$. This is

because the Ω_g measures the discrepancy between the empirical distribution of the sample individual benefits and the theoretical distribution obtained under the normality assumption for the random effects. We expect larger values of R to be associated with greater deviations from normality. The SAS procedures MIXED and IML were used to implement the simulations (SAS Institute Inc. Cary, NC).

5. Simulation results

5.1. Scenario 1: symmetric mixtures of bivariate normal distributions

As expected, larger distances between the two components of the mixture distribution determined more apparent asymmetric departures of the points on the BQQ plot from the $y=x$ line (Figure 3). By comparison, the BQQ plots for data simulated from the corresponding reference normal distributions did not show any asymmetric deviations from the diagonal line. Figure 7 shows that the R ratios comparing CVM discrepancies between no-normal versus comparable normal distributions were always > 1 and increased with the distance between the components of the mixture. In general, R increased with both the number of patients N and the number of repeated measures n, suggesting that the sample size contributes positively to the sensitivity of BQQ plots. Table S1 in the Supplementary Material shows the average CVM discrepancies $\overline{\overline{\Omega}}$ for all cases of Scenario 1.

5.2. Scenario 2: asymmetric mixtures of two bivariate normal distributions

For the investigated mixtures of normal distributions, quantiles of EB benefits from a patient sample tended to be larger than their corresponding theoretical quantiles that assume normality for the random effects (Figure 4). Moreover, this pattern was more apparent with smaller variances for the components of the mixture. The pattern was not observed in the BQQ plots corresponding to the reference normal distributions. Figure 8 shows that the ratios comparing CVM discrepancies of non-normal to reference normal distributions decreased with the variance of the mixture components, suggesting that BQQ plots are sensitive to deviations from normality. The ratios also increased with sample size N and the number of repeated measures n, suggesting that larger sample sizes increase the likelihood that BQQ plots capture normality violations. Table S2 in the Supplementary Material shows average CVM discrepancies for the non-normal and normal cases.

5.3. Scenario 3: trivariate t distribution

As the degrees of freedom increased, the BQQ plots for data simulated with the t distribution became more similar to the BQQ plots for data simulated with a comparable normal distribution (Figure 5). The theoretical quantiles of individual benefits under the normality assumption tended to be larger than the quantiles for EB sample benefits when the tails of the t distribution became heavier. Figure 9 shows that the ratios R comparing CVM discrepancies under the t distribution versus the reference normal distribution increased as the degrees of freedom decreased, suggesting that BQQ plots can reliably capture the heaviness of the tails of the t distribution. The ratios tended to increase as both N and n increased, implying that the larger the sample size is the more powerful the proposed graphical approach is for detecting tail heaviness.

Table S3 in the Supplementary Material shows average CVM discrepancies $\overline{\overline{\Omega}}$ for the non-normal and normal cases.

5.4. Scenario 4: symmetric mixture of two trivariate normal distributions

Analogous to scenario 1, marked departures in the appearance of BQQ plots from what is expected under comparable normal distributions are observed when the random effects are distributed as a mixture of normal distributions (Figure 6). Greater distances between the two mean vectors of the mixture components tended to be associated with larger asymmetric deviations from the $y = x$ line. This trend can also be inferred from Figure 10, which shows that, compared with the reference normal distribution, CVM discrepancies under a mixture of normal distributions increased as the distance between the mixture components increased. Table S4 in Supplementary Material shows average CVM discrepancies for the non-normal and normal cases.

6. Discussion and conclusions

This paper proposes a graphical approach to evaluate the goodness-of-fit of random effects models for continuous responses when the purpose of the model is to estimated individual benefits of medical treatments. In our approach, empirical quantiles of individual benefits estimated with an empirical Bayes approach are plotted against the quantiles of the distribution of individual benefits calculated under a normality assumption on the random effects. The rationale underlying our approach is that EB estimates of the random effects are robust to violations of the normality

assumption (McCulloch and Neuhaus, 2011a; 2011b). In fact, they are also estimates of the BLUPs whose optimality properties do not require the normality assumption (Frees, 2004).

If the normality assumption is valid, we expect empirical quantiles to be close to the theoretical quantiles. Thus, we can infer the goodness-of-fit of the 2-PM model if the BQQ plot does not show obvious asymmetric deviations from the $y = x$ diagonal line. We evaluated the performance of this approach using CVM discrepancies which measure the discrepancy between an empirical and a theoretical probability distribution. CVM discrepancies confirmed that our graphical approach captures accurately deviations from the normality assumption. Importantly, we found that the ratios R of average CVM discrepancies ($\overline{\overline{\Omega}}$), which compared non-normal distributions with closely comparable normal distributions, were not smaller than 1 in all simulations. This suggests that BQQ plots are powerful tools to detect deviations from normality for the distribution of the random effects in 2-PM models. Our simulations also showed that larger sample sizes give greater sensitivity to BQQ plots for detecting non-normality. Relatively moderate sample sizes, however, were enough to detect moderate deviations from normality.

Normal quantile-quantile (QQ) plots are routinely used in statistical practice to examine the assumption of normality for a variety of models (Aldor-Noiman et al., 2013; García Ben and Yohai, 2004; Stine, 2017). For mixed models, it is frequent practice to explore the normality assumption for the random effects by plotting separate normal QQ plots for each random effect in the model. In this alternative approach, the EB predictors of the random effects are used to compute both the empirical distribution of the random effects and the mean and variance of the normal distribution used to obtain the theoretical quantiles. This circularity limits the interpretation of the resultant plots because the EB predictors estimates themselves. Moreover, research shows that the shape of the empirical distribution of the EB predictors of the random effects does not necessarily

reflect the random-effects distribution (Verbeke and Lesaffre, 1997; Mcculloch and Neuhaus, 2011b). Thus, direct normal QQ plots calculated with only EB predictors of the random effects may be misleading as a tool to examine normality. In contrast, in our proposed approach, the theoretical quantiles given in Eq. (3) are estimated directly using the MLEs or RMLEs of model parameters without the mediation of the EB predictors. In addition, our simulations show that BQQ plots are reliable. Thus, if the model will be used to make decisions related with personalized medicine, BQQ plots are recommendable as a complementary tool for the exploration of the normality assumption for the random effects.

A limitation of our approach is that it requires that continuous or ordinal covariates be categorized before implementing Eq (3). Future research must examine how to incorporate continuous covariates to BQQ plots. The extension of BQQ plots to 2-PM models with non-continuous responses also needs further research.

References

Abad, A.A., Litière, S., Molenberghs, G., 2010. Testing for misspecification in generalized linear mixed models. Biostatistics 11, 771–786.

Agresti, A., Caffo, B., Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. Comput. Stat. Data Anal. 47, 639–653.

Aldor-Noiman, S., Brown, L.D., Buja, A., Rolke, W., Stine, R.A., 2013. The power to see: A new graphical test of normality. Am. Stat. 67, 249–260.

Alonso, A., Litière, S., Molenberghs, G., 2008. A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. Comput. Stat. Data Anal. 52, 4474–4486.

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis, 3rd ed. Wiley-Interscience.

Anderson, T.W., 1962. On the Distribution of the Two-Sample Cramer-von Mises Criterion. Ann. Math. Stat. 33, 1148–1159.

CSöRgő, S., Faraway, J.J., 1996. The Exact and Asymptotic Distributions of Cramér-Von Mises Statistics. J. R. Stat. Soc. Ser. B 58, 221–234.

Darling, D.A., 1957. The Kolmogorov-Smirnov, Cramer-von Mises Tests. Ann. Math. Stat. 28, 823–838.

Diaz, F.J., 2019. Estimating individual benefits of medical or behavioral treatments in severely

ill patients. Stat. Methods Med. Res. 28, 911–927.

Diaz, F.J., 2016. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat. Med. 35, 4077–4092.

Diaz, F.J., Cogollo, M.R., Spina, E., Santoro, V., Rendon, D.M., de Leon, J., 2012a. Drug dosage individualization based on a random-effects linear model. J. Biopharm. Stat. 22, 463–484.

Diaz, F.J., de Leon, J., 2013. The Mathematics of Drug Dose Individualization Should Be Built With Random-Effects Linear Models. Ther. Drug Monit. 35, 276–277.

Diaz, F.J., Rivera, T.E., Josiassen, R.C., de Leon, J., 2007. Individualizing drug dosage by using a random intercept linear model. Stat. Med. 26, 2052–2073.

Diaz, F.J., Santoro, V., Spina, E., Cogollo, M., Rivera, T.E., Botts, S., De Leon, J., 2008. Estimating the size of the effects of co-medications on plasma clozapine concentrations using a model that controls for clozapine doses and confounding variables. Pharmacopsychiatry 41, 81–91.

Diaz, F.J., Yeh, H.W., de Leon, J., 2012b. Role of Statistical Random-Effects Linear Models in Personalized Medicine. Curr. Pharmacogenomics Person. Med. 10, 22–32.

Drikvandi, R., Verbeke, G., Molenberghs, G., 2017. Diagnosing misspecification of the random-effects distribution in mixed models. Biometrics 73, 63–71.

Efendi, A., Drikvandi, R., Verbeke, G., Molenberghs, G., 2017. A goodness-of-fit test for the random-effects distribution in mixed models. Stat. Methods Med. Res. 26, 970–983.

Fitzmaurice, G.M., Davidian, M., Verbeke, G., Molenberghs, G., 2009. Longitudinal data

analysis, Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Boca Raton.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2011. Applied longitudinal analysis, 2nd ed, Wiley series in probability and statistics. Wiley, Hoboken, N.J.

Frees, E.W., 2004. Longitudinal and panel data : analysis and applications in the social sciences. Cambridge University Press, Cambridge, UK ; New York.

García Ben, M., Yohai, V.J., 2004. Quantile-quantile plot for deviance residuals in the generalized linear model. J. Comput. Graph. Stat. 13, 36–47.

Grady, J.J., Helms, R.W., 1995. Model selection techniques for the covariance matrix for incomplete longitudinal data. Stat. Med. 14, 1397–1416.

Gregoire, T.G., Schabenberger, O., Barrett, J.P., 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Can. J. For. Res. 25, 137–156.

Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. Wiley-Interscience, Hoboken, N.J.

Litière, S., Alonso, A., Molenberghs, G., 2008. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. Stat. Med. 27, 3125–3144.

Litière, S., Alonso, A., Molenberghs, G., 2007. Type I and type II error unde random-effects misspecification in generalized linear mixed models. Biometrics 63, 1038–1044.

McCulloch, C., Searle, S., Neuhaus, J., 2008. Generalized, Linear, and Mixed Models, 2nd ed. Wiley Series in Probability and Statistics, Hoboken, N.J.

Mcculloch, C.E., Neuhaus, J.M., 2011a. Prediction of Random Effects in Linear and Generalized Linear Models under Model Misspecification. Biometrics 67, 270–279.

McCulloch, C.E., Neuhaus, J.M., 2011b. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Stat. Sci. 26, 388–402.

Pan, Z.Y., Lin, D.Y., 2005. Goodness-of-fit methods for generalized linear mixed models. Biometrics 61, 1000–1009.

Reisby, N., Gram, L.F., Bech, P., Nagy, A., Petersen, G.O., Ortmann, J., Ibsen, I., Dencker, S.J., Jacobsen, O., Krautwald, O., Søndergaard, I., Christiansen, J., 1977. Imipramine: Clinical effects and pharmacokinetic variability. Psychopharmacology (Berl). 54, 263–272.

Stine, R.A., 2017. Explaining Normal Quantile-Quantile Plots Through Animation: The Water-Filling Analogy. Am. Stat. 71, 145–147.

Verbeke, G., Lesaffre, E., 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Comput. Stat. Data Anal. 23, 541–556.

Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. J. Am. Stat. Assoc. 91, 217–221.

Verbeke, G., Molenberghs, G., 2013. The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. Biostatistics 14, 477–490.

Verbeke, G., Molenberghs, G., 2000. Linear mixed models for longitudinal data, Springer series

in statistics. Springer, New York.

Table 1. Summary of simulation scenarios for evaluating the performance of BQQ plots.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Model:	1 (Eq. 4)	1 (Eq. 4)	2 (Eq. 5)	2 (Eq. 5)
Number of patients (N)	$\{50, 100, 150, 200, 300, 500\}$	$\{60, 100, 160, 200, 300, 500\}$	$\{50, 100, 150, 200, 300, 500\}$	$\{50, 100, 150, 200, 300, 500\}$
# of measurements per patient, $n = k_{0,\omega} + k_{1,\omega}$	When $k_{1,\omega} = 2$, $n = 4$; when $k_{1,\omega} = 4$, $n = 6$.			
Binary covariate	$x_\omega \sim$ Bernoulli(0.6)			
Measurement errors	i. i. d., $\varepsilon_{\omega,j} \sim N(0, \sigma_\varepsilon^2 = 10)$			
Fixed effects ($\boldsymbol{\psi}$)	$(21, 2, -5, 0.5)^T$	$(21, 2, -5, 0.5)^T$	$(21.4, 1.92, -3.97, 0.35)^T$	$(24, 1.92, 0.97, -0.35)^T$
Non-normal Random effects	$\boldsymbol{\tau}_\omega \sim \frac{1}{2}N(\boldsymbol{m}_1^* = w \times \boldsymbol{m}_1, V) + \frac{1}{2}N(\boldsymbol{m}_2^* = w \times \boldsymbol{m}_2, V)$ $\boldsymbol{m}_1 = (0, -1)^T,\ \boldsymbol{m}_2 = (0, 1)^T$ $V = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ $w \in \{1, 2, 3, 4, 5\}$	$\boldsymbol{\tau}_\omega \sim \frac{3}{4}N(\boldsymbol{m}_1, V) + \frac{1}{4}N(\boldsymbol{m}_2, V)$ $\boldsymbol{m}_1 = (0, -1)^T,\ \boldsymbol{m}_2 = (0, 3)^T$ $V = \begin{bmatrix} \sigma_1^2 & 0.9 \\ 0.9 & \sigma_2^2 \end{bmatrix}$ $\sigma_1^2 = \sigma_2^2 \in \{1, 2, 3, 4, 5\}$	$\boldsymbol{\tau}_\omega \sim t_v(\boldsymbol{m}, \Gamma)$ $v \in \{3, 5, 7, 9, 11, 13\}$ $\boldsymbol{m} = (0, 0, 0)^T$ $\Gamma = \begin{bmatrix} 10.4 & 0.279 & -0.341 \\ 0.279 & 13.06 & -2.466 \\ -0.341 & -2.466 & 0.581 \end{bmatrix}$	$\boldsymbol{\tau}_\omega \sim \frac{1}{2}N(\boldsymbol{m}_1^* = w \times \boldsymbol{m}_1, V) + \frac{1}{2}N(\boldsymbol{m}_2^* = w \times \boldsymbol{m}_2, V)$ $\boldsymbol{m}_1 = (0, -1, 1)^T$ $\boldsymbol{m}_2 = (0, 1, -1)^T$ $V = \begin{bmatrix} 10.4 & 0.279 & -0.341 \\ 0.279 & 13.06 & -2.466 \\ -0.341 & -2.466 & 0.581 \end{bmatrix}$ $w \in \{0.5, 1, 1.5, 2, 2.5, 3\}$
Reference normal random effects[a]	$\boldsymbol{\tau}_\omega \sim N(\boldsymbol{m}, D^*)$ $\boldsymbol{m} = (0, 0)^T$ $D^* = \frac{1}{2}\boldsymbol{m}_1^*\boldsymbol{m}_1^{*T} + \frac{1}{2}\boldsymbol{m}_2^*\boldsymbol{m}_2^{*T} + V$	$\boldsymbol{\tau}_\omega \sim N(\boldsymbol{m}, D^*)$ $\boldsymbol{m} = (0, 0)^T$ $D^* = \frac{3}{4}\boldsymbol{m}_1\boldsymbol{m}_1^T + \frac{1}{4}\boldsymbol{m}_2\boldsymbol{m}_2^T + V$	$\boldsymbol{\tau}_\omega \sim N(\boldsymbol{m}, D^*)$ $\boldsymbol{m} = (0, 0, 0)^T$ $D^* = \frac{v}{v-2} \times \Gamma$	$\boldsymbol{\tau}_\omega \sim N(\boldsymbol{m}, D^*)$ $\boldsymbol{m} = (0, 0, 0)^T$ $D^* = \frac{1}{2}\boldsymbol{m}_1^*\boldsymbol{m}_1^{*T} + \frac{1}{2}\boldsymbol{m}_2^*\boldsymbol{m}_2^{*T} + V$

[a]The reference normal distribution is a distribution with the same mean and variance-covariance matrix as the corresponding non-normal distribution. BQQ plots and CVM discrepancies computed with a non-normal distribution were compared with its reference distribution.

Figure 1. Selected $p \times 100\%$ percentiles of the probability distribution of individual benefits of imipramine treatment as functions of treatment duration in weeks, for patients with non-endogenous diagnosis with p =0.1, 0.25, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.90, 0.95. (A) Percentiles from Eq. (3) which assumes normality for the random effects. (B) Sample percentiles of EB predictors of individual benefits.

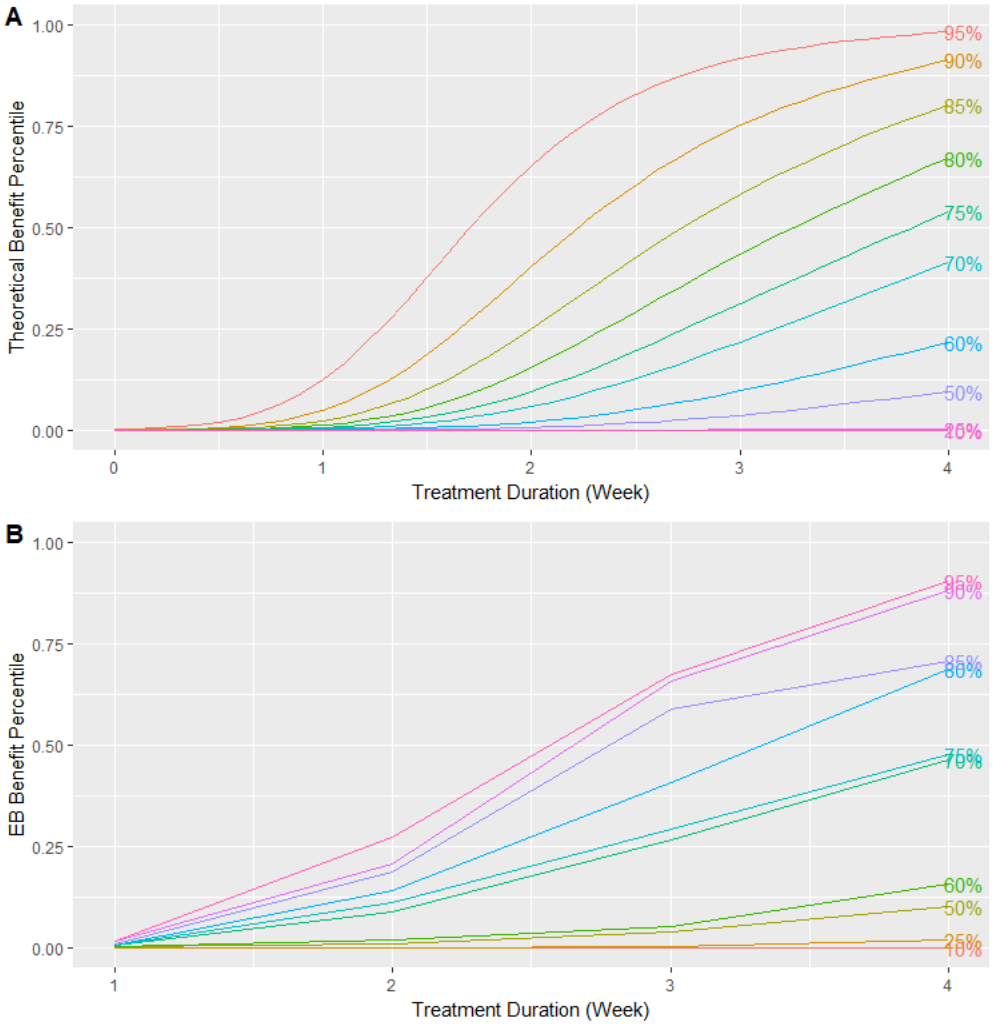


Figure 2. BQQ plot of individual treatment benefits after 4 weeks of imipramine treatment for patients with depression.

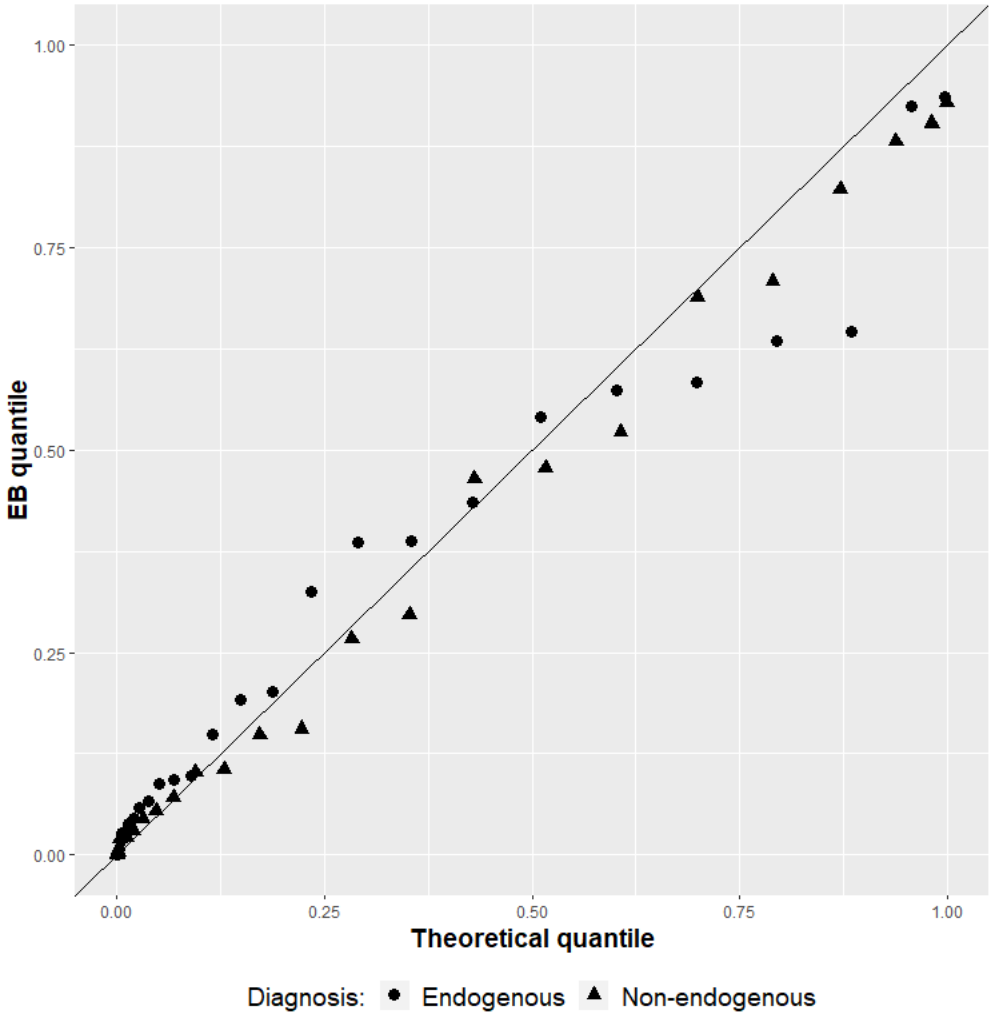


Figure 3 (Scenario 1). Benefit quantile-quantile (BQQ) plots of simulated treatment benefits at $t = 4$ for $N = 100$ patients with $n = 6$ measures per patient. The plots on the right panel correspond to random effects simulated from mixtures of two bivariate normal distributions whose mean vectors were separated by distances of 2, 4, 6, 8 or 10. The left panels correspond to random effects simulated from bivariate normal distributions with the same mean and variance-covariance matrix as the distribution for the right panel on the same row.

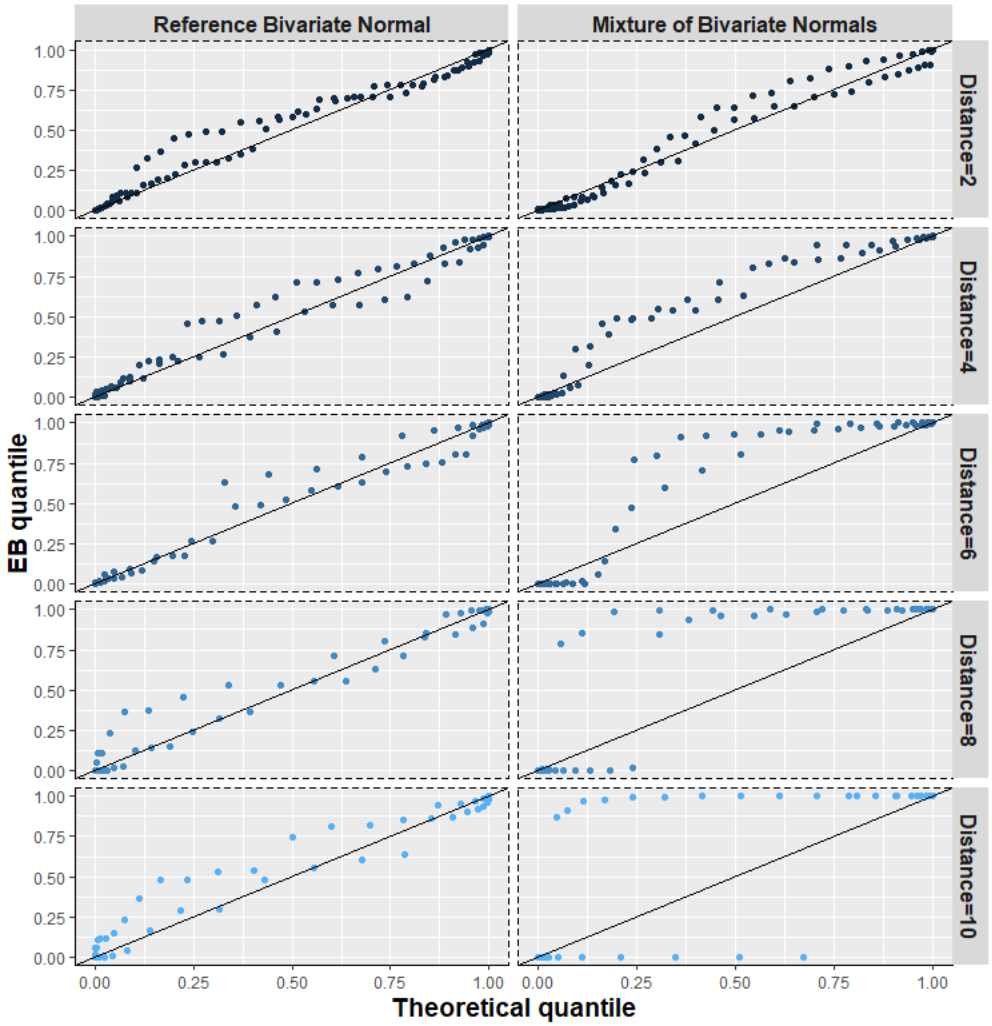


Figure 4 (Scenario 2). Benefit quantile-quantile (BQQ) plots of simulated treatment benefits at $t = 4$ for $N = 100$ patients with $n = 6$ measures per patient. The plots on the right panel correspond to random effects simulated from mixtures of two bivariate normal distributions. Either bivariate component had variances $\sigma_1^2 = \sigma_2^2$ with values 1, 2, 3, 4 or 5. The left panels correspond to random effects simulated from bivariate normal distributions with the same mean and variance-covariance matrix as the distribution for the right panel on the same row.

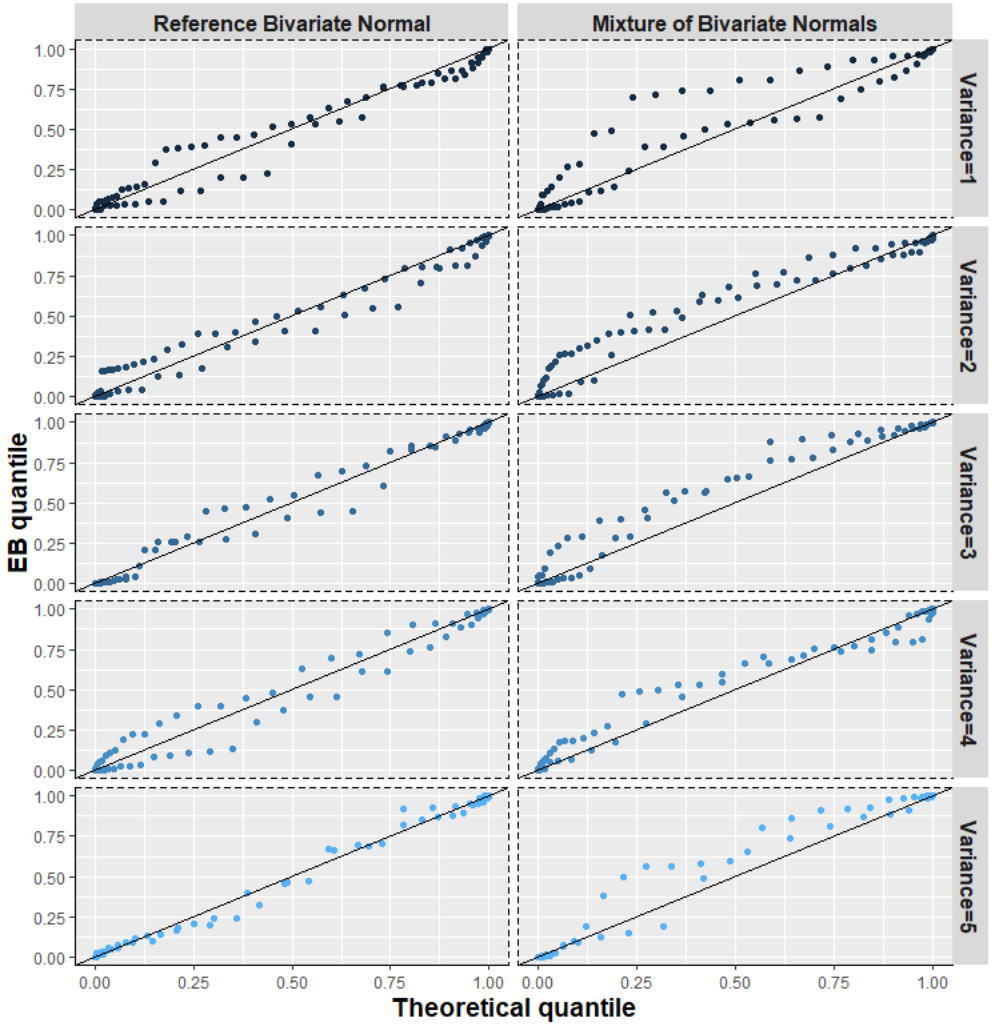


Figure 5 (Scenario 3). Benefit quantile-quantile (BQQ) plots of simulated treatment benefits at $t = 4$ for $N = 100$ patients with $n = 6$ measures per patient. The plots on the right panel correspond to random effects simulated from trivariate t distributions with degrees of freedom (df) of 3, 5, 7, 9, 11 or 13. The left panels correspond to random effects simulated from trivariate normal distributions with the same mean and variance-covariance matrix as the distribution for the right panel on the same row.

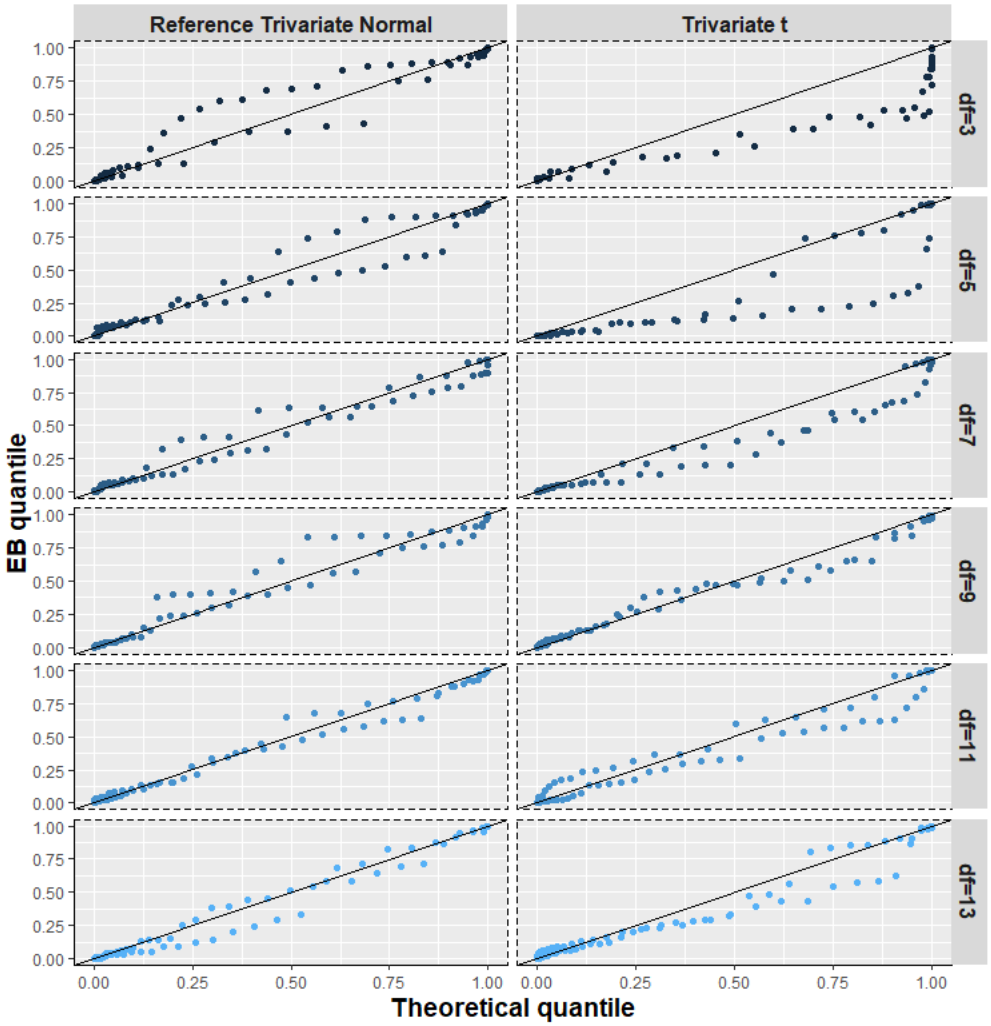


Figure 6 (Scenario 4). Benefit quantile-quantile (BQQ) plots of simulated treatment benefits at $t = 4$ for $N = 100$ patients with $n = 6$ measures per patient. The plots on the right panel correspond to random effects simulated from mixtures of two trivariate normal distributions whose mean vectors were separated by distances of 1.4, 2.8, 4.2, 5.7, 7.1 or 8.5. The left panels correspond to random effects simulated from trivariate normal distributions with the same mean and variance-covariance matrix as the distribution for the right panel on the same row.

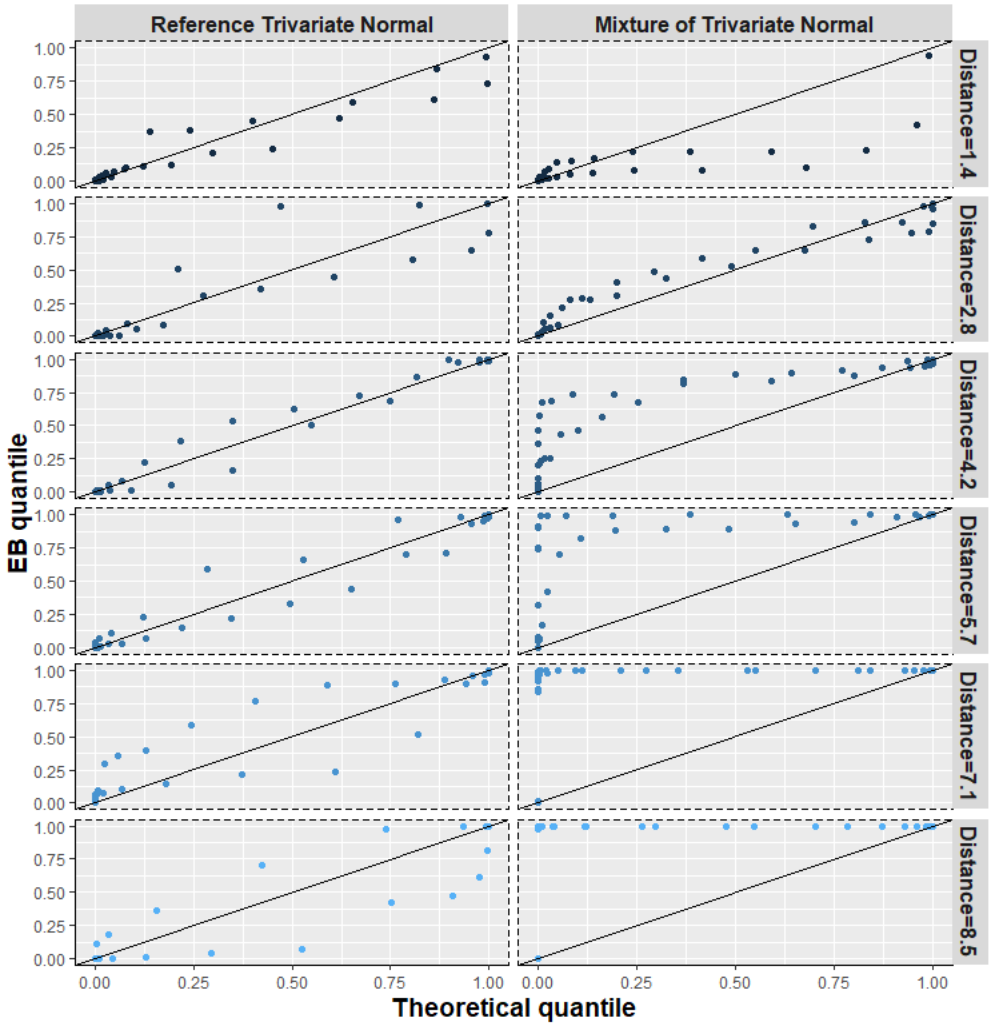


Figure 7 (Scenario 1). Ratios R comparing averages of CVM discrepancies under mixtures of two bivariate normal distributions versus comparable normal distributions with the same mean and variance-covariance matrix for $N = 50$, 100, 150, 200, 300 and 500 as a function of distance between mean vectors of the mixture components. (A) $n = 6$. (B) $n = 4$.

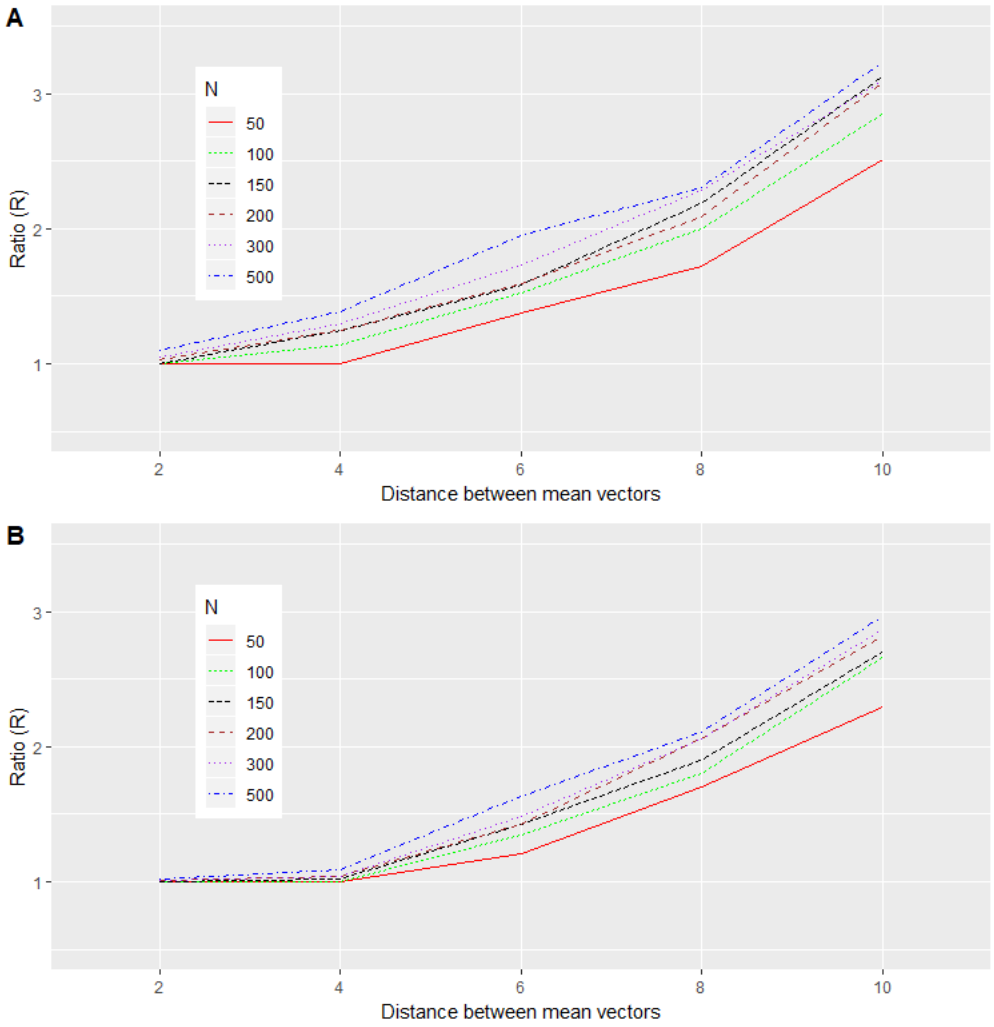


Figure 8 (Scenario 2). Ratios R comparing averages of CVM discrepancies under mixtures of two bivariate normal distributions versus comparable normal distributions with the same mean and variance-covariance matrix for $N = 50$, 100, 150, 200, 300 and 500 as a function of the variance $\sigma_1^2 = \sigma_2^2$. (A) $n = 6$. (B) $n = 4$.

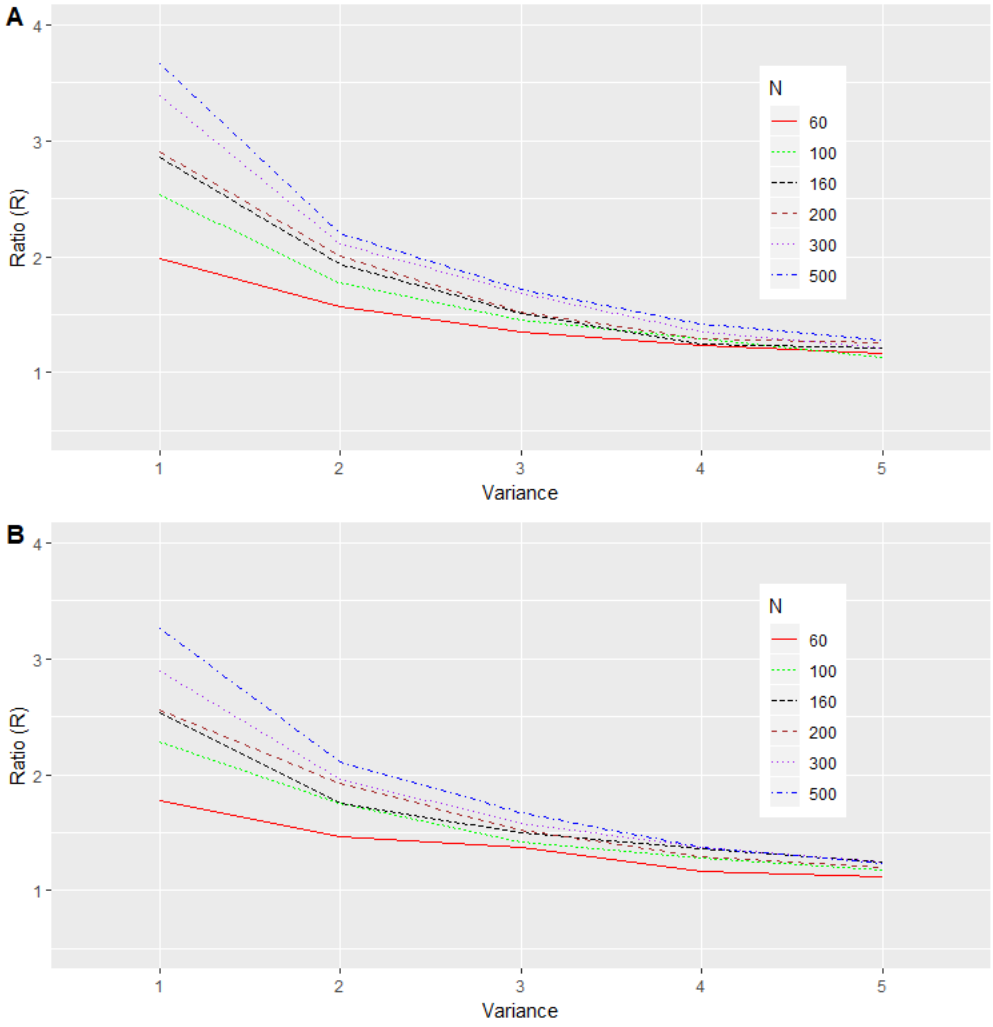


Figure 9 (Scenario 3). Ratios R comparing averages of CVM discrepancies under mixtures of two trivariate normal distributions versus comparable normal distributions with the same mean and variance-covariance matrix for $N = 50$, 100, 150, 200, 300 and 500 as a function of the degrees of freedom v. (A) $n = 6$. (B) $n = 4$.

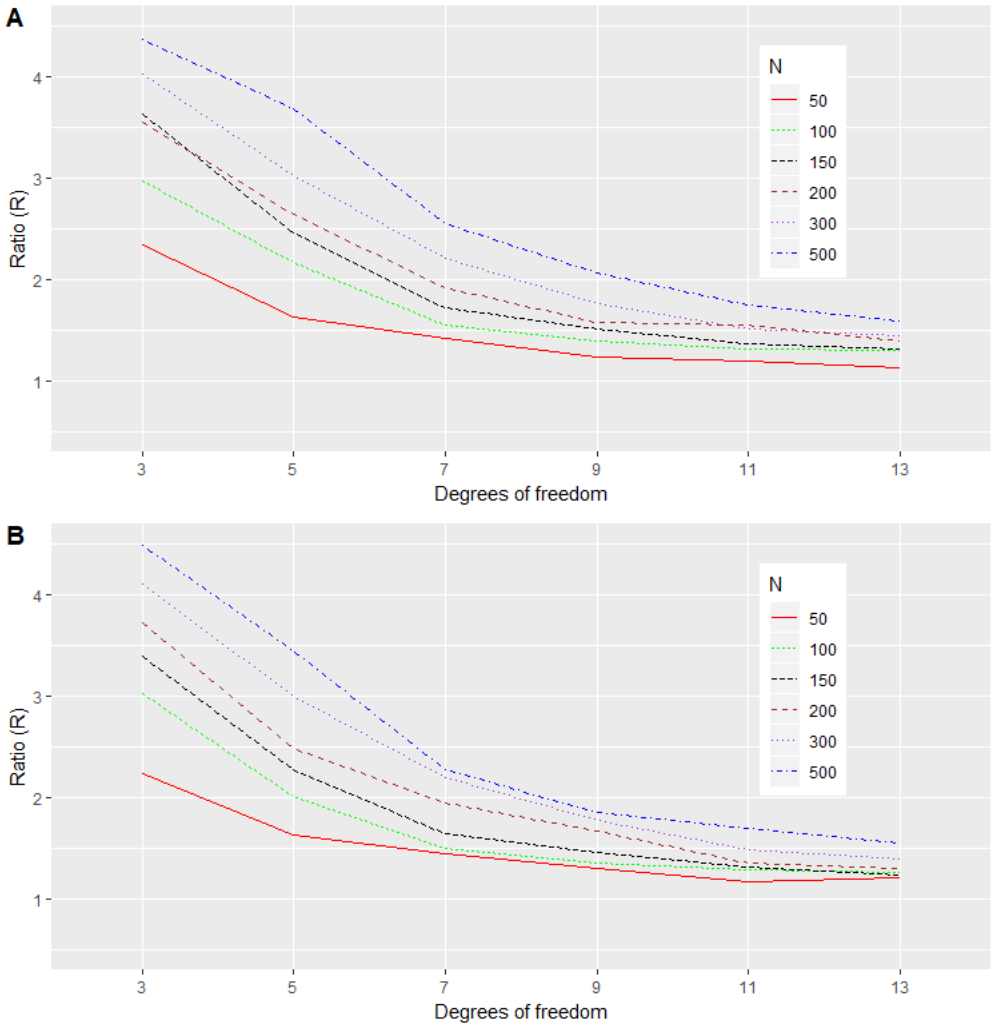
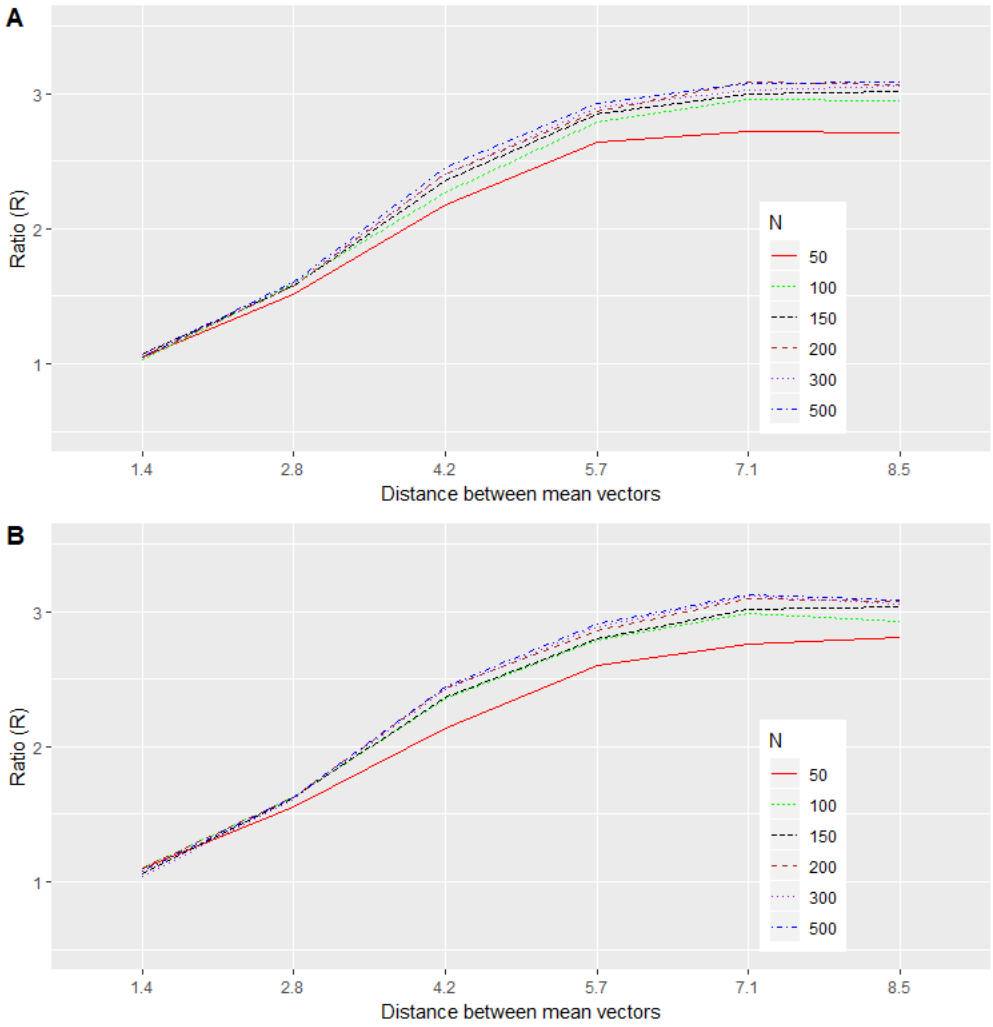


Figure 10 (Scenario 4). Ratios R comparing averages of CVM discrepancies under mixtures of two trivariate normal distributions versus comparable normal distributions with the same mean and variance-covariance matrix for $N = 50$, 100, 150, 200, 300 and 500 as a function of distance between mean vectors of the mixture components. (A) $n = 6$. (B) $n = 4$.



Chapter V: Summary and future directions

In this dissertation, mixed effects models for continuous and binary responses were investigated in terms of random effects prediction accuracy and robustness to the normality assumption of the random effects, via the well-known empirical Bayes approach and an alternative approach based on quadratic inference functions (QIFs) (Chapters II and III). We used mean square prediction errors as the comparison criteria and the performance was explored for a variety of non-normal distributions for the random effects. In linear mixed effects models, we concluded that the empirical Bayes approach generates more accurate predictors of random effects when the random effects follow a non-normal distribution and when the error variances are relatively not large (say less than 10). The approach based on QIFs outperforms the empirical Bayes approach in the presence of extremely large error variances. For logistic mixed effects models, when there is only a random intercept, the empirical Bayes approach is superior to the QIF approach with respect to the prediction accuracy not only when the random effects follow a normal distribution, but also when this assumption is violated. In general, in the presence of a random intercept and a random slope, the prediction performances of the two approaches are comparable. However, the QIF approach produces slightly more accurate predictors of random effects when the random effects follow a mixture of normal distributions and this is especially true for the random slope.

A message from the first two chapters is that the empirical Bayes approach is relatively robust to misspecifications of the normality assumption of the random effects. The approach based on QIFs is mathematically more complicated and can be very time consuming in both implementation and computation. Thus, the empirical Bayes approach is more recommendable in statistical practice, even when the normality assumption is questionable. Future research should explore the predictive performance of these approaches for repeated count data using Poisson mixed effects models or negative binomial mixed effects models. Comparisons can also be made

in the presence of missing data or with other approaches proposed in the literature (Ghidey et al., 2004; Ten Have and Localio, 1999; Zhang and Davidian, 2001).

Finally, in Chapter IV, we proposed a graphical approach to examine the goodness-of-fit of linear mixed effects models with emphasis in the normality assumption for the random effects, when the goal is to estimate or predict individual benefits from medical or behavioral treatments for severely ill patients. Simulation studies in which the random effects were assumed to follow a variety of non-normal distributions show that our approach successfully captures violations of the normality assumption. The approach is more powerful when the sample size is large, but still provides valid conclusions under relatively small sample sizes. As future research, our method may be extended to logistic mixed effects models for binary responses. The sensitivity of the graphical approach to violations of the normality assumption other than the ones investigated in this dissertation can also be explored; for instance, under mixtures of normal distributions whose variances and covariances depend on some of the model covariates.

References

Abad, A.A., Litière, S., Molenberghs, G., 2010. Testing for misspecification in generalized linear mixed models. Biostatistics 11, 771–786.

Agresti, A., Caffo, B., Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. Comput. Stat. Data Anal. 47, 639–653.

Alonso, A., Litière, S., Molenberghs, G., 2008. A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. Comput. Stat. Data Anal. 52, 4474–4486.

Breslow, N.E., Clayton, D.G., 1993. Approximate Inference in Generalized Linear Mixed Models. J. Am. Stat. Assoc. 88, 9–25.

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

Cho, H., Wang, P., Qu, A., 2017. Personalize Treatment for Longitudinal Data Using Unspecified Random-Effects Model. Stat. Sin. 27, 187–205.

Diaz, F.J., 2018. Construction of the Design Matrix for Generalized Linear Mixed-Effects Models in the Context of Clinical Trials of Treatment Sequences. Rev. Colomb. Estadística 41, 191–233.

Diaz, F.J., 2017. Estimating individual benefits of medical or behavioral treatments in severely ill patients. Stat. Methods Med. Res. 962280217739033.

Diaz, F.J., 2016. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat. Med. 35, 4077–4092.

Diaz, F.J., Cogollo, M.R., Spina, E., Santoro, V., Rendon, D.M., de Leon, J., 2012a. Drug dosage individualization based on a random-effects linear model. J. Biopharm. Stat. 22, 463–484.

Diaz, F.J., de Leon, J., 2013. The Mathematics of Drug Dose Individualization Should Be Built With Random-Effects Linear Models. Ther. Drug Monit. 35, 276–277.

Diaz, F.J., Yeh, H.W., de Leon, J., 2012b. Role of Statistical Random-Effects Linear Models in Personalized Medicine. Curr. Pharmacogenomics Person. Med. 10, 22–32.

Drikvandi, R., Verbeke, G., Molenberghs, G., 2017. Diagnosing misspecification of the random-effects distribution in mixed models. Biometrics 73, 63–71.

Efendi, A., Drikvandi, R., Verbeke, G., Molenberghs, G., 2017. A goodness-of-fit test for the random-effects distribution in mixed models. Stat. Methods Med. Res. 26, 970–983.

Fitzmaurice, G.M., Davidian, M., Verbeke, G., Molenberghs, G., 2009. Longitudinal data analysis, Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Boca Raton.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2011. Applied longitudinal analysis, 2nd ed, Wiley series in probability and statistics. Wiley, Hoboken, N.J.

Frees, E.W., 2004. Longitudinal and panel data : analysis and applications in the social sciences. Cambridge University Press, Cambridge, UK ; New York.

Ghidey, W., Lesaffre, E., Eilers, P., 2004. Smooth random effects distribution in a linear mixed

model. Biometrics 60, 945–953.

Gillies, C.S., Hebblewhite, M., Nielsen, S.E., Krawchuk, M.A., Aldridge, C.L., Frair, J.L., Saher, D.J., Stevens, C.E., Jerde, C.L., 2006. Application of random effects to the study of resource selection by animals. J. Anim. Ecol.

Have, T.R. Ten, Kunselman, A.R., Pulkstenis, E.P., Landis, J.R., 1998. Mixed Effects Logistic Regression Models for Longitudinal Binary Response Data with Informative Drop-Out. Biometrics 54, 367–383.

Hedeker, D., 2003. A mixed-effects multinomial logistic regression model. Stat. Med. 22, 1433–1446.

Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. Wiley-Interscience, Hoboken, N.J.

Horrocks, J., van Den Heuvel, M.J., 2009. Prediction of pregnancy: A joint model for longitudinal and binary data. Bayesian Anal. 4, 523–538.

Jiang, J., 1999. Conditional inference about generalized linear mixed models. Ann. Stat. 27, 1974–2007.

Kleinman, K., Lazarus, R., Platt, R., 2004. A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism. Am. J. Epidemiol. 159, 217–224.

Lin, X., Breslow, N.E., 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. J. Am. Stat. Assoc. 91, 1007–1016.

Liu, D., Ghosh, D., Lin, X., 2008. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. BMC Bioinformatics.

Mann, R.P., Spaiser, V., Hedman, L., Sumpter, D.J.T., 2018. Choice modelling with Gaussian processes in the social sciences: A case study of neighbourhood choice in stockholm. PLoS One.

Marquart, L., Haynes, M., 2019. Misspecification of multimodal random-effect distributions in logistic mixed models for panel survey data. J. R. Stat. Soc. Ser. A Stat. Soc. 182, 305–321.

McCulloch, C.E., Neuhaus, J.M., 2011. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Stat. Sci. 26, 388–402.

Pan, Z.Y., Lin, D.Y., 2005. Goodness-of-fit methods for generalized linear mixed models. Biometrics 61, 1000–1009.

Qu, A., Lindsay, B.G., Li, B., 2000. Improving generalised estimating equations using quadratic inference functions. Biometrika 87, 823–836.

Sashegyi, A.I., Brown, K.S., Farrell, P.J., 2000. Application of a generalized random effects regression model for cluster-correlated longitudinal data to a school-based smoking prevention trial. Am. J. Epidemiol. 152, 1192–1200.

Shen, W., Louis, T.A., 1999. Empirical Bayes estimation via the smoothing by roughening approach. J. Comput. Graph. Stat. 8, 800–823.

Skrondal, A., Rabe-Hesketh, S., 2003. Some Applications of Generalized Linear Latent and Mixed Models in Epidemiology. Nor. Epidemiol. 13, 265–278.

Tchetgen, E.J., Coull, B.A., 2006. A diagnostic test for the mixing distribution in a generalised linear mixed model. Biometrika 93, 1003–1010.

Ten Have, T.R., Kunselman, A.R., Tran, L., 1999. A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. Stat. Med. 18, 947—960.

Ten Have, T.R., Localio, A.R., 1999. Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. Biometrics 55, 1022–1029.

Van Den Noortgate, W., De Boeck, P., Meulders, M., 2003. Cross-Classification Multilevel Logistic Models in Psychometrics. J. Educ. Behav. Stat. 28, 369–386.

Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. J. Am. Stat. Assoc. 91, 217–221.

Verbeke, G., Molenberghs, G., 2013. The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. Biostatistics 14, 477–490.

Waagepetersen, R., 2006. A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. Scand. J. Stat. 33, 721–731.

Wang, P., Tsai, G.F., Qu, A.N., 2012. Conditional Inference Functions for Mixed-Effects Models With Unspecified Random-Effects Distribution. J. Am. Stat. Assoc. 107, 725–736.

Zhang, D.W., Davidian, M., 2001. Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics 57, 795–802.

Zhu, X.L., Qu, A.N., 2016. Individualizing drug dosage with longitudinal data. Stat. Med. 35,

4474–4488.

Appendix:

A. Chapter II: Prediction accuracy and robustness to non-normality of two methods of predicting random effects in linear mixed effects models for longitudinal data: empirical Bayes versus quadratic inference functions

Zhiwen Wang, John Keighley, Jianghua He, Jo Wick, Francisco J. Diaz*

Department of Biostatistics, The University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, United States.

*Corresponding author: Tel: +1-913-945-7006.

Email address: zwang3@kumc.edu (Z. Wang), Jkeighle@kumc.edu (J. Keighley), jhe@kumc.edu (J. He), jwick@kumc.edu (J. Wick), fdiaz@kumc.edu (F.J. Diaz).

Figure S1. Scatterplots of random effects for the depression data, predicted by QIF versus EB approaches under an exchangeable error correlation structure. (A) Random intercept predictors. (B) Predictors of the random coefficient of time. (C) Predictors of the random coefficient of time square. The solid line is the y=x line.

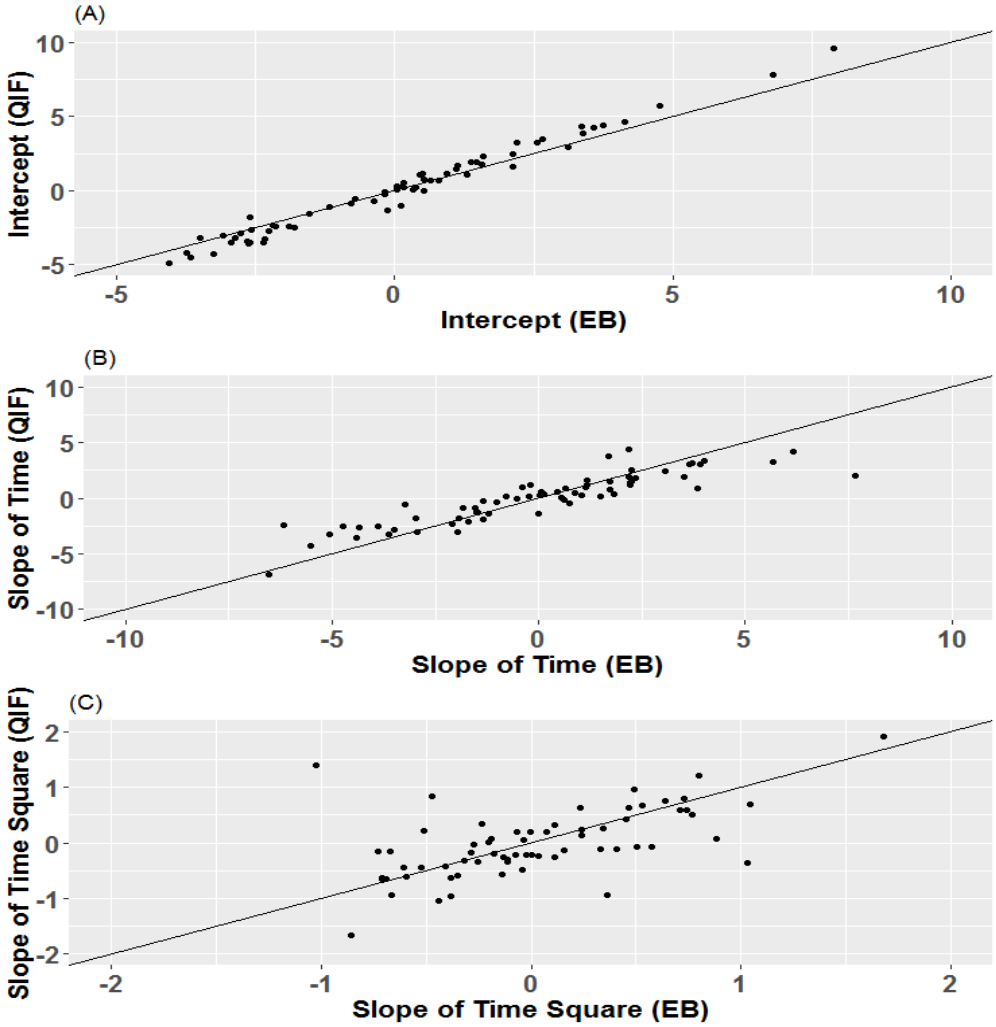


Figure S2. Scatterplots of random effects for the depression data, predicted by QIF versus EB approaches under an AR(1) error correlation structure. (A) Random intercept predictors. (B) Predictors of the random coefficient of time. (C) Predictors of the random coefficient of time square. The solid line is the y=x line.

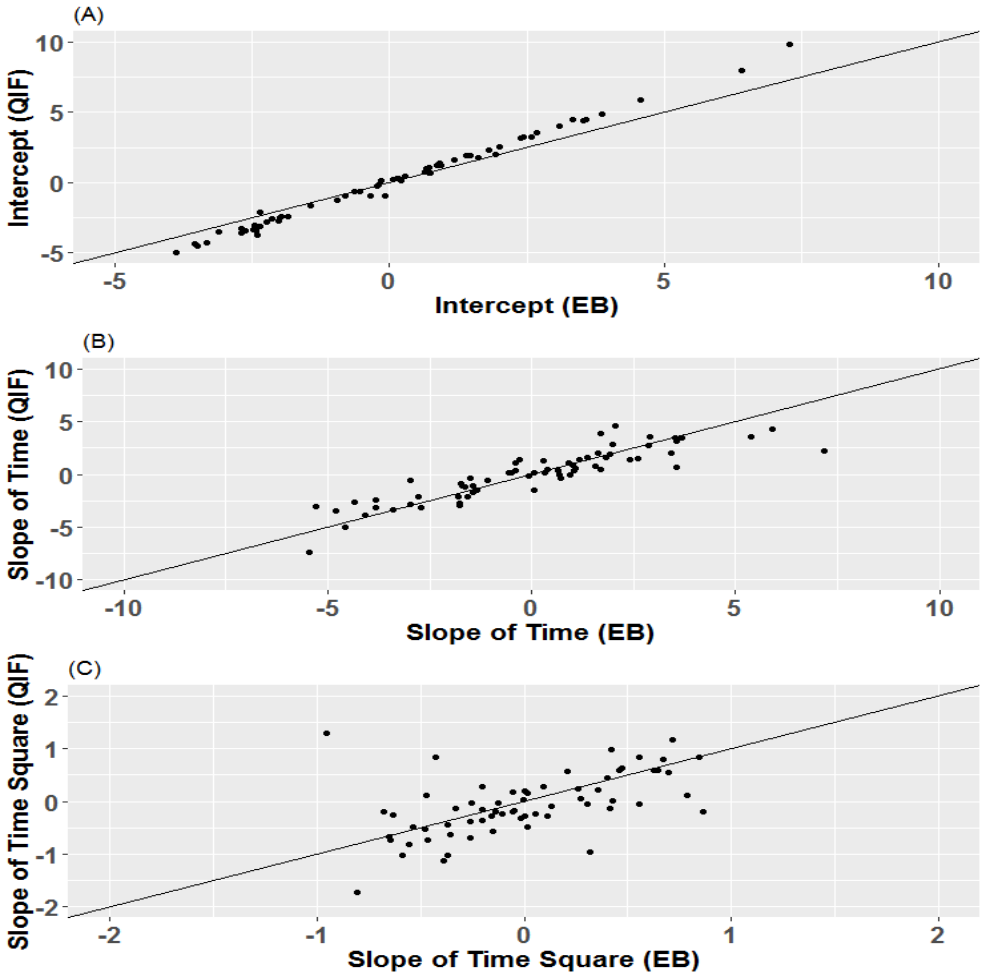
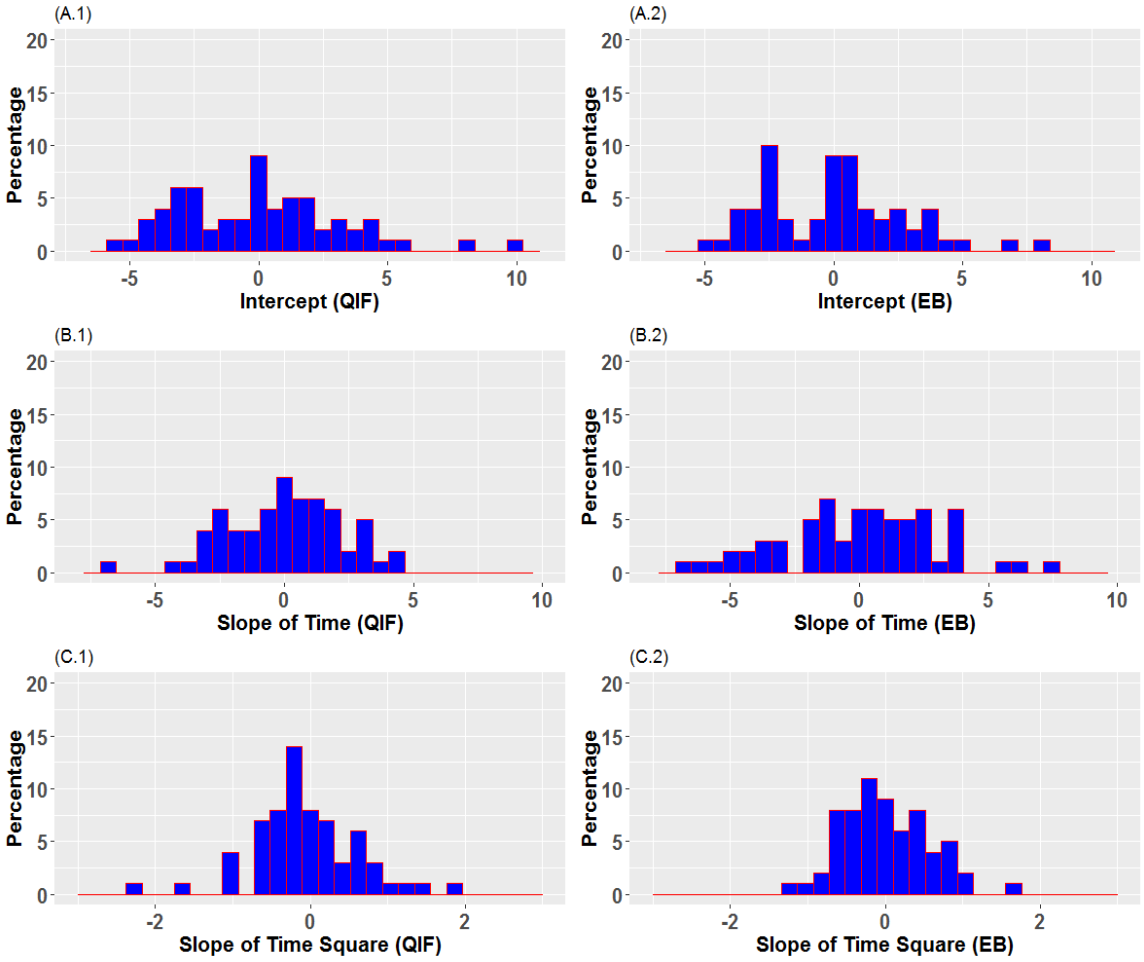


Figure S3. Histograms of predictors of random effects for the depression data using QIF and EB approaches under homoscedastic independent errors. (A1-A2) Random intercept predictors. (B1-B2) Predictors of the random slope of time. (C1-C2) Predictors of the random slope of time square.



SAS/IML code implementing the QIF approach.

```
/*
Programmers:
Zhiwen Wang, Francisco J. Diaz
Department of Biostatistics,
The University of Kansas Medical Center,
Mail Stop 1026, 3901 Rainbow Blvd.,
Kansas City, KS 66160, United States.
Emails: zwang3@kumc.edu (Z. Wang), fdiaz@kumc.edu (F.J. Diaz).


This code implements the QIF approach used for the application and simulations in the paper:
Wang, Z., Keighley, J., He, J., Wick, J., Diaz, F.J. "Comparison of the accuracy and robustness to non-normality
of two methods of random effects prediction for linear mixed effects models of longitudinal data:
empirical Bayes versus quadratic inference functions".
*/
/*************************************************************************************************
**********/


/*************************************************************************************************
**********/

/* Definition of global parameters:*/

/*n: the total number of unique time points recorded in the dataset, this is K in the paper;  */
/*SampSize: total number of subjects; */
/*nfe: number of fixed-effect parameters; */
/*nre: number of random-effect parameters; */
/*lambda1: tuning parameter lambda1;*/
/*lambda2: tuning parameter lambda2;*/
/*maxiter: maximum number of iterations for the dataset, it needs to be large enough, say 10^5;*/
/*loc_missing: location vector of missing responses in Y (this can be obtained using the "loc" function in
SAS/IML);*/
/*obs_ind: vector of observation indicator for all subjects. It has the same dimension as Y; it is 1 if the response if
non-missing, 0 if it is missing (this vector can be obtained using the "missing" function in SAS/IML).*/
/*Y: the response matrix with dimension SampSize*n by 1;*/
/*X: the design matrix for the fixed effects, it's dimension is SampSize*n by nfe;*/
/*Z: the block diagonal design matrix for the random effects, it's dimension is SampSize*n by nre;*/


/*CorrStruc = 1 : Independent structure;*/
/*CorrStruc = 2 : Exchangeable (EX) correlation structure;*/
/*CorrStruc = 3 : AR1 correlation structure;*/
/*CorrStruc = other numbers : WRONG INDEX;*/

/*************************************************************************************************
**********/



proc iml;
```

```
/*****************************Part 1. Calculate projection matrix
Pj*********************************/
/*This module is used to calculate the projection matrix Pj for a given dataset*/

/*Input: design matrix X, and design matrix Z;*/
/*Output: Projection Matrix Pj;*/

start projection(Pj, X,Z);

        /* find generalized inverse */
        Xinv = ginv(X);
        /* form projection matrix onto (left) column space */
        Px = X*Xinv;
        A = (I(ncol(Px))-Px)*Z;

        /*using QR decomposition to obtain null space of A*/
        CALL QR(Q,R,PIV,LINDEP,t(A));
        /*rank of matrix A*/
        ranka = round(trace(ginv(A)*A));
        /* obtain the basis of the null space of A*/
        Jnull = Q[ , (ranka+1):ncol(Q)];

        /* calculate projection matrix Pj on Jnull */
        Pj = Jnull*ginv(Jnull);

finish projection;

/*****************************************************************************************/




/*****************************************Part 2. Calculate Basis matrix M
************************************************/
/*This module is used to calculate the second basis matrix M2 for an assumed correlation structure. */
/*Note that the first basis matrix M1 is always the identity matrix of dimension n by n;*/

/*Input: */
/*      CorrStruc: correlation structure indicator (1, 2 or 3);*/
/*      n: the total number of unique time points recorded in the dataset;*/
/*Output: */
/*      M2: The second basis matrix*/


start BaseM (M2, n, CorrStruc);


/* independent structure needs just one M matrix */
if CorrStruc = 1 then do;
        M2 = diag(j(1,n));              /* create diagonal matrix M2*/
end;
```

```
/* EX structure */
if CorrStruc = 2 then do;

        M2 = j(n,n,1);                                  /* create matrix M2*/
        /* create diagonal matrix of M2 to be 0*/
        start SetDiag(A, V);
                diagIdx = do(1,nrow(A)*ncol(A), ncol(A)+1);
                A[diagIdx] = V;
        finish;
        run SetDiag(M2, 0);

end;



/* AR(1) structure */
if CorrStruc = 3 then do;

        M2 = j(n,n,0);                  /* create matrix M2*/
        supDiag = T(1:ncol(M2)-1) || T(2:ncol(M2)); /* subscripts for superdiagonal */
        subDiag = T(2:ncol(M2)) || T(1:ncol(M2)-1); /* subscripts for subdiagonal */
        dim = dimension(M2);                    /* find index of all super- and subdiagonal elements */
        idxM2 = sub2ndx(dim, supDiag//subDiag);
        M2[idxM2] = 1;                  /* assign 1 to sub- and superdiagonal */
end;


/* Message for unsupported structures*/
if CorrStruc ^= {1 2 3} then do;
print "not defined, wrong correlation index";  M2=0;
end;

finish BaseM;

/*****************************************************************************************************
******************/




/***********************************Part 3: Fixed effect QIF (given
b)*************************************/
/*This module is used to define and minimize the fixed effects QIF for a given */
/*random-effect estimates “b” and a given initial vector of fixed effects “beta” that needs to be updated. This "beta"
can be obtained from previous iteration.*/


/*Input: */
/*        b: a given random effects estimates for all subjects, this is a vector of dimension (SampSize*nre) by 1;*/
/*        beta: initial fixed effects estimates( to be updated), this is a vector of dimension nfe by 1;*/
/*Output:*/
/*        beta_new: updated fixed effects estimates; this is a vector of dimension nfe by 1;*/
```

```
start betafunction (beta_new, b, beta)
        global (CorrStruc, M2, n, SampSize, nfe, nre, maxiter, loc_missing, obs_ind, Y, X, Z);


M_large = I(SampSize) @ M2;

vu2 = I(SampSize*n);
diagIdxvu2 = do(1,nrow(vu2)*ncol(vu2), ncol(vu2)+1);
vu2[diagIdxvu2] = obs_ind;          /* set diagonal elements to 1 or 0*/


if CorrStruc = 1 then do;
                gi_f = j(nfe,SampSize,0);
end;
else do;
                gi_f = j(2*nfe,SampSize,0);
end;


do i = 2 to maxiter until (diff < 0.00001);

        beta_old = beta;


        /*Normal mean response*/
        mu = X*beta + Z*b;

        mu[loc_missing,] = 0;

        residual = Y - mu;


        /* Independent structure */
        if CorrStruc = 1 then do;
        do i=1 to SampSize;
                gi_f[,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
        end;
        sum_g_first = -t(X)*X;
        end;



        /* EX/AR1 structure */
        if CorrStruc ^= 1 then do;
        do i=1 to SampSize;
                gi_f[1:nfe,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
                gi_f[(nfe+1) : (2*nfe),i] = t(X[(1+(i-1)*n):(i*n),])*M2*(residual[(1+(i-1)*n):(i*n),]);
        end;
        sum_g_first1 = -t(X)*X;
        sum_g_first2 = -t(X)*vu2*M_large*vu2*X;
        sum_g_first = sum_g_first1//sum_g_first2;
        end;
```

```
        C2 = gi_f*t(gi_f);

        sum_g_score = gi_f[,+];

        invC = ginv(C2);

        /*QIF for beta*/
        QIF_beta       = t(sum_g_score)*invC*sum_g_score;
        QIF_first_dev  = t(sum_g_first)*invC*sum_g_score;
        QIF_second_dev = t(sum_g_first)*invC*sum_g_first;


        delta = ginv(QIF_second_dev)*QIF_first_dev;
/*      delta = solve(QIF_second_dev, QIF_first_dev);*/

        beta = beta - delta;

        diff = sum(abs(beta_old - beta));

end;

/* Return missing if does not converge */
if i = maxiter  then  do;
beta_new = j(nfe,1,.); cov2=j(nfe,nfe,.);
end;
else  do;
beta_new = beta; cov2=inv(QIF_second_dev);
end;



finish betafunction;
/*****************************************************************************************************
******************/


/*****************************************************************************************************
********************/



/*************************************Part 4: Random effects QIF (given
beta)*********************************/
/*This module is used to define and minimize the random effects QIF for a given */
/*fixed effects estimate “beta” and a given initial vector of random effects “b” that needs to be updated. This "b" can
be obtained from previous iteration.*/


/*Input: */
/*      beta: a given fixed effects estimates; this is a vector of dimension nfe by 1;*/
/*      b: initial random effects estimates (to be updated) for all subjects; this is a vector of dimension
(SampSize*nre) by 1;*/
/*Output:*/
/*      b_new: updated random effects estimates for all subjects; this is a vector of dimension (SampSize*nre) by
1;*/
```

```
start bfunction(b_new, b, beta)
            global(n, SampSize, nfe, nre, Pj, lambda1, lambda2, maxiter, loc_missing, Y, X, Z);


do j = 2 to maxiter until (diff < 0.00001);

            b_old = b;


            /*normal mean*/
            mu = X*beta + Z*b;

    mu[loc_missing,] = 0;

            residual = Y - mu;

            Gr = t(Z)*residual;

            Gr1 = -t(Z)*Z;

            /*Extended score, the QIF for b*/
            extend_score = Gr // lambda1*b // lambda2*Pj*b;
            extend_first_deriv = Gr1 // lambda1*I(SampSize*nre) // lambda2*Pj;
            /*QIF for b*/
            QIF_b = t(extend_score)*extend_score;
            dh = t(extend_first_deriv)*extend_score;
            ddh= t(extend_first_deriv)*extend_first_deriv;

            delta = ginv(ddh)*dh;
/*          delta = solve(ddh,dh);*/

            b = b - delta;

            diff = sum(abs(b_old - b));

end;

/*b_new = shape(b, SampSize,nre);*/
/* return missing if not converge */
if j = maxiter then  b_new = j(SampSize*nre,1,.);
else  b_new = b;

finish bfunction;
/*****************************************************************************************************
******************/

/*****************************************************************************************************
******************/


/***************************Part 5. Iterative module to minimize fixed and random effects QIFs defined in parts
4 and 5***************************/
/*This module implements the iterative minimization process of the fixed and random effects QIFs.*/
```

```
/*The iteration gives the optimal fixed and random effects estimates when the stopping criteria (see the detail in the
paper) is satisfied.*/
/*The initial vector of random effects is set to be a vector of zeros.*/


/*Input: */
/*        beta_initial: initial vector of fixed effects; this vector can be obtained with the GLM procedure in SAS;*/

/*Output:*/
/*        QIFbeta: optimal fixed effects estimates; this is a vector of dimension nfe by 1;*/
/*        QIFb: optimal random effects estimates for all subjects; this is of dimension (SampSize*nre) by 1*/
/*        k: number of iterations required for this minimization;*/



start Robust (QIFbeta,QIFb,k,  beta_initial)
          global(CorrStruc, n, SampSize, nfe, nre, maxiter, Pj, M2, lambda1, lambda2, loc_missing, obs_ind, Y, X,
Z);


        beta_new = beta_initial;

        /*initial random effects is set to be a zero vector*/
        b_initial = j(SampSize*nre,1,0);
        b_new = b_initial;

        do k = 2 to maxiter until (diff_sum < 0.00001);
                beta_old = beta_new;
                b_old = b_new;

                call bfunction (b_new, b_new,beta_new);
                call betafunction (beta_new, b_new,beta_new);

                diff_sum = sum(abs(beta_new-beta_old)) + sum(abs(b_new-b_old));
        end;

        QIFb = shape(b_new, SampSize, nre);
        QIFbeta = beta_new;
        QIFVarCov = cov2;


finish Robust;
/*************************************************************************************************
*************************/


/*********************************Part 6. Cross-validation module used to choose the optimal tuning
parameter lambda1*********************************/
/*This module selects the optimal tuning parameter lambda1 that gives the smallest cross-validation error for a range
of */
/*pre-specified candidates. The candidates are specified by an interval [mini, maxi], with an increment
"increment".*/

/**/
/*Input:*/
/*        n_measure: the total number of unique time points recorded in the dataset, this is K in the paper;*/
```

```
/*          mini: lower bound of the tuning parameter candidates interval;*/
/*          maxi: upper bound of the tuning parameter candidates interval; */
/*          increment: increment of the tuning parameter candidates interval [mini, maxi];*/
/*          Y_design: the response matrix of dimension (SampSize*n) by 1;*/
/*          X_design: the design matrix for the fixed effects; it is of dimension (SampSize*n) by nfe;*/
/*          Z_design: the block diagonal design matrix for the random effects; it is of dimension (SampSize*n) by
nre;*/
/*          beta_initial: initial vector of fixed effects; this can be obtained with the GLM procedure in SAS;*/

/*Output:*/
/*          CVE: a matrix of cross-validation errors for all lambda1 candidates; */
/*                    the first column contains all lambda1 candidates; the second column contains the cross-validation
errors corresponding to a specified candidate in the same row;*/
/*          lambda1_optimal: optimal tuning parameter lambda1;*/



start CV (CVE,lambda1_optimal,  n_measure, mini,maxi,increment, Y_design,X_design,Z_design, beta_initial)
     global(CorrStruc, n, SampSize, nfe, nre, maxiter, Pj, M2, lambda1, lambda2, loc_missing, obs_ind, Y, X, Z);

          n = n_measure-1;
          call BaseM (M2, n, CorrStruc);

          PE = j(n_measure,1,0);

          lamdba1value = do(mini,maxi,increment);

          CVE = j(ncol(lamdba1value),2,0);
          CVE[,1] = t(lamdba1value);



          do l = 1 to ncol(lamdba1value) until  (stop = 1);

                    lambda1 = lamdba1value[,l];


                    do r = 1 to n_measure;

                              i=1:SampSize;
                              cvk = r+(i-1)*n_measure;

                              /*Split the response vector for training and testing */
                              Y_cvk = Y_design[setdif(1:SampSize*n_measure, cvk)];
                              Y_pred = Y_design[cvk];

                              /*Split the design matrix X for training and testing */
                              X_cvk = X_design[setdif(1:SampSize*n_measure, cvk),];
                              X_pred = X_design[cvk,];

                              /*Split the design matrix Z for training and testing */
                              Z_cvk = Z_design[setdif(1:SampSize*n_measure, cvk),];
                              Z_pred = Z_design[cvk,];
```

```
                        obs_ind = 1-missing(Y_cvk);
                        loc_missing = loc(Y_cvk=.);
                        Y_cvk[loc_missing] = 0;

                        loc_missing_pred = loc(Y_pred=.);
                        Y_pred[loc_missing_pred] = 0;


                        Y = Y_cvk;
                        X = X_cvk;
                        Z = Z_cvk;

                        call projection(Pj, X_cvk,Z_cvk);

                        call Robust (QIFbeta,QIFb,k,  beta_initial);
                        Pred_Ran_Efft = QIFb;
                        Pred_Fixed_Efft = QIFbeta;

                        mu_pred = X_pred*Pred_Fixed_Efft + Z_pred*shape(Pred_Ran_Efft, nre*SampSize, 1);

                        mu_pred[loc_missing_pred] = 0;

                        PE[r] = t(Y_pred - mu_pred)*(Y_pred - mu_pred);

                end;

/*Cross-validation error*/
                CVE[l,2] = sum(PE)/(n_measure*SampSize);

                if l = 1 then do;
                        diff = 1; stop = 0;
                end;
                else do;
                        diff = abs(CVE[l,2]-CVE[(l-1),2]);
                        if (diff <= 0.00001)  then stop = 1; else stop = 0;
                end;

        end;

if l = ncol(lamdba1value)+1 then ll = ncol(lamdba1value); else ll = l;
/*Choose the optimal lambda1*/
idx = loc( CVE = min(CVE[1:ll,2]) );
lambda1_optimal = CVE[idx-1];

finish CV;



/*************************************************************************************/
/*********************************Fit the depression data*****************************/
/*************************************************************************************/


/*See Diaz (2017) for description of the 2-PM model fitted to the depression data and the way the time variable was
built.*/
/*The dataset is available in Hedeker and Gibbons (2006).*/
```

```
/*In this last part of the code, the data is handled to make n equal for all subjects,*/
/*and then the Parts 1 through 6 of this file are called to estimate the fixed effects and predict the random effects.*/

/*R3 is the dataset*/

/*hamd: Hamilton score*/
/*EDG: Diagnosis (1=endogenous, 0=nonendogenous)*/
/*time: Weeks under treatment*/



use R3;
  read all var {hamd} into Y_design;
  read all var {EDG}  into EDG;
  read all var {time} into TIME;


n=6; SampSize = 66;

CorrStruc = 1;
/*CorrStruc = 2;*/
/*CorrStruc = 3;*/

nfe = 4;
nre = 3;

maxiter = 1000000;

lambda2 = log(SampSize);
/*cs/ind=1, ar1=0.9  */
/*lambda1 = 0.9;*/
lambda1 = 1;


beta_initialvalue = {21.50012019, 1.86645495, -3.97668046, 0.33981694};


/*design matrix X*/
intcp = j(n*SampSize,1,1);
TIMESQR = TIME##2;
week = TIME;
X_design = intcp || EDG || TIME || TIMESQR;



/* create block diagonal design matrix Z */
Z_old = X_design[,{1,3,4}];
Z_design = Z_old[1:n,];
do i = 2 to SampSize;
        Z_design = block(Z_design, Z_old[1+(i-1)*n:i*n ,  ]);
end;



/*Insert zero to replace missing observations in design matrices and response vector*/
```

```
loc_missing_all = loc(Y_design=.);
loc_missing = loc_missing_all;
obs_ind = 1-missing(Y_design);

X_design[loc_missing_all,] = 0;
Z_design[loc_missing_all,] = 0;
Y_design[loc_missing_all,] = 0;


X = X_design;
Z = Z_design;
Y = Y_design;


call projection(Pj, X,Z); /* This calls Part 1 of this file*/
call BaseM (M2, n, CorrStruc); /*This calls Part 2 of this file*/
call Robust (QIFbeta,QIFb,k, beta_initialvalue); /*This calls Part 5 of this file; this routine calls Parts 3 and 4*/

print QIFbeta,QIFb,k;

quit;
```

B. Chapter III: The impact of violations of the normality assumption for the random effects in the logistic mixed effects model for longitudinal data on two methods of random-effect prediction: empirical Bayes versus quadratic inference functions

Zhiwen Wang and Francisco J. Diaz*

University of Kansas Medical Center, 3901 Rainbow Blvd, Kansas City, KS 66160

*Corresponding author: Francisco J. Diaz, Department of Biostatistics & Data Science, The University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, United States. Email: fdiaz@kumc.edu. Phone: 913-945-7006.

SAS/IML code implementing the QIF approach.

```
proc import datafile =
'/panfs/pfs.local/home/zwang3/paper3/Simulations/n6/intslpmodel2/N100rho4/t/true_response_y.csv'      out =
work.true_response_y     dbms = CSV; run;

proc import datafile = '/panfs/pfs.local/home/zwang3/paper3/Simulations/n6/intslpmodel2/N100rho4/t/trtdata.csv'
out = work.true_trt           dbms = CSV; run;



proc iml;


/*******************************Calculate projection matrix A***********************************/

start projection(Pj, X,Z);

        /* find (generalized) inverse */
        Xinv = ginv(X);
        /* form projection matrix onto (left) column space */
        Px = X*Xinv;
        A = (I(ncol(Px))-Px)*Z;

        /*using QR decomp to obtain null space of A*/
        CALL QR(Q,R,PIV,LINDEP,t(A));
        /*rank of matrix A*/
        ranka = round(trace(ginv(A)*A));
        /*OBTAIN THE BASIS OF NULL SPACE OF MATRIX a: m-r columns of Q*/
        Jnull = Q[ , (ranka+1):ncol(Q)];
        /* dimJ=dimension(Jnull); print dimJ;*/
        /* calculate projection matrix Pj on J */
        Pj = Jnull*ginv(Jnull);

finish projection;

/*****************************************************************************************************/




/*****************************************Base M
matrtix*************************************************/

start BaseM (M2, n, CorrStruc);

/*CorrStruc = 1 : Independent structure;*/
/*CorrStruc = 2 : EX;*/
/*CorrStruc = 3 : AR1;*/
/*CorrStruc = others : WRONG INDEX;*/
```

```
/* independent structure needs 1 M matrix */
if CorrStruc = 1 then do;
        M2 = diag(j(1,n));              /* create diagnol matrix of M2*/
end;


/* exchangeble structure */
if CorrStruc = 2 then do;

        M2 = j(n,n,1);                                  /* create matrix of M2*/
        /* create diagnol matrix of M2 to be 0*/
        start SetDiag(A, V);
                diagIdx = do(1,nrow(A)*ncol(A), ncol(A)+1);
                A[diagIdx] = V;
        finish;
        run SetDiag(M2, 0);

end;



/* AR(1) structure */
if CorrStruc = 3 then do;

        M2 = j(n,n,0);                  /* create matrix of M2*/
        supDiag = T(1:ncol(M2)-1) || T(2:ncol(M2)); /* subscripts for superdiagonal */
        subDiag = T(2:ncol(M2)) || T(1:ncol(M2)-1); /* subscripts for subdiagonal */
        dim = dimension(M2);                    /* find index of all super- and subdiagonal elements */
        idxM2 = sub2ndx(dim, supDiag//subDiag);
        M2[idxM2] = 1;                  /* assign sub- and superdiagonal to 1 */
end;


/* other structure not defined*/
if CorrStruc ^= {1 2 3} then do;
print "not defined, wrong correlation index";  M2=0;
end;

finish BaseM;

/*****************************************************************************************
******************/




/**********************Part 2: module to define random effect QIF function(given
beta)*****************************/


start bfunction(b_new,j, b, beta)
          global(n, SampSize, nfe, nre, Pj, lambda1, lambda2, maxiter, Y, X, Z);


do j = 2 to maxiter until (diff < 0.00001);
```

```
        b_old = b;


        /*vector of mean*/
        mu = 1/(1+exp(-(X*beta + Z*b)));

   /*vector of residual*/
        residual = Y - mu;

        /*vector of variance*/
        varu = mu#(1-mu);


        Gr = t(Z)*residual;

        Gr1 = -t(Z)*diag(varu)*Z;

        /*extended score*/
        extend_score = Gr // lambda1*b // lambda2*Pj*b;
        extend_first_deriv = Gr1 // lambda1*I(SampSize*nre) // lambda2*Pj;

        dh = t(extend_first_deriv)*extend_score;
        ddh= t(extend_first_deriv)*extend_first_deriv;

        delta = ginv(ddh)*dh;
/*      delta = solve(ddh,dh);*/
        b = b - delta;

        diff = sum(abs(b_old - b));

end;

/*b_new = shape(b, SampSize,nre);*/
/* return missing if no convergence */
if j = maxiter then  b_new = j(SampSize*nre,1,.);
else  b_new = b;

finish bfunction;
/*****************************************************************************************************
******************/

/*test */
/*beta0 = beta_initialvalue;*/
/*b0 = j(SampSize*nre,1,1.5); */
/*call  bfunction (b_new,j,b0,beta0) ;*/
/*print b_new, j;*/
/*****************************************************************************************************
******************/


/*****************************************************************************************************
******************/
```

```
/**********************************Part 3: module to define fixed effect QIF (given
b)*************************************/

start betafunction (beta_new,cov2, i, b, beta)
        global (CorrStruc, M2, n, SampSize, nfe, nre, maxiter, Y, X, Z);


M_large = I(SampSize) @ M2;

if CorrStruc = 1 then do;
                gi_f = j(nfe,SampSize,0);
end;
else do;
                gi_f = j(2*nfe,SampSize,0);
end;


do i = 2 to maxiter until (diff < 0.00001);

        beta_old = beta;


        /*vector of mean*/
        mu = 1/(1+exp(-(X*beta + Z*b)));

   /*vector of residual*/
        residual = Y - mu;

        /*vector of variance*/
        varu = mu#(1-mu);

        /*A_i matrix*/
        vu1 = diag(mu#(1-mu));
        /*A_i^(-1/2) matrix*/
        vu2 = diag(1/sqrt(mu#(1-mu)));



        /* independent structure */
        if CorrStruc = 1 then do;
        do i=1 to SampSize;
                gi_f[,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
        end;
        sum_g_first = -t(X)*diag(varu)*X;
        end;




        /* EX/AR1 structure */
        if CorrStruc ^= 1 then do;
        do i=1 to SampSize;
                gi_f[1:nfe,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
                gi_f[(nfe+1) : (2*nfe),i] = t(X[(1+(i-1)*n):(i*n),])*vu1[(1+(i-1)*n):(i*n),(1+(i-
1)*n):(i*n)]*vu2[(1+(i-1)*n):(i*n),(1+(i-1)*n):(i*n)]*M2*vu2[(1+(i-1)*n):(i*n),(1+(i-1)*n):(i*n)]*(residual[(1+(i-
1)*n):(i*n),]);
```

```
        end;
        sum_g_first1 = -t(X)*diag(varu)*X;
        sum_g_first2 = -t(X)*diag(varu)*vu2*M_large*vu2*diag(varu)*X;
        sum_g_first = sum_g_first1//sum_g_first2;
        end;




        C2 = gi_f*t(gi_f);

        sum_g_score = gi_f[,+];

        invC = ginv(C2);

        QIF_beta       = t(sum_g_score)*invC*sum_g_score;
        QIF_first_dev  = t(sum_g_first)*invC*sum_g_score;
        QIF_second_dev = t(sum_g_first)*invC*sum_g_first;


        delta = ginv(QIF_second_dev)*QIF_first_dev;
/*      delta = solve(QIF_second_dev, QIF_first_dev);*/

        beta = beta - delta;

        diff = sum(abs(beta_old - beta));

end;

/* return missing if no convergence */
if i = maxiter  then  do;
beta_new = j(nfe,1,.); cov2=j(nfe,nfe,.);
end;
else  do;
beta_new = beta; cov2=ginv(QIF_second_dev);
end;



finish betafunction;
/*****************************************************************************************
******************/

/*test*/
/*b0 = j(SampSize*nre,1,0.1);*/
/*beta0 = beta_initialvalue;*/
/*call betafunction (beta_new,cov2, i, b0, beta0);*/
/*print beta_new,cov2,i;*/
/*****************************************************************************************
******************/



/*********************New Part 4. module iterative minimization to calculate fixed and random effects based on
Robust QIF*****************************/
```

```
start Robust (QIFbeta,QIFVarCov,QIFb,k,  beta_initial)
          global(CorrStruc, n, SampSize, nfe, nre, maxiter, Pj, M2, lambda1, lambda2, Y, X, Z);


        beta_new = beta_initial;

        /*initial randox effect 0*/
        b_initial = j(SampSize*nre,1,0);
        b_new = b_initial;

        do k = 2 to maxiter until (diff_sum < 0.00001);
                beta_old = beta_new;
                b_old = b_new;

                call bfunction (b_new,j, b_new,beta_new);
                call betafunction (beta_new,cov2,i, b_new,beta_new);

                diff_sum = sum(abs(beta_new-beta_old)) + sum(abs(b_new-b_old));
        end;

        QIFb = shape(b_new, SampSize, nre);
        QIFbeta = beta_new;
        QIFVarCov = cov2;


finish Robust;
/*****************************************************************************************************
**************************/


/*****************************************************************************************************
****************/
start BIC_Cal (BIC,PART_beta,PART_b, b,beta)
          global(CorrStruc, n, SampSize, nfe, nre, maxiter, Pj, M2, lambda1, lambda2, Y, X, Z);


        M_large = I(SampSize) @ M2;

        /*vector of mean*/
        mu = 1/(1+exp(-(X*beta + Z*b)));

        /*vector of residual*/
        residual = Y - mu;

        /*vector of variance*/
        varu = mu#(1-mu);

        /*A_i matrix*/
        vu1 = diag(mu#(1-mu));
        /*A_i^(-1/2) matrix*/
        vu2 = diag(1/sqrt(mu#(1-mu)));


        /* independent structure */
        if CorrStruc = 1 then do;
        gi_f = j(nfe,SampSize,0);
```

```
        do i=1 to SampSize;
                gi_f[,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
        end;
        end;

        /* EX/AR1 structure */
        if CorrStruc ^= 1 then do;
        gi_f = j(2*nfe,SampSize,0);
        do i=1 to SampSize;
                gi_f[1:nfe,i] = t(X[(1+(i-1)*n):(i*n),])*(residual[(1+(i-1)*n):(i*n),]);
                gi_f[(nfe+1) : (2*nfe),i] = t(X[(1+(i-1)*n):(i*n),])*vu1[(1+(i-1)*n):(i*n),(1+(i-
1)*n):(i*n)]*vu2[(1+(i-1)*n):(i*n),(1+(i-1)*n):(i*n)]*M2*vu2[(1+(i-1)*n):(i*n),(1+(i-1)*n):(i*n)]*(residual[(1+(i-
1)*n):(i*n),]);
        end;
        end;


        C2 = gi_f*t(gi_f);

        sum_g_score = gi_f[,+];

        invC = ginv(C2);

        PART_beta = t(sum_g_score)*invC*sum_g_score;

        /*PJB = Pj*b; print PJB; print CovPjb;*/

        CovPjb = cov(Pj*b);
        PART_b = lambda2*( t(Pj*b)*ginv(CovPjb)*(Pj*b) );

        BIC = PART_beta + PART_b;

        /*print BIC, PART_beta, PART_b;*/


finish BIC_Cal;



/***********************************fit QIF************************************/

use true_response_y;
  read all var _ALL_ into Y_design_all;

use true_trt;
  read all var _ALL_ into true_trt;



n=7;

SampSize=100;
```

```
/*total simulation number*/
w=ncol(Y_design_all);

/*2. fit QIF*/

CorrStruc = 1;

nfe = 4;
nre = 2;

maxiter = 1000000;

lambda2 = log(SampSize);



/*form dataset and fit logistic fixed models for initial values of beta*/

/*time and sqrt_time*/
time = {0,1,2,3,4,5,6};
time_vec = shape(repeat(time,SampSize),SampSize*n,1);

sqrt_time = sqrt(time);
sqrt_time_vec = shape(repeat(sqrt_time,SampSize),SampSize*n,1);;

ID = shape(t(repeat(1:SampSize,n)),SampSize*n,1);

dataset1 = ID||time_vec||sqrt_time_vec;
/*print dataset1;*/

dataset2 = repeat(dataset1,w);
/*print dataset2;*/

sim = shape(t(repeat(1:w,SampSize*n)),SampSize*n*w,1);
/*print sim;*/

Y_vec = colvec(t(Y_design_all));

trt_vec = colvec(t(true_trt));

dataset = sim||Y_vec||trt_vec||dataset2;
/*print dataset;*/

/** create data set **/
varNames = {"sim","Y","trt","ID","time","sqrt_time"};
create MyData from dataset [colname=varNames];
append from dataset;
close MyData;


/* 1. fixed-effects logistic regression model for initial value of beta*/
submit MyData;
ods output ParameterEstimates=GLM_fit;
PROC LOGISTIC DATA=MyData DESCENDING;
BY sim;
```

```
MODEL Y = trt sqrt_time trt*sqrt_time ;
RUN;
endsubmit;


/*output fiexed effects as initial beta*/
use GLM_fit;
        read all var {Estimate} into beta_initial;



beta_initial_vec = t(shape(t(beta_initial),w,nfe));

print beta_initial, beta_initial_vec;



/*****************************************************************************/

/*simulation index*/
/*sim_idx=1;*/


/*store optimal lambda1*/
lambda1_optimal_vec = j(w,1);

/*store random effects*/
QIF_itp = j(SampSize,w,1);
QIF_slp = j(SampSize,w,1);


do sim_idx=1 to w;



        /*************************************/
        /*design matrix X*/

        intcp = j(n*SampSize,1,1);

        trt_col = true_trt[,sim_idx];

        txswk = trt_col#sqrt_time_vec;

        X_design = intcp || trt_col || sqrt_time_vec || txswk;


        /*design matrix Z*/
   Zi = repeat(1,n)||sqrt_time;
        Z_design = I(SampSize)@Zi;


        beta_initialvalue = beta_initial_vec[,sim_idx];

        Y_design = Y_design_all[,sim_idx];
```

```
        /*print X_design, Z_design, Y_design, beta_initialvalue;*/


        X = X_design;
        Z = Z_design;
        Y = Y_design;


        /*********select lambda1 by the BIC*********/

        /*candidate lambda1 values*/
/*      lambda1_values = do(0.1,lambda2,0.3);*/
   lambda1_values = log(n)|| (log(n)+lambda2)/2 ||lambda2;

        BIC_MATRIX = j(ncol(lambda1_values),2,0);


        do l = 1 to ncol(lambda1_values);

                lambda1 = lambda1_values[,l];

                call projection(Pj, X,Z);
                call BaseM (M2, n, CorrStruc);
                call Robust (QIFbeta,QIFVarCov,QIFb,k,  beta_initialvalue);

                b_test = shape(QIFb, SampSize*nre,1);
                beta_test = QIFbeta;
                call BIC_Cal(BIC,PART_beta,PART_b, b_test,beta_test);


                BIC_MATRIX[l,1] = lambda1;
                BIC_MATRIX[l,2] = BIC;

/*              print QIFbeta,QIFVarCov,QIFb,k,BIC;*/


        end;

        print BIC_MATRIX;

        /*choose the optimal lambda1*/
        idx = loc( BIC_MATRIX = min(BIC_MATRIX[,2]) );
        lambda1_optimal = BIC_MATRIX[idx-1];

/*      print lambda1_optimal;*/

        lambda1_optimal_vec[sim_idx] = lambda1_optimal;


        /*fit QIF with the optimal lambda1*/
        lambda1 = lambda1_optimal;

        call projection(Pj, X,Z);
        call BaseM (M2, n, CorrStruc);
        call Robust (QIFbeta,QIFVarCov,QIFb,k,  beta_initialvalue);
```

```
/*          b_test = shape(QIFb, SampSize*nre,1);*/
/*          beta_test = QIFbeta;*/
/*          call BIC_Cal(BIC,PART_beta,PART_b, b_test,beta_test);*/

/*          print QIFbeta,QIFVarCov,QIFb,k,BIC;*/


            /*output QIF_random effects*/
            QIF_itp[,sim_idx] = QIFb[,1];
            QIF_slp[,sim_idx] = QIFb[,2];



end;



/*print QIF_itp,QIF_slp;*/
print lambda1_optimal_vec;


/***********************************************/

create QIF_itp from QIF_itp;
append from QIF_itp;
close QIF_itp;


create QIF_slp from QIF_slp;
append from QIF_slp;
close QIF_slp;



proc export data=QIF_itp
outfile="/panfs/pfs.local/home/zwang3/paper3/Simulations/n6/intslpmodel2/N100rho4/t/qif_rnd_intercept.csv"
dbms=csv replace; run;

proc export data=QIF_slp
outfile="/panfs/pfs.local/home/zwang3/paper3/Simulations/n6/intslpmodel2/N100rho4/t/qif_rnd_slope.csv"
dbms=csv replace; run;
```

C. Chapter IV: A Graphical Approach to Assess the Goodness-of-Fit of Random Effects Linear Models When the Goal is to Measure Individual Benefits of Medical Treatments in Severely Ill Patients

Zhiwen Wang and Francisco J. Diaz*

University of Kansas Medical Center, 3901 Rainbow Blvd, Kansas City, KS 66160

*Corresponding author: Francisco J. Diaz, Department of Biostatistics, The University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, United States. Email: fdiaz@kumc.edu. Phone: 913-945-7006.

Figure S1. Scatterplot, histogram and normal Q-Q plot of conditional residuals from the random effects linear model for the depression data.

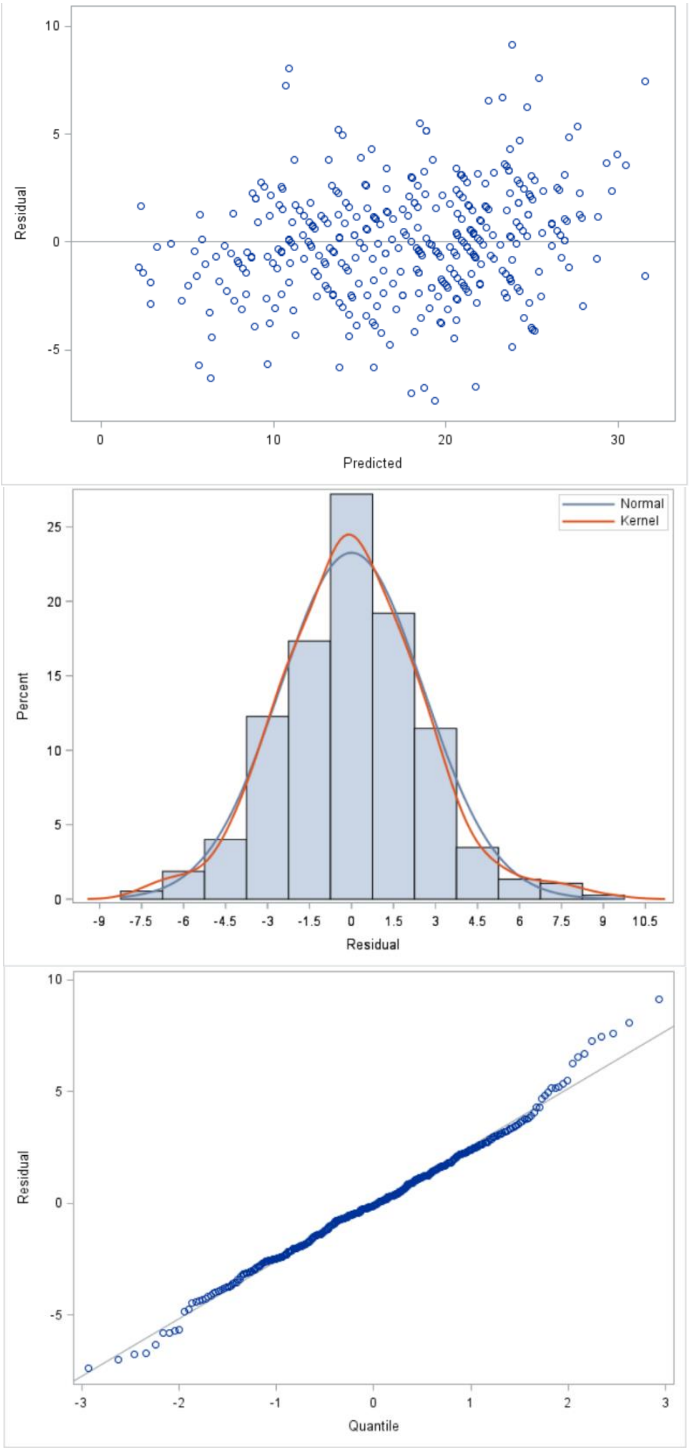


Figure S2. Histograms and kernel densities of EB predictors of random effects for 66 patients with depression under imipramine treatment. (A) Random intercept predictors. (B) Predictors of the random slope of time. (C) Predictors of the random slope of time square.

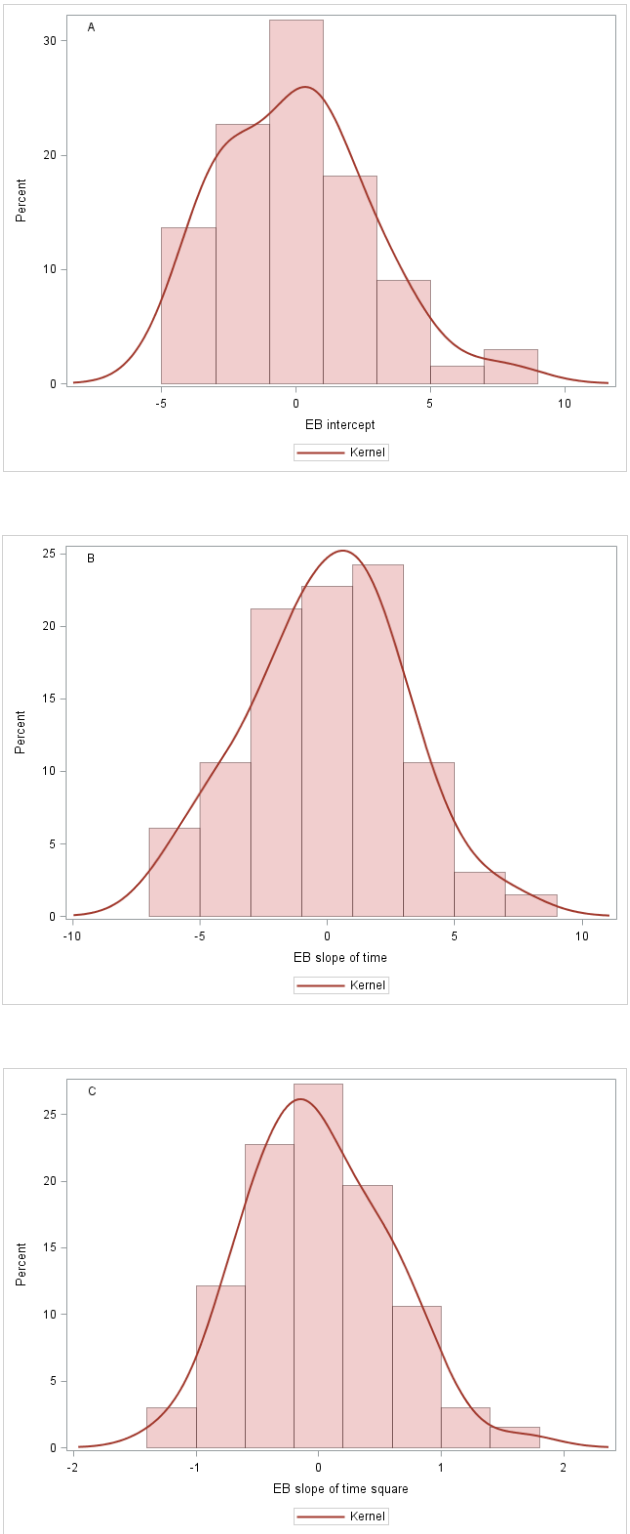


Figure S3. Selected $p \times 100\%$ percentiles of the probability distribution of individual benefits of imipramine treatment as functions of treatment duration in weeks, for patients with endogenous diagnosis with p =0.1, 0.25, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.90, 0.95. (A) Percentiles from Eq. (3) which assumes normality for the random effects. (B) Sample percentiles of EB predictors of individual benefits.

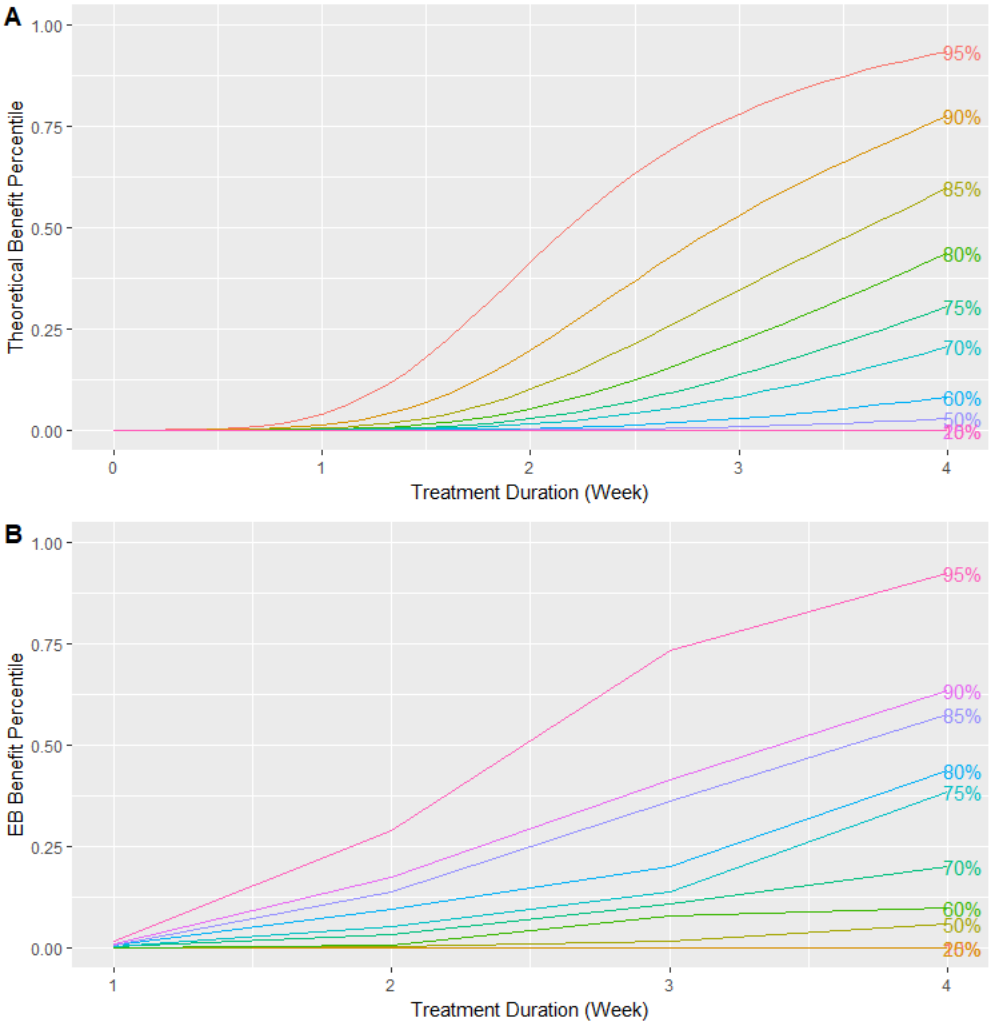


Table S1 (Scenario 1). Average Cramer-von Mises discrepancies $\bar{\bar{\Omega}}$ between the empirical distribution of EB predicted benefits and the theoretical distribution of benefits assuming normality for the random effects, for simulated reference bivariate normal (RBN) distributions and mixtures of two bivariate normal distributions with the same mean and variance-covariance matrix, by distance between mean vectors of the mixture components.

$n = 6$	Distance between means									
	2		4		6		8		10	
N	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture
50	0.00434	0.00435	0.00552	0.00555	0.00818	0.0113	0.0135	0.0233	0.0189	0.0475
100	0.00216	0.00216	0.00300	0.00343	0.00530	0.00807	0.00990	0.0198	0.0147	0.0418
150	0.00137	0.00138	0.00197	0.00246	0.00434	0.00686	0.00847	0.0185	0.0132	0.0414
200	0.00113	0.00116	0.00172	0.00215	0.00408	0.00648	0.00828	0.0173	0.0129	0.0398
300	0.000711	0.000746	0.00127	0.00165	0.00348	0.00604	0.00735	0.0168	0.0125	0.0386
500	0.000527	0.000579	0.000962	0.00133	0.00298	0.00579	0.00701	0.0161	0.0118	0.0381

$n = 4$	Distance between means									
	2		4		6		8		10	
N	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture
50	0.00455	0.00455	0.00530	0.00532	0.00870	0.0106	0.0140	0.0238	0.0194	0.0445
100	0.00245	0.00245	0.00290	0.00290	0.00566	0.00761	0.0101	0.0182	0.0149	0.0396
150	0.00168	0.00168	0.00218	0.00222	0.00452	0.00646	0.00895	0.0170	0.0135	0.0364
200	0.00133	0.00135	0.00169	0.00175	0.00419	0.00599	0.00797	0.0164	0.0127	0.0358
300	0.00100	0.00102	0.00135	0.00140	0.00365	0.00540	0.00761	0.0157	0.0123	0.0353
500	0.000713	0.000730	0.00100	0.00109	0.00312	0.00510	0.00707	0.0149	0.0118	0.0352

Table S2 (Scenario 2). Average Cramer-von Mises discrepancies $\overline{\overline{\Omega}}$ between the empirical distribution of EB predicted benefits and the theoretical distribution of benefits assuming normality for the random effects, for simulated reference bivariate normal (RBN) distributions and mixtures of two bivariate normal distributions with the same mean and variance-covariance matrix, for selected variances of the mixture components $\sigma_1^2 = \sigma_2^2$.

$n = 6$	σ_1^2, σ_2^2									
	1		2		3		4		5	
N	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture
60	0.00660	0.0131	0.00808	0.0127	0.00932	0.0126	0.0110	0.0136	0.0119	0.0138
100	0.00484	0.0123	0.00621	0.0110	0.00746	0.0108	0.00886	0.0114	0.0104	0.0118
160	0.00373	0.0107	0.00498	0.00965	0.00622	0.00941	0.00793	0.00992	0.00887	0.0108
200	0.00352	0.0103	0.00468	0.00936	0.00610	0.00931	0.00730	0.00947	0.00827	0.0104
300	0.00289	0.00982	0.00424	0.00897	0.00541	0.00909	0.00678	0.00917	0.00789	0.00955
500	0.00261	0.00958	0.00384	0.00846	0.00494	0.00849	0.00611	0.00865	0.00733	0.00939

$n = 4$	σ_1^2, σ_2^2									
	1		2		3		4		5	
N	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture	RBN	Mixture
60	0.00707	0.0126	0.00814	0.0119	0.00914	0.0125	0.0114	0.0132	0.0124	0.0139
100	0.00473	0.0108	0.00600	0.0105	0.00753	0.0106	0.00859	0.0110	0.0103	0.0121
160	0.00374	0.00946	0.00521	0.00911	0.00617	0.00927	0.00747	0.0102	0.00879	0.0110
200	0.00356	0.00910	0.00466	0.00896	0.00592	0.00902	0.00729	0.00940	0.00839	0.0101
300	0.00303	0.00877	0.00423	0.00828	0.00539	0.00854	0.00657	0.00903	0.00768	0.00957
500	0.00257	0.00839	0.00380	0.00801	0.00491	0.00822	0.00613	0.00841	0.00746	0.00918

Table S3 (Scenario 3). Average Cramer-von Mises discrepancies $\bar{\bar{\Omega}}$ between the empirical distribution of EB predicted benefits and the theoretical distribution of benefits assuming normality for the random effects, for simulated reference trivariate normal (RTN) distributions and trivariate t distributions with the same mean and variance-covariance matrix, for selected degrees of freedom v.

$n = 6$	Degree of freedom (v)											
	3		5		7		9		11		13	
N	RTN	t	RTN	t	RTN	t	RTN	t	RTN	t	RTN	t
50	0.00842	0.0197	0.00528	0.00859	0.00473	0.00671	0.00469	0.00579	0.00460	0.00548	0.00446	0.00505
100	0.00567	0.0169	0.00288	0.00626	0.00268	0.00416	0.00248	0.00347	0.00245	0.00322	0.00235	0.00306
150	0.00463	0.0168	0.00219	0.00541	0.00190	0.00328	0.00174	0.00263	0.00170	0.00232	0.00165	0.00216
200	0.00436	0.0155	0.00182	0.00480	0.00149	0.00286	0.00142	0.00224	0.00131	0.00204	0.00129	0.00178
300	0.00384	0.0155	0.00143	0.00432	0.00117	0.00259	0.00107	0.00189	0.00103	0.00155	0.000996	0.00143
500	0.00345	0.0151	0.00108	0.00400	0.000833	0.00213	0.000768	0.00158	0.000728	0.00128	0.000685	0.00109

$n = 4$	Degree of freedom (v)											
	3		5		7		9		11		13	
N	RTN	t	RTN	t	RTN	t	RTN	t	RTN	t	RTN	t
50	0.00888	0.0198	0.00551	0.00897	0.00489	0.00704	0.00471	0.00611	0.00471	0.00551	0.00456	0.00553
100	0.00557	0.0168	0.00316	0.00635	0.00276	0.00412	0.00257	0.00350	0.00249	0.00320	0.00244	0.00308
150	0.00479	0.0162	0.00232	0.00527	0.00204	0.00335	0.00179	0.00261	0.00189	0.00249	0.00179	0.00222
200	0.00419	0.0156	0.00191	0.00474	0.00158	0.00307	0.00145	0.00242	0.00147	0.00199	0.00143	0.00186
300	0.00381	0.0157	0.00147	0.00442	0.00121	0.00266	0.00109	0.00194	0.00112	0.00167	0.00105	0.00147
500	0.00342	0.0154	0.00117	0.00405	0.000952	0.00217	0.000847	0.00157	0.000816	0.00138	0.000789	0.00122

Table S4 (Scenario 4). Average Cramer-von Mises discrepancies $\overline{\overline{\Omega}}$ between the empirical distribution of EB predicted benefits and the theoretical distribution of benefits assuming normality for the random effects, for simulated reference trivariate normal (RTN) distributions and mixtures of two trivariate normal distributions with the same mean and variance-covariance matrix, by distance between mean vectors of the mixture components.

$n = 6$	Distance between means											
	1.4		2.8		4.2		5.7		7.1		8.5	
N	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture
50	0.0282	0.0295	0.0439	0.0665	0.0528	0.115	0.0563	0.149	0.0582	0.158	0.0590	0.160
100	0.0246	0.0253	0.0394	0.0626	0.0502	0.114	0.0525	0.146	0.0530	0.157	0.0536	0.158
150	0.0233	0.0250	0.0386	0.0608	0.0484	0.114	0.0507	0.144	0.0521	0.156	0.0517	0.156
200	0.0228	0.0239	0.0388	0.0611	0.0471	0.113	0.0504	0.145	0.0505	0.156	0.0513	0.157
300	0.0222	0.0237	0.0379	0.0603	0.0466	0.112	0.0498	0.144	0.0512	0.155	0.0510	0.156
500	0.0221	0.0233	0.0374	0.0600	0.0459	0.113	0.0495	0.145	0.0505	0.155	0.0504	0.156

$n = 4$	Distance between means											
	1.4		2.8		4.2		5.7		7.1		8.5	
N	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture	RTN	Mixture
50	0.0303	0.0333	0.0442	0.0686	0.0556	0.119	0.0569	0.148	0.0575	0.159	0.0576	0.162
100	0.0255	0.0281	0.0398	0.0645	0.0495	0.116	0.0525	0.147	0.0527	0.157	0.0540	0.158
150	0.0243	0.0258	0.0394	0.0638	0.0484	0.114	0.0519	0.145	0.0520	0.157	0.0522	0.159
200	0.0222	0.0243	0.0378	0.0613	0.0470	0.114	0.0508	0.145	0.0506	0.156	0.0512	0.157
300	0.0228	0.0237	0.0379	0.0610	0.0471	0.114	0.0502	0.145	0.0502	0.156	0.0512	0.156
500	0.0215	0.0231	0.0375	0.0608	0.0466	0.114	0.0498	0.145	0.0500	0.156	0.0509	0.157

SAS/IML code implementing the graphical approach and calculating Cramer-von Mises statistic for simulation scenario 3.

```
proc iml;

start cal (CvM,Y,EDG, SampSize);

/*1. simulate data*/

/*SampSize = 100; */
n=6;  target = 7;  nfe = 4;  nre = 3;  TIMEPOINT = 4;  df=13;



/*binary x_i*/
x_EDG = randfun(SampSize, "Bernoulli", 0.6);
/*print x_EDG;*/

EDG = colvec(repeat(x_EDG,1,n));
/*print EDG;*/



/*fixed effect*/
beta = {21.4,1.92,-3.97,0.35};
/*beta = {21.4,1.92,-6.97,0.35}; */

/*Mixture*/
/*beta = {22.4,1.92,-3.97,0.35}; */



time_i = {0,0,1,2,3,4};
timesqr_i = time_i##2;
/*print timesqr_i;*/

time = shape(repeat(time_i,SampSize),SampSize*n,1);
timesqr = shape(repeat(timesqr_i,SampSize),SampSize*n,1) ;

/*design matrix X*/
intcp = j(SampSize*n,1,1);
X = intcp || EDG || time || timesqr;
X_design = X;
/*print X;*/

/*design matrix Z*/
Zi = repeat(1,n)||time_i||timesqr_i;
/*Zi = repeat(1,n)||time_i; */
Z = I(SampSize)@Zi;
```

```
Z_design = Z;
/*print Z;*/



/*error term*/
vc = 10*I(n);
/*print vc;*/

Mean= repeat(0,n);
eps = RandNormal(SampSize, Mean, vc);
eps = shape(eps,SampSize*n,1);
/*print eps;*/

/*****************************************************sim
1*****************************************************/

/*MVN/T*/
/*mu = {0  0  0};*/
/*cov = {10.4  0.279  -0.341, 0.279  13.06  -2.466, -0.341  -2.466  0.581}; */
/*rnd = RANDNORMAL(SampSize, mu, cov);*/

/*cov_t = ((df-2)/df)*cov;*/
/*rnd = RandMVT(SampSize, df, mu, cov_t);*/
/*print cov_t;*/


/*****************************************************************************************************
***********/


/*****************************************************sim
2*****************************************************/

/*MVN/T*/
mu = {0  0  0};
cov = {10.4  0.279  -0.341, 0.279  13.06  -2.466, -0.341  -2.466  0.581};
cov_t = (df/(df-2))*cov;
/*print cov_t;*/

/*rnd = RandMVT(SampSize, df, mu, cov_t);*/

rnd = RANDNORMAL(SampSize, mu, cov_t);



/*****************************************************************************************************
***********/

rndA = shape(rnd, SampSize*nre,1);


Y = X*beta+Z*rndA+eps;
/*print Y;*/
```

```
/*Create ID vectors*/
ID = colvec(repeat(T(1:SampSize),1,n)) ;
/*print ID;*/


dataset = ID||Y||EDG||time;
/*print dataset;*/



create MyData from dataset [colname={"id", "Y", "EDG", "time"}];
append from dataset;
close MyData;

/*************************************************************/

/*2. fit mixed*/
submit MyData;
ods listing close;
ods output SolutionF = fixed  SolutionR = rnds  CovParms=CovP;
proc mixed data=MyData method=ml COVTEST;
            class id;
     model Y = EDG time time*time / solution cl covb;
            random intercept time time*time / SUB=id TYPE=un SOLUTION G GCORR V;
run;
ods listing;
endsubmit;
/*************************************************************/



/*3. calculate CvM statistic*/

use rnds;
   read all into rnds[colname=varNames];
close;

intercept_idx= do(1,SampSize*nre,nre);
/*print intercept_idx;*/
time_idx     = do(2,SampSize*nre,nre);
/*print time_idx;*/
timesqr_idx  = do(3,SampSize*nre,nre);
/*print timesqr_idx;*/

EB_intercept= rnds[intercept_idx,"Estimate"];
EB_time     = rnds[time_idx,"Estimate"];
EB_timesqr  = rnds[timesqr_idx,"Estimate"];

/*print EB_intercept,EB_time,EB_timesqr;*/

use Fixed;
           read all var {Estimate} into EB_FIXED;
close;

use CovP;
           read all var {Estimate} into CovPars;
```

```
close;



resid_var = CovPars[nrow(CovPars)];
/*print resid_var;*/


EB_RND = EB_intercept||EB_time||EB_timesqr;

shape_EB_RND = shape(EB_RND,SampSize*nre,1);

mu = X_design*EB_FIXED + Z_design*shape_EB_RND;

/*print mu;*/

Benefit = probnorm( (target - (X_design*EB_FIXED + Z_design*shape_EB_RND)) / sqrt(resid_var) ) ;

Benefit_all = id || EDG || time || Benefit;

/*print Benefit_all;*/

/*************************************************************************************************
************************/
/*************************************GROUP THE BENEFITS by x_i and at
TIMEPOINT*********************************************/

edg_rows_t   = loc((Benefit_all[,2]=1) & (Benefit_all[,3]=TIMEPOINT));
noedg_rows_t = loc((Benefit_all[,2]=0) & (Benefit_all[,3]=TIMEPOINT));


e_ct   = countn(edg_rows_t) ;
NOe_ct = countn(noedg_rows_t) ;
/*print e_ct, NOe_ct;*/

/*obtain EB benefits at TIMEPOINT*/
edg_t   = Benefit_all[edg_rows_t,ncol(Benefit_all)];
noedg_t = Benefit_all[noedg_rows_t,ncol(Benefit_all)];
/*print edg_t;*/
/*print noedg_t;*/

/*****************************************exclude benefits 0 and
1***********************************************/

smallCutoff = 1e-10;

LargeCutoff = 1 - smallCutoff;

idx_e =  loc((edg_t < smallCutoff ) | (edg_t > LargeCutoff));
idx_ne = loc((noedg_t < smallCutoff ) | (noedg_t > LargeCutoff));

/*print idx_e,idx_ne ;*/
```

```
if ncol(idx_e)>0 then do;
   edg_t = edg_t[setdif(1:nrow(edg_t), idx_e)];
   e_ct  = countn(edg_t) ;
end;
else print "There are no positive values";


if ncol(idx_ne)>0 then do;
   noedg_t = noedg_t[setdif(1:nrow(noedg_t),idx_ne)];
   NOe_ct  = countn(noedg_t) ;
end;
else print "There are no positive values";



/*print e_ct, NOe_ct;*/
/*print edg_t, noedg_t;*/


/*************************************************************************************************
************************/



/*mean function of time and binary covariates*/
start Mut (LB, x_EDG, t, Fixed_Efft);

            tsqr = t*t;
            X_LB = 1 || x_EDG || t || tsqr;
            LB = X_LB*Fixed_Efft;

finish Mut;




/*variance function of time and binary covariates*/
start Vart (LB_VAR, t, Varparms);

            tsqr = t*t;
            tcub = t*t*t;
            tfor = t*t*t*t;
            Xvar_LB = 1 || 2*t || tsqr || 2*tsqr || 2*tcub || tfor;
            LB_VAR = Xvar_LB*Varparms[1:(nrow(Varparms)-1)];

finish Vart;




/*CDF of benefits*/
start CDF_b (F_z_cdf,  x_EDG,Fixed_Efft,Varparms, t, z, target);
```

```
        call Mut (LB, x_EDG, t, Fixed_Efft);
        sigma_sqr = Varparms[nrow(Varparms)];
        mean_t = (target - LB)/sqrt(sigma_sqr);
/*      print mean_t;*/

        call Vart (LB_VAR, t,Varparms);
/*      print LB_VAR;*/

        gamma_sqr = LB_VAR/sigma_sqr;
/*      print gamma_sqr;*/

        F_z = quantile('NORMAL',z);
        F_z_t = (F_z - mean_t)/sqrt(gamma_sqr);
        F_z_cdf = cdf('NORMAL',F_z_t);
/*      print F_z, F_z_t, F_z_cdf;*/

finish CDF_b;




/*Cramer-von Mises statistic*/
start CRM_stat (CRM, dataset);

/*obtain order statistics by sorting the CDFs*/
        call sort(dataset, 1);

        cons = j(nrow(dataset),1,0);
        do i = 1 to nrow(dataset);
                cons[i] = (2*i-1)/ (2*nrow(dataset));
        end;
/*      cons1 = dataset - cons;*/
/*      cons2 = (dataset - cons)##2;*/
/*      cons3 = sum((dataset - cons)##2);*/

        CRM = 1/(12*nrow(dataset)*nrow(dataset)) + 1/(nrow(dataset)) * sum((dataset - cons)##2);
/*      print cons, cons1, cons2, cons3, CRM;*/

finish CRM_stat;



/*************************************************************************************************
************************/



edg_t_cdf = j(nrow(edg_t), 1, 0);
do i = 1 to nrow(edg_t);
        call CDF_b (F_z_cdf,  1,EB_FIXED,CovPars, TIMEPOINT, edg_t[i],target);
        edg_t_cdf[i] = F_z_cdf;
```

```
end;


noedg_t_cdf = j(nrow(noedg_t), 1, 0);
do i = 1 to nrow(noedg_t);
        call CDF_b (F_z_cdf,  0,EB_FIXED,CovPars, TIMEPOINT, noedg_t[i],target);
        noedg_t_cdf[i] = F_z_cdf;
end;


/*print edg_t_cdf, noedg_t_cdf;*/



/*****************************************************************************************************
************************/
/*Cramer-von Mises statistic*/

call CRM_stat (edg_CRM, edg_t_cdf);
call CRM_stat (noedg_CRM, noedg_t_cdf);
/*print edg_CRM, noedg_CRM;*/


CvM = (e_ct*edg_CRM + NOe_ct*noedg_CRM)/(e_ct+NOe_ct);
/*print CvM;*/

finish cal;


/*****************************************************************************************************
************************/

sim=500;

store = j(sim,1);
Y_matrix = j(100*6,sim);
EDG_matrix = j(100*6,sim);


do j=1 to sim;

        call cal(CvM,Y,EDG,100);

        store[j]=CvM;
        Y_matrix[,j]=Y;
        EDG_matrix[,j]=EDG;

end;

mean_t = mean(store);

print store,mean_t;

/*print Y_matrix;*/
/*print EDG_matrix;*/
```

```
/*choose the row/column of Y that has the smallest CvM*/
dd = abs(store-mean_t);

lo = loc (dd = min(dd));

if ncol(lo)>1 then do;
   location = lo[1];
end;
else location = lo;

diff = dd[location];

CvM_diff = store[location];
print CvM_diff;

Y = Y_matrix[,location];
EDG = EDG_matrix[,location];


/*************************************************************************************************
*************************/
/*************************************************************************************************
*************************/
/*********************************************************Plot QQ plot
*********************************************************/
/*************************************************************************************************
*************************/
/*************************************************************************************************
*************************/



/*1. formulate data*/

SampSize = 100; n=6;  target = 7;  nfe = 4;  nre = 3;  TIMEPOINT = 4;



time_i = {0,0,1,2,3,4};
timesqr_i = time_i##2;
/*print timesqr_i;*/

time = shape(repeat(time_i,SampSize),SampSize*n,1);
timesqr = shape(repeat(timesqr_i,SampSize),SampSize*n,1) ;

/*design matrix X*/
intcp = j(SampSize*n,1,1);
X = intcp || EDG || time || timesqr;
X_design = X;
/*print X;*/
```

```
/*design matrix Z*/
Zi = repeat(1,n)||time_i||timesqr_i;
/*Zi = repeat(1,n)||time_i; */
Z = I(SampSize)@Zi;
Z_design = Z;
/*print Z;*/



/*Create ID vectors*/
ID = colvec(repeat(T(1:SampSize),1,n)) ;
/*print ID;*/


dataset = ID||Y||EDG||time;
/*print dataset;*/



create MyData from dataset [colname={"id", "Y", "EDG", "time"}];
append from dataset;
close MyData;

/*************************************************************/

/*2. fit mixed*/
submit MyData;
ods listing close;
ods output SolutionF = fixed  SolutionR = rnds  CovParms=CovP;
proc mixed data=MyData method=ml COVTEST;
            class id;
     model Y = EDG time time*time / solution cl covb;
            random intercept time time*time / SUB=id TYPE=un SOLUTION G GCORR V;
run;
ods listing;
endsubmit;
/*************************************************************/



/*3. calculate CvM statistic*/

use rnds;
   read all into rnds[colname=varNames];
close;

intercept_idx= do(1,SampSize*nre,nre);
/*print intercept_idx;*/
time_idx     = do(2,SampSize*nre,nre);
/*print time_idx;*/
timesqr_idx  = do(3,SampSize*nre,nre);
/*print timesqr_idx;*/

EB_intercept= rnds[intercept_idx,"Estimate"];
EB_time     = rnds[time_idx,"Estimate"];
EB_timesqr  = rnds[timesqr_idx,"Estimate"];
```

```
/*print EB_intercept,EB_time,EB_timesqr;*/

use Fixed;
        read all var {Estimate} into EB_FIXED;
close;

use CovP;
        read all var {Estimate} into CovPars;
close;




resid_var = CovPars[nrow(CovPars)];
/*print resid_var;*/


EB_RND = EB_intercept||EB_time||EB_timesqr;

shape_EB_RND = shape(EB_RND,SampSize*nre,1);

mu = X_design*EB_FIXED + Z_design*shape_EB_RND;

/*print mu;*/

Benefit = probnorm( (target - (X_design*EB_FIXED + Z_design*shape_EB_RND)) / sqrt(resid_var) ) ;

Benefit_all = id || EDG || time || Benefit;

/*print Benefit_all;*/

/*************************************************************************************************
************************/



/*mean function of time and binary covariates*/
start Mut (LB, x_EDG, t, Fixed_Efft);

        tsqr = t*t;
        X_LB = 1 || x_EDG || t || tsqr;
        LB = X_LB*Fixed_Efft;

finish Mut;




/*variance function of time and binary covariates*/
start Vart (LB_VAR, t, Varparms);
```

```
        tsqr = t*t;
        tcub = t*t*t;
        tfor = t*t*t*t;
        Xvar_LB = 1 || 2*t || tsqr || 2*tsqr || 2*tcub || tfor;
        LB_VAR = Xvar_LB*Varparms[1:(nrow(Varparms)-1)];

finish Vart;



/*CDF of benefits*/
start CDF_b (F_z_cdf,  x_EDG,Fixed_Efft,Varparms, t, z, target);


        call Mut (LB, x_EDG, t, Fixed_Efft);
        sigma_sqr = Varparms[nrow(Varparms)];
        mean_t = (target - LB)/sqrt(sigma_sqr);
/*      print mean_t;*/

        call Vart (LB_VAR, t,Varparms);
/*      print LB_VAR;*/

        gamma_sqr = LB_VAR/sigma_sqr;
/*      print gamma_sqr;*/

        F_z = quantile('NORMAL',z);
        F_z_t = (F_z - mean_t)/sqrt(gamma_sqr);
        F_z_cdf = cdf('NORMAL',F_z_t);
/*      print F_z, F_z_t, F_z_cdf;*/

finish CDF_b;



/*Cramer-von Mises statistic*/
start CRM_stat (CRM, dataset);

/*obtain order statistics by sorting the CDFs*/
        call sort(dataset, 1);

        cons = j(nrow(dataset),1,0);
        do i = 1 to nrow(dataset);
                cons[i] = (2*i-1)/ (2*nrow(dataset));
        end;
/*      cons1 = dataset - cons;*/
/*      cons2 = (dataset - cons)##2;*/
/*      cons3 = sum((dataset - cons)##2);*/

        CRM = 1/(12*nrow(dataset)*nrow(dataset)) + 1/(nrow(dataset)) * sum((dataset - cons)##2);
/*      print cons, cons1, cons2, cons3, CRM;*/

finish CRM_stat;
```

```
/*quantile function of benefits*/
start quantile_B (B_p_t,  x_EDG,Fixed_Efft,Varparms, t, p, target);

        call Mut (LB, x_EDG, t, Fixed_Efft);

        sigma_sqr = Varparms[nrow(Varparms)];

        mean_t = (target - LB)/sqrt(sigma_sqr);
/*      print mean_t;*/

        call Vart (LB_VAR, t,Varparms);
/*      print LB_VAR;*/

        gamma_sqr = LB_VAR/sigma_sqr;
/*      print gamma_sqr;*/

        q_p   = quantile('NORMAL', p);
        q_p_t = sqrt(gamma_sqr)*q_p + mean_t;
        B_p_t = cdf('NORMAL',q_p_t);

/*      print q_p, q_p_t, B_p_t;*/

finish quantile_B;

/*************************************************************************************************
************************/

/**********************************GROUP THE BENEFITS by x_i and at
TIMEPOINT*****************************************/

edg_rows_t   = loc((Benefit_all[,2]=1) & (Benefit_all[,3]=TIMEPOINT));
noedg_rows_t = loc((Benefit_all[,2]=0) & (Benefit_all[,3]=TIMEPOINT));


e_ct   = countn(edg_rows_t) ;
NOe_ct = countn(noedg_rows_t) ;
/*print e_ct, NOe_ct;*/

/*obtain EB benefits at TIMEPOINT*/
edg_t   = Benefit_all[edg_rows_t,ncol(Benefit_all)];
noedg_t = Benefit_all[noedg_rows_t,ncol(Benefit_all)];
/*print edg_t;*/
/*print noedg_t;*/

/*****************************************exclude benefits 0 and
1***********************************************/

smallCutoff = 1e-10;

LargeCutoff = 1 - smallCutoff;

idx_e =  loc((edg_t < smallCutoff ) | (edg_t > LargeCutoff));
idx_ne = loc((noedg_t < smallCutoff ) | (noedg_t > LargeCutoff));
```

```
/*print idx_e,idx_ne ;*/


if ncol(idx_e)>0 then do;
   edg_t = edg_t[setdif(1:nrow(edg_t), idx_e)];
   e_ct  = countn(edg_t) ;
end;
else print "There are no positive values";


if ncol(idx_ne)>0 then do;
   noedg_t = noedg_t[setdif(1:nrow(noedg_t),idx_ne)];
   NOe_ct  = countn(noedg_t) ;
end;
else print "There are no positive values";



/*print e_ct, NOe_ct;*/
/*print edg_t, noedg_t;*/


edg_t_cdf = j(nrow(edg_t), 1, 0);
do i = 1 to nrow(edg_t);
            call CDF_b (F_z_cdf,  1,EB_FIXED,CovPars, TIMEPOINT, edg_t[i],target);
            edg_t_cdf[i] = F_z_cdf;
end;


noedg_t_cdf = j(nrow(noedg_t), 1, 0);
do i = 1 to nrow(noedg_t);
            call CDF_b (F_z_cdf,  0,EB_FIXED,CovPars, TIMEPOINT, noedg_t[i],target);
            noedg_t_cdf[i] = F_z_cdf;
end;


/*print edg_t_cdf, noedg_t_cdf;*/



/*************************************************************************************************
************************/
/*Cramer-von Mises statistic*/

call CRM_stat (edg_CRM, edg_t_cdf);
call CRM_stat (noedg_CRM, noedg_t_cdf);
/*print edg_CRM, noedg_CRM;*/


CvM = (e_ct*edg_CRM + NOe_ct*noedg_CRM)/(e_ct+NOe_ct);
print CvM;
/*************************************************************************************************
************************/
```

```
/*****************************************************************************************************
**************************/
/*Part 2. QQplot of benefits at TIMEPOINT*/
/*********************************************GROUP THE BENEFITS by
x_i*********************************************/


e_rows   = loc((Benefit_all[,2]=1));
NOe_rows = loc((Benefit_all[,2]=0));


/*count # of subjects in each subgroups*/
e_ct1   = countn(e_rows)/n;
NOe_ct1 = countn(NOe_rows)/n;
/*print e_ct1, NOe_ct1;*/


edg_pt    = Benefit_all[e_rows,];
nonedg_pt = Benefit_all[NOe_rows,];
/*print edg_pt,nonedg_pt;*/

/*****************************************************************************************************
************************/



/*****************************calculate empirical quantiles of the
benefit*************************************/


start cal_quant (q_EB,  data,samp);

        week = {0,0,1,2,3,4};
        p = t(do((1-0.5)/Samp,(Samp-0.5)/Samp,1/Samp));

        data_q_EB  = j(nrow(week),nrow(p),0);
        nrowdata = nrow(data);

        do i = 1 to nrow(week);

                time_index = do(i,nrowdata,nrow(week));
/*              obtain benefits for time "time_index"*/
                temp_EB  = data[time_index,ncol(data)];

                call qntl(temp1, temp_EB, p);
                data_q_EB[i,]  = t(temp1);

                q_EB = week || data_q_EB;

        end;

finish cal_quant;
```

```
/*****************************************************************************************/
/*1*/
call cal_quant (edg_pt_EB, edg_pt,e_ct1);
/*print edg_pt_EB;*/


/*****************************************************************************************/
/*2*/
call cal_quant (nonedg_pt_EB, nonedg_pt,NOe_ct1);
/*print nonedg_pt_EB;*/



/*****************************************************************************************/
/*****************************************************************************************/
/*empirical quantiles at TIMEPOINT*/

edg_pt_EBq     = t(edg_pt_EB[TIMEPOINT+2,2:ncol(edg_pt_EB)]);

nonedg_pt_EBq  = t(nonedg_pt_EB[TIMEPOINT+2,2:ncol(nonedg_pt_EB)]);

/*print edg_pt_EBq, nonedg_pt_EBq;*/


/*****************************calculate theoretical quantiles of the benefit using quantile
function*****************************************/
/************************theoretical benefit quantiles for combinations of
schiz_canna_percentile_time*****************************/



tii = t(do(1,4,1));

/*combinations for EDG_percentile_time*/
percent_EDG  = t(do((1-0.5)/e_ct1,(e_ct1-0.5)/e_ct1,1/e_ct1));
combo_infor_EDG = expandGrid(1, percent_EDG, tii);
B_q_EDG = j(nrow(combo_infor_EDG),1,0);

/*combinations for NOEDG_percentile_time*/
percent_NOEDG = t(do((1-0.5)/NOe_ct1,(NOe_ct1-0.5)/NOe_ct1,1/NOe_ct1));
combo_infor_NOEDG = expandGrid(0, percent_NOEDG, tii);
B_q_NOEDG = j(nrow(combo_infor_NOEDG),1,0);



do i=1 to nrow(combo_infor_EDG);

        EDG_point  = combo_infor_EDG[i,1];
        time_point1= combo_infor_EDG[i,3];
        p_point1   = combo_infor_EDG[i,2];

        call quantile_B(B_p_t_E,  EDG_point,EB_FIXED,CovPars,time_point1,p_point1,target);
        B_q_EDG[i] = B_p_t_E;

end;
```

```
do j=1 to nrow(combo_infor_NOEDG);

        NOEDG_point = combo_infor_NOEDG[j,1];
        time_point2 = combo_infor_NOEDG[j,3];
        p_point2    = combo_infor_NOEDG[j,2];

        call quantile_B(B_p_t_NOE,  NOEDG_point,EB_FIXED,CovPars,time_point2,p_point2,target);
        B_q_NOEDG[j] = B_p_t_NOE;

end;


quan_df_E   = combo_infor_EDG   || B_q_EDG;
quan_df_NOE = combo_infor_NOEDG || B_q_NOEDG;

/*theoretical quantiles for each subgroup AT TIME POINT TIMEPOINT*/
EDG_tqrow   = loc(quan_df_E[,3]=TIMEPOINT);
noEDG_tqrow = loc(quan_df_NOE[,3]=TIMEPOINT);


EDG_tq   = quan_df_E[EDG_tqrow,ncol(quan_df_E)];
noEDG_tq = quan_df_NOE[noEDG_tqrow,ncol(quan_df_NOE)];


/*****************************************************************************************/
/*******************************************QQplot*****************************************/
/*********************************Combine all into one*************************************/
df=13;

EDG_quantile   = repeat(0,nrow(EDG_tq))  || repeat(df,nrow(EDG_tq))  || repeat(1,nrow(EDG_tq))  || EDG_tq   ||
edg_pt_EBq;
noEDG_quantile = repeat(0,nrow(noEDG_tq))|| repeat(df,nrow(noEDG_tq))|| repeat(2,nrow(noEDG_tq))||
noEDG_tq || nonedg_pt_EBq;

all_quantile = EDG_quantile // noEDG_quantile;
/*print all_quantile ;*/

create all_quantile from all_quantile[colname={"t", "df", "group", "Theoretical", "EB"}];
append from all_quantile;
close all_quantile;

submit all_quantile;
proc sgplot data=all_quantile;
  scatter x=Theoretical  y=EB /
    group=group groupdisplay=cluster clusterwidth=0.5;
    lineparm x=0 y=0 slope=1; /** intercept, slope **/
        XAXIS LABEL = 'Theoretical quantile';
        YAXIS LABEL = 'EB quantile';
run;
endsubmit;
```

```
proc export data=all_quantile outfile="C:\Users\zwang3\Desktop\PAPER2\no_trt\Simulation\plot
QQ\mtmvn\mvn13.csv" dbms=csv replace; run;
```