Conditional Estimation for Generalized Linear Models When Covariates Are Subject-specific Parameters in a Mixed Model for Longitudinal Measurements

Erning Li, Daowen Zhang*, **and Marie Davidian**

Department of Statistics, Box 8203, North Carolina State University,

Raleigh, North Carolina 27695-8203, U.S.A.

**email:* dzhang2@stat.ncsu.edu

Summary. The relationship between a primary endpoint and features of longitudinal profiles of a continuous response is often of interest, and a relevant framework is that of a generalized linear model with covariates that are subject-specific random effects in a linear mixed model for the longitudinal measurements. Naive implementation by imputing subject-specific effects from individual regression fits yields biased inference, and several methods for reducing this bias have been proposed. These require a parametric (normality) assumption on the random effects, which may be unrealistic. Adapting a strategy of Stefanski and Carroll (1987 *Biometrika* 74:703–716), we propose estimators for the generalized linear model parameters that require no assumptions on the random effects and yield consistent inference regardless of the true distribution. The methods are illustrated via simulation and by application to a study of bone mineral density in women transitioning to menopause.

The definitive version of this article is available at `www.blackwell-synergy.com` at `http://www.blackwell-synergy.com/doi/abs/10.1111/j.0006-341X.2004.00170.x`

Key words: Conditional score; Longitudinal data; Measurement error; Mixed effects model; Regression calibration; Semiparametric.

1. Introduction

In many studies, a primary endpoint and longitudinal measures of a continuous response are collected for each participant along with other covariates, and the association between the primary endpoint and features of the longitudinal profiles is of interest. For example, in the Study of Women's Health Across the Nation (SWAN, Sowers et al., 2003), investigators wished to understand the association between osteopenia, characterized by bone mineral density (BMD) at or below the 33rd percentile, and features of hormonal patterns over the menstrual cycle in peri-menopausal women. For one cycle, longitudinal progesterone levels derived from urine (PDG) and a single total hip BMD measurement were obtained from 624 women. Because cycle lengths vary among women, it is common (e.g., Zhang et al., 1998) to standardize lengths to a reference of 28 days. In Figure 1a, profiles of log PDG versus standardized cycle time are relatively constant for the first 14 days, with an approximately symmetric rise and fall over days 14–28. Assessing whether log PDG level over the initial 14 days or the later rate of rise and fall are associated with osteopenia would provide insight on hormonal mechanisms that could guide interventions to maintain bone mass and reduce risk of osteoporotic fracture in these women. However, these features are observed only through the log PDG measurements, which are subject to assay error and other variation.

A framework that accounts for this variation is the "joint model" studied by Wang, Wang, and Wang (2000), which assumes that the longitudinal data follow a linear mixed model whose random effects are covariates in a generalized linear model for the primary endpoint. For SWAN, a piecewise linear model for log PDG with random effects representing baseline-to-14 days level and rate of symmetric rise/fall for days 14–28 captures the salient features of woman-specific profiles, and, for the primary endpoint presence/absence of osteopenia, a logistic model incorporating these effects and other covariates is a natural choice.

Wang et al. (2000) showed that "naive" implementation of this model by substituting subject-specific ordinary least squares estimates of the random effects in the primary generalized linear model yields biased inference on its parameters. Thus, viewing this as a measurement error problem, they considered regression calibration (RC; Carroll, Ruppert, and Stefanski, 1995, Ch. 3), where the random effects are replaced in the primary model by "estimated" best linear unbiased predictors from the fit of the mixed model, which reduces but does not eliminate bias. Wang et al. (2000) also proposed a pseudo-expected estimating equation (EEE) approach based on the average of conditional expectations of the primary-model score given the observed data, which requires numerical integration to compute the conditional expectations. To circumvent this integration for logistic or probit regression, they developed an approximate pseudo-EEE that may be viewed as "refined" RC.

These approaches depend on the usual assumptions of normal random effects and intra-subject errors. The latter is often reasonable, perhaps on a transformed scale, but the assumption of normal random effects may be be unrealistic. Several authors (e.g., Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001) have shown that its violation can compromise inference, raising this concern for the joint model. For SWAN, we fitted the above mixed model for log PDG with pre-14-day "intercept" and post-14-day "rise/fall 'slope" random effects by the approach of Zhang and Davidian (2001), which assumes the random effects have a "smooth" but unspecified density that is estimated along with other model parameters. Flexibility of the density estimate is controlled by a tuning parameter K, where $K = 0$ corresponds to normality, chosen via standard model selection criteria, allowing assessment of departure from normality. All criteria reject $K = 0$, and the chosen density estimate in Figure 1b provides visual evidence of a departure, particularly for the "slope."

These considerations suggest that methods for the joint model that do not require a parametric random effects assumption are needed. For a generalized linear model with normal covariate measurement error, Stefanski and Carroll (1987) derived estimating equations by conditioning on certain sufficient statistics that are unbiased regardless of the true covariate distribution. Tsiatis and Davidian (2001) adapted this to the proportional hazards model with longitudinal covariates and demonstrated consistency of inferences on hazard parameters across a range of true distributions.

In this paper, we exploit this approach to develop two strategies for inference in the joint model that require no assumptions on the random effects distribution. In Section 2, we formulate the joint model, we derive the proposed estimators in Section 3, and we explicate popular special cases in Section 4. In Section 5, we compare the methods to those predicated on normality via simulation, and apply them to the SWAN data in Section 6.

2. Joint Model

Denote the observed data for subject i, $i = 1, \ldots, n$, by Y_i, the primary response; $\boldsymbol{Z}_i$, a p-vector of explanatory variables; and longitudinal measurements of a continuous response $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{im_i})^T$ at times $t_{i1}, \ldots, t_{im_i}$. Assume $\boldsymbol{W}_i$ follows the linear mixed model $\boldsymbol{W}_i = \boldsymbol{D}_i \boldsymbol{X}_i + \boldsymbol{U}_i$, where $\boldsymbol{D}_i$ is a full-rank $(m_i \times q)$ design matrix; and $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{im_i})^T$ are within-subject errors reflecting uncertainty in measuring $\boldsymbol{W}_i$, independently and identically with mean zero and variance σ_u^2, i.e., $\boldsymbol{U}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{m_i})$ for $(\ell \times \ell)$ identity matrix $\boldsymbol{I}_\ell$, independent of $\boldsymbol{X}_i$. The $\boldsymbol{X}_i$ are $(q \times 1)$ random effects representing unobserved subject-specific features of the longitudinal profiles; e.g., $\boldsymbol{D}_i$ with rows $(1, t_{ij})$, $j = 1, \ldots, m_i$, and $\boldsymbol{X}_i = (X_{1i}, X_{2i})^T$ yields a linear random coefficient model for the longitudinal data with subject-specific intercept X_{1i} and slope X_{2i}. The primary endpoint and profile features are assumed to be related via a generalized linear model in canonical form; i.e., the conditional density of Y_i given $\boldsymbol{X}_i$ (and $\boldsymbol{Z}_i$; conditioning on $\boldsymbol{Z}_i$ is suppressed throughout) is

$$f(Y_i|\boldsymbol{X}_i; \boldsymbol{\beta}, \phi) = \exp\left\{ \frac{Y_i(\boldsymbol{\beta}_0^T \boldsymbol{Z}_i + \boldsymbol{\beta}_1^T \boldsymbol{X}_i) - b(\boldsymbol{\beta}_0^T \boldsymbol{Z}_i + \boldsymbol{\beta}_1^T \boldsymbol{X}_i)}{a(\phi)} + c(Y_i, \phi) \right\}, \qquad (1)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T)^T$ are regression parameters, ϕ is a dispersion parameter, and $a(\cdot)$, $b(\cdot)$, and $c(\cdot, \cdot)$ are known functions. All variables are independent across i. We assume that Y_i and $\boldsymbol{W}_i$ are conditionally independent given $\boldsymbol{X}_i$ (e.g., Carroll et al, 1995, sec. 1.6).

Interest focuses on estimating $\boldsymbol{\theta} = (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \phi, \sigma_u^2)^T$. Although normality of the within-subject errors $\boldsymbol{U}_i$ may be reasonable, the usual assumption of normal $\boldsymbol{X}_i$ may be a poor representation of the population variation in longitudinal profiles. Consequences for inference on $\boldsymbol{\theta}$ if methods depending on this assumption are used when it is violated are not known. We now develop estimators that do not require any assumption on the distribution of $\boldsymbol{X}_i$.

3. Proposed Methods

3.1 *Sufficiency Estimator*

Because $f(Y_i|\boldsymbol{X}_i;\boldsymbol{\beta},\phi)$ is an exponential family, re-arranging (1) shows that $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$ is a sufficient and complete "statistic" for $\boldsymbol{X}_i$ when all parameters are known. Let $\boldsymbol{E}_i$ $\{m_i \times (m_i - q)\}$ be orthogonal to $\boldsymbol{D}_i$ (i.e., $\boldsymbol{D}_i^T\boldsymbol{E}_i = 0$) with $\boldsymbol{E}_i^T\boldsymbol{E}_i = \boldsymbol{I}_{m_i-q}$, so $\boldsymbol{E}_i$ is determined by $\boldsymbol{D}_i$. Let $\widehat{\boldsymbol{X}}_i = (\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i^T\boldsymbol{W}_i$, the least squares estimator for $\boldsymbol{X}_i$, and $\boldsymbol{R}_i = \boldsymbol{E}_i^T\boldsymbol{W}_i$, a "residual-type" vector. It is straightforward to deduce that $\widehat{\boldsymbol{X}}_i|\boldsymbol{X}_i \sim \mathcal{N}\{\boldsymbol{X}_i, \sigma_u^2(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\}$, $\boldsymbol{R}_i|\boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_u^2\boldsymbol{I})$, and $\boldsymbol{R}_i$ and $\widehat{\boldsymbol{X}}_i$ are independent given $\boldsymbol{X}_i$. Thus, $\boldsymbol{R}_i$ and $\boldsymbol{S}_i$ are independent given $\boldsymbol{X}_i$. As the transformation from $\boldsymbol{W}_i$ to $(\widehat{\boldsymbol{X}}_i, \boldsymbol{R}_i)$ is free of unknown parameters, $f(Y_i, \boldsymbol{W}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) \propto f(Y_i|\boldsymbol{X}_i;\boldsymbol{\beta},\phi)f(\widehat{\boldsymbol{X}}_i|\boldsymbol{X}_i;\sigma_u^2)f(\boldsymbol{R}_i|\boldsymbol{X}_i;\sigma_u^2)$. Also, because the Jacobian of the transformation from $(Y_i, \widehat{\boldsymbol{X}}_i, \boldsymbol{R}_i)$ to $(Y_i, \boldsymbol{S}_i, \boldsymbol{R}_i)$ does not involve unknown parameters, the conditional distribution of data $(Y_i, \boldsymbol{W}_i)$ given $\boldsymbol{S}_i$ is

$$f(Y_i, \boldsymbol{W}_i|\boldsymbol{X}_i, \boldsymbol{S}_i;\boldsymbol{\theta}) = f(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) \propto f(Y_i|\boldsymbol{S}_i)f(\boldsymbol{R}_i). \tag{2}$$

By arguments similar to those leading to (2.5) of Stefanski and Carroll (1987), we have that $f(Y_i|\boldsymbol{S}_i) = \exp[Y_i\eta_i - Y_i^2\sigma_u^2\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\{2a^2(\phi)\} + c(Y_i,\phi) - \log\{H(\eta_i, \boldsymbol{\beta}_1, \phi, \sigma_u^2)\}]$, where $H(\eta_i, \boldsymbol{\beta}_1, \phi, \sigma_u^2) = \int \exp\{y\eta_i - y^2\sigma_u^2\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\{2a^2(\phi)\} + c(y,\phi)\}dm(y)$, $m(y)$ is the dominating measure for Y_i, and $\eta_i = \{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/a(\phi)$. Moreover, because $f(\boldsymbol{R}_i)$ is proportional to the likelihood of σ_u^2 for data $\boldsymbol{R}_i$, $f(\boldsymbol{R}_i) \propto (2\pi\sigma_u^2)^{-(m_i-q)/2}\exp\{-\boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^2)\}$. For simplicity, $\boldsymbol{R}_i$ may be represented by the degenerate version $\boldsymbol{R}_i = \{\boldsymbol{I} - \boldsymbol{D}_i(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{D}_i^T\}\boldsymbol{W}_i$, the least squares residuals. Substituting in (2), we obtain an expression proportional to the conditional density $f(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta})$ and hence the log conditional likelihood for $\boldsymbol{\theta}$.

Following Stefanski and Carroll (1987), $f(Y_i|\boldsymbol{S}_i)$ is also an exponential family; thus $E(Y_i|\boldsymbol{S}_i) = \partial/\partial\eta_i(\log H)$ evaluated at $\eta_i = \{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/a(\phi)$ and $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$. Thus, we can construct an estimating equation for $\boldsymbol{\theta}$ that is unbiased regardless of the distribution of $\boldsymbol{X}_i$ of the form $\sum_{i=1}^n \boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) = \boldsymbol{0}$, where $\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) = \partial/\partial\boldsymbol{\theta}\{\log f(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i)\}$ evaluated at $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$. In particular, the sufficiency score function $\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta})$ is

$$\begin{pmatrix} \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\boldsymbol{Z}_i/a(\phi) \\ \{Y_i - E(Y_i|\boldsymbol{S}_i)\}(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{S}_i/a(\phi) - \{Y_i^2 - E(Y_i^2|\boldsymbol{S}_i)\}(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\sigma_u^2\boldsymbol{\beta}_1/a^2(\phi) \\ -\{Y_i - E(Y_i|\boldsymbol{S}_i)\}\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}a'(\phi)/a^2(\phi) \\ +\{Y_i^2 - E(Y_i^2|\boldsymbol{S}_i)\}\sigma_u^2\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1 a'(\phi)/a^3(\phi) + \partial/\partial\phi\{c(Y_i;\phi)\} - E[\partial/\partial\phi\{c(Y_i;\phi)\}|\boldsymbol{S}_i] \\ -(m_i - q)/(2\sigma_u^2) + \boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^4) - \{Y_i^2 - E(Y_i^2|\boldsymbol{S}_i)\}\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\{2a^2(\phi)\} \end{pmatrix}$$

evaluated at $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$. Following Stefanski and Carroll (1987), we call the solution to such an estimating equation $\widehat{\boldsymbol{\theta}}_S$ a sufficiency estimator.

3.2 *Conditional Estimator*

Alternatively, we consider the conditional score approach (McCullagh and Nelder, 1989, sec. 7.2.2) based on the bias-corrected score function. Treating $\boldsymbol{X}_i$ as a nuisance parameter, the bias-corrected score, $\partial/\partial\boldsymbol{\theta}\log\{f(Y_i, \boldsymbol{W}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\} - E[\partial/\partial\boldsymbol{\theta}\log\{f(Y_i, \boldsymbol{W}_i|\boldsymbol{X}_i;\boldsymbol{\theta})\}|\boldsymbol{S}_i]$, is

$$\begin{pmatrix} \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\boldsymbol{Z}_i/a(\phi) \\ \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\boldsymbol{X}_i/a(\phi) \\ -\{Y_i - E(Y_i|\boldsymbol{S}_i)\}(\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{\beta}_1^T\boldsymbol{X}_i)a'(\phi)/a^2(\phi) + \partial/\partial\phi\{c(Y_i;\phi)\} - E[\partial/\partial\phi\{c(Y_i,\phi)\}|\boldsymbol{S}_i] \\ -(m_i - q)/(2\sigma_u^2) + \boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^4) + \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\{\boldsymbol{\beta}_1^T\boldsymbol{X}_i - \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/\{a(\phi)\sigma_u^2\} \\ \{Y_i^2 - E(Y_i^2|\boldsymbol{S}_i)\}\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\{2a^2(\phi)\} \end{pmatrix}$$

evaluated at $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$. As this involves the unknown $\boldsymbol{X}_i$, we again follow Stefanski and Carroll (1987) and replace $\boldsymbol{X}_i$ by a q-dimensional function $\boldsymbol{t}(\boldsymbol{S}_i)$, which yields the conditional score function $\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta})$ given by

$$\left(\begin{array}{c} \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\boldsymbol{Z}_i/a(\phi) \\ \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\boldsymbol{t}(\boldsymbol{S}_i)/a(\phi) \\ -\{Y_i - E(Y_i|\boldsymbol{S}_i)\}\{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{\beta}_1^T\boldsymbol{t}(\boldsymbol{S}_i)\}a'(\phi)/a^2(\phi) + \partial/\partial\phi\{c(Y_i;\phi)\} - E[\partial/\partial\phi\{c(Y_i,\phi)\}|\boldsymbol{S}_i] \\ -(m_i-q)/(2\sigma_u^2) + \boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^4) + \{Y_i - E(Y_i|\boldsymbol{S}_i)\}\{\boldsymbol{\beta}_1^T\boldsymbol{t}(\boldsymbol{S}_i) - \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/\{a(\phi)\sigma_u^2\} \\ \{Y_i^2 - E(Y_i^2|\boldsymbol{S}_i)\}\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\{2a^2(\phi)\} \end{array}\right)$$

evaluated at $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1/a(\phi)$. As long as $\boldsymbol{t}(\boldsymbol{S}_i)$ depends on $(Y_i, \boldsymbol{W}_i)$ only through $\boldsymbol{S}_i$, $\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta})$ has mean zero; thus, we may form an unbiased estimating equation $\sum_{i=1}^n \boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i;\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$. As in Stefanski and Carroll (1987), we call $\widehat{\boldsymbol{\theta}}_C$ solving this equation a conditional estimator. Because $\widehat{\boldsymbol{X}}_i$ is unbiased for $\boldsymbol{X}_i$ and $\boldsymbol{S}_i$ is complete and sufficient for $\boldsymbol{X}_i$, a natural choice is $\boldsymbol{t}(\boldsymbol{S}_i) = E(\widehat{\boldsymbol{X}}_i|\boldsymbol{S}_i) = (\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\{\boldsymbol{S}_i - E(Y_i|\boldsymbol{S}_i)\sigma_u^2\boldsymbol{\beta}_1/a(\phi)\}$, a uniformly minimum variance unbiased estimator (UMVUE) for $\boldsymbol{X}_i$.

3.3 *Inference*

Both the sufficiency and conditional score estimators effectively "condition away" dependence on $\boldsymbol{X}_i$ so require no assumption on the distribution of $\boldsymbol{X}_i$. In general, $\boldsymbol{\psi}_S$ and $\boldsymbol{\psi}_C$ differ regardless of the choice of $\boldsymbol{t}(\boldsymbol{S}_i)$. Solution of either equation may be implemented via the Newton-Raphson algorithm without much computational burden. Similar to Stefanski and Carroll (1987) and Tsiatis and Davidian (2001), the equations may have multiple roots, not all of which are consistent for $\boldsymbol{\theta}$. We have found that by using a naive estimator as the starting value, as in Tsiatis and Davidian (2001), a consistent solution may be identified, demonstrated in Section 5. Moreover, as both are M-estimators, under regularity conditions, such consistent solutions should be asymptotically normal, and standard errors for either estimator may thus be obtained by using the usual empirical sandwich approach (e.g., Carroll et al., 1995, sec. A.3.1). We show that this technique yields reliable inferences in Section 5.

4. Normal and Logistic Models

We elucidate the form of the sufficiency and conditional scores in two popular cases.

4.1 *Normal Primary Model*

Suppose $Y_i|\boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{\beta}_1^T\boldsymbol{X}_i, \phi)$. Then the dominating measure $m(\,\cdot\,)$ for Y_i is Lebesgue measure; and, in (1), $a(\phi) = \phi$, $b(\rho) = \rho^2/2$, and $c(y, \phi) = -\{y^2/\phi + \log(2\pi\phi)\}/2$. Hence $\eta_i = \{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/\phi$. Substituting in the expression for $f(Y_i|\boldsymbol{S}_i)$ in Section 3.1, it follows that $Y_i|\boldsymbol{S}_i \sim \mathcal{N}(\mu_i, v_i)$, where $\mu_i = \{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}/\{1 + \sigma_u^2\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^t\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\phi\}$, and $v_i = \phi/\{1 + \sigma_u^2\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^t\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/\phi\}$, so that $E(Y_i|\boldsymbol{S}_i) = \mu_i$ and $E(Y_i^2|\boldsymbol{S}_i) = \mu_i^2 + v_i$. Substituting these results, $a(\phi) = \phi$, and $a'(\phi) = 1$ in $\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta})$ yields the sufficiency score estimating equation for $\boldsymbol{\theta}$; similarly, substitution in $\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta})$ with $\boldsymbol{t}(\boldsymbol{S}_i) = (\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}(\boldsymbol{S}_i - \mu_i\sigma_u^2\boldsymbol{\beta}_1/\phi)$ gives the conditional score equation.

4.2 *Logistic Primary Model*

Suppose Y_i is binary (=0,1) and $P(Y_i = 1|\boldsymbol{X}_i) = [1 + \exp\{-(\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{\beta}_1^T\boldsymbol{X}_i)\}]^{-1}$. Then $a(\phi) = 1$, $b(\rho) = \log(1 + e^\rho)$, and $c(y, \phi) = 1$ in (1); $m(\,\cdot\,)$ is counting measure on (0,1); and $\eta_i = \{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + \boldsymbol{S}_i^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}$. Substitution in $f(Y_i|\boldsymbol{S}_i)$ in Section 3.1 yields $P(Y_i = 1|\boldsymbol{S}_i) = [1 + \exp\{-\boldsymbol{\beta}_0^T\boldsymbol{Z}_i + (\boldsymbol{S}_i - \frac{1}{2}\sigma_u^2\boldsymbol{\beta}_1)^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1\}]^{-1} = \mu_i$, say, which is also a logistic model. Hence, $E(Y_i|\boldsymbol{S}_i) = E(Y_i^2|\boldsymbol{S}_i) = \mu_i$. As ϕ is known, the third entry in each of $\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta})$ and $\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta})$ may be ignored; combining with the foregoing results, we obtain

$$\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta}) = \begin{pmatrix} (Y_i - \mu_i)\boldsymbol{Z}_i \\ (Y_i - \mu_i)(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}(\boldsymbol{S}_i - \sigma_u^2\boldsymbol{\beta}_1) \\ -(m_i - q)/(2\sigma_u^2) + \boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^4) - (Y_i - \mu_i)\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1/2 \end{pmatrix},$$

$$\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i|\boldsymbol{S}_i; \boldsymbol{\theta}) = \begin{pmatrix} (Y_i - \mu_i)\boldsymbol{Z}_i \\ (Y_i - \mu_i)(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}(\boldsymbol{S}_i - \mu_i\sigma_u^2\boldsymbol{\beta}_1) \\ -(m_i - q)/(2\sigma_u^2) + \boldsymbol{R}_i^T\boldsymbol{R}_i/(2\sigma_u^4) + (Y_i - \mu_i)\boldsymbol{\beta}_1^T(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\boldsymbol{\beta}_1(1/2 - \mu_i) \end{pmatrix},$$

with $\boldsymbol{t}(\boldsymbol{S}_i) = (\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}(\boldsymbol{S}_i - \mu_i\sigma_u^2\boldsymbol{\beta}_1)$, both evaluated at $\boldsymbol{S}_i = \boldsymbol{D}_i^T\boldsymbol{W}_i + Y_i\sigma_u^2\boldsymbol{\beta}_1$.

5. Simulation Studies

We carried out several simulations for the logistic primary model $P(Y_i = 1|\boldsymbol{X}_i) = [1 + \exp\{-(\beta_0 + \boldsymbol{\beta}_1^T\boldsymbol{X}_i)\}]^{-1}$ with $\boldsymbol{X}_i = (X_{1i}, X_{2i})$ and $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T$. In all cases, $E(X_{1i}) =$

$E(X_{2i}) = 0.5$, $\text{var}(X_{1i}) = 1.0$, $\text{var}(X_{2i}) = 0.64$, $\text{cov}(X_{1i}, X_{2i}) = -0.2$, $\beta_0 = -2.5$, and $\boldsymbol{\beta}_1 = (3.0, 2.0)^T$; longitudinal data were generated according to $W_{ij} = X_{1i} + X_{2i}t_{ij} + U_{ij}$, where $t_{ij} \sim \mathcal{N}(j-1, 0.1)$ for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, and $U_{ij} \sim \mathcal{N}(0, \sigma_u^2)$. In the simulations reported here, $n = 500$; $\sigma_u^2 = 0.5$, and $m_i \leq 5$; five W_{ij} were generated for each subject i, with each arbitrarily missing with probability 0.05 to create imbalance. Each scenario below involved 1000 Monte Carlo data sets. Viewing this as a measurement error problem, the magnitude of the measurement error may be gauged informally by the "reliability ratio" defined by $\lambda_k = \text{var}(X_{ki})/E\{\text{var}(\widehat{X}_{ki})\}$ for $k = 1, 2$. For all scenarios, $(\lambda_1, \lambda_2) = (0.72, 0.92)$, suggesting moderate (mild) measurement error associated with $\widehat{X}_{1i}$ ($\widehat{X}_{2i}$).

To investigate performance under a range of true $\boldsymbol{X}_i$ distributions, we considered four scenarios: (a) $\boldsymbol{X}_i$ bivariate normal; (b) $\boldsymbol{X}_i$ distributed as a bimodal 50-50 mixture of normals; (c) $\boldsymbol{X}_i$ following the bivariate skew-normal (Azzalini and Dalla Valle, 1996), with coefficients of skewness -0.10 and 0.85 for X_{1i} and X_{2i}, respectively; and (d) $\boldsymbol{X}_i$ bivariate t with 5 degrees of freedom. In each case, the distribution was scaled to have the moments above. For each data set under each scenario, $\boldsymbol{\theta} = (\beta_0, \beta_{11}, \beta_{12}, \sigma_u^2)^T$ was estimated five ways: (i) by the naive approach in Section 1 (N); (ii) using regression calibration (RC) (Wang et al., 2000, sec. 4.1); (iii) by refined regression calibration (Wang et al., 2000, sec. 4.2), which applies a probit approximation to the logistic to approximate the pseudo-EEE (RR); (iv) via the sufficiency score method (SS); and (v) via the conditional score method (CS). Methods (ii) and (iii) require normality of $\boldsymbol{X}_i$, so their performance under violations of this assumption is of interest. Standard errors for (i) were obtained from usual logistic regression formulæ, ignoring least-squares imputation of $\boldsymbol{X}_i$; those for (ii)–(v) were derived from the empirical sandwich approach. 95% Wald confidence intervals were constructed using the standard normal critical value of 1.96. The naive estimate was used as the starting value for (iii)–(v).

From Table 1, in all cases, the naive estimator shows considerable bias, while bias of the proposed estimators (iv) and (v) is negligible. The RC estimator is also unacceptably biased except for bimodal $\boldsymbol{X}_i$; for normal $\boldsymbol{X}_i$, this estimator shows appreciable bias, verifying concerns of Wang et al. (2000). Under normality, the RR and proposed estimators exhibit negligible bias and similar performance. However, under departures from normality, RC and RR can show degradation of performance. Under bimodality, although the RC estimator is remarkably unbiased for β_0 and β_{11}, coverages are well below the nominal level. Interestingly, here, the RR estimator (iii) shows considerable bias, and standard errors vastly under-represent true sampling variation. For scenario (c), with only mild skewness in the X_{2i} dimension, the RR estimator achieves good performance; however, for the "heavy-tailed" bivariate t, although bias of the RR estimator is not large (although larger than that of SS and CS), inferences are flawed, with coverage falling far short of nominal.

We obtained similar results for other choices of n, $\max m_i$, and $(\beta_0, \boldsymbol{\beta}_1)$; for larger σ_u^2, bias and failure to achieve nominal coverage of the naive, RC, and RR estimators worsen and are less pronounced for smaller σ_u^2. The sufficiency and conditional score estimators always show negligible bias and attain nominal coverage regardless of scenario. In all cases, the starting value strategy encountered no problems identifying the apparent consistent solution.

Overall, under the nonlinear (in $\boldsymbol{X}_i$) logistic model, RC and RR show poor performance. On the other hand, for the linear (in $\boldsymbol{X}_i$), normal primary model with ϕ estimated, for which there is no RR estimator, simulations not reported show that while the naive estimator is still biased, the RC estimator shows little bias in all cases and achieves slightly better performance than SS and CS under normality but is inefficient otherwise. For a related linear model, results of Buonaccorsi, Demidenko, and Tosteson (2000) suggest that, under normal random effects, RC is a pseudo-likelihood estimator for $\boldsymbol{\beta}$, as also noted by Wang et al. (2000), so should be consistent and nearly efficient. Buonaccorsi et al. (2000) further show for their model that the RC estimator is consistent but with likely loss of precision, coinciding with our simulation results and suggesting that a similar property holds here.

6. Application to SWAN

In addition to osteopenia status, where $Y_i = 1$ denotes absence of osteopenia (BMD above the 33rd percentile) and $Y_i = 0$ denotes presence for woman i, $i = 1, \ldots, n = 624$, a number of baseline covariates and longitudinal log PDG measurements were collected; numbers of longitudinal measurements m_i ranged from 8 to 25. For woman i, let $Z_{\mathrm{BMI},i}$ denote body mass index (BMI, kg/m^2), $Z_{\mathrm{age},i}$ denote age (years), and $Z_{\mathrm{B},i}$, $Z_{\mathrm{W},i}$ and $Z_{\mathrm{C},i}$ denote ethnicity indicator variables $= 1$ as woman i as woman i is Black, White, or Chinese; $Z_{\mathrm{B},i} = Z_{\mathrm{W},i} = Z_{\mathrm{C},i} = 0$ indicates woman i is Japanese. Let $\boldsymbol{Z}_i = (1, Z_{\mathrm{BMI},i}, Z_{\mathrm{age},i}, Z_{\mathrm{B},i}, Z_{\mathrm{W},i}, Z_{\mathrm{C},i})^T$.

From Figure 1a, a reasonable representation of woman i's log PDG profile is

$$W_{ij} = X_{1i} + X_{2i}(t_{ij} - 1.4)_+ - 2X_{2i}(t_{ij} - 2.1)_+ + U_{ij}, \qquad (3)$$

where $u_+ = u$ if $u > 0$ and 0 otherwise, time is in units of 10 days, U_{ij} is a $\mathcal{N}(0, \sigma_u^2)$ measurement error, X_{1i} is the woman-specific "true" (error-free) log PDG up to day 14, and X_{2i} is the woman-specific "slope" of the symmetric rise (days 14–21) and fall (days 21–28) depicted in Figure 1a. To describe the relationship between propensity for osteopenia and $\boldsymbol{Z}_i$ and $\boldsymbol{X}_i = (X_{1i}, X_{2i})^T$, we considered the primary logistic regression model

$$P(Y_i = 1|\boldsymbol{X}_i) = [1 + \exp\{-(\boldsymbol{\beta}_0^T \boldsymbol{Z}_i + \boldsymbol{\beta}_1^T \boldsymbol{X}_i)\}], \qquad (4)$$

where $\boldsymbol{\beta}_0 = (\beta_0, \beta_{0,\mathrm{BMI}}, \beta_{0,\mathrm{age}}\beta_{0,\mathrm{B}}, \beta_{0,\mathrm{W}}, \beta_{0,\mathrm{C}})^T$ and $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T$.

Table 2 shows results from fitting the joint model (3), (4) by the methods in Section 5. Inferences on the coefficients of baseline covariates are similar across all methods and indicate strong association between absence of osteopenia and BMI after taking into account age, ethnicity, and hormone profile, consistent with observations in previous studies. Inferences on the relationship between osteopenia and pre-14-day log PDG level and post-14-day rise/fall are also similar, although point estimates are more disparate across methods. The proposed estimators suggest that absence of osteopenia may be positively associated with post-14-day "slope" of log PDG profile, after taking into account other factors.

The consistency across methods here may not be unexpected. The estimated "reliability ratio" defined in Section 5 is $\lambda = (0.88, 0.84)$ for these data, so the magnitude of "measurement error" in both components is not great. From Figure 1, although bimodal, the estimated density for $\boldsymbol{X}_i$ does not deviate considerably from normality; coefficients of skewness based on the density estimate are $(-0.13, -0.87)$ for $(X_{1i}, X_{2i})^T$, comparable in magnitude to the bivariate skew-normal scenario (c) in Section 5, where RR, SS, and CS performed similarly. To investigate further, we carried out a small simulation based on the estimates and design for the data, taking $\boldsymbol{U}_i$ and $\boldsymbol{X}_i$ to be bivariate normal, and found that all methods yielded similar inferences. Evidently, the design and variation/correlation configuration for these data leads to only minimal differences among the procedures. Nonetheless, use of the proposed methods, which, unlike the competitors are consistent under any conditions, offers the analyst assurance of credible estimation of the relationship.

7. Discussion

We have proposed inferential strategies for generalized linear models for a primary outcome with covariates that are underlying subject-specific random effects in a linear mixed model for a longitudinal response. The methods need no assumptions on the distribution of the random effects and are straightforward and fast to implement, requiring only a few seconds for data on 500 subjects. In contrast to methods predicated on a parametric (normality) assumption for the random effects, the methods yield valid inferences under departures from this assumption and are competitive when the assumption holds. While the refined regression calibration estimator of Wang et al. (2000), which performed the best in our simulations among competing estimators, is only available for the logistic and probit models, our approach applies to any generalized linear model formulation. For all methods studied, normality of within-subject longitudinal data measurement errors, which is often reasonable, is assumed, so the analyst should take care to explore the relevance of this assumption.

The proposed estimators are semiparametric in that the random effects distribution is completely unspecified. Results of Stefanski and Carroll (1987, sec. 3) and Lindsay and Lesperance (1995) suggest that the conditional score estimator for our model, with $\boldsymbol{t}(\boldsymbol{S}_i) = E(\boldsymbol{X}_i|\boldsymbol{S}_i)$, achieves the semiparametric efficiency bound. A modification along the lines suggested by Lindsay (1985) may offer further improvment over our approach. We focused on methods that are straightforward to implement; an alternative would be a full likelihood procedure, which involves increased computational burden. One possibility is to assume that the random effects have a "smooth" but unspecified density (e.g., Zhang and Davidian, 2001); if this assumption is valid, one would expect efficiency gains over the methods proposed here, which do not restrict the class of random effects distributions. Further research on formal methods for assessing possible departures from normal random effects and evaluating the veracity of apparent multimodality as in Figure 1b (e.g., using Chaudhuri and Marron, 1999) would be useful. These topics are the subject of current research to be reported elsewhere.

Acknowledgements

This work was supported by NIH grants R01-CA085848 and R37-AI31789. The authors are grateful to MaryFran Sowers for providing the SWAN data.

References

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726.

Buonaccorsi, J., Demidenko, E., and Tosteson, T. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica* **10**, 885–903.

Carroll, R.J., Ruppert. D., and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**, 807–823.

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

Lindsay, B.G. (1985). Using empirical partially Bayes inference for increasing efficiency. *Annals of Statistics* **13**, 914–931.

Lindsay, B.G. and Lesperance, M. (1995). A review of semiparametric mixture models. *Journal of Statistical Planning and Inference* **47**, 29–39.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.

Sowers, M.R., Finkelstein, J., Ettinger, B. Bondarenko, I., Neer, R., Cauley, J., Sherman, S., and Greendale, G. (2003). The association of endogenous hormone concentrations and bone mineral density measures in pre- and peri-menopausal women of four ethnic groups: SWAN. *Osteoporosis International* **14**, 44–52.

Stefanski, L.A. and Carroll, R.J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703–716.

Tsiatis, A.A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random effects distribution in linear mixed effects models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541–556.

Wang, C.Y., Wang, N., and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* **56**, 487–495.

Zhang, D., Lin, X., Raz, J., and Sowers, M.F. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93**, 710–719.

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

Table 1

Simulation results for four underlying random effect distributions for $P(Y_i = 1|\boldsymbol{X}_i) = 1/\{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}$, with true values $\beta_0 = -2.5$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T$, $\sigma_u^2 = 0.5$, $n = 500$, RB, relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score.

	Method	RB (%)	SD	SE	CP	RB (%)	SD	SE	CP
		(a) Normal				(b) Bimodal mixture			
β_0	N	−25.3	0.22	0.22	0.20	−14.0	0.25	0.25	0.69
	RC	−17.3	0.23	0.23	0.51	−4.4	0.27	0.26	0.89
	RR	1.4	0.38	0.37	0.95	39.5	1.05	0.72	0.95
	SS	3.0	0.41	0.42	0.97	2.8	0.37	0.36	0.96
	CS	2.6	0.38	0.41	0.97	3.0	0.38	0.37	0.95
β_{11}	N	−32.0	0.18	0.19	0.01	−18.3	0.22	0.21	0.30
	RC	−17.0	0.23	0.24	0.41	0.1	0.28	0.26	0.92
	RR	2.1	0.44	0.43	0.96	46.3	1.32	0.91	0.97
	SS	3.7	0.49	0.50	0.97	2.5	0.36	0.36	0.95
	CS	3.2	0.44	0.50	0.96	2.6	0.37	0.37	0.95
β_{12}	N	−14.5	0.20	0.20	0.67	−8.0	0.23	0.23	0.87
	RC	−17.0	0.21	0.21	0.58	−11.6	0.23	0.23	0.80
	RR	1.7	0.30	0.30	0.97	28.6	0.74	0.52	0.96
	SS	3.1	0.32	0.33	0.98	2.6	0.33	0.32	0.95
	CS	2.8	0.30	0.32	0.97	2.7	0.33	0.33	0.96
		(c) Skewed				(d) Bivariate t_5			
β_0	N	−25.0	0.21	0.22	0.20	−28.9	0.21	0.21	0.11
	RC	−17.1	0.23	0.23	0.53	−21.3	0.23	0.23	0.35
	RR	1.8	0.40	0.37	0.95	−6.6	0.37	0.33	0.84
	SS	3.3	0.45	0.43	0.96	3.5	0.47	0.45	0.97
	CS	3.1	0.42	0.42	0.95	3.4	0.44	0.45	0.96
β_{11}	N	−32.1	0.19	0.19	0.01	−36.5	0.18	0.19	0.00
	RC	−17.1	0.24	0.24	0.40	−22.5	0.23	0.23	0.19
	RR	1.9	0.47	0.43	0.94	−7.9	0.41	0.37	0.81
	SS	3.4	0.56	0.51	0.95	3.9	0.57	0.56	0.97
	CS	3.2	0.51	0.51	0.94	3.7	0.52	0.55	0.95
β_{12}	N	−14.5	0.21	0.20	0.68	−17.0	0.20	0.21	0.59
	RC	−17.2	0.21	0.21	0.60	−19.3	0.21	0.21	0.53
	RR	1.6	0.33	0.30	0.94	−4.4	0.30	0.28	0.90
	SS	2.9	0.35	0.33	0.96	3.6	0.35	0.35	0.97
	CS	2.8	0.33	0.33	0.95	3.4	0.34	0.35	0.97

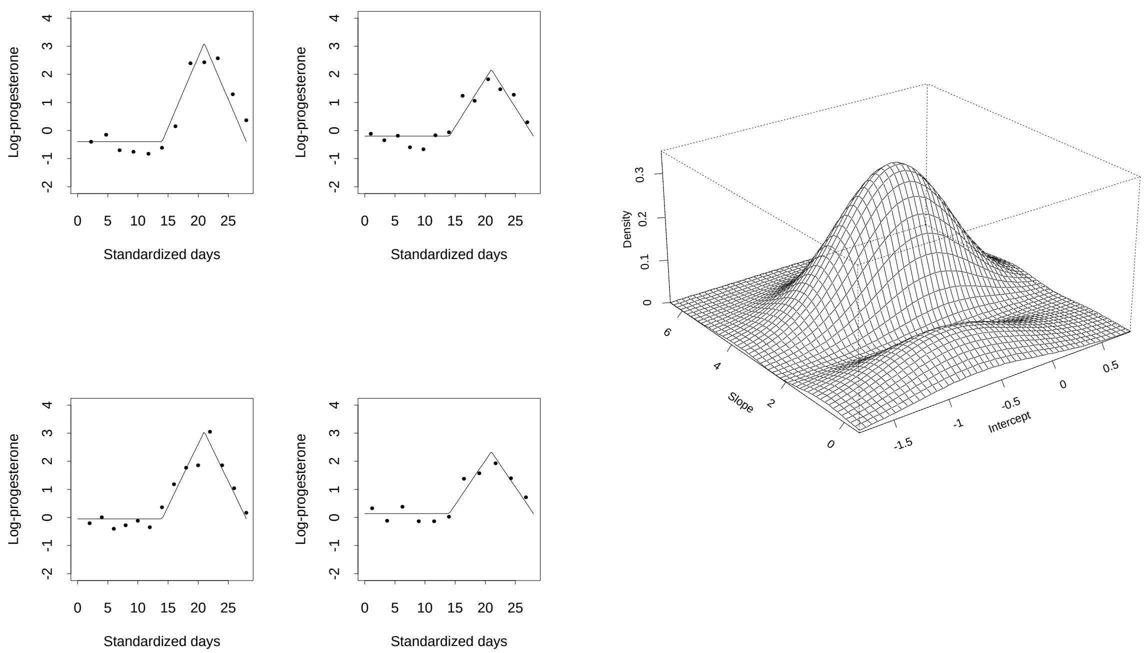
Table 2

Parameter estimates for the SWAN data analysis via several methods; abbreviations for methods are as in Table 1. Estimated standard errors are in parentheses below each estimate. Standard errors for estimates of σ_u^2 are multiplied by 10. For the naive method, σ_u^2 was estimated by mean square of pooled individual least squares residuals.

	N		RC		RR		SS		CS	
	Est.	p-val.	Est.	p-val.	Est.	p-val.	Est.	p-val.	Est.	p-val.
β_0	−7.12	0.00	−7.28	0.00	−7.28	0.00	−7.32	0.00	−7.32	0.00
	(2.16)		(2.18)		2.21)		(2.25)		(2.25)	
$\beta_{0,\text{BMI}}$	0.29	0.00	0.29	0.00	0.29	0.00	0.29	0.00	0.29	0.00
	(0.03)		(0.03)		(0.04)		(0.04)		(0.04)	
$\beta_{0,\text{age}}$	0.01	0.75	0.02	0.73	0.02	0.73	0.02	0.70	0.02	0.70
	(0.04)		(0.04)		(0.04)		(0.04)		(0.04)	
$\beta_{0,B}$	0.54	0.39	0.54	0.39	0.55	0.39	0.55	0.39	0.55	0.39
	(0.17)		(0.16)		(0.16)		(0.16)		(0.16)	
$\beta_{0,W}$	−0.21	0.43	−0.21	0.44	−0.21	0.45	−0.20	0.46	−0.20	0.46
	(0.27)		(0.27)		(0.28)		(0.28)		(0.28)	
$\beta_{0,C}$	−0.68	0.01	−0.68	0.01	−0.68	0.01	−0.68	0.01	−0.68	0.01
	(0.25)		(0.25)		(0.25)		(0.25)		(0.25)	
β_{11}	0.16	0.37	0.13	0.56	0.13	0.56	0.15	0.50	0.15	0.50
	(0.19)		(0.22)		(0.22)		(0.22)		(0.22)	
β_{12}	0.11	0.08	0.14	0.07	0.14	0.08	0.13	0.11	0.13	0.11
	(0.06)		(0.08)		(0.08)		(0.08)		(0.08)	
σ_u^2	0.33	–	0.33	0.00	0.33	0.00	0.33	0.00	0.33	0.00
	(–)		(0.05)		(0.05)		(0.12)		(0.12)	

FIGURE CAPTIONS

Figure 1. **a.** Natural logarithm of progesterone levels measured in urine (PDG) for four randomly selected women plotted against days in standardized menstrual cycle. Individual ordinary least squares fits of the piecewise linear model are superimposed. **b.** Estimated joint density of woman-specific intercepts (level up to day 14) and slopes (rate of change for days 14–21 and negative rate of change for days 21–28) in a linear mixed model for log PDG using the approach of Zhang and Davidian (2001).


Log-progesterone
Standardized days
Density
Slope
Intercept

(a) (b)

Figure 1.