STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:221–239
Published online 14 November 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3478

Fixed effects, random effects and GEE: What are the differences?

Joseph C. Gardiner[1,*,†], Zhehui Luo[1,2] and Lee Anne Roman[3,4]

[1]*Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, MI 48824, U.S.A.*
[2]*RTI International, Behavioral Health Economics Program, Research Triangle Park, NC 27709, U.S.A.*
[3]*Department of Obstetrics and Gynecology, College of Human Medicine, Michigan State University, East Lansing, MI 48824, U.S.A.*
[4]*Institute for Healthcare Studies, Michigan State University, East Lansing, MI 48824, U.S.A.*

SUMMARY

For analyses of longitudinal repeated-measures data, statistical methods include the random effects model, fixed effects model and the method of generalized estimating equations. We examine the assumptions that underlie these approaches to assessing covariate effects on the mean of a continuous, dichotomous or count outcome. Access to statistical software to implement these models has led to widespread application in numerous disciplines. However, careful consideration should be paid to their critical assumptions to ascertain which model might be appropriate in a given setting. To illustrate similarities and differences that might exist in empirical results, we use a study that assessed depressive symptoms in low-income pregnant women using a structured instrument with up to five assessments that spanned the pre-natal and post-natal periods. Understanding the conceptual differences between the methods is important in their proper application even though empirically they might not differ substantively. The choice of model in specific applications would depend on the relevant questions being addressed, which in turn informs the type of design and data collection that would be relevant. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: linear mixed model; generalized linear mixed model; random effects; fixed effects; robust variance; conditional maximum likelihood; Hausman test; CES-D

1. INTRODUCTION

In longitudinal studies each subject is assessed on the same qualitative or quantitative response at several points in time, and the objective is often to characterize the changes in the outcome over

*Correspondence to: Joseph C. Gardiner, Division of Biostatistics, Department of Epidemiology, B629 West Fee Hall, Michigan State University, East Lansing, MI 48824, U.S.A.
†E-mail: jgardiner@epi.msu.edu, gardine3@msu.edu

Contract/grant sponsor: The Agency for Healthcare Research & Quality; contract/grant number: 1R01 HS14206
Contract/grant sponsor: The Maternal and Child Health Bureau (title V, Social Security Act), Health Resources and Services Administration, Department of Health and Human Services; contract/grant number: MCJ-260743

Received 27 July 2007
Accepted 15 September 2008

time and assess the significant determinants or predictors of the change. The outcomes in the ith subject are a vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ with time-ordered components Y_{ij} denoting the outcome assessed at time t_{ij}. In repeated-measures designs the same type of response is evaluated under different conditions. For example, in a 3-period crossover study, (Y_{i1}, Y_{i2}, Y_{i3}) are the responses to treatments in periods 1, 2 and 3, respectively, where 'period' has the same meaning for all subjects. The objective in growth curve analyses is to model the expected response as a function of time. In the classic example of growth assessments in boys and girls by Potthoff and Roy [1], Y_{ij} is the distance in millimeters from the center of the pituitary to the pterygomaxillary fissure at ages $j = 8, 10, 12$ and 14 years. Panel data models [2, 3] also fall within the same general purview where in standard notation Y_{it} is the outcome measure at time t and one has in a balanced panel a fixed grid of time points $t = 1, \ldots, T$ that represent T years or months or some other time unit.

In these examples dependence exists between the responses Y_{ij} within the same subject and one needs to apply methods that take into account the correlation in statistical analyses. However, the extent to which this dependence should be acknowledged will depend upon the objectives of the analysis. For example, if interest lies primarily on the population response means and the impact of covariates on these means, then a very detailed consideration of the dependence structure might be unnecessary and one could opt for robust inference, which does not depend on specification of the covariance of $\mathbf{Y}_i$. On the other hand when subject-specific inference is desired, for instance in estimating the growth trajectories of individual children, a careful evaluation is warranted for deciding upon an appropriate covariance structure. In practice paucity of data would also preclude use of an unstructured covariance for $\mathbf{Y}_i$ because the available data might not support estimation of the $\frac{1}{2}n_i(n_i+1)$ different variances and covariances.

This paper discusses the structural similarities and dissimilarities of the random effects (RE) model [2, 4], the linear mixed model [5, 6], the fixed effects (FE) model [2, 3] and the method of generalized estimating equations (GEE) [7, 8] in addressing correlation in longitudinal data. The motivation of this paper stems from our analysis of a randomized trial of a nurse–community health worker team intervention in low-income pregnant women to reduce depressive symptoms and stress and improve their psychosocial resources during and after their pregnancy [9]. The control condition was community standard of care that included professional care coordination and home visiting in the context of a state-sponsored maternal and infant support program. Depressive symptoms, stress and mastery were assessed using structured instruments at 6 and $1\frac{1}{2}$ months prior to the estimated delivery date, and subsequently at $1\frac{1}{2}$, 6 and 12 months after delivery. Although effort was made to adhere to these protocol times there was some variation in the time of actual assessments. Our objective is to deal with this variation and use all available measurements to assess changes in outcomes in the intervention and control groups. In this paper we focus on depressive symptoms as assessed by the CES-D scale (Center for Epidemiologic Studies—Depression) [10]. First, we analyze the responses Y_{ij} at the 5 waves $j = 1, 2, 3, 4, 5$ as a continuous measure using linear RE, FE and GEE models. Second, we use the cutoff of 16 on the CES-D scale (range 0–60) to define a dichotomous response of indication for depression at each wave and compare results from the nonlinear RE, FE and GEE models.

Section 2 introduces notation and provides a succinct description of the RE, FE and GEE models followed by a discussion of their similarities and differences in applications. In Section 3 we describe our application to the aforementioned nurse–community health worker team study and a comparison of results of fitting different models. The last section is devoted to discussion and conclusions.

DOI: 10.1002/sim

2. MODELS

The models described in this paper are for a random draw $(\mathbf{Y}_i, \mathbf{X}_i)$ from the population of interest, where typically the index i denotes the sampling unit, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ the time-ordered $n_i \times 1$ vector of responses and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$ an $n_i \times p$ matrix of explanatory variables with $\mathbf{x}_{ij}$ a $p \times 1$ vector associated with the response Y_{ij}. The conditional mean vector and covariance matrix are, respectively, $\boldsymbol{\mu}_i = E(\mathbf{Y}_i|\mathbf{X}_i)$ and $\mathbf{V}_i = E[(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)'|\mathbf{X}_i]$.

In this notation each component of the conditional mean $\mu_{ij} = E(Y_{ij}|\mathbf{X}_i)$ is a function of all the covariates. The total number of observations in the sample is $N = \sum_{i=1}^{n} n_i$. Let g be a known link function such that $g(\mu_{ij}) = \mathbf{x}_{ij}'\beta$ where $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of unknown parameters. Whereas the mean $\boldsymbol{\mu}_i$ depends on β, the covariance matrix $\mathbf{V}_i$ may depend on β and perhaps additional parameters α ($p_1 \times 1$ vector) so that the total number of parameters is $p + p_1$.

The models compared in this paper are summarized in Table I. There are three general categories of models—the marginal model, RE model and FE model. Each broad category could have linear and nonlinear cases. We describe briefly the model specification, underlying assumptions and estimation methods used in each model.

2.1. Marginal model

The marginal model specifies only the conditional mean $\boldsymbol{\mu}_i = E(\mathbf{Y}_i|\mathbf{X}_i)$ but treats parameters in $\mathbf{V}_i$ as nuisance parameters. A distribution function in the exponential family [11] usually suggests the form of the mean and variance of Y_{ij}. If the conditional mean is correctly specified, the method of GEE yields a consistent estimator $\hat{\beta}$ of β by solving the equation $\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0$ where $\mathbf{D}_i$ is the $n_i \times p$ matrix of derivatives of $\boldsymbol{\mu}_i$ with respect to β. The asymptotic normality of $\hat{\beta}$ and the sandwich estimator of its asymptotic variance matrix $\mathrm{Var}(\hat{\beta})$ are robust to misspecification of $\mathbf{V}_i$ and the underlying distribution of $(\mathbf{Y}_i, \mathbf{X}_i)$. Although these results for $\hat{\beta}$ hold asymptotically, there can be gains in efficiency if an appropriate covariance structure for $\mathbf{V}_i$ can be assumed [12].

2.2. Random effects (RE) model

The marginal model does not make explicit the sources of correlation in the observed data. In the RE model, correlation is induced through an unobserved heterogeneity ζ_i ($q \times 1$ vector) in the conditional mean specification $\mu_{ij} = E(Y_{ij}|\mathbf{x}_{ij}, \zeta_i)$. The random coefficient model [13], the linear or generalized linear mixed effects model [14] and the hierarchical model [15] can all fall under this umbrella, allowing one for example, to acknowledge dependencies at different levels of a hierarchy. The key assumptions of the RE model (A1–A3) in Table I allow us to express the conditional likelihood of $\mathbf{Y}_i$ given $\mathbf{X}_i$ in the form $\ell_i = \int f(\mathbf{Y}_i|\mathbf{X}_i, \zeta_i = \zeta) h(\zeta)\, \mathrm{d}\zeta$ where h is the joint density of ζ_i and $f(\mathbf{Y}_i|\mathbf{X}_i, \zeta_i = \zeta) = \prod_{j=1}^{n_i} f(Y_{ij}|\mathbf{x}_{ij}, \zeta_i = \zeta)$.

For example, in the case of binary responses Y_{ij} and logit link function g, we model $\mu_{ij} = E[Y_{ij}|\mathbf{x}_{ij}, \zeta_i]$ by $\log(\mu_{ij}/(1-\mu_{ij})) = \mathbf{x}_{ij}'\beta + \mathbf{z}_{ij}'\zeta_i$ where $\mathbf{z}_{ij}$ ($q \times 1$ vector) is a subset of the covariates in $\mathbf{X}_i$. From assumptions (A2) and (A3) in Table I, $f(\mathbf{Y}_i|\mathbf{X}_i, \zeta_i) = \exp(\sum_{j=1}^{n_i} [Y_{ij}(\mathbf{x}_{ij}'\beta + \mathbf{z}_{ij}'\zeta_i) - \log(1 + \exp(\mathbf{x}_{ij}'\beta + \mathbf{z}_{ij}'\zeta_i))])$. Assuming that h is a multivariate normal density, numerical integration (e.g. adaptive Gaussian quadrature) is needed to evaluate ℓ_i [4]. The parameters are β and the variances and covariances in h, which together total $p + \frac{1}{2}q(q+1)$ components.

DOI: 10.1002/sim

Table I. Random effects, fixed effects and marginal models with correlated errors.

Model/specification	Assumptions	Estimation of β
1. Marginal model		
$\boldsymbol{\mu}_i = E(\mathbf{Y}_i \vert \mathbf{X}_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta$ $\mathbf{V}_i(\alpha) = \mathrm{Var}(\mathbf{Y}_i \vert \mathbf{X}_i)$	Distribution in exponential family informs mean and variance functions of Y_{ij}	GEE, weighted least squares (quasi-likelihood) $\sum_{i=1}^{n} \mathbf{D}'_i \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0$ $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta'}$
2. Random effects model *Generalized linear mixed model (GLMM)*		
$\mu_{ij} = E(Y_{ij} \vert \mathbf{x}_{ij}, \zeta_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\zeta_i$	(A1) ζ_i's are independent across subjects and independent of $\mathbf{X}_i$ (A2) Conditional on $(\mathbf{X}_i, \zeta_i)$, the $Y_{i1}, \ldots, Y_{in_i}$ are independent with density $f(Y_{ij} \vert \mathbf{X}_i, \zeta_i)$ (A3) Strict exogeneity: $E(Y_{ij} \vert \mathbf{X}_i, \zeta_i) = E(Y_{ij} \vert \mathbf{x}_{ij}, \zeta_i)$	Maximum likelihood estimator (MLE) $\ell = \prod_{i=1}^{n} \ell_i$
Linear mixed models—including hierarchical linear models, random coefficient models		
$E(\mathbf{Y}_i \vert \mathbf{X}_i, \zeta_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\zeta_i$ $\mathrm{Var}(\zeta_i \vert \mathbf{X}) = \mathbf{G}$ $\mathrm{Var}(\mathbf{Y}_i \vert \mathbf{X}_i, \zeta_i) = \mathbf{R}_i$	(B1) ζ_i's are independent across subjects and conditional on $\mathbf{X}_i$, $\zeta_i \sim \mathrm{N}(0, \mathbf{G})$	MLE, (feasible) generalized least squares (GLS) and generalized method of moments (GMM)
$\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i + \mathbf{R}_i$	(B2) Conditional on $(\mathbf{X}_i, \zeta_i)$, $\mathbf{Y}_i \vert \mathbf{X}_i, \zeta_i \sim \mathrm{N}(\mathbf{X}_i\beta + \mathbf{Z}_i\zeta_i, \mathbf{R}_i)$ (B3) Rank $E(\mathbf{X}'_i\mathbf{V}_i^{-1}\mathbf{X}_i) = p$	$\hat{\beta}_{\mathrm{GLS}} = (\sum_{i=1}^{n} \mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1} \times \sum_{i=1}^{n} \mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{Y}_i$
Linear random intercept model		
$\mathbf{Y}_i = \mathbf{X}_i\beta + \zeta_i\mathbf{1}_i + \varepsilon_i$ $\mathrm{Var}(\zeta_i \vert \mathbf{X}_i) = \sigma_c^2$ $\mathrm{Var}(\mathbf{Y}_i \vert \mathbf{X}_i, \zeta_i) = \sigma_e^2\mathbf{I}_i$ $\mathbf{V}_i = \sigma_c^2\mathbf{1}_i\mathbf{1}'_i + \sigma_e^2\mathbf{I}_i$	(B1′) $E(\zeta_i \vert \mathbf{X}_i) = 0$ (B2′) $E(\varepsilon_i \vert \mathbf{X}_i, \zeta_i) = 0$ (B3) Rank $E(\mathbf{X}'_i\mathbf{V}_i^{-1}\mathbf{X}_i) = p$	GLS or method of moments
3. Fixed effects models *Linear fixed effects (FE) models*		
$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\zeta_i + \varepsilon_i$ $\mathrm{Var}(\varepsilon_i \vert \mathbf{X}_i, \zeta_i) = \sigma_e^2\mathbf{I}_i$	(B2′) $E(\varepsilon_i \vert \mathbf{X}_i, \zeta_i) = 0$ (B3) Rank $E(\mathbf{X}'_i\mathbf{M}_i\mathbf{X}_i) = p$	Transformation by $\mathbf{M}_i = \mathbf{I}_i - \mathbf{Z}_i(\mathbf{Z}'_i\mathbf{Z}_i)^{-1}\mathbf{Z}'_i$ and OLS on transformed data
Linear FE models with $\mathbf{Z}_i = \mathbf{1}_i$		
$\mathbf{Y}_i = \mathbf{X}_i\beta + \zeta_i\mathbf{1}_i + \varepsilon_i$ $\mathrm{Var}(\varepsilon_i \vert \mathbf{X}_i, \zeta_i) = \sigma_e^2\mathbf{I}_i$	(B2′) $E(\varepsilon_i \vert \mathbf{X}_i, \zeta_i) = 0$ (B3) Rank $E(\mathbf{X}'_i\mathbf{M}_i\mathbf{X}_i) = p$	The covariance, or the dummy variable estimator, $\hat{\beta}_{\mathrm{CV}}$ $\mathbf{M}_i = \mathbf{I}_i - n_i^{-1}\mathbf{1}_i\mathbf{1}'_i$
$Y_{ij} = \mathbf{x}'_{ij}\beta + \zeta_i + \varepsilon_{ij}$, $\mathbf{x}_{ij}$ includes lagged Y_{ij}	(C1) $\mathbf{w}_{ij}$ correlated with Δy_{ij-1} but uncorrelated with $\Delta\varepsilon_{ij}$ (C2) Rank $E(\sum_{j=2}^{n_i} \mathbf{w}_{ij}(\Delta\mathbf{x}_{ij})') = p$ (C3) $E(\varepsilon_{ij} \vert \mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij}, \zeta_i) = 0$	Instrumental variables ($\mathbf{w}_{ij}$) approach [3, p. 85]
Nonlinear FE models with binary outcomes		
$\mu_{ij} = E(Y_{ij} \vert \mathbf{x}_{ij}, \zeta_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \zeta_i$	(F1) ζ_i's are independent across subjects	Conditional maximum likelihood estimator (CMLE)

DOI: 10.1002/sim

Table I. *Continued.*

Model/specification	Assumptions	Estimation of β
g the logit link	(F2) Conditional on $(\mathbf{X}_i, \zeta_i)$, $Y_{i1}, \ldots, Y_{in_i}$ are independent	
$\mu_{ij} = E(Y_{ij}\|\mathbf{x}_{ij}, \zeta_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \zeta_i$ g the probit link	(F3) Conditional on $\mathbf{X}_i$, $\zeta_i \sim \mathrm{N}(\varphi + \bar{\mathbf{x}}_i\delta, \sigma_c^2)$	Chamberlain's estimator [2, p. 487]
Nonlinear FE models with count outcomes		
$\mu_{ij} = E(Y_{ij}\|\mathbf{x}_{ij}, \zeta_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \zeta_i$ g the log link	(F2) Conditional on $(\mathbf{X}_i, \zeta_i)$, $Y_{i1}, \ldots, Y_{in_i}$ are independent (F4) $Y_{ij}\|\mathbf{X}_i, \zeta_i \sim POISSON(\mu_{ij})$	CMLE or MLE with dummy variables
$\mu_{ij} = E(Y_{ij}\|\mathbf{x}_{ij}, \zeta_i)$ $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \zeta_i$ g the log link	(F2) Conditional on $(\mathbf{X}_i, \zeta_i)$ $Y_{i1}, \ldots, Y_{in_i}$ are independent (F5) $Y_{ij}\|\mathbf{X}_i, \zeta_i \sim NEGBIN(\mu_{ij})$	MLE with dummy variables

For a count response with log link, $\log \mu_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\zeta_i$ and a Poisson distribution specified under (A2) give $f(\mathbf{Y}_i|\mathbf{X}_i, \zeta_i) = \exp(\sum_{j=1}^{n_i}[Y_{ij}(\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\zeta_i) - \exp(\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\zeta_i) - \log Y_{ij}!])$. A distribution conjugate to the Poisson is the Gamma distribution. With a single random effect ($q = 1$) and $\exp(\zeta_i)$ having the one-parameter Gamma distribution with mean 1 and variance γ the marginal density is explicitly

$$f(\mathbf{Y}_i|\mathbf{X}_i) = \left(\prod_{j=1}^{n_i} \frac{\lambda_{ij}^{Y_{ij}}}{Y_{ij}!}\right) \frac{\Gamma(Y_{i.} + \gamma^{-1})}{\Gamma(\gamma^{-1})} \gamma^{Y_{i.}} (1 + \gamma\lambda_{i.})^{-(Y_{i.} + \gamma^{-1})}$$

where $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$, $\lambda_{i.} = \sum_{j=1}^{n_i} \lambda_{ij}$ and $\lambda_{ij} = \exp(\mathbf{x}'_{ij}\beta)$. The distribution of Y_{ij} (given $\mathbf{x}_{ij}$) is a negative binomial distribution with $E(Y_{ij}|\mathbf{x}_{ij}) = \lambda_{ij}$ and $\mathrm{Var}(Y_{ij}|\mathbf{x}_{ij}) = \lambda_{ij} + \gamma\lambda_{ij}^2$.

2.3. Linear mixed model

For continuous response with the identity link, we can relax the conditional independence assumption (A2) and derive the unconditional likelihood. Assumptions (B1–B3) in Table I complete the specification in the linear mixed model. Thus, $\mathbf{Y}_i|\mathbf{X}_i \sim \mathrm{N}(\mathbf{X}_i\beta, \mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i + \mathbf{R}_i)$ where $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})'$ is a $n_i \times q$ matrix of a subset of variables in $\mathbf{X}_i$. Estimation of parameters in $\mathbf{G}$ and $\mathbf{R}_i$ can be carried out by maximizing the profile likelihood that first substitutes for β its generalized least-squares (GLS) estimator, or by restricted maximum likelihood, which constructs a likelihood that does not depend on β using certain linear transformations of $\mathbf{Y}_i$. This leads to an estimator $\hat{\mathbf{V}}_i$ and then to the feasible GLS estimator of β given by $\hat{\beta}_{\mathrm{GLS}} = (\sum_{i=1}^{n} \mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}(\sum_{i=1}^{n} \mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{Y}_i)$ (see Appendix A.1). A linear mixed model is a special form of RE models where the normality is assumed for RE. However, in a linear mixed model with only random intercepts, for inference on β the normality assumptions in (B1) and (B2) are not needed because the GLS estimator $\hat{\beta}_{\mathrm{GLS}}$ can be derived under moment conditions (see Appendix A.2) and assumptions (B1′, B2′, B3).

DOI: 10.1002/sim

2.4. Fixed effects (FE) model

The strong assumption of independence of $\mathbf{X}_i$ and ζ_i in the RE model is often implausible in empirical research using observational data. This assumption is relaxed in the FE model, allowing the distribution of ζ_i to depend on $\mathbf{X}_i$. The term 'fixed' perhaps stems from the econometrics literature on panel data where the unobserved heterogeneity is time-invariant; hence, 'fixed'. Unfortunately, this leads to considerable confusion of nomenclature in the statistical literature where the term 'fixed' usually refers to observed characteristics of the sample. Traditional treatments of panel data models used the term 'random effect' when ζ_i is viewed as a random variable, and the term 'fixed effect' when ζ_i is treated as a parameter. We follow the current exposition [2, 16] regarding ζ_i as random and emphasize the important distinction between FE and RE models is whether or not ζ_i are correlated with the regressors $\mathbf{X}_i$.

2.4.1. Linear FE model. The estimation strategies for the FE model vary depending on additional assumptions on the distribution of the outcome and the link functions. For continuous responses with the identity link, we can eliminate ζ_i in the model $\mathbf{Y}_i=\mathbf{X}_i\beta+\mathbf{Z}_i\zeta_i+\varepsilon_i$ via transformation by the projection $\mathbf{M}_i=\mathbf{I}_i-\mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'$ and consistently estimate β via ordinary least-squares (OLS) estimation from the transformed data $\mathbf{M}_i\mathbf{Y}_i=\mathbf{M}_i\mathbf{X}_i\beta+\mathbf{M}_i\varepsilon_i$ under the assumptions (B2′ and B3) in Table I. This is called the FE estimator $\hat{\beta}_{\mathrm{FE}}=(\sum_{i=1}^n\mathbf{X}_i'\mathbf{M}_i\mathbf{X}_i)^{-1}(\sum_{i=1}^n\mathbf{X}_i'\mathbf{M}_i\mathbf{Y}_i)$, see Appendix A.3 for details on implementation.

An important special case is when ζ_i is univariate so that $\mathbf{Z}_i=\mathbf{1}_i$. Then $\mathbf{M}_i$ is a demeaning transformation, i.e. $\mathbf{M}_i\mathbf{Y}_i=\mathbf{Y}_i-\bar{\mathbf{Y}}_i$ is the $n_i\times 1$ vector with components $\{Y_{ij}-\bar{Y}_i:1\leqslant j\leqslant n_i\}$ where $\bar{Y}_i=n_i^{-1}\sum_{j=1}^{n_i}Y_{ij}$ and $\bar{\mathbf{Y}}_i=\bar{Y}_i\mathbf{1}_i$. Because $\mathbf{M}_i\mathbf{X}_i$ has demeaned components $\{\mathbf{x}_{ij}-\bar{\mathbf{x}}_i:1\leqslant j\leqslant n_i\}$ only covariates that vary within subjects at their observational level should be used in the model. For instance $\mathbf{x}_{ij}$ should not have an intercept. The FE estimator then reduces to the covariance estimator $\hat{\beta}_{\mathrm{CV}}=(\sum_{i=1}^n\sum_{j=1}^{n_i}(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)')^{-1}(\sum_{i=1}^n\sum_{j=1}^{n_i}(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)(Y_{ij}-\bar{Y}_i))$. Its asymptotic properties are derived from assumptions B2′ and $\mathrm{Var}(\varepsilon_i|\mathbf{X}_i,\zeta_i)=\sigma_e^2\mathbf{I}_i$ in Table I.

If the ζ_i are regarded as unknown constants, the FE estimator is obtained by regressing Y_{ij} on $\mathbf{x}_{ij}$ (without intercept) and the n dummy variables $d_{1i},d_{2i},\ldots,d_{ni}$, where $d_{si}=1$ if $s=i$ and $d_{si}=0$ if $s\neq i$. For this reason $\hat{\beta}_{\mathrm{CV}}$ is sometimes called the dummy variable estimator. The parameter ζ_i is estimated by $\hat{\zeta}_i=\bar{Y}_i-\bar{\mathbf{x}}_i'\hat{\beta}_{\mathrm{CV}}$. Although it is an unbiased estimator of ζ_i, it is not consistent (as $n\to\infty$).

Another important special case is when there are lagged-dependent variables among the regressors. Then strict exogeneity cannot hold and the FE estimator is not consistent. However, under sequential exogeneity, $E(u_{ij}|\mathbf{x}_{ij},\ldots,\mathbf{x}_{i1},\zeta_i)=0$, by the strategy of differencing between adjacent time points to eliminate ζ_i and using an instrumental variable for $\Delta\mathbf{x}_{ij}=\mathbf{x}_{ij}-\mathbf{x}_{ij-1}$, a consistent estimator of β can be derived using an instrumental variables (IV) approach under assumptions (C1–C3). For example, $\mathbf{w}_{ij}=\mathbf{x}_{ij-1}$ can serve as an IV for $\Delta\mathbf{x}_{ij}$ but there are several other choices. The literature in this area is too vast to cover in this paper, see Hsiao [3] or Wooldridge [2].

2.4.2. Nonlinear FE model. In nonlinear FE models with $\mathbf{Z}_i=\mathbf{1}_i$, the dummy variable estimation strategy implemented in linear FE models may lead to inconsistent estimation because of the incidental parameters problem. Instead, for some special nonlinear FE models a conditional

DOI: 10.1002/sim

maximum likelihood estimator (CMLE) of β can be obtained [17, 18]. The CMLE, also called the FE estimator, is derived from a conditional likelihood that removes ζ_i by conditioning on a sufficient statistic for ζ_i. This approach is adopted in conditional logistic regression in epidemiologic designs for matched studies.

For a binary outcome Y_{ij} with logit link, the distribution of $\mathbf{Y}_i$ given $\mathbf{X}_i, \zeta_i$ and $m_i = \sum_{j=1}^{n_i} Y_{ij}$, does not depend on ζ_i. The conditional log likelihood function from assumptions (F1–F2) is $\ell_i = \log\{\exp(\sum_{j=1}^{n_i} Y_{ij}\mathbf{x}_{ij}'\beta)[\sum_{a\in R_i}\exp(\sum_{j=1}^{n_i} a_j\mathbf{x}_{ij}'\beta)]^{-1}\}$, where R_i is defined as the set $\{\mathbf{a}\in\mathbf{R}^{n_i} : a_j \in \{0,1\}$ and $\sum_{j=1}^{n_i} a_j = m_i\}$. The disadvantages of the CMLE are that observations with no variation in Y_{ij} for individual i are not used because they drop out of the likelihood, and the effects of time-invariant covariates cannot be estimated either.

As emphasized in the econometrics literature [2, 3, 16], the fundamental distinction between RE and FE is not whether the ζ_i's are parameters or random variables, but whether they are correlated with the observables $\{\mathbf{x}_{ij} : 1 \leqslant j \leqslant n_i\}$. A specific form of the relationship between the two, such as assumption (F3), along with (F1 and F2) and a probit link allows for consistent estimation of model parameters. In principle a logit link and a logistic distribution for (F3) can also be used but because of properties of the normal distribution the probit model provides relatively easy computations. For example, a closed-form expression for the marginal probabilities can be obtained for the probit, but not the logit model.

For a count outcome Y_{ij} with the Poisson distribution with conditional mean $E(Y_{ij}|\mathbf{x}_{ij}, \zeta_i) = \exp(\mathbf{x}_{ij}'\beta + \zeta_i)$, the incidental parameters ζ_i do not pose a problem for consistent estimation of β. In addition, a CMLE exists for the Poisson and Type I negative binomial distribution under assumptions (F2 and F4) and (F2 and F5), respectively [16]. The sufficient statistic for ζ_i is $m_i = \sum_{j=1}^{n_i} Y_{ij}$. Some moment-based estimators with transformed data also exist and are more efficient under some assumptions [19].

2.5. Comparisons between GEE, RE and FE models

2.5.1. Estimation and diagnosis. In the linear GEE and RE models, the estimator of β has the same structural form as the GLS estimator. The methods of estimation of the variance $\mathbf{V}_i = \mathbf{V}_i(\alpha)$ are of course different. In the GEE method $\mathbf{V}_i$ is specified through a working correlation whose parameters α are estimated by the method of moments. The true variance is not known, but even though it may be misspecified, the asymptotic variance of the GEE estimator of β can be made robust to this misspecification by using the empirical variance estimator [7]. However, some loss of efficiency could result if the assumed working correlation is far from the true correlation. In practice, infeasible estimates of α could result if the data do not support the correlation structure [12]. Recent evidence has shown that for GEE estimates to converge properly, the estimates of α need to be within the ranges of feasible values [12]. Only in this case does the GEE ensure consistent estimation of effects of covariates on the marginal expectation of outcome.

The quasi-likelihood information criterion (QIC) [20] has been advocated with GEE for choosing a reasonable working correlation and for selecting covariates. The GEE method structures the covariance matrix in the form $\mathbf{V}_i = \phi\mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$ where $\mathbf{A}_i$ is a diagonal matrix of the variance functions $\upsilon(\mu_{ij})$ in $\mathrm{Var}(Y_{ij}|\mathbf{x}_i) = \phi\upsilon(\mu_{ij})$ and ϕ a constant. In the linear case the choice of working correlation matrix $\mathbf{R}_i$ is suggested by examining the QIC given by $\mathrm{QIC} = -2Q(\hat{\boldsymbol{\mu}}, \mathbf{I}) + 2\,\mathrm{trace}(\hat{\boldsymbol{\Omega}}_I^{-1}\hat{\boldsymbol{\Sigma}}_e)$, where $\hat{\boldsymbol{\Omega}}_I$ is the variance of $\hat{\beta}_I = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ assuming that $\mathbf{R}_i = \mathbf{I}_i$, $-2Q(\hat{\boldsymbol{\mu}}, \mathbf{I}) = (\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathrm{GEE}})'(\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathrm{GEE}})$ for the GEE estimator $\hat{\beta}_{\mathrm{GEE}}$ of β under a

DOI: 10.1002/sim

specified structure $\mathbf{R}_i$, $\hat{\boldsymbol{\mu}}=\mathbf{X}\hat{\beta}_{\mathrm{GEE}}$ and $\hat{\boldsymbol{\Sigma}}_e$ is the robust variance of $\hat{\beta}_{\mathrm{GEE}}$. For binary responses, $Q(\hat{\boldsymbol{\mu}},\mathbf{I})=\sum_{i=1}^{n}\sum_{j=1}^{n_i}(Y_{ij}\log(\hat{\pi}_{ij})+(1-Y_{ij})\log(1-\hat{\pi}_{ij}))$, where $\hat{\pi}_{ij}$ estimates $\pi_{ij}(\mathbf{x}_{ij})=P[Y_{ij}=1|\mathbf{x}_{ij}]=(1+\exp(-\mathbf{x}'_{ij}\beta))^{-1}$ under a logit link function. Here $\hat{\boldsymbol{\Omega}}_I$ is the model-based covariance of the estimator of β under the independence working correlation structure, while $\hat{\boldsymbol{\Sigma}}_e$ is the robust variance of the GEE estimator $\hat{\beta}_{\mathrm{GEE}}$ under the specified working correlation structure.

In practice there are some caveats in the calculation of QIC [21, 22]. For unbalanced panels, where the intervals between consecutive time points are unequal, Stata and SAS can force the estimation assuming equal spacing with the choice of some working covariance structures, for example, AR(1). The correlation is estimated using subjects with two or more observations (SAS GENMOD), but all records contribute to the estimation of β. Stata on the other hand restricts the estimation of β to subjects with two or more observations. To calculate $Q(\hat{\boldsymbol{\mu}},\mathbf{I})$ these dropped observations should be brought back to make the comparison between models fair. To calculate $\mathrm{trace}(\hat{\boldsymbol{\Omega}}_I^{-1}\hat{\boldsymbol{\Sigma}}_e)$ we again face the issue that $\hat{\boldsymbol{\Omega}}_I^{-1}$ could be estimated using more observations than when $\hat{\boldsymbol{\Sigma}}_e$ is estimated when the panels are unbalanced.

The linear normal mixed model explicitly incorporates subject-specific RE ζ_i and a residual error ε_i that combine the between-subject and within-subject variance to give $\mathbf{V}_i=\mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i+\mathbf{R}_i$. Being likelihood-based, the parameters α in $\mathbf{G}$ and $\mathbf{R}_i$ can be consistently estimated and are asymptotically unbiased. Several forms for $\mathbf{G}$ and $\mathbf{R}_i$ are available in software packages such as SAS and Stata. Information criteria such as Akaike (AIC) and Bayes–Schwarz (BIC) could be used to guide the selection of an appropriate form for $\mathbf{G}$ and $\mathbf{R}_i$. Unlike GEE, the mixed model makes feasible subject-specific inference using the empirical Bayes estimator $\hat{\zeta}_i$ (see Appendix A.1). Hence, the mixed model allows both marginal and subject-specific inference, for example, on the subject-specific mean $E(\mathbf{Y}_i|\mathbf{X}_i,\zeta_i)=\mathbf{X}_i\beta+\mathbf{Z}_i\zeta_i$ and the population mean $E(\mathbf{Y}_i|\mathbf{X}_i)=\mathbf{X}_i\beta$. Recent evidence [23] suggests that the traditional AIC may not be appropriate to select models for subject-specific inferences. Vaida and Blanchard [23] propose the conditional AIC (cAIC) and marginal AIC (mAIC) for model selections when the focus of inference is subject-specific and population average, respectively. Use of cAIC and mAIC might lead to different model specifications.

Nested RE models can be tested using the likelihood ratio (LR) or score test. However, when the null hypothesis is on the boundary of the parameter space standard asymptotic results for the LR test statistic do not hold. This typically happens when testing the null hypothesis that one or more of the RE (variance components) are zero. For example, when testing for a single random effect versus none the correct asymptotic distribution of the LR statistic is a 50:50 mixture of a degenerate-at-zero χ^2 distribution and a χ^2 distribution with 1 degree of freedom [4, 5, 24].

Differences in the above models are not only in the estimation methods but also in their interpretations. As such, the selection of models depends not only on statistical diagnosis but also on the goal of the analysis. If one is interested in the average effect of covariates on the response in a population, then marginal models are the choice, which is why marginal models are also called population-average models. If one is interested in subject-specific effects of variables then the RE models are more appropriate. From RE models the marginal mean estimates can be obtained by averaging across the distribution of the subject-specific RE. When there is reason to suspect that unobserved heterogeneity is correlated with explanatory variables then the FE models are more appropriate because the RE model would yield inconsistent estimates. Further discussion is given in Section 2.5.3.

DOI: 10.1002/sim

2.5.2. Missing data in GEE, RE and FE models. Under some assumptions on the distribution of missingness in the responses, using only the available data in estimation will lead to valid inference. Let $\mathbf{s}_i$ denote the binary selection indictors for non-missing responses $\mathbf{Y}_i$, i.e. $s_{ij}=1$ if Y_{ij} is observed, and $s_{ij}=0$, otherwise. The standard GEE assumes that missingness is completely at random (MCAR) [5], $E(\mathbf{Y}_i|\mathbf{X}_i,\mathbf{s}_i)=E(\mathbf{Y}_i|\mathbf{X}_i)$. In the econometrics literature selection is said to be ignorable, or that selection is exogenous [2].

In the linear RE and FE models similar assumptions apply to estimation from the selected sample. For example, the conditional mean and variance of the error ε_i in the FE model would be modified with $(\mathbf{X}_i,\mathbf{s}_i,\zeta_i)$ in the conditioning set. Although in linear and nonlinear models likelihood-based estimation on the selected sample is valid under the assumption $f(\mathbf{Y}_i|\mathbf{X}_i,\mathbf{s}_i,\zeta_i)=f(\mathbf{Y}_i|\mathbf{X}_i,\zeta_i)$, we can also obtain valid inference under the missing at random (MAR) assumption, also called selection on observables. Here $f(\mathbf{s}_i|\mathbf{Y}_i,\mathbf{X}_i,\zeta_i)=f(\mathbf{s}_i|\mathbf{Y}_i^0,\mathbf{X}_i)$ where $\mathbf{Y}_i^0$ is the observed component of the response. It is also assumed that the selection distribution $f(\mathbf{s}_i|\mathbf{Y}_i^0,\mathbf{X}_i)$ does not share parameters with $f(\mathbf{Y}_i|\mathbf{X}_i)$; hence, the inference can be based only on $f(\mathbf{Y}_i^0|\mathbf{X}_i)$ [25]. For example, estimation in the linear normal mixed model is valid under the MAR assumption. Other approaches, such as the inverse-probability of selection weighting scheme [26, 27], are useful to accommodate missing data patterns that are neither MCAR nor MAR as in estimation of medical costs where missingness is due to informative censoring [28, 29]. Specific schemes are available in the econometrics literature that take into consideration the type of missing pattern, such as attrition in panel data models, which is a case of the monotone missingness pattern [2, 27].

2.5.3. Interpretation of regression coefficients. In some situations the distinction between the marginal model and the RE model is not important in that the parameter estimates have both the population-average and the subject-specific interpretation and there is an explicit connection between the two models. For example, a linear marginal model with compound symmetry $\mathbf{V}_i=\sigma^2((1-\rho)\mathbf{I}_i+\rho\mathbf{J}_i)$, $\mathbf{J}_i=\mathbf{1}_i\mathbf{1}_i'$, is equivalent to a model with random intercept $\zeta_i\sim\mathrm{N}(0,\sigma_c^2)$ and error $\varepsilon_{ij}\sim\mathrm{N}(0,\sigma_e^2)$. Another example is a linear marginal model with an exponential temporal correlation, which is derivable from a random intercept model where the error term follows a first-order autoregressive process. For the linear model, the conditional effects and marginal effects β are the same.

With binary outcomes and the probit link, the comparison between the coefficients β_{RE} in the RE model with a random intercept $\zeta_i\sim\mathrm{N}(0,\sigma_c^2)$, with the coefficients β_{M} in the marginal model, is $\beta_{\mathrm{RE}}/\beta_{\mathrm{M}}=(1+\sigma_c^2)^{1/2}$. This is obtained by comparison of the function form of the mean response in the two models and should not be viewed as comparison of numerical estimates. In the same context with the logit link, there is an approximate relationship [24], $\beta_{\mathrm{RE}}/\beta_{\mathrm{M}}\approx(1+a^2\sigma_c^2)^{1/2}$, where $a^2=16\sqrt{3}/15\pi$.

With count outcomes with the log link and normally distributed RE $\zeta_i\sim\mathrm{N}(0,\mathbf{G})$, the marginal mean is $E(Y_{ij}|\mathbf{X}_i)=\exp(\mathbf{x}_{ij}'\beta+\frac{1}{2}\mathbf{z}_{ij}'\mathbf{G}\mathbf{z}_{ij})$, which reduces to $\exp(\mathbf{x}_{ij}'\beta+\frac{1}{2}\sigma_c^2)$ for a single random intercept. This shows that β_{RE} and β_{M} differ only in the intercept. With count outcomes it is common to use $\zeta_i\sim\mathrm{N}(-\frac{1}{2}\sigma_c^2,\sigma_c^2)$ or generally $E(\exp\zeta_i)=1$ in order to have the same exponential form of the response means in the marginal and RE models [30]. We emphasize that these comparisons should not be construed as comparisons between parameter estimates. For example, the Poisson count model without RE and the RE model with a log-Gamma distributed random effect (leading to a marginal negative binomial count model) has the same function form of the mean response, but the estimated coefficients can be very different.

DOI: 10.1002/sim

2.5.4. Testing between FE and RE models—the Hausman test. The major distinction between the FE and RE models is whether or not the RE are correlated with covariates $\mathbf{X}_i$. When $\mathbf{X}_i$ is correlated with the RE, i.e. $\mathbf{X}_i$ is endogenous, the RE estimators are no longer consistent.

In the simple RE and FE models with a single random intercept ζ_i the estimates $\hat{\beta}_{\mathrm{RE}}$ are closer to $\hat{\beta}_{\mathrm{FE}}$ when the proportion of variance due to the intercept is large [2]. A formal test of endogeneity of $\mathbf{X}_i$ proposed by Hausman [31] is based on the difference between $\hat{\beta}_{\mathrm{RE}}$ and $\hat{\beta}_{\mathrm{FE}}$, where the null hypothesis is that $\mathbf{X}_i$ is exogenous (and so the RE assumptions hold). As $\hat{\beta}_{\mathrm{FE}}$ is consistent when ζ_i is correlated with $\mathbf{X}_i$, but $\hat{\beta}_{\mathrm{RE}}$ is inconsistent, a statistically significant difference between the two estimates is interpreted as evidence against the RE assumption (B1′). Assuming (B2′) holds under the null and the alternative hypotheses and the RE variance structure $\mathrm{Var}(\zeta_i|\mathbf{X}_i)=\sigma_c^2$ and $\mathrm{Var}(\varepsilon_i|\mathbf{X}_i,\zeta_i)=\sigma_e^2\mathbf{I}_i$ holds under the null, the Hausman statistic $H=(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})'[\mathrm{Var}(\hat{\beta}_{\mathrm{FE}})-\mathrm{Var}(\hat{\beta}_{\mathrm{RE}})]^{-}(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})$ under the null is distributed asymptotically as χ_k^2 where $k=\mathrm{rank}[\mathrm{Var}(\hat{\beta}_{\mathrm{FE}})-\mathrm{Var}(\hat{\beta}_{\mathrm{RE}})]$.

The above form of the Hausman statistic does not hold if the relative efficiency of RE estimators is not established under other variance structures on (ζ_i,ε_i) [2]. Specifically, when ζ_i or ε_i are not identically distributed, $\mathrm{Var}(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})\neq\mathrm{Var}(\hat{\beta}_{\mathrm{FE}})-\mathrm{Var}(\hat{\beta}_{\mathrm{RE}})$. If the homoskedasticity assumption for the RE model is violated a robust form of the Hausman test can be devised via an auxiliary regression or via bootstrap [16].

There are several caveats in using the Hausman test. First, the strict exogeneity assumption (B2′) is maintained under both the null and the alternative hypotheses. Second, standard statistical packages usually implement the test under the compound-symmetry form $\mathbf{V}_i=\sigma_e^2\mathbf{I}_i+\sigma_c^2\mathbf{J}_i$. Third, the test is sensitive to misspecification of the model and the power of the test is low for typically encountered sample sizes [32]. The sampling distribution of the test statistic may not be well approximated by a χ^2 distribution.

3. THE NURSE–COMMUNITY HEALTH WORKER TEAM STUDY

We demonstrate the application of linear and nonlinear GEE, RE and FE models with data from a community-based, multi-site randomized controlled trial for Medicaid pregnant women conducted in Kent County, Michigan. The intervention—a nurse and Community Health Worker team home visiting program—was compared with the standard Community Care that included a home visiting program from nurses or other professionals. Following an enrollment interview (<24 weeks gestation), psychosocial outcomes were measured at 34–36 weeks gestation and 6 weeks, six months, and 12 months after delivery, providing up to 5 assessments on each women. For our analysis we use 530 women (out of 613 in the main study) who had a live birth. The excluded women include those who had a spontaneous or elective abortion, fetal loss, or had agreed to adoption or lost custody of their infant.

The treatment and control groups were balanced with respect to demographic and baseline psychosocial variables, including CES-D, perceived stress and mastery. At enrollment, over half of the women screened positive for depressive symptoms (CES-D $\geqslant$16) and nearly a third had scores that indicated probable depression (CES-D $\geqslant$24). Perceived stress was measured with Cohen's Perceived Stress 14-item scale and mastery with the Pearlin 7-item Sense of Mastery scale [33, 34]. These two scales were transformed to have a range 0–100. Higher scores for mastery are favorable but higher scores for stress are unfavorable.

DOI: 10.1002/sim

The CES-D was available on all 530 women at baseline and on 79–87 per cent of these women at subsequent waves. Using the CES-D as a continuous variable Y_{ij} we compare the differences in GEE, RE and FE estimates in Table II. Time is measured in months anchored at the participant's infant birth date. As there was some variation across participants in their actual assessment times at each wave, we regarded time as a continuous variable. In addition to variables involving time, stress and mastery were time-dependent leaving only the indicator variable for control group as time invariant.

Table II. Linear model for CES-D: GEE, RE, FE estimates.

Parameter	GEE* (model SE) [robust SE]	RE† (model SE) [robust SE]	FE‡ (model SE) [robust SE]
Intercept β_0	−9.5397§	−5.7156§	···
	(2.1234)	(2.1251)	
	[2.3846]	[2.2513]	
Control β_1	−0.2371	−0.2198	···
	(0.3770)	(0.4726)	
	[0.4637]	[0.4698]	
Time β_2	−0.2393§	−0.2618§	···
	(0.0428)	(0.04311)	
	[0.0433]	[0.04448]	
Time × control β_3	−0.0581	−0.06148	···
	(0.0606)	(0.06085)	
	[0.0606]	[0.06055]	
Time × time β_4	0.0007	0.001261	···
	(0.0042)	(0.00413)	
	[0.0041]	[0.00423]	
Time × time × control β_5	0.0084	0.009096	···
	(0.0059)	(0.00586)	
	[0.0057]	[0.00583]	
Stress β_6	0.8781§	0.7819§	0.6907§
	(0.0450)	(0.04456)	(0.07409)
	[0.0521]	[0.04859]	[0.09556]
Mastery β_7	0.1307§	0.1105§	0.1356§
	(0.0283)	(0.02817)	(0.04808)
	[0.0290]	[0.02738]	[0.05110]
Stress × mastery β_8	−0.0075§	−0.00665§	−0.00592§
	(0.0006)	(0.00064)	(0.00106)
	[0.0007]	[0.00065]	[0.00127]
Number of observations	2240	2240	1981
Number of subjects	530	530	412

CES-D = Center for Epidemiologic Studies-Depression scale (range 0–60).
SE = Standard error.
Entries are: estimate, model-based SE (.), robust SE [.].
*GEE under independence working correlation.
†Normal model, $\mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$ with $\mathbf{Z}_i = [\mathbf{1}_i\ \mathbf{t}_i\ \mathbf{t}_i^2]$, $\mathbf{G}$ unstructured and $\mathbf{R}_i$ single banded.
‡Within-subject transformation, $\mathbf{M}_i = \mathbf{I}_i - \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i'$, $\mathbf{Z}_i = [\mathbf{1}_i\ \mathbf{t}_i\ \mathbf{t}_i^2]$.
§Significant at 1 per cent using model-based SE.

DOI: 10.1002/sim

For the marginal GEE analysis we estimated several models with different mean and variance specifications. The covariance structures compared using QIC included unstructured, compound symmetry, auto-regressive, 2- and 3-dependent structures. An m-dependent covariance structure has only the first m diagonal bands, with all other entries set to zero. The covariate specifications included various combinations of the control group indicator, time, time squared, stress and mastery. The smallest QIC was attained for the independence model with eight covariates as shown in Table II (column 2).

The linear mixed (a special RE) model uses the covariance $\mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$, where $\mathbf{Z}_i$ has three columns, intercept, time and time squared. The choice of RE was made by examining the evolution of $\mathbf{Y}_i$ over time. Information criteria were computed from $\text{AIC} = -2\ell + 2d$ and $\text{BIC} = -2\ell + d \log n$, where ℓ is the restricted maximum log likelihood, d the number of covariance parameters in the model and n the number of subjects. On the basis of AIC, BIC and LR tests for nested RE models, we chose $\mathbf{G}$ unstructured (six covariance terms) and $\mathbf{R}_i$ as a single-banded diagonal matrix (five variance terms). Thus, in contrast with the GEE this approach indicated the necessity for addressing correlation. The LR test of the independence model (one variance term) versus the mixed model with $\mathbf{Z}_i = [\mathbf{1}_i \, \mathbf{t}_i]$ was highly significant ($p < 0.0001$) based on a 50:50 mixture of χ^2 distributions [5] with degrees of freedom 1 and 2. The LR test comparing models with $\mathbf{Z}_i = [\mathbf{1}_i \, \mathbf{t}_i]$ and $\mathbf{Z}_i = [\mathbf{1}_i \, \mathbf{t}_i \, \mathbf{t}_i^2]$ RE was also statistically significant ($p = 0.010$). Table II (column 3) shows that RE estimates of β are very similar to the estimates from GEE. Robust standard errors were calculated by the standard sandwich formulae using the appropriate residuals (see Appendixes A.1–A.3).

The test of the hypothesis $H_0 : L\beta = 0$, where L is a known $r \times p$ matrix of rank r is based on the statistic $F = (L\hat{\beta})' [L' \text{Var}(\hat{\beta}) L]^{-1} (L\hat{\beta})/r$, which under H_0 has an approximate F distribution with degrees of freedom (r, ddfm). The appropriate ddfm is a matter of some controversy. By default SAS uses the CONTAIN method with several other options (e.g. Kenward–Roger and Satterthwaite) to account for the downward bias in standard errors due to the estimation of parameters in $\mathbf{G}$ and $\mathbf{R}_i$ and improve the accuracy of the F-test [35, 36]. With the large sample size of this study the conclusions remain the same under these different options.

Table II suggests that effects involving the treatment group could be eliminated. The mean difference between control and treatment group at a fixed time t (with stress and mastery also held fixed) is $\beta_1 + \beta_3 t + \beta_5 t^2$. The β-parameters are subscripted as shown in Table II (column 1). A test of $H_0 : (\beta_1, \beta_3, \beta_5) = 0$ via a F-test did not reveal significance ($p = 0.477$). The corresponding chi-square score test under the GEE analysis is also not significant ($p = 0.515$).

The signs on the estimated coefficients for the scales stress and mastery are plausible and in the expected direction, where CES-D and stress are negatively scored, meaning higher scores indicate worse outcomes and the mastery scale is positively scored. The gradient of mean CES-D relative to stress is $\beta_6 + \beta_8$ MASTERY, which is positive, but decreasing with increasing mastery. The gradient with respect to MASTERY is $\beta_7 + \beta_8$ STRESS, which is positive for STRESS < 16.6 and then negative for STRESS > 16.6 (Table II, column 3). Thus, higher stress has negative impact on CES-D, whereas higher mastery has generally a positive impact.

The FE estimates $\hat{\beta}_{\text{FE}}$ are shown in Table II (column 4). The within-subject transformation $\mathbf{M}_i = \mathbf{I}_i - \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i'$ eliminates from consideration covariates that are in the column space of $\mathbf{Z}_i = [\mathbf{1}_i \, \mathbf{t}_i \, \mathbf{t}_i^2]$, where $\mathbf{t}_i$ is the $n_i \times 1$ time vector. For estimation we require $n_i > 3$, which results in 412 subjects (79 with 4 assessments and 333 with 5 assessments). This data set was also used in the RE analysis and yielded estimates similar to those from the mixed model using all records. When $E(\zeta_i | \mathbf{X}_i) \neq 0$, the RE estimates are inconsistent whereas the FE estimates are consistent.

DOI: 10.1002/sim

The robust version of the Hausman test statistic was calculated using an auxiliary regression. The test was highly significant (3-df χ^2 distribution, $p<0.0001$) suggesting that the FE estimates are perhaps more reliable.

Although the GEE and RE estimates were comparable, the independence working correlation suggested by QIC for the former was deemed inappropriate. All preliminary analyses of the data pointed to the need to address correlations. For the goals of the analysis (effects of stress, mastery on CES-D), estimates from the RE model are preferred (Table II, column 3). As there are concerns about the RE assumptions, the FE analysis would be our choice. However, from the FE analysis of the adopted model, a treatment effect cannot be estimated.

Table III summarizes the results of analyses of CES-D as a binary outcome. The CES-D scale (0–60) was dichotomized at 16, with $Y_{ij}=1$ if CES-D $\geqslant 16$. With the GEE method, the model for Y_{ij} is the logit model:

$$\log\left(\frac{\pi_{ij}(\mathbf{x}_{ij})}{1-\pi_{ij}(\mathbf{x}_{ij})}\right)=\mathbf{x}'_{ij}\beta$$

where $\pi_{ij}(\mathbf{x}_{ij})=P[Y_{ij}=1|\mathbf{x}_{ij}]$. The structure of the covariance matrix is $\mathbf{V}_i=\mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$ where $\mathbf{A}_i$ is a diagonal matrix of the variance functions $\upsilon(\pi_{ij})=\pi_{ij}(\mathbf{x}_{ij})(1-\pi_{ij}(\mathbf{x}_{ij}))$.

We used QIC to compare some covariance structures in all models with the covariates shown in Table III (column 1). The smallest QIC was attained for the exponential temporal correlation structure $\mathbf{R}_i=\{\exp(-d_{jk}/\rho):1\leqslant j,k\leqslant n_i\}$, where d_{jk} is the temporal distance between the jth and kth observations. The estimates are shown in column 2. The joint test of $H_0:(\beta_3,\beta_4,\beta_5)=0$ is not significant. Starting with the main effects model (control, time, stress, mastery) the QIC was used to select among super models with interactions of the main effects model. This also resulted in the main effects model.

The RE model is

$$\log\left(\frac{\pi_{ij}(\mathbf{x}_{ij},\zeta_i)}{1-\pi_{ij}(\mathbf{x}_{ij},\zeta_i)}\right)=\mathbf{x}'_{ij}\beta+\mathbf{z}'_{ij}\zeta_i$$

where $\pi_{ij}(\mathbf{x}_{ij},\zeta_i)=P[Y_{ij}=1|\mathbf{x}_{ij},\zeta_i]$. Estimation of β is based on maximum marginal likelihood of $\mathbf{Y}_i=(Y_{i1},\ldots,Y_{in_i})$ given $\mathbf{X}_i$ as described in Section 2.2. This is computationally quite challenging with several RE. With a single normally distributed random effect ($\zeta_i\sim\mathrm{N}(0,\sigma^2)$) two computational methods were used (1) optimization of the marginal likelihood based on integral approximation such as adaptive quadrature and (2) approximating the nonlinear mixed model by a linear mixed model (linearization method). The linearization is accomplished by an expansion of $\pi_{ij}(\mathbf{x}_{ij})=(1+\exp(-(\mathbf{x}'_{ij}\beta+\zeta_i)))^{-1}$ about a current estimate of (β,ζ_i) and then using the ensuing pseudodata to estimate a linear mixed model. The process is iterative until some convergence criterion is satisfied. Results from these two estimation schemes were used to inform a model with $\mathbf{Z}_i=[\mathbf{1}_i\,\mathbf{t}_i]$ and ζ_i bivariate normal, $\zeta_i\sim\mathrm{N}(0,\mathbf{G})$. The model with $\mathbf{G}$ having compound symmetry (two parameters) was adequate (Table III, column 3). More complex forms for $\mathbf{G}$ required it to be non-positive definite to assure convergence.

The LR test comparing the model without RE with the model with a single random intercept was highly significant. This test is based on a 50:50 mixture of a degenerate-at-zero distribution and a 1-degree of freedom χ^2 distribution. There was no significant difference between the intercept plus slope model over the intercept only model, based on the LR test of a 50:50 mixture of χ^2 distributions with 1 and 2 degrees of freedom. Computations were carried out using the

DOI: 10.1002/sim

Table III. Nonlinear model for the probability (CES-D $\geqslant$16): GEE, RE, FE estimates.

Parameter	GEE* (model SE) [robust SE]	RE† (model SE) [robust SE]	FE‡ (model SE) [robust SE]
Intercept β_0	−0.4183 (0.4554) [0.5270]	−0.4471 (0.5033) [0.5336]	···
Control β_1	−0.2535 (0.1502) [0.1596]	−0.2829 (0.1639) [0.1653]	···
Time β_2	−0.0362§ (0.0153) [0.0156]	−0.0410§ (0.0169) [0.0160]	−0.0841¶ (0.0020) [0.0214]
Time×control β_3	−0.0297 (0.0221) [0.0222]	−0.0274 (0.0244) [0.0229]	−0.0101 (0.0285) [0.0295]
Time×time β_4	−0.0029 (0.0017) [0.0017]	−0.0031 (0.0018) [0.0017]	−0.0011 (0.0021) [0.0023]
Time×time×control β_5	0.0041 (0.0023) [0.0022]	0.0041 (0.0025) [0.0023]	0.0037 (0.0031) [0.0031]
Stress β_6	0.0901¶ (0.0052) [0.0055]	0.0928¶ (0.0057) [0.0057]	0.0734¶ (0.0081) [0.0083]
Mastery β_7	−0.0542¶ (0.0053) [0.0061]	−0.0553¶ (0.0058) [0.0062]	−0.0621¶ (0.0091) [0.0087]
Number of observations	2240	2240	1366
Number of individuals	530	530	303

CES-D = Center for Epidemiologic Studies-Depression scale (range 0–60).
SE = Standard error.
Entries are: estimate, model-based SE (.), robust SE [.].
*Under exponential temporal working correlation.
†Random effects, $\zeta_i \sim \mathrm{N}(0, \mathbf{G}), \mathbf{Z}_i = [\mathbf{1}_i \mathbf{t}_i]$ with $\mathbf{G}$ compound symmetry.
‡Single random intercept model was estimated using conditional maximum likelihood.
§Significant at 5 per cent, using model-based standard error.
¶Significant at 1 per cent.

GLIMMIX and NLMIXED procedures in SAS [37] and the xtmelogit and gllamm routines in Stata [38, 39].

The FE estimates in column 4 of Table III are derived under CMLE with a single random effect. There is an inevitable loss of sample size because in the conditional model subject strata with $\sum_{j=1}^{n_i} Y_{ij} = 0$ or n_i are non-informative. In addition, the effects of time-constant covariates cannot be estimated because they do not appear in the conditional likelihood function.

DOI: 10.1002/sim

4. DISCUSSION

In this paper we described the structural assumptions that underlie the method of generalized estimating equations (GEE), random effects (RE) and fixed effects (FE) approaches to estimating covariate effects in both linear and nonlinear models. The versatility of these models has spawned several monographs and books addressing both theory and practice in many applications that encompass the biomedical disciplines [5, 14, 40–44], social and behavioral sciences [4, 13, 15, 45–47] and econometrics [2, 3, 48]. The context we used in this paper is repeated response measures $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ associated with explanatory variables $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$ with inference focused on the regression parameter β in specification of the conditional mean $\mu_{ij} = E(Y_{ij}|\mathbf{X}_i)$ through the link function $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta$. In the presence of unobserved heterogeneity ζ_i we model $\mu_{ij} = E(Y_{ij}|\mathbf{X}_i, \zeta_i)$ by $g(\mu_{ij}) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\zeta_i$. For valid inference we need a consistent estimator $\hat{\beta}$ of β, which is generally achieved in the GEE, RE and FE settings by correct specification of μ_{ij} under their corresponding assumptions, e.g. whether or not ζ_i is uncorrelated with $\mathbf{X}_i$. Appropriately accounting for the correlation between the repeated responses through specification of the variance $\mathbf{V}_i = E(\mathbf{Y}_i|\mathbf{X}_i)$ is also important because it can affect the validity of the standard errors of the estimated β. When the assumptions of a specific model hold, the model-based standard errors will be efficient. However, to guard against possible misspecification of $\mathbf{V}_i$ a robust form of the estimated variance matrix of $\hat{\beta}$ should be used.

Does this mean that serious consideration should not be given to selecting an appropriate structure for $\mathbf{V}_i$? In the GEE method valid inference based on $\hat{\beta}$ is feasible if μ_{ij} is correctly specified even though the variance may be misspecified. We do not need to introduce an unobserved (random) heterogeneity ζ_i together with its attendant assumptions as in the RE model. However, under the assumptions of the RE, our estimator will be more efficient than the corresponding GEE estimator. In the linear setting with normally distributed ζ_i and residual errors ε_i, we can use likelihood-based methods to estimate both β and the covariance parameters in $\mathbf{V}_i$. Consideration can be given to different competing structures for $\mathbf{V}_i$ and empirical Bayes estimators $\hat{\zeta}_i$ of the unobserved heterogeneity derived from the posterior distribution (of ζ_i given $\mathbf{Y}_i$). This allows for subject-specific inference, for example, on the conditional response means $E(\mathbf{Y}_i|\mathbf{X}_i, \zeta_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\zeta_i$, as well as on the population average $E(\mathbf{Y}_i|\mathbf{X}_i) = \mathbf{X}_i\beta$. Access to statistical software (e.g. MIXED and GLIMMIX in SAS, xtmixed in Stata) for analysis of this (normal) linear mixed model has produced numerous applications in several disciplines.

Without the strong normality assumptions of the linear mixed model, we can still achieve consistent estimation of β and robust standard errors using essentially moment assumptions on (ζ_i, ε_i). For practical reasons only simple structural forms for $\mathbf{V}_i$ are useful such as the variance component structure, which is ubiquitous in the econometrics literature [2, 16] where it is called the RE model. A critical assumption of the RE model is that the unobserved heterogeneity ζ_i is uncorrelated with the covariates (specifically, $E(\zeta_i|\mathbf{X}_i) = 0$). Without this assumption the RE estimator in this setting would be inconsistent. Removing this restriction leads to FE models, but it comes with some cost. As FE estimation uses a within-subject transformation or conditioning to eliminate ζ_i, inference on the effects of time-constant covariates is not possible. Several approaches to circumvent this deficiency are available in the econometrics literature, for example, the Hausman and Taylor hybrid linear model [2, 16] and techniques of quasi-differencing in some nonlinear models [16]. The standard FE analysis also removes observations that have too few repeated components. For example, with a single random intercept to account for unobserved heterogeneity,

DOI: 10.1002/sim

subjects with a single record will be eliminated. In contrast the RE analysis would retain these records. If one is interested in evaluating a treatment effect this might seem to be a serious drawback of the FE analysis. We argue that due consideration should be given at the design stage of the study so that an appropriate model for assessing a treatment effect can be applied. Crossing the treatment indicator with the time variable is one way to retain the treatment variable in the FE analysis. However, under the full RE assumptions, the RE estimator is more efficient than the FE estimator. A comparison of the RE and FE estimators can be carried out using a Hausman-type χ^2 test.

Our illustration of the GEE, RE and FE analyses for estimating covariate effects on CES-D in the Nurse–Community Health Worker team intervention study has shown that empirically the methods might not differ substantively. Generally, statistical significance remained the same across the analyses, although some differences in effect size were noticed. As noted above fundamentally different assumptions underlie these models and careful consideration must be given to these assumptions in subject-matter applications.

In nonlinear RE models practical considerations may force use of a relatively small number of normally distributed RE, for instance in the normal-logistic model for correlated binary outcomes. In our application, however, estimation of $P[\text{CES-D} \geqslant 16]$ by a RE normal-logistic model did not require long run-times for 1 and 2 RE with available software, but convergence problems arose with 3 RE. More elaborate variance structures that incorporate more RE are perhaps better handled via linearization methods within the generalized linear mixed model (GLMM) framework. Although the GEE method is easily implemented in many nonlinear models, care must be exercised in choosing a working variance structure for $\mathbf{V}_i$ due to parameter constraints forced by the functional dependence of between means μ_{ij} and covariances [49].

Under full likelihood specification the RE model for linear and some nonlinear outcomes, we can derive both subject-specific and population-average estimates of the β coefficients but only the latter is possible with the GEE method. Finally, the FE model with a single unobserved heterogeneity is based on conditional maximum likelihood and leads to estimates of β for covariates that are time varying. The interpretation of β as population-average log-odds ratios cannot be made because $\text{logit}(P[Y_{ij}=1|\mathbf{x}_{ij}, \zeta_i]) = \mathbf{x}'_{ij}\beta + \zeta_i$ contains the unknown ζ_i. However, a within-subject interpretation can be made of a covariate that changes within subject (e.g. time varying). In conclusion, understanding the conceptual differences between GEE, RE and FE methods is important even though empirically they might not differ substantively as we see in our application. The choice of model for a particular application would depend on the relevant questions being addressed, which in turn informs the type of design and data collection that would be relevant. We should not rely on statistical tools to mitigate deficiencies in design and data acquisition.

APPENDIX A

A.1. Linear mixed effects model

In the linear mixed effects model, minimization with respect to β of the sum of squares $\sum_{i=1}^{n}(\mathbf{Y}_i - \mathbf{X}_i\beta)'\hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\beta)$ yields the feasible generalized least-squares (GLS) estimator $\hat{\beta}_{\text{GLS}} = (\sum_{i=1}^{n}\mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}(\sum_{i=1}^{n}\mathbf{X}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{Y}_i)$. A consistent estimator $\hat{\mathbf{V}}$ of $\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i + \mathbf{R}_i$ is obtained by estimating the parameters in $\mathbf{G}$ and $\mathbf{R}_i$ under the normal assumptions (B1–B2). This is achieved by full maximum likelihood, or restricted maximum likelihood after transforming

DOI: 10.1002/sim

the log-likelihood to a function of only $\mathbf{G}$ and $\mathbf{R}_i$. Consistency and asymptotic normality follow directly from the expression for $\hat{\beta}_{\mathrm{GLS}}$ and the asymptotic variance is estimated by $\mathrm{Var}(\hat{\beta}_{\mathrm{GLS}}) = (\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)^{-1}$. This is generally a slight underestimate of the true variance because the variability in $\hat{\mathbf{V}}_i$ is ignored, but for most practical purposes the bias is small. In testing the hypothesis $H_0 : L\beta = 0$ where L is a matrix with full row rank r, this bias is usually accounted for by using the approximate F-statistic $F = (L\hat{\beta}_{\mathrm{GLS}})'[L'\mathrm{Var}(\hat{\beta}_{\mathrm{GLS}})L]^{-1}(L\hat{\beta}_{\mathrm{GLS}})/r$ with degrees of freedom (r, d) where d is calculated by inflating the variance matrices of both $\hat{\beta}_{\mathrm{GLS}}$ and $\hat{\zeta}_i - \zeta_i$. Several choices are available for the degrees of freedom d [35, 36], which are now available in many statistical software.

To make inference robust to choice of the variance structure of $\mathbf{V}_i$, we might use the empirical Huber–White [50, 51] heteroscedasticity consistent estimator of $\mathrm{Var}(\hat{\beta}_{\mathrm{GLS}})$ given by $(\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \hat{\varepsilon}_i \hat{\varepsilon}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i) (\sum_{i=1}^{n} \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)^{-1}$, where $\hat{\varepsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{GLS}}$ are the GLS residuals. The corresponding F-statistic for testing $H_0 : L\beta = 0$ has degrees of freedom (r, d).

For subject-specific analyses the empirical Bayes estimates $\hat{\zeta}_i = \hat{\mathbf{G}} \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{GLS}})$ of the RE are obtained by substitution of estimates for $\beta, \mathbf{G}$ and $\mathbf{V}_i$ in the conditional means $E(\zeta_i | \mathbf{Y}_i)$. The variance is $\mathrm{Var}(\hat{\zeta}_i - \zeta_i) = \hat{\mathbf{G}} - \hat{\mathbf{G}} \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} [\hat{\mathbf{V}}_i - \mathbf{X}_i \mathrm{Var}(\hat{\beta}_{\mathrm{GLS}}) \mathbf{X}_i'] \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \hat{\mathbf{G}}$.

An alternative derivation of $(\hat{\beta}_{\mathrm{GLS}}, \hat{\zeta}_i)$ solves the mixed model equations of Henderson [52, 53], which result from minimizing $\sum_{i=1}^{n} ((\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \zeta_i)' \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \zeta_i) + \zeta_i' \mathbf{G}^{-1} \zeta_i)$ with respect to β and ζ_i regarded as parameters. This objective function is justified from the likelihood $f(\mathbf{Y}_i, \zeta_i | \mathbf{X}_i) = f(\mathbf{Y}_i | \mathbf{X}_i, \zeta_i) f(\zeta_i | \mathbf{X}_i)$ and the assumptions (B1–B2).

A.2. Linear random intercept model

Under assumptions (B1′, B2′, B3) use the pooled ordinary least-squares (POLS) estimator $\hat{\beta}_{\mathrm{POLS}} = (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i)^{-1} (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{Y}_i)$ to get the POLS residuals $\tilde{\mathbf{v}}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{POLS}}$. Then $E(\tilde{\mathbf{v}}' \tilde{\mathbf{v}}) = \mathrm{N}(\sigma_e^2 + \sigma_c^2) - p\sigma_e^2 - \sigma_c^2 \,\mathrm{trace}\, [\mathbf{J}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{J}]$ where $\tilde{\mathbf{v}}$ $(N \times 1)$ and X $(N \times p)$ are the stacked $\{\tilde{\mathbf{v}}_i : 1 \leqslant i \leqslant n\}$ and $\{\mathbf{X}_i : 1 \leqslant i \leqslant n\}$, respectively, and $\mathbf{J} = \mathrm{diag}\{\mathbf{1}_i : 1 \leqslant i \leqslant n\}$. Next, obtain the residuals $\hat{\varepsilon}_i = \mathbf{M}_i (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{CV}})$ from the demeaned model $\mathbf{M}_i \mathbf{Y}_i = \mathbf{M}_i \mathbf{X}_i \beta + \mathbf{M}_i \varepsilon_i$, where $\mathbf{M}_i = \mathbf{I}_i - \mathbf{1}_i (\mathbf{1}_i' \mathbf{1}_i)^{-1} \mathbf{1}_i'$. If $\hat{\varepsilon}$ is the stacked $\{\hat{\varepsilon}_i : 1 \leqslant i \leqslant n\}$ we get $E(\hat{\varepsilon}' \hat{\varepsilon}) = \sigma_e^2 (N - n - (p-1))$. Hence, we get estimators $(\hat{\sigma}_e^2, \hat{\sigma}_c^2)$ and $\hat{\mathbf{V}}_i = \hat{\sigma}_e^2 \mathbf{I}_i + \hat{\sigma}_c^2 \mathbf{1}_i \mathbf{1}_i'$ is used to define the RE estimator $\hat{\beta}_{\mathrm{RE}}$, which has the same expression as $\hat{\beta}_{\mathrm{GLS}}$ in A.1. As such, $\mathrm{Var}(\hat{\beta}_{\mathrm{RE}})$ also has the same form and expression for the robust variance, but uses the RE residuals $\hat{\varepsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{RE}}$.

A.3. Linear FE model

Under (B2′ and B3) the FE estimator $\hat{\beta}_{\mathrm{FE}} = (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i)^{-1} (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \mathbf{Y}_i)$ is consistent asymptotically normal with $\mathrm{Var}(\hat{\beta}_{\mathrm{FE}}) = \sigma_e^2 (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i)^{-1}$. To obtain a consistent estimator of σ_e^2 use the FE residuals $\hat{\varepsilon}_i = \mathbf{M}_i (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{\mathrm{FE}})$ where $\mathbf{M}_i = \mathbf{I}_i - \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i'$. Denoting the stacked $\{\hat{\varepsilon}_i : 1 \leqslant i \leqslant n\}$, $\{\mathbf{Y}_i : 1 \leqslant i \leqslant n\}$, $\{\mathbf{X}_i : 1 \leqslant i \leqslant n\}$ and $M = \mathrm{diag}\{\mathbf{M}_i : 1 \leqslant i \leqslant n\}$ by $\hat{\varepsilon}$, $\mathbf{Y}$ and $\mathbf{X}$, respectively, we have $\hat{\varepsilon} = [\mathbf{M} - \mathbf{MX}(\mathbf{X}'\mathbf{MX})^{-} \mathbf{X}'\mathbf{M}]\mathbf{Y} = \mathbf{QY}$ where $\mathbf{A}^{-}$ is a generalized inverse of $\mathbf{A}$. An estimator of σ_e^2 is motivated by $E(\hat{\varepsilon}' \hat{\varepsilon}) = \sigma_e^2 \,\mathrm{trace}(\mathbf{Q}) = \sigma_e^2 (\mathrm{N} - \mathrm{nq} - \mathrm{trace}(\mathbf{MX}(\mathbf{X}'\mathbf{MX})^{-} \mathbf{X}'\mathbf{M}))$. Note that taking $\mathbf{Z}_i = \mathbf{1}_i$ we get the same expression as in A.2 because $q = 1$ and the demeaned $\mathbf{MX}$ would exclude

DOI: 10.1002/sim

the intercept term only assuming that no time-invariant variables are in $\mathbf{X}$. The robust asymptotic variance of $\hat{\beta}_{\mathrm{FE}}$ is $(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i)^{-1} (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \hat{\varepsilon}_i \hat{\varepsilon}_{ii}' \mathbf{M}_i \mathbf{X}_i) (\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i)^{-1}$.

ACKNOWLEDGEMENTS

The authors wish to thank the Associate Editor and two anonymous referees, who provided valuable comments and suggestions that have improved the presentation of the paper. We also thank Drs Oliver Schabenberger and Jan Chovsta of the SAS Institute for clarifying the technical details of some SAS procedures.

This study was supported by the Agency for Healthcare Research and Quality under grant 1R01 HS14206 and by grant MCJ-260743 from the Maternal and Child Health Bureau (title V, Social Security Act), Health Resources and Services Administration, Department of Health and Human Services.

REFERENCES

1. Potthoff RF, Roy SN. Generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 1964; **51**(3–4):313–326.
2. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, 2002.
3. Hsiao C. *Analysis of Panel Data* (2nd edn). Cambridge University Press: Cambridge, U.K., 2003.
4. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling*: *Multilevel*, *Longitudinal*, *and Structural Equation Models*. Chapman & Hall, CRC: Boca Raton, FL, 2004.
5. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, NY, 2000.
6. Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley: Chichester, England, 1999.
7. Diggle PJ, Heagerty PK, Liang KY, Zeger SL. *Analysis of Longitudinal Data* (2nd edn). Oxford University Press: New York, NY, 2002.
8. Hardin JW, Hilbe JM. *Generalized Estimating Equations*. Chapman & Hall, CRC: Boca Raton, FL, 2002.
9. Roman LA, Lindsay JK, Moore JS, Duthie PA, Peck C, Barton LR, Gebben MR, Baer LJ. Addressing mental health and stress in Medicaid-insured pregnant women using a nurse–community health worker home visiting team. *Public Health Nursing* 2007; **24**(3):239–248.
10. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977; **1**(3):385–401.
11. McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman & Hall: New York, NY, 1989.
12. Shults J, Ratcliffe SJ, Leonard M. Improved generalized estimating equation analysis via xtqls for quasi-least squares in Stata. *Stata Journal* 2007; **7**(2):147–166.
13. Longford NT. *Random Coefficient Models*. Oxford University Press: Oxford, 1993.
14. McCulloch CE, Searle SR. *Generalized*, *Linear*, *and Mixed Models*. Wiley: New York, NY, 2001.
15. Gelman A, Hill J. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press: New York, NY, 2007.
16. Cameron AC, Trivedi PK. *Microeconomics*: *Methods and Applications*. Cambridge University Press: New York, NY, 2005.
17. Chamberlain G. Analysis of covariance with qualitative data. *Review of Economic Studies* 1980; **47**(1):225–238.
18. McFadden D. Econometric analysis of qualitative response models. In *Handbook of Econometrics*, Griliches Z, Intriligator MD (eds), vol. 2. North-Holland: Amsterdam, 1984; 1395–1457.
19. Wooldridge JM. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 1999; **90**(1):77–97.
20. Pan W. Model selection in estimating equations. *Biometrics* 2001; **57**(2):529–534.
21. Cui J, Qian G. Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Communications in Statistics—Simulation and Computation* 2007; **36**(5):987–996.
22. Cui J. QIC program and model selection in GEE analyses. *Stata Journal* 2007; **7**(2):209–220.
23. Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. *Biometrika* 2005; **92**(2):351–370.
24. Molenberghs G, Verbeke G. Meaningful statistical model formulations for repeated measures. *Statistica Sinica* 2004; **14**(3):989–1020.

DOI: 10.1002/sim

25. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995; **90**(431):1112–1121.
26. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression-models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**(429):106–121.
27. Wooldridge JM. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 2007; **141**(2):1281–1301.
28. Gardiner JC, Liu L, Luo Z. Estimation of medical costs from a transition model. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor PK Sen*, Balakrishnan N, Silvapulle M, Pena E (eds). Institute of Mathematical Statistics: Beachwood, OH, 2008; 350–363.
29. Baser O, Gardiner JC, Bradley CJ, Yuce H, Given C. Longitudinal analysis of censored medical cost data. *Health Economics* 2006; **15**(5):513–525.
30. Winkelmann R. *Econometric Analysis of Count Data* (5th edn). Springer: Berlin-Heidelberg, 2008.
31. Hausman JA. Specification tests in econometrics. *Econometrica* 1978; **46**(6):1251–1271.
32. Long JS, Trivedi PK. Some specification tests for the linear regression model. In *Testing Structural Equation Models*, Bollen KA, Long JS (eds). Sage Publications, Inc.: Thousand Oaks, CA, 1993.
33. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior* 1983; **24**(4):385–396.
34. Pearlin LI, Menaghan EG, Lieberman MA, Mullan JT. The stress process. *Journal of Health and Social Behavior* 1981; **22**(4):337–356.
35. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**(3):983–997.
36. Fai AHT, Cornelius PL. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation* 1996; **54**(4):363–378.
37. SAS/STAT User's Guide [program]. *9.1.3 Version*. SAS Institute Inc.: Cary, NC, 2006.
38. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modelling using Stata*. Stata Press: College Station, TX, 2005.
39. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 2005; **128**(2):301–323.
40. Jiang J. *Linear and Generalized Linear Mixed Models and their Applications*. Springer: New York, NY, 2007.
41. Demidenko E. *Mixed Models*: *Theory and Applications*. Wiley: New York, NY, 2006.
42. Brown H, Prescott R. *Applied Mixed Models in Medicine* (2nd edn). Wiley: Chichester, England, 2006.
43. Pinheiro JC, Bates DM. *Mixed Effects Models in S and S-Plus*. Springer: New York, NY, 2000.
44. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley: Hoboken, NJ, 2004.
45. Goldstein H. *Multilevel Statistical Models* (3rd edn). Arnold: London, 2003.
46. Frees EW. *Longitudinal and Panel Data*: *Analysis and Applications in the Social Sciences*. Cambridge University Press: Cambridge, U.K., 2004.
47. Raudenbush SW, Bryk AS. *Hierarchical Linear Models*. Sage Publishing: Thousand Oaks, CA, 2002.
48. Baltagi BH. *Econometric Analysis of Panel Data* (3rd edn). Wiley: West Sussex, England, 2005.
49. Gilliland D, Schabenberger O. Limits on pairwise association for equi-correlated binary variables. *Journal of Applied Statistical Sciences* 2001; **10**:279–285.
50. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of Fifth Berkeley Symposium in Mathematical Statistics*, vol. 1. University of California Press: Berkeley, CA, 1967; 221–233.
51. White H. *Estimation*, *Inference and Specification Analysis*. Cambridge University Press: Cambridge, U.K., 1994.
52. Henderson CR. *Applications of Linear Models in Animal Breeding*. University of Guelph: Guelph, CN, 1984.
53. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 1975; **31**(2):423–447.

DOI: 10.1002/sim