Statistical Methods in Medical Research 2004; **13**: 309–323

Equivalence of conditional and marginal regression models for clustered and longitudinal data

John Ritz and **Donna Spiegelman** Departments of Biostatistics and Epidemiology, Harvard School of Public Health, Boston, MA, USA

Certain statistical models specify a conditional mean function, given a random effect and covariates of interest. On the other hand, one may instead model a marginal mean only in terms of the covariates. We discuss some common situations where conditional and marginal means coincide. In a Gaussian linear mixed effects model we have equivalent interpretations of the conditional and marginal regression parameter estimates. Similar results exist for more general link functions. In this paper we give a short overview of some models, where conditional and marginal results are equivalent and we illustrate this with some examples. When the conditional mean is additive in a random effect on the log scale, it is seen that the marginal mean equals the conditional mean plus a constant, such that slope parameters have the same interpretation in both formulations. No further distributional assumptions are needed in either of these cases. With a logit link and a double exponential random effect, a closed form marginal link function is derived from the conditional model. When a logit or probit link is used with a normal random effect, the marginal mean parameters become attenuated by a factor which depends on parameters of the distribution of the covariates. In a conditional Weibull proportional hazards model with a positive stable frailty, the marginal hazards are again Weibull but with slope parameters attenuated towards zero.

1 Introduction

Consider a random effects model where we specify the conditional mean, given the random effect. The response variable is an $n_i \times 1$ vector denoted by $\mathbf{Y}_i$, $i = 1, \ldots, n$, and the random effect by the q-dimensional vector $\mathbf{U}_i$ with distribution F_u and density f_u. Vectors and matrices are in bold, and random quantities are in upper case letters. In generalized linear models, the mean function μ_i of $\mathbf{Y}_i$ is often modelled through a link function g, such that the conditional mean function of the component Y_{ij}, $j = 1, \ldots, n_i$, of $\mathbf{Y}_i$ is modelled as

$$g\{E[Y_{ij}|\mathbf{U}_i]\} = K[\mathbf{z}_i'\mathbf{U}_i, \mu(\mathbf{X}_i, \boldsymbol{\beta})] \tag{1}$$

for some function K, a $q \times 1$ design vector $\mathbf{z}_i$ and where μ is some function of a $p \times 1$ vector $\boldsymbol{\beta}$, with corresponding $p \times 1$ covariate vector, $\mathbf{X}_i$, perhaps the linear function $\mathbf{X}_i'\boldsymbol{\beta}$.

Throughout this paper, models and expectations are considered conditional on the random effect $\mathbf{U}_i$ or marginal with respect to the random effect, $\mathbf{U}_i$, and these

Address for correspondence: John Ritz, 401 N. Washington Street, Suite 700, Rockville, MD 20850, USA. E-mail: stjor@channing.harvard.edu

 10.1191/0962280204sm368ra

conditional and marginal models are always conditional on the covariate vector $\mathbf{X}_i$. In addition, as is typical of the random and mixed effects model literature, we assume that the covariate is independent of the random effect. In Sections 2 and 3, we consider $\mathbf{Y}_i$ as a scalar, as well as multivariate, with several covariance structures.

Without loss of generality, we may consider only the distribution of $\mathbf{U}$, which could perhaps, represent some linear transformation $\mathbf{Z}'\mathbf{U}$ for a design matrix $\mathbf{Z}$. If further restrictions are placed on Equation (1), the entire conditional distribution of the response, $\mathbf{Y}_i$ (the index i is suppressed here), can be specified as

$$\mathbf{Y}|\mathbf{U},\mathbf{X} \sim F_c(\mathbf{y}|\mathbf{U},\mathbf{X}) \tag{2}$$

The marginal mean of $\mathbf{Y}_i$ is

$$\begin{aligned} E[\mathbf{Y}|\mathbf{X}] &= \int_{\mathcal{U}} E[\mathbf{Y}|\mathbf{U},\mathbf{X}]\, \mathrm{d}F_u(\mathbf{u}|\mathbf{X}) \\ &= \int_{\mathcal{U}} E[\mathbf{Y}|\mathbf{U},\mathbf{X}]\, \mathrm{d}F_u(\mathbf{u}) \end{aligned} \tag{3}$$

with the marginal density given by

$$\begin{aligned} f(\mathbf{y}|\mathbf{X}) &= \int_{\mathcal{U}} f_c(\mathbf{y}|\mathbf{U},\mathbf{X})\, \mathrm{d}F_u(\mathbf{u}|\mathbf{X}) \\ &= \int_{\mathcal{U}} f_c(\mathbf{y}|\mathbf{U},\mathbf{X})\, \mathrm{d}F_u(\mathbf{u}) \end{aligned} \tag{4}$$

Models (1) and (2) are often referred to as subject-specific (SS)[1] or cluster-specific models.[2] The marginal models in Equations (3) and (4) are also called the population-averaged (PA) model.[2]

In general, the parameters of SS and PA models do not have the same interpretations and it is important to understand the distinction. When dealing with clustered or longitudinal data analysis, it is necessary to know how to appropriately choose between the SS and PA model on the basis of the particular data given. The papers by Crouchley and Davis,[3] Lindsey and Lambert[1] and Neuhaus *et al.*[2] give sound guidance on these issues. The purpose of this paper is not to discuss how to select the SS over the PA model. The emphasis here is to show that in some common cases key parameters in these model are identical or have the same interpretation, and in other situations, they do not. When the parameters have the same interpretation, the choice between the SS or PA model does not have to be made.

In many applications, the parameters of the conditional model are the ones of interest; a marginal model would then be usefully applied only when the parameters of these models have the same interpretation as those of the conditional model. Detailed consideration of the advantages of the conditional modelling approach over the marginal for scientific studies, including those in epidemiology, are given in Lindsey and Lambert.[1] An interesting question is thus for what conditional distributions f_c of $\mathbf{Y}$ given ($\mathbf{X}$, $\mathbf{U}$), for what link functions g, and for what densities f_u do the parameters of the marginal mean derived from the conditional mean in Equation (1) for $\mathbf{Y}$ equal those of the marginal

mean function in Equation (3). It is of interest to know when the parameters of these two models have the same interpretation and under what conditions we will obtain marginal estimates which are unbiased for the conditional parameters of interest.

In Section 2, we review and consolidate previous results, and then provide several examples on the equivalence or lack thereof of marginal and conditional mean functions and densities for the four most common link functions: identity, log, logit and probit. In Section 3, we discuss some further generalizations of these issues, in particular when $f(\mathbf{u}|\mathbf{X}) \neq f(\mathbf{u})$.

2 Equivalence between conditional and marginal models

2.1 Marginal regression equals conditional regression for regressions that are additive in the random effect

Neuhaus *et al.*[2] and Jewell[4] considered a linear mixed effects model of the form given in Equation (2) where f_c was normal. The model was

$$Y_{ij}|U_i = \mu + \beta_0 + X_{ij}\beta + U_i + \varepsilon_{ij} \tag{5}$$

where ε_{ij} are independent $N(0, \sigma_e^2)$, U_i are independent $N(0, \sigma_A^2)$ variates and i and j denote the cluster and within-cluster index, respectively. The parameter β is the regression coefficient of interest and X_{ij} are fixed covariates. This model is more restrictive than Equation (1) because it models the entire distribution of the response Y conditional on the random effect. In this case, $\mu(X, \beta) = X\beta$, $g(E[Y|U]) = K(U, X\beta) = U + X\beta$. That is, $\mu(X, \beta)$ is linear in the parameters of interest, β, and the covariates and U is additive with respect to μ.

It is clear from the previously mentioned model of Neuhaus *et al.*[2] that when we have the identity link g and U additive in Equation (1), model (5) implies model (1). Consequently, the unconditional marginal model of Y with density f_m is again normal, with the same linear mean $X\beta$, but with a different variance.

In fact, Neuhaus' case is a special case of a more general result. For the mean model given in Equation (1), where the conditional mean is additive in $\mathbf{U}$ and $\mu(\mathbf{X}, \boldsymbol{\beta})$, it follows directly that for any distribution of $\mathbf{Y}$ (whose mean exists), the marginal mean of $\mathbf{Y}$ is given by the iterated mean formula

$$\begin{aligned} E[\mathbf{Y}] &= E_{\mathbf{U}}[E[\mathbf{Y}|\mathbf{U}]] \\ &= E_{\mathbf{U}}[\mu + \mathbf{U}] = \mu + E[\mathbf{U}] \end{aligned} \tag{6}$$

where $\mu = \mu(\mathbf{X}, \boldsymbol{\beta})$, if the random effect is independent of $\mathbf{X}$. Note importantly that Equation (6) is true for *any* function $h(\mathbf{X}, \boldsymbol{\beta}) = \mu$, including but not restricted to linear h, and for *any* distribution F_u of $\mathbf{U}$ whose mean exists.

Also, the marginal variance is given by the well known variance formula

$$\begin{aligned} \mathrm{Var}[\mathbf{Y}] &= E_{\mathbf{U}}[\mathrm{Var}[\mathbf{Y}|\mathbf{U}]] + \mathrm{Var}_{\mathbf{U}}[E[\mathbf{Y}|\mathbf{U}]] \\ &= E_{\mathbf{U}}[\mathrm{Var}[\mathbf{Y}|\mathbf{U}]] + \mathrm{Var}[\mathbf{U}] \end{aligned} \tag{7}$$

Hence, for any conditional distribution of **Y** given **X** and **U**, for any random effect **U** with mean zero, independent of the covariate **X**, if the conditional mean Equation (1) is additive in μ and **U**, the marginal mean exactly equals the function μ. If $E[\mathbf{U}] \neq 0$, the marginal mean equals the μ plus an offset equal to $E[\mathbf{U}]$. The key requirement for this identity is additivity on the original scale of the random effect to the fixed effect mean function μ. Please note that neither normality nor a linear fixed effect mean function is required. This same result was given previously by Graubard and Korn.[5]

2.1.1 *Some more results of the additive random effect case*

Let the conditional distribution of Y given U be distributed as normal with mean $\beta_0 + X\beta + U$ and variance σ^2, where X is a covariate and the random effect U is normal with mean 0 and variance τ^2. This is a model of the form (2) and is a case considered by Neuhaus *et al.*[2] As shown by Neuhaus, the marginal mean of Y is $\beta_0 + X\beta$ and the marginal variance of Y is $\sigma^2 + \tau^2$. In this case, the full marginal distribution can be derived as well, and is normal.

An example in Diggle *et al.*[6] considered Y given U as Poisson with mean U, where U is gamma with mean γ and variance $\phi\gamma^2$. In this case, it can be shown that the marginal distribution of Y can be derived as a negative binomial, where Y has mean γ and variance $\gamma + \phi\gamma^2$, as given by Equations (6) and (7).

In a final example, we let the conditional distribution of Y given U be Poisson with mean $\mu + U$ and the identity link, $\mu = \beta_0 + X\beta$, where U is $N(0, \sigma^2)$. The marginal mean of Y is then again equal to μ.

2.1.2 *An example*

In the first example, we look at Poisson-distributed data with an identity link function in the conditional model and a normal random effect. In the marginal model, we also have an identity link. The example is a seven month crossover study investigating the effect of administering a soy supplement versus placebo on menopausal symptoms, hot flashes, hormones and cholesterol levels in women at increased risk for breast cancer from Woods *et al.*[7] Here, the response variable, y_{ij}, is the number of hot flashes per day and is assumed Poisson, where $i = 1, \ldots, 80$ patients, $j = 1, \ldots, n_i$, and n_i is the number of days for patient i, ranging from 25 to 221 (median 176). The within-subject correlation matrix is assumed to be exchangeable, that is, the covariance matrix $\Sigma_{\mathbf{w}}$ is σ^2 on the diagonal, with $\rho\sigma^2$ on the off-diagonals. Here, the random effect $\mathbf{U}_i = (u_{i1}, \ldots, u_{in_i})$ is assumed to be distributed as normal with mean 0 and covariance matrix $\Sigma_{\mathbf{w}}$. The model was given by

$$\begin{aligned} E[y_{ij}|u_{ij}] &= \mu(Z_{ij}, T_{ij}) + u_{ij} \\ \mu(Z_{ij}, T_{ij}) &= \beta_0 + \beta_1' T_{ij} + \beta_2' Z_{ij} \end{aligned} \tag{8}$$

where T_{ij} is an indicator for treatment or placebo and Z_{ij} is a time-varying covariate vector. The results from fitting the conditional and marginal models using PROC NLMIXED and PROC GENMOD in SAS[8] are presented in Table 1. As discussed in Section 2.1 and Table 1, the estimates from both methods should be similar. The estimated intraclass correlation coefficient is $\hat{\sigma}_{\mathrm{b}}^2/(\hat{\sigma}_{\mathrm{b}}^2 + \hat{\sigma}_{\mathrm{w}}^2) = 0.57$, where $\hat{\sigma}_{\mathrm{b}}^2$ and $\hat{\sigma}_{\mathrm{w}}^2$ are

Table 1 Soy bar study: conditional and marginal model fits with the identity link function

	Variable				
	Conditional		Marginal		
Treatment	$\hat{\beta}$ (flashes/day) (95% CI)	$\widehat{\text{SE}}$ *P*-value	$\hat{\beta}$ (95% CI)	Robust $\widehat{\text{SE}}$ *P*-value	Naive $\widehat{\text{SE}}$ *P*-value
Intercept	6.8	0.32	6.6	0.37	0.33
	(6.17,7.43)	<0.001	(5.87–7.32)	<0.001	<0.001
Soy bar	−1.7	0.080	−1.3	0.34	0.09
	(−1.86, −1.54)	<0.001	(−1.97, −0.63)	<0.001	<0.001
Placebo bar	−1.8	0.080	−1.6	0.38	0.09
	(−1.95, −1.64)	<0.001	(−2.34, −0.86)	<0.001	<0.001

the between- and within-variances. The naive (or model-based) estimated standard errors correspond to those which would be obtained if the matrix of derivatives of the score equations derived from Equation (8) corresponded to the inverse of the information matrix of a well defined likelihood which correctly specifies the full distribution of **Y**. The robust (or empirical) estimated standard errors uses a sandwich covariance estimator of Var[$\hat{\boldsymbol{\beta}}$] that is consistent even if the working correlation matrix is misspecified,[9] as long as model (8) is correct. The robust- and naive (model-based) estimated standard errors (SE in Table 1) for the marginal and conditional model are given in Table 1, together with Wald-type *P*-values (Table 1, *P*-values for the marginal model are robust). From Table 1, we see that both the soy bar and the placebo bar led to a significant decrease in the number of hot flashes experienced per day, compared with baseline, as estimated by both the conditional and marginal approaches. The model-based standard errors of the marginal model are similar to the standard errors from the conditional model. This is not necessarily expected because a conditional Poisson model with a normal random effect is not marginally Poisson. Thus, the naive standard errors are unlikely to be valid, and the robust standard errors should be relied upon, if the marginal approach is taken.

2.2 When conditional mean is additive in a random effect on the log scale, marginal mean equals conditional mean plus a constant

Here, we assume that $\log(E[Y|U]) = \mu(X,\beta) + U$ and the random effect, U, is independent of the covariate X. Then, the marginal mean for Y is given by

$$E[Y] = \int_{\mathcal{U}} e^{\mu(X,\beta)+u}\, dF_u(u)$$

As long as the integral of $e^u\, dF_u(u)$ exists and is finite, the marginal mean will equal the conditional mean plus a constant which depends on the parameters indexing the distribution of the random effect. This constant will be absorbed into the intercept of the marginal mean model, in the case where $\mu(X,\beta)$ is a linear function. Note importantly that these results hold for *any* distribution F_u, including but not restricted

to the normal, and for *any* distribution of Y given (X, U) for which the conditional mean is additive on the log scale.[10]

Suppose that given U is normal with distribution $F_u(u)$, Y is Poisson. Then the marginal mean of Y is

$$E[Y] = \int_0^\infty e^{\mu+u}\, dF_u(u) = e^{\mu+\tau^2/2}$$

where μ is the conditional mean in Equation (1) and U has mean 0 and variance τ^2. If $\mu(X, \beta) = \beta_0 + X\beta$, then the marginal mean function is $\beta_0^\star + X\beta$, where $\beta_0^\star = \beta_0 + \tau^2/2$. This Poisson-normal model is described in Hinde[11] and is given as an example in Diggle *et al.*[6]

Breslow and Clayton[12] considered the model where the responses Y_i are conditionally independent, given a $q \times 1$ dimensional random effect $\mathbf{U}$, the random effect is assumed multivariate normal with mean 0 and covariance matrix $\mathbf{D}$. By introducing design vectors $\mathbf{z_i}$ on the random effect $\mathbf{U}$, several Y_is may share the same components of $\mathbf{U}$ thereby allowing for more general covariance structures among the Y_is. They modelled the conditional means as $E[Y_i|\mathbf{U}] = \mu_i^{(u)}$, which is then linked through h as $\mu_i^{(u)} = h(\mathbf{X}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{U})$, where h is the inverse of the link function g and the conditional variance is modelled as $\mathrm{Var}[Y_i|\mathbf{U}] = \phi a_i v(\mu_i^{(u)})$, where $v(\cdot)$ is a specified variance function, a_i is a known constant and ϕ is a dispersion parameter. As discussed earlier, with a log link, one finds the exact mean as $E[Y_i] = \exp(\mathbf{X}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{D}\mathbf{z}_i/2)$ with the random effect variance adding an offset to the marginal mean. However, note again that no distributional assumption is needed for the random effect $\mathbf{U}$ when we have the identity or log link. The only requirement is that the fixed effect and the random effect are additive and the expected value of $h(\mathbf{z}_i'\mathbf{U})$ exists and is finite.

For another example, let U be gamma with parameters γ and δ and density $f_u(u) = \Gamma(\delta)\gamma^{-\delta}u^{\delta-1}e^{-u/\gamma}$. It is easily shown that the marginal mean is then $\log(E[Y]) = \mu(X, \beta) - \delta\log(1+\gamma)$. If $\mu(X, \beta) = \beta_0 + X\beta$, then $\log(E[Y]) = \beta_0^\star + X\beta$, where $\beta_0^\star = \beta_0 - \delta\log(1+\gamma)$. Crouchley and Davis[3] considered Y given U as Poisson and U gamma. The authors show that when the conditional mean model follows a linear form in the covariates with the log link, the marginal and conditional models give identical slope parameters.

2.2.1 *An example*

We use the same data set as in Section 2.1.2, but now we assume a log link function for the conditional model. As shown earlier, the marginal model will also have a log link function. The results from using PROC NLMIXED and PROC GENMOD in SAS,[8] for this model and the data described in Section 2.1.2 are presented in Table 2. As discussed in Section 2.2, the estimates from either method will be approximately the same; as in Table 2. Again Wald-type *P*-values are presented in Table 2 (Table 1, *P*-values for the marginal model are robust). As in the previous example, the soy bar and the placebo bar led to a significant percent decrease in the number of hot flashes experienced per day, compared with baseline, regardless of whether conditional or marginal model was fit. Again as in Table 1, the model-based standard errors from the

Table 2 Soy bar study: conditional and marginal model fits with the log link function

	Variable				
	Conditional		Marginal		
Treatment	$\hat{\beta}$ [log(flashes/day)]	$\widehat{\text{SE}}$	$\hat{\beta}$	Robust $\widehat{\text{SE}}$	Naive $\widehat{\text{SE}}$
Soy bar	−0.22	0.014	−0.23	0.060	0.016
	(−0.25, −0.19)	<0.001	(−0.35, −0.11)	<0.001	<0.001
Placebo bar	−0.28	0.014	−0.28	0.069	0.017
	(−0.31, −0.25)	<0.001	(−0.41, −0.14)	<0.001	<0.001

marginal analysis agree with the standard errors from the conditional analysis, and the standard errors from the conditional analysis are smaller than the robust standard errors from the marginal analysis.

2.3 Relationship between the marginal and conditional mean functions when *U* is additive to $\mu(\mathbf{X}, \beta)$ on the probit scale

As another example we consider the paper by Heagerty and Zeger.[13] Their model is given by $\text{probit}(E[Y|U]) = \mu(\mathbf{X}, \boldsymbol{\beta}) + U$ as given in Equation (1), where the link function is $g = \text{probit}$. They note that for normal random effects, if the random effects variance is independent of $\mathbf{X}$, the marginal parameters become attenuated by a factor depending on $\mathbf{X}$. The attenuation now takes the form $E[Y] = h[\mu(\mathbf{X}, \boldsymbol{\beta})/c]$, where $c = \sqrt{1 + \text{Var}(U)}$ and $h = g^{-1}$.

2.4 Conditions for the equivalence of the marginal and conditional mean functions when *U* is additive to $\mu(\mathbf{X}, \beta)$ on the logit scale

The mixed-effects logistic model considered by Neuhaus *et al.*[2] is an example of Equation (1) with logit link g and linear function μ in the covariates X_{ij} and slope β. In this model, there are Bernoulli responses Y_{ij}, and the random effect intercept U_i varies between clusters. In this case, these authors showed that the marginal model is not equivalent to the conditional one. These authors gave a Taylor series expansion about $\beta = 0$ on the population average effect β_P in the log odds scale, where $E[Y_{ij}|X_{ij}] = f(X_{ij};\beta_P)$, and $f(X_{ij};\beta_P) = X_{ij}\beta_P$. In a first-order Taylor series expansion, these authors showed that if U is nondegenerate, then β_P differs approximately from the β_c by a factor that depends on the variance of the inverse logit of the random effect U. The same result holds true for an additional random effect on the slope.

For the logit link $g = \text{logit}$ such that $\text{logit}(E[Y|U])$ in Equation (2) with response Y and any additive random effect U, model (3) takes the form

$$E[Y] = \int_{\mathcal{U}} (1 + \mathrm{e}^{-\mu-u})^{-1} \, \mathrm{d}F_u(u)$$

In the case of a normal random effect, there does not appear to be a simple explicit form for the distribution of Y and there is no equivalence between the marginal and

conditional model.[2] Crouch and Spiegelman[14] also gave a numerical algorithm for calculating this integral to a prespecified degree of accuracy. The integral may also be estimated by adaptive quadrature, which is superior to ordinary quadrature.[15] In Knox *et al.*[9] the logit link was also considered. They show that the $\boldsymbol{\beta}$ coefficients become attenuated:

$$E[Y_i] \approx \frac{\exp(a_i \mathbf{X}_i' \boldsymbol{\beta})}{1 + \exp(a_i \mathbf{X}_i' \boldsymbol{\beta})}$$

where $a_i = |\mathbf{I} + b^2 \mathbf{D} \mathbf{z}_i \mathbf{z}_i'|^{-1/2}$ and $b = 16\sqrt{3}/(15\pi)$, a_i ranges between 0 and 1.

A special case occurs when U is exponentially distributed with mean 1, for which a closed form expression for the expected value of a binary response exists, namely

$$E[Y] = \int_0^\infty \frac{e^{-u}}{1 + e^{-\mu - u}} \, du = e^{\mu} \log(1 + e^{-\mu})$$

Thus, for an exponential random effect with mean 1, a conditional logit link function gives a closed form marginal link function $h(\mu) = e^{\mu} \log(1 + e^{-\mu})$. If the random effect is exponentially distributed with mean λ^{-1}, then the conditional logit link function

$$g(\mu) = (1 + e^{-\mu - \lambda u})^{-1} \tag{9}$$

also gives the marginal link h. Note importantly that λ disappears from the marginal link function. There is no explicit expression for the inverse of the link h, but h is a strictly increasing function of μ from 0 to 1 and hence the inverse link of h exists. Figure 1 shows the logit link together with the link h. The link h has the same shape as the logit but is shifted somewhat to the left of the logit link. (straight line). As a second case, let the random effect have a double exponential distribution with density $f_u(u) = \lambda \exp(-\lambda|u|)/2$, it's shape having heavier tails than that of the normal. Here, the variance of the double exponential distribution is $\sigma_b^2 = 1/\lambda^2$. If the conditional link function is given by Equation (9), the marginal mean can then be shown to equal

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} \frac{1}{2} \frac{e^{-|u|}}{1 + e^{-\mu - u}} \, du \\ &= \frac{1}{2}[1 + e^{\mu} \log(1 + e^{-\mu}) - e^{-\mu} \log(1 + e^{\mu})] \end{aligned} \tag{10}$$

The shape of this link function (here denoted by $h^\star$) is very similar to h (Figure 1). An assumed double exponential random effects distribution has the desirable properties of symmetry and mean 0. In addition, the nuisance parameter λ does not appear in the marginal mean function. Unfortunately, these new link functions cannot be fit in SAS PROC GENMOD,[8] because an explicit form for the inverse link function cannot be written. Although there has been some research on methods to investigate the validity of assumptions about the form of the distribution for U in some settings, further work is

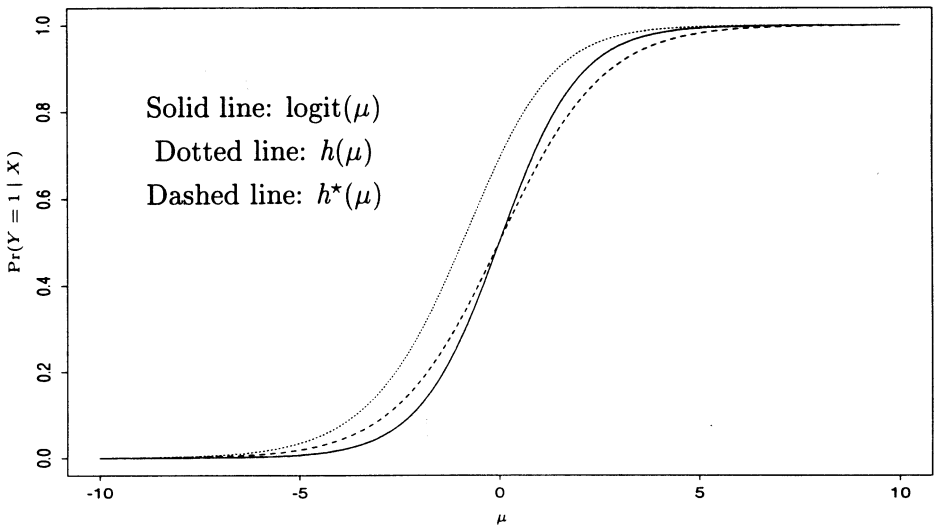

Figure 1 A comparison of the logit link to the h and $h^{\star}$ links for a Bernoulli model with an exponential random effect and a double exponential random effect, respectively.

needed.[16] In addition, research on the extent of bias in estimating $\boldsymbol{\beta}$ is needed, when U is assumed double exponential but is in fact normal or some other symmetric or asymmetric distribution.[17]

2.4.1 An example

This example is illustrated by data from a longitudinal study of HIV-infected adults in the greater Boston area, the Nutrition for healthy living study.[18] In this example, we look at correlates of current use of highly active anti-retroviral therapy (HAART) in relation to demographic characteristics, such as race (white/nonwhite), sex (male/ female), highest level of educational achievement (2–8 corresponding to grades 1–6, grades 7–11, high school/GED, trade/technical, some college, college degree and graduate school), absolute cd4 cell count (100 cells/mm^3) and date of protocol (years since 2/17/1995).[19] There are 650 participants, with between 1 and 11 repeated measures (median = 4). Subjects reported using HAART during 54% of the person-visits. The response is current use of HAART (yes/no) and is modelled by a logistic model for a Bernoulli random variable, with a normal random effect in a logit link function, that is

$$\text{logit}(E[Y_{it}|u_i]) = \mu_{it}, \quad \mu_{it} = \beta_0 + \mathbf{X}_{it}\boldsymbol{\beta} + u_i \tag{11}$$

where Y_{it} is either 1 for patient using HAART at time t and 0 otherwise and $\mathbf{X}_{it}$ are the demographic characteristic for HAART usage. The results of fitting the linear logistic marginal model and the linear logistic conditional model with a Gaussian random effect to these data in SAS PROC NLMIXED and PROC GENMOD[8] are given in Table 3. This time, the marginal model estimates are dramatically attenuated as expected. When the point estimates lead to such different interpretations (e.g., an OR of 0.54 and 0.80

Table 3 Nutrition for healthy living study of correlates of HAART use: conditional, $h^\star$, and marginal model fits with logit link function

	Variable					
	Conditional		$h^\star$-link		Marginal	
Treatment	Odds ratio (95% CI)	$\widehat{\mathrm{SE}}(\hat{\beta})$ *P*-value	Odds ratio (95% CI)	$\widehat{\mathrm{SE}}(\hat{\beta})$ *P*-value	Odds ratio (95% CI)	Robust $\widehat{\mathrm{SE}}(\hat{\beta})$ *P*-value
Nonwhite race	1.61	0.16	1.41	0.11	1.26	0.08
	(1.11–2.11)	0.003	(1.11–1.71)	0.001	(1.06–1.46)	0.004
Female sex	0.54	0.34	0.96	0.28	0.80	0.17
	(0.18–0.90)	0.07	(0.42–1.49)	0.44	(0.53–0.1.07)	0.21
Education	1.61	0.09	1.41	0.06	1.28	0.04
	(1.31–1.91)	10^{-4}	(1.24–1.58)	10^{-8}	(1.17–1.39)	10^{-4}
CD4	1.15	0.05	1.06	0.03	1.07	0.02
(100 cells/mm^3 increase)	(1.05–1.25)	0.003	(0.99–1.13)	0.051	(1.03–1.11)	0.003
Calendar time	2.60	0.07	2.01	0.08	1.61	0.04
(years since 2/17/95)	(2.24–2.96)	10^{-4}	(1.69–2.32)	10^{-4}	(1.47–1.75)	10^{-4}
Mean squared error	0.222		0.210		0.203	

for women in the conditional and marginal models, respectively), efficiency considerations are less important. Also, Table 3 presents the results of fitting a linear logistic conditional model with a double exponential random effect leading to the $h^\star$ link function given in Equation (10) marginally. Here, the estimates for the $h^\star$ link model lie between the conditional and marginal fitted values. In the table we see that the mean squared error (MSE) of the marginal and the $h^\star$ fit are equivalent with the conditional a slightly poorer fit. We fit this model to the data in FORTRAN 77 using the FSQP algorithm[20] to minimize the function $\mathbf{S}'\mathbf{S}$, where $\mathbf{S} = \sum_{i=1}^{n} \mathbf{D}_i \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mu_i)$, where $\mathbf{D}_i$ is the matrix of derivatives of the link function with respect to $\boldsymbol{\beta}$, $\mathbf{R}_i$ is the working correlation matrix with estimated intraclass correlation 0.49 (PROC GENMOD) in the off-diagonals. Automatic differentiation by Adifor[21] was used to obtain the gradient of $\mathbf{S}'\mathbf{S}$. Interpretation of the results thus depends on whether a SS interpretation is desired or a PA one. Like most analytic epidemiology studies, NHFL is not designed to be a random sample from any population of interest; hence the PA effect has uncertain applicability.

2.5 Conditions for the equivalence of the Cox proportional hazards model and the Cox frailty model

The Cox proportional hazards model by Cox[22] is specified through the hazard function, $\lambda(t)$:

$$\lambda_i(t) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta}) \tag{12}$$

where $i = 1, \ldots, n$. A random effect,or frailty, can be introduced into a model for the conditional hazard function[23] by

$$\lambda_i(t|u) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta} + u_i) \tag{13}$$

where u_i is the random effect. Define the random effect or frailty W as $W = \exp(U)$. As previously stated, the question that then arises is under what conditions do the parameters $\boldsymbol{\beta}$ in model (12) have the same interpretation as the parameters which arise from the marginal hazard derived from the conditional given by Equation (13)? It turns out that when the conditional hazard has the form given by the Weibull distribution, and $\exp(U)$ follows a positive stable frailty distribution indexed by the parameter $\alpha(0 < \alpha < 1)$,[24] the marginal hazard also has the Weibull form.[23] To be more specific, let

$$\lambda_i(t|u) = \kappa t^{\kappa-1} \exp(\mathbf{x}_i'\boldsymbol{\beta} + u_i) = \lambda_0(t) \exp(\mathbf{x}_i'\boldsymbol{\beta} + u_i) \tag{14}$$

where the frailty $\exp(U)$ has a positive stable distribution with parameter α, κ is the index parameter and $\exp(\mathbf{x}_i'\boldsymbol{\beta})$ is the shape parameter. The conditional survivor function $S(t|w)$ is then $\exp[-\exp(\mathbf{x}_i'\boldsymbol{\beta})t^{\kappa}w]$ and the marginal survivor function is $S(t) = \exp[-(\exp(\mathbf{x}_i'\boldsymbol{\beta}\alpha)t)^{\kappa\alpha}]$. The marginal hazard function is then given by

$$\lambda_i(t) = \kappa\alpha t^{\kappa\alpha-1} \exp(\mathbf{x}_i'\boldsymbol{\beta}\alpha) = \lambda_0^{\star}(t) \exp(\mathbf{x}_i'\boldsymbol{\beta}\alpha) \tag{15}$$

Hence, the marginal hazard is again Weibull proportional hazards but with index $\kappa\alpha$. With a conditional Weibull distribution and a positive stable frailty, the estimate of $\boldsymbol{\beta}$ will be attenuated by α because α ranges between 0 and 1.[25]

3 Discussion

3.1 Marginal model when the random effect and the covariate are dependent

The models in Equations (1) and (2) assume that the random effect is independent of the covariates. When this is not the case we instead have

$$f(\mathbf{y}|\mathbf{X}) = \int_{\mathrm{U}} f(\mathbf{y}|\mathbf{u}, \mathbf{X}) f(\mathbf{u}|\mathbf{X})\, \mathrm{d}\mathbf{u}$$

where the conditional distribution of U given X now depends on the covariates. Heagerty and Zeger[13] proposed a marginalized random effects model. They showed that if one incorrectly assumed that the random effects variance did not depend on the covariates, bias would be introduced in estimates of the parameters of interest. An alternative method is proposed when the random effect is correlated with the covariate[2,26] This situation is identical to the omitted covariate problem,[27,28] where $\mathbf{U}$ can be regarded as the omitted covariate in the marginal model.

These authors show that, unless the omitted covariates are uncorrelated with the variables of interest, the estimates for the effects of the variables of interest will be biased. Gail *et al.*[28] studied the model with a covariate U independent of a binary treatment variable X. The conditional mean of the response Y given U and X is given by $E[Y|U, X] = h(\mu + X\beta + U\alpha)$, where h is a known function and as previously stated,

β is the parameter of interest. For the incorrect assumption that $E[Y|U, X] = E[Y|X]$, that is, omitting the covariate U, Gail *et al.*[28] gave an expression for the magnitude of the bias of β in estimation of the effect of the included covariate because of an omitted covariate. The distribution of Y given (U,X) is assumed to belong to the exponential family.

Pepe and Anderson[29] gave an example of a simple linear auto-regression model with covariates. The model was given by

$$Y_{it} = \alpha Y_{i(t-1)} + X_{it}\beta + \varepsilon_{it}, \quad t = 1, \ldots, n_i$$

where X_{it} and ε_{it} $(i = 1, \ldots, K)$ are all normally distributed with mean 0 and unit variance. Here $Y_{i0} = 0$ and the covariates X_{it} are correlated within each series with $\rho_x = \text{Corr}(X_{it}, X_{i(t-1)})$ and independent of ε_{it}. They showed that estimates of the effect of the included covariate, because of an omitted covariate, is biased when there is a correlation between the omitted covariates $(X_{it}, t \neq t')$ and the included covariate $(X_{it'})$. Emond *et al.*[30] derived an expression for this bias of β, which agreed with simulations in Pepe and Anderson.[29]

The problem of omitted covariates has been discussed in the epidemiologic literature as the problem of confounding.[31] If U is now recast as the unmeasured confounder, many interesting results from this literature apply to the problem considered here. For example, for binomial and Bernoulli models for Y with a logit link in (X, U), if $\text{Corr}(X, U|Y = 0) = 0$, the marginal slope for X will equal the conditional slope for X. For binomial and Bernoulli models for Y with log or identity link functions, the marginal slope for X will equal the conditional slope for X if $\text{Corr}(X, U) = 0$ marginally. These are conditions on the correlation between X and U rather than on the entire distribution, somewhat less restrictive conditions than that given previously, that is $f(u|x) = f(u)$.

4 Conclusions

The problem addressed in this paper can be formulated generally as follows: given a certain conditional mean function $g(\mathbf{u};\mu) = E[\mathbf{Y}|\mathbf{U} = \mathbf{u}]$, for what distributions $F_u(\mathbf{u})$ of the random effect $\mathbf{U}$ will the integral

$$\int_{\mathrm{U}} g(\mathbf{u};\, \mu)\, \mathrm{d}F_u(\mathbf{u})$$

have an expression that is exactly linear in the mean function μ plus some other term not involving μ? In the model given by Equation (1), we have seen that the conditional and marginal models are equivalent when we have a regression that is additive (linear) in a mean 0 random effect. The conditional and marginal slopes are also the same when a random effect is additive in the conditional mean on the log scale. However, both of these results only hold when the random effect and the covariate are independent. The conditional and marginal slopes are also the same in a logistic model of Bernoulli data when the covariate and random effect are uncorrelated given $Y = 0$.[31] In a proportional

hazards frailty model with a positive stable frailty and conditional Weibull hazard, the marginal hazard is also Weibull. However, the regression coefficient becomes attenuated by a factor determined by the frailty parameter. The regression coefficient from the marginal model will always be smaller in magnitude than the regression coefficient from the conditional model.

These results are useful in practice. Marginal models popularly fit by generalized estimating equations (GEEs), as implemented, for example, in PROC GENMOD in SAS[8] and in the GEE function in S-PLUS,[32] can be validly applied for estimating marginal model parameters induced by their conditional model counterparts as long as the log or identity link function is of interest. In the documentation for PROC NLMIXED, there is an example (Example 44.4) using Poisson counts with normal random effects. As shown in Section 2.2, for estimating covariate effects, for the nonlinear mixed model the same estimates would be obtained by fitting a GEE in PROC GENMOD. Several further examples of this use of nonlinear mixed models, using the Poisson distribution with log link, can also be found in Breslow and Clayton.[12] Of course, nonlinear mixed models[12,33,34] will be needed to estimate conditional model parameters for the logit link function validly, unless U is reasonably assumed to be distributed as double exponential and the implicit link function in Equation (9) is applied to the induced marginal model.

When equivalence of the two regressions is established, it is of interest in further research to determine circumstances when either estimator from methods such as penalized quasi-likelihood (PQL) or the GEE approach for the marginal model have a uniform efficiency advantage over each other. When estimation and inference is based on the likelihood, except when the conditional distribution of $\mathbf{Y}|\mathbf{U}$ and $\mathbf{U}$ are normal and the identity link applies, it is unclear under what other distributions for $\mathbf{Y}|\mathbf{U}$ and for $\mathbf{U}$ and with what other link functions will the induced marginal distribution equal an assumed marginal distribution. Future software development permitting routine estimation and inference for $\boldsymbol{\beta}$ under more numerically intractable forms for $\mathbf{Y}|\mathbf{U}$ and $\mathbf{U}$ will make some of the issues discussed in this paper less important.

Acknowledgements

We would like to thank the editor and referee for their valuable comments and suggestions in preparing this article.

References

1 Lindsey JK, Lambert P. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* 1998; **17**: 447–69.

2 Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 1991; **59**: 25–35.

3 Crouchley R, Davis RB. A comparison of population average and random-effect models for the analysis of longitudinal count data with base-line information. *Journal of the Royal Statistical Society, Series A* 1999; **162** (Part 3): 331–47.

4 Neuhaus JM, Jewell N. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrics* 1993; **80**: 807–15.

5 Graubard BI, Korn EL. Regression analysis with clustered data. *Statistics in Medicine* 1994; **13**: 509–22.

6 Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford University Press, 1994.

7 Woods M, Spiegelman D, LaBrode A, Hertzmark E, Longcope C. Effect of soy on hormones and hot flashes. In *Proceedings of 'Era of Hope'*, Department of Defense Research Program Meetings, 8–11 June 2000, Atlanta, GA; **II**: 506.

8 SAS Institute Inc. *SAS/STAT user's guide, version 8*. Cary, NC: SAS Institute Inc., 1999, 3884 pp.

9 Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**: 1049–60.

10 Grömping U. A note on fitting a marginal model to mixed effects log-linear regression data via GEE. *Biometrics* 1996; **52**: 280–85.

11 Hinde J. Compound Poisson regression models. In Gilchrist R, ed. *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*. Berlin, Germany: Springer Verlag, 1983.

12 Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**: 9–25.

13 Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference. *Statistical Science* 2000; **14**: 29–46.

14 Crouch EAC, Spiegelman D. The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t) \exp(-t^2)\, dt$: application to logistic-normal models. *Journal of the American Statistical Association* 1990; **85**: 464–69.

15 Rabe-Hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* 2002; **2**: 1–21.

16 Lange N, Ryan L. Assessing normality in random effects models. *The Annals of Statistics* 1989; **17**: 624–42.

17 Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* 1997; **23**: 541–56.

18 Silva M, Skolnik PR, Gorbach SG, Spiegelman D, Wilson IB, Fernandez-deFranco MG, Knox TA. The effect of protease inhibitors on weight and body composition in HIV-infected patients. *AIDS* 1998; **12**: 1645–51.

19 Knox TA, Spiegelman D, Skinner SC *et al.* Diarrhea and abnormalities of gastrointestinal function in a cohort of men and women with HIV Infection. *American Journal of Gastroenterology* 2000; **95**: 3482–89.

20 Panier ER, Tits AL. On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming* 1993; **59**: 261–76.

21 Bischof C, Carle A, Corliss G, Griewank A, Hovland P. Adifor: generating derivative codes from Fortran programs. *Scientific Programming* 1992; **1**: 11–29.

22 Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**: 187–220.

23 Hougaard P. Life table methods for heterogenous populations: distributions describing the heterogeneity. *Biometrika* 1984; **71**: 75–83.

24 Feller W. *An introduction to probability theory and its applications 2*. New York, NY: John Wiley and Sons, 1971.

25 Hougaard P. A class of multivariate failure time distributions. *Biometrika* 1986; **73**: 671–78.

26 Ten Have TR, Landis JR, Weaver S. Association models for periodontal disease progression: a comparison of methods for clustered binary data. *Statistics in Medicine* 1996; **15**: 1227–29.

27 Lee LF. Specification error in multinomial logit models: analysis of the omitted variable bias. *Journal of Econometrics* 1982; **20**: 197–209.

28 Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with non-linear regressions and omitted covariates. *Biometrika* 1984; **71**: 431–44.

29 Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulations and Computations* 1994; **23**: 939–51.

30 Emond MJ, Ritz J, Oakes D. Bias in GEE estimates from misspecified models for longitudinal data. *Communications in Statistics, Theory and Methods* 1997; **26**: 15–32.
31 Boivin JF, Wacholder S. Conditions for confounding of the risk ratio and the odds ratio. *American Journal of Epidemiology* 1985; **121**: 152–58.
32 Mathsoft. *S-PLUS user's manual, version 6.0*. Data Analysis Division, Seattle, WA: Mathsoft, 2000.
33 Lin X, Breslow NE. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* 1996; **91**: 1007–16.
34 Wolfinger RD. Laplace's approximation for nonlinear mixed models. *Biometrika* 1993; **80**: 791–95.