Expected Estimating Equations to Accommodate Covariate Measurement Error

C. Y. Wang and Margaret Sullivan Pepe

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Abstract

Estimating equations which are not necessarily likelihood–based score equations are becoming increasingly popular for estimating regression model parameters. This paper is concerned with estimation based on general estimating equations when true covariate data are missing for all study subjects, but surrogate or mismeasured covariates are available instead. The method is motivated by the covariate measurement error problem in marginal or partly conditional regression of longitudinal data. We propose to base estimation on the expectation of the complete data estimating equation (EEE) conditioned on available data. The regression parameters and other nuisance parameters are estimated simultaneously by solving the resulting estimating equations. The EEE estimator is equal to the maximum likelihood estimator if the complete data scores are likelihood scores and conditioning is with respect to all available data. A pseudo–EEE estimator, which requires less computation, is also investigated. Asymptotic distribution theory is derived. Small sample simulations are conducted when the error process is an order–1 autoregressive model. Regression calibration is extended to this setting and compared with the EEE approach. We demonstrate the methods on data from a longitudinal study of the relationship between childhood growth and adult obesity.

Some key words: Missing Data; Generalized Estimating Equation; Longitudinal Data; Logistic Regression; Maximum Likelihood; Regression Calibration.

Abbreviated title: Expected Estimating Equations

1 INTRODUCTION

Errors in variables is a problem which arises in many scientific fields. Some examples are: the measurement of radiation exposure (Prentice, 1982), the measurement of systolic blood pressure (Carroll *et al.*, 1984) and the measurement of dietary intake (Rosner *et al.*, 1989). Measurement error may be due to inadequacies in measuring instruments or due to human error. Errors in variables can also arise even when the measurement process per se is accurate. Consider, for example, when the variable of interest is the long-term average of a quantity which varies over time. Because in reality only a finite number of observations can be made, only an estimate of the variable of interest is obtained, which therefore has error associated with it. In the literature, the term measurement error is used synonymously with errors in variables and we will use the former term here.

It is well known that if measurement error in a covariate is ignored, then the associated regression coefficient is attenuated in the simple linear regression model when the classical additive error model holds (Carroll *et al.*, 1995, Chapter 2). In more general settings the effects of covariate measurement error can be complex, for example when the measurement error has a systematic bias (Prentice, 1996). Various strategies for dealing with measurement error have been proposed. Two important monographs have been published on statistical methods which accommodate measurement error. Fuller (1987) considered inference for the classical linear model. Carroll *et al.* (1995) more recently considered inference in nonlinear models. In this paper we will assume that true covariate data are not available for any study subject but that repeated measurements of a mismeasured variable or a surrogate variable are available. This is sometimes called a reliability sample. This contrasts with a validation sample (Armstrong *et al.*, 1992, Chapter 4) where the true covariate of interest is available for a subsample of observations termed the validation set (Pepe, 1992; Robins *et al.*, 1994; Reilly and Pepe, 1995).

Consider first the classic setting where the parameters in the regression model relating the outcome to the true covariate data of interest could be estimated with maximum likelihood techniques if true covariate data were observed. In the presence of covariate measurement error, one approach to data analysis involves modelling the true covariate and its association with the observed covariate data. Maximum likelihood estimation can then be implemented in order to estimate all of the model parameters simultaneously. The EM algorithm (Dempster *et al.*, 1977) is often invoked to find the maximum likelihood estimates. If S_β(complete data) denotes the likelihood score function for the entire set of parameters if complete data were observed, then EM seeks to solve the expected estimating equation

$$E_\beta \left\{ S_\beta \left(\text{complete data}\right) \mid \text{observed data} \right\} = 0. \tag{1}$$

The purpose of this paper is to propose the use of analogous expected estimating equations for settings where the complete data score function S_β(complete data) is not necessarily a likelihood based score function.

Our motivation for this technique derives from application of a marginal regression analysis to longitudinal data. Let Y_{ij} denote the outcome measure at time t_j for subject i and X_{ij} denote the true covariate. Let $\widehat{X}_{ij}$ denote an estimate of X_{ij}. With true covariate data available, marginal or partly conditional regression analysis can be based on estimating equations of the form

$$S(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} S_\beta \left(Y_{ij}, X_{ij}\right) = 0, \tag{2}$$

(Pepe and Couper, 1997). In our application the covariate X_{ij} is an average taken over some time interval which is estimated by a sample average $\widehat{X}_{ij} = \sum_L w_{ijL}/m_{ij}$ where $\{w_{ijL}, L = 1, \ldots, m_{ij}\}$ are the m_{ij} observed covariate values for individual i in the jth interval. Since the distributions of X_{ij} and its association with $\widehat{X}_{ij}$ are easily characterized, if the estimating equation (2) were a likelihood score equation, maximum likelihood estimates which solve the EM equation (1) could be calculated. We propose that an EM type equation can still be used even when the complete data estimating function is of the form (2), or more generally if it is an unbiased estimating equation.

Our idea of applying expected estimating equations to the measurement error problem is different from the two other main approaches to this problem, namely, regression calibration and inverse selection probability weighting. For quasi-likelihood models with small measurement error, Carroll and Stefanski (1990) used regression calibration, where a missing or mismeasured covariate is replaced by its conditional expectation given surrogates. Inverse selection probability weighting was used by Rotnitzky *et al.* (1998), for example. Indeed they showed that the optimal estimator in their class attains the semi-parametric variance bound for the missing data model they considered. However, such optimal estimators are often difficult to calculate explicitly.

In order to motivate our proposed EEE we first consider the full likelihood setting in detail in Section 2. We derive the maximum likelihood estimator of parameters based on observed data as a solution to an EEE. The more general EEE estimator is proposed in Section 3 along with two important modifications. The first allows conditioning on subsets of available data, thereby reducing the number of modelling assumptions required. This is particularly important for the aforementioned marginal regression problem. The second modification concerns increasing the flexibility with which parameters in the covariate and error distributions are estimated. Substantial computational advantages arise in our example for certain choices of simple estimating procedures. Asymptotic distribution theory and technical considerations regarding computation are discussed in Section 3.

We consider logistic regression with correlated covariate measurement error as the illustrative example in this paper. This is an important example since logistic regression is commonly used in biomedical research and covariate measurement error is a common problem. The error structure we consider is more general than the classic 'independent identically distributed errors' model which has received a lot of attention in the literature. A correlated rather than independent error structure is likely to hold in longitudinal data applications such as the one motivating our work. Such an error structure has previously been considered for a generalized linear model by Wang, Carroll and Liang (1996). They assumed balanced equally–spaced data. In this paper, we allow unbalanced or unequally spaced data. In Section 4 the models and estimating methods are defined for this example. Regression calibration has been shown to perform well for the logistic regression problem with classic covariate measurement error structure when the relative risk parameter for the unobserved covariate is moderate. We define a regression calibration procedure for the correlated error structure in Section 4 as well. Results of a simulation study are presented in Section 5. Finally, analyses of data from the Childhood Predictors of Adult Obesity Study (Whitaker *et al.*, 1997) which motivated our work are presented in Section 6.

2 MAXIMUM LIKELIHOOD

2.1 *Notation*

Let Y_i be a possibly vector-valued outcome variable for the i–th of n subjects. Let X_i be a corresponding covariate vector which is measured with error. In this section, which is concerned with likelihood based inference, the model of interest is written as $P_\beta(Y|X)$ for some unknown parameter β. Instead of X_i, mismeasured or surrogate values denoted by $\tilde{W}_i = (W_{i1}, \ldots, W_{ik_i})$ are available. We assume that $P_\beta(Y|X, \tilde{W}) = P_\beta(Y|X)$. This has been called the surrogacy condition and means that $\tilde{W}$ offers no additional information regarding the outcome Y given data on the true covariate X. The measurement error process $P_{\eta_w}(\tilde{W}|X)$ is parameterized up to an unknown vector η_w, and the true covariate distribution $P_{\eta_x}(X)$ is modeled up to an unknown vector η_x. Let $\eta = (\eta_w^t, \eta_x^t)^t$ which is a vector of nuisance parameters since we are mainly interested in the coefficients contained in β. For notational simplicity, we suppress notation for the subindex β in $P_\beta(Y|X)$, η_w in $P_{\eta_w}(\tilde{W}|X)$ and η_x in $P_{\eta_x}(X)$ throughout.

The vector of all unknown parameters is $\Theta = (\beta^t, \eta^t)^t$. If X is not measured with error, estimation of Θ can be based on the likelihood derived from $P(Y|X)$, $P(\tilde{W}|X)$, and $P(X)$. Let the corresponding likelihood scores be $\phi_\beta(Y_i, X_i) = (\partial/\partial\beta)\log P(Y|X)$, $\phi_{\eta_w}(\tilde{W}, X) = (\partial/\partial\eta_w)\log P(\tilde{W}|X)$ and $\phi_{\eta_x}(X) = (\partial/\partial\eta_x)\log P(X)$. Then the resulting score function estimating equations are:

$$\begin{cases} \sum_{i=1}^n \phi_\beta(Y_i, X_i) = 0; \\ \sum_{i=1}^n \phi_\eta(\tilde{W}_i, X_i) = 0, \end{cases} \tag{3}$$

where $\phi_\eta(\tilde{W}, X) = \{\phi_{\eta_w}^t(\tilde{W}, X), \phi_{\eta_x}^t(X)\}^t$. Note that the second equation of (3) is superfluous here since the first equation is enough to yield an estimate of β. We keep it for the development of the EEE method.

2.2 *Maximum Likelihood EEE*

Equation (3), the complete data estimating equation, involves the true covariate, X_i. Because X_i is not available we now consider likelihoods based on the available data. First note that the observed data likelihood components for β and η are $P(Y|\tilde{W})$ and $P(Y|\tilde{W})P(\tilde{W})$, respectively. Under the reasonable assumption that the distribution of $(\tilde{W}, X)$ does not involve β, then

$$\begin{aligned} \frac{\partial}{\partial\beta}\log P(Y|\tilde{W}) &= \frac{(\partial/\partial\beta)\int P(Y|X)P(\tilde{W}|X)P(X)dX}{P(Y,\tilde{W})} \\ &= E_\Theta\{\frac{\partial}{\partial\beta}\log P(Y|X)|Y,\tilde{W}\}. \end{aligned}$$

Similarly, $(\partial/\partial\eta_w)\log P(Y|\tilde{W}) = E_\Theta\{(\partial/\partial\eta_w)\log P(\tilde{W}|X)|Y,\tilde{W}\}$; $(\partial/\partial\eta_w)\log P(\tilde{W}) = E_\eta\{(\partial/\partial\eta_w)\log P(\tilde{W}|X)|\tilde{W}\}$; $(\partial/\partial\eta_x)\log P(Y|\tilde{W}) = E_\Theta\{(\partial/\partial\eta_x)\log P(X)|Y,\tilde{W}\}$; and $(\partial/\partial\eta_x)\log P(\tilde{W}) = E_\eta\{(\partial/\partial\eta_x)\log P(X)|\tilde{W}\}$. Hence,

$$\begin{aligned} \frac{\partial}{\partial\eta_w}\log P(Y,\tilde{W}) &= E_\Theta\{\frac{\partial}{\partial\eta_w}\log P(\tilde{W}|X)|Y,\tilde{W}\} + E_\eta\{\frac{\partial}{\partial\eta_w}\log P(\tilde{W}|X)|\tilde{W}\}; \\ \frac{\partial}{\partial\eta_x}\log P(Y,\tilde{W}) &= E_\Theta\{\frac{\partial}{\partial\eta_x}\log P(X)|Y,\tilde{W}\} + E_\eta\{\frac{\partial}{\partial\eta_x}\log P(X)|\tilde{W}\}. \end{aligned}$$

Therefore, the maximum likelihood estimator of Θ solves

$$\left\{ \begin{array}{l} \sum_{i=1}^{n} E_\Theta\{\phi_\beta(Y_i, X_i)|Y_i, \tilde{W}_i\} = 0; \\ \\ \sum_{i=1}^{n} \left[E_\Theta\{\phi_\eta(\tilde{W}_i, X_i)|Y_i, \tilde{W}_i\} + E_\eta\{\phi_\eta(\tilde{W}_i, X_i)|\tilde{W}_i\}\right] = 0. \end{array} \right. \tag{4}$$

Observe that the estimating equation (4) is derived from the complete data estimating equation (3) by taking conditional expectations with respect to observed data.

3 EXPECTED ESTIMATING EQUATION METHOD

3.1 *Expected Estimating Equations*

Motivated in part by the representation of the likelihood score equations in (4) we now consider a general approach to estimation when the complete data estimating equation is an unbiased estimating equation. Let $S_\beta(Y, X)$ and $S_\eta(\tilde{W}, X)$ be the components of the complete data estimating equation which correspond to β and η and which are assumed to satisfy the condition of unbiasedness, $E_\Theta\{S_\beta(Y, X)\} = 0$, $E_\eta\{S_\eta(\tilde{W}, X)\} = 0$. Thus in the absence of measurement error the estimating equations for β and η are $\sum_{i=1}^{n} S_\beta(Y_i, X_i) = 0$ and $\sum_{i=1}^{n} S_\eta(\tilde{W}_i, X_i) = 0$. We define the expected estimating equation estimator of Θ as that which solves

$$\left\{ \begin{array}{ll} \sum_{i=1}^{n} E_\Theta\{S_\beta(Y_i, X_i)|Y_i, \tilde{W}_i\} = 0; & (5.1) \\ \\ \sum_{i=1}^{n} \left[E_\Theta\{S_\eta(\tilde{W}_i, X_i)|Y_i, \tilde{W}_i\} + E_\eta\{S_\eta(\tilde{W}_i, X_i)|\tilde{W}_i\}\right] = 0. & (5.2) \end{array} \right. \tag{5}$$

This equation is the direct analogue of (4) for general estimating equations in that it is calculated by taking conditional expectations of complete data scores given observed data.

3.2 *Marginal Regression for Longitudinal Data*

Our original motivation for considering the EEE approach derived from our interest in making inference for marginal or partly conditional regression models for longitudinal data (Pepe and Couper, 1997) when some covariates are measured with error. Let $\{(Y_{ir}, X_{ir}), r = 1, \cdots, n_i\}$ be outcome and covariate pairs for subject $i, i = 1, \cdots, n$. Suppose that the marginal regression model of interest is written as $P_\beta(Y_{ir}|X_{ir})$ and instead of X_{ir} a surrogate or estimate of X_{ir} is measured. If X_{ir} were observed then inference could be based on the estimating equation

$$\sum_{i=1}^{n} \sum_{r=1}^{n_i} S_\beta(Y_{ir}, X_{ir}) = 0 \tag{6}$$

where $S_\beta(Y_{ir}, X_{ir}) = (\partial/\partial\beta)\log\ P_\beta(Y_{ir}|X_{ir})$. Generalized estimating equations with non-diagonal working covariance matrices can be used if certain conditions are satisfied (Pepe and Anderson, 1994) but we do not consider those here for reasons which will be apparent below. Calculation of an expected estimating equation by directly applying the conditioning approach of (4) or (5) yields terms of the form $E\{S_\beta(Y_{ir}, X_{ir})|Y_i, \tilde{W}_i\}$ where $Y_i = (Y_{i1}, \cdots, Y_{in_i})$ and $\tilde{W}_i = (\tilde{W}_{i1}, \cdots, \tilde{W}_{in_i})$. This requires specification of the distributions which condition on the vector $(Y_{i1}, \cdots, Y_{in_i})$ and hence which involve the joint distribution of the outcome variables $Y_{i1}, \cdots, Y_{in_i}$. Since a key attribute of the marginal regression approach is that it does not require specification of such a joint distribution, we modify the EEE approach here for clustered data. Instead of $E_\Theta\{S_\beta(Y_{ir}, X_{ir})|Y_i, \tilde{W}_i\}$ we employ terms of the form $E_\Theta\{S_\beta(Y_{ir}, X_{ir})|Y_{ir}, \tilde{W}_{ir}\}$ in the EEE. These involve only marginal distributions.

Thus, in addition to the marginal (or partly conditional) model for $P(Y_{ir}|X_{ir})$ we specify a marginal model for $P(W_{ir}|X_{ir})$ and a marginal model for $P(X_{ir})$. Let $S_\eta(W_{ir}, X_{ir}) = \phi_\eta(\tilde{W}_{ir}, X_{ir})$ where $\phi_{\eta_w}(\tilde{W}_{ir}, X_{ir}) = (\partial/\partial_{\eta_w})\ \log P(\tilde{W}_{ir}|X_{ir})$ and $\phi_{\eta_x}(X_{ir}) = (\partial/\partial\eta_x)\log P(X_{ir})$. In summary then we propose that the EEE for marginal regression be

$$\begin{cases} \sum_{i=1}^n \sum_{r=1}^{n_i} E_\Theta\{S_\beta(Y_{ir}, X_{ir})|Y_{ir}, \tilde{W}_{ir}\} = 0; \\ \sum_{i=1}^n \sum_{r=1}^{n_i} \Big[E_\Theta\{S_\eta(\tilde{W}_{ir}, X_{ir})|Y_{ir}, \tilde{W}_{ir}\} + E_\eta\{S_\eta(\tilde{W}_{ir}, X_{ir})|\tilde{W}_{ir}\}\Big] = 0. \end{cases} \tag{7}$$

3.3 *Pseudo–EEE estimator*

The key step in extending (5) to the EEE in (7) is to calculate expectations conditioning on $(Y_{ir}, \tilde{W}_{ir})$ rather than conditioning on $(Y_i, \tilde{W}_i)$. That is, instead of conditioning on the entire set of available data for the i–th subject we condition on a subset of the data. This principle can be used more generally. In equation (5) one could condition on a subset of $(Y_i, \tilde{W}_i)$ in either the first or second equations, and the resultant EEE would still be unbiased. For example, in the second equation one could condition on $\tilde{W}_i$ only rather than on $(Y_i, \tilde{W}_i)$. Indeed if the true association between X and Y given $\tilde{W}$ is weak then the conditional expectations, $E_\Theta\{S_\eta(\tilde{W}_i, X_i)|Y_i, \tilde{W}_i\}$ and $E_\eta\{S_\eta(\tilde{W}_i, X_i)|\tilde{W}_i\}$ will be approximately the same so that little will be lost in terms of efficiency. The computational burden however in solving the reduced equation is likely to be much less. If the association between Y and X is strong then ignoring the information in Y about X in the conditional expectation of $S_\eta(\tilde{W}_i, X_i)$ will reduce the efficiency in estimating the measurement error parameters η from the second equations in (5). This may in turn reduce efficiency in estimating β_1, though in our simulation studies we found this to be minimal in large samples.

A method which avoids calculation of the conditional expectations in (5.2) is sometimes feasible. In fact, if an estimating function Ψ for η exists which is only a function of the observed covariate data then the second component of the complete data estimating equation can be written as $\sum_{i=1}^{n} \Psi(\tilde{W}_i, \eta) = 0$. In this case, the conditioning in (5.2) is irrelevant and the expected estimating equation is

$$\begin{cases} \sum_{i=1}^{n} E_{\Theta}\{S_{\beta}(Y_i, X_i)|Y_i, \tilde{W}_i\} = 0; \\ \\ \sum_{i=1}^{n} \Psi(\tilde{W}_i, \eta) = 0. \end{cases} \tag{8}$$

Furthermore, the equations can be solved for Θ, by first finding $\widehat{\eta}$ to solve $\sum_{i=1}^{n} \Psi(\tilde{W}_i, \eta) = 0$ and then finding $\widehat{\beta}$ to solve $\sum_{i=1}^{n} E\{S_{\beta}(Y_i, X_i)|Y_i, \tilde{W}_i\} = 0$ at $\eta = \widehat{\eta}$. This bears a resemblance to pseudolikelihood methods (Gong and Samaniego, 1981) and hence we call (8) a pseudo–EEE. Pseudo-EEE estimates based on (7) for clustered data can also be defined. In the example, we employ a method of moments type estimator for η which uses an estimating function of the form $\Psi(\tilde{W}, \eta)$. Although the resultant pseudo–EEE estimator is less efficient than alternative EEE estimators, it requires far less computation time.

3.4 *Asymptotic Distribution Theory*

Asymptotic distribution theory for the estimator which solves the EEE is summarized in the following result.

Proposition 1. Assume complete data scores $S_{\beta}(Y, X), S_{\eta_w}(\tilde{W}, X)$, $S_{\eta_x}(X)$ satisfy $E_{\Theta}\{S_{\beta}(Y, X)\} = 0, E_{\eta}\{S_{\eta_w}(\tilde{W}, X)\} = 0$ and $E_{\eta}\{S_{\eta_x}(X)\} = 0$. Write (5), (7) or (8) as $\sum_{i=1}^{n} g_i(\Theta) = 0$ and denote the estimator of Θ which solves (5), (7) or (8) by $\widehat{\Theta}$. If $n^{-1}\sum_{i=1}^{n}(\partial/\partial\Theta^t)g_i(\Theta)$ converges in probability to $G(\Theta)$ for some negative definite matrix $G(\Theta)$ and $n^{-1}\sum_{i=1}^{n} g_i(\Theta)g_i^t(\Theta)$ converges in probability to $V(\Theta)$ for some nonnegative definite matrix $V(\Theta)$, then $n^{1/2}(\widehat{\Theta} - \Theta)$ is asymptotically normally distributed with mean 0 and variance $G^{-1}(\Theta)V(\Theta)\{G^{-1}(\Theta)\}^t$.

By noting that the estimating equations (5), (7) and (8) are themselves unbiased and are sums of independent components, the above result can be proven using steps analogous to standard proofs for generalized estimating equations. The asymptotic variance of $n^{1/2}(\widehat{\Theta} - \Theta)$ can be consistently estimated by $G_n^{-1}(\widehat{\Theta})V_n(\widehat{\Theta})\{G_n^{-1}(\widehat{\Theta})\}^t$, where $G_n(\Theta) = n^{-1}\sum_{i=1}^{n}(\partial/\partial\Theta^t)g_i(\Theta)$ and $V_n(\Theta) = n^{-1}\sum_{i=1}^{n} g_i(\Theta)g_i^t(\Theta)$.

3.5 CALCULATION OF EEE AND PSEUDO-EEE ESTIMATORS

Consider now evaluation of the expected scores in (5) or (8). These are of the form $E_\Theta\{S_\beta(Y,X)| Y,\tilde{W}\}$, $E\{S_\eta(\tilde{W},X)|Y,\tilde{W}\}$ and $E\{S_\eta(\tilde{W},X)|\tilde{W}\}$. It can be shown that

$$E_\Theta\{S_\beta(Y,X)|Y,\tilde{W}\} = \frac{\int S_\beta(Y,x)P(Y|x)P(\tilde{W}|x)P(x)dx}{\int P(Y|x)P(\tilde{W}|x)P(x)dx};$$
$$E_\Theta\{S_\eta(\tilde{W},X)|Y,\tilde{W}\} = \frac{\int S_\eta(\tilde{W},x)P(Y|x)P(\tilde{W}|x)P(x)dx}{\int P(Y|x)P(\tilde{W}|x)P(x)dx}; \quad \text{and}$$
$$E_\eta\{S_\eta(\tilde{W},X)|\tilde{W}\} = \frac{\int S_\eta(\tilde{W},x)P(\tilde{W}|x)P(x)dx}{\int P(\tilde{W}|x)P(x)dx}.$$

At any particular value of Θ, evaluation of the expected scores requires integration with respect to the distribution of X. The joint distribution of W and X needs to be specified. In our simulations in Section 4, numerical integration was used to approximate the expected scores. Alternatively, one may apply an approximation based on Monte-Carlo simulations (McFadden and Ruud, 1994).

4 LOGISTIC REGRESSION WITH AR(1) COVARIATE ERROR

4.1 *Models and EEE*

As an illustrative example, we consider logistic regression with correlated measurement error in covariates. Thus, the outcome variable Y is binary, X is a univariate covariate which is measured with error, and the error follows an order-1 autoregressive model (AR(1)). Let $P(Y=1|X) = H(\beta_0+\beta_1 X)$, where $H(z) = \{1+\exp(-z)\}^{-1}$. For each subject i, instead of observing the true covariate X we observe $\tilde{W}_i = \{W_i(t_{ij}), j=1,\cdots,k_i\}$, where

$$W_i(t_{ij}) = X_i + U(t_{ij}); \qquad j=1,\ldots,k_i.$$

Such data would arise, for example in a longitudinal study, where the covariate X_i is an average over some time interval and observations on the measurement are available at k_i discrete points in the interval. Suppose that the covariate X_i is normally distributed with mean μ_x and variance σ_x^2. The "error" $\tilde{U}_i \equiv \{U(t_{i1}),\ldots,U(t_{ik_i})\}^t$ which is normally distributed and independent of X_i, satisfies $E(\tilde{U}_i|X_i) = 0$, $\text{var}(\tilde{U}_i) = \Sigma_{U_i}$ and the (j,j')th element of Σ_{U_i} is equal to $\sigma_u^2\rho^{|t_{ij}-t_{ij'}|}$ for $j,j' \in \{1,\ldots,k_i\}$ where ρ is a nonnegative autocorrelation parameter. The correspondence between the parameters in this model and $\Theta = (\beta^t,\eta_w^t,\eta_x^t)^t$, is $\beta = (\beta_0,\beta_1)^t$, $\eta_w = (\sigma_u,\rho)^t$ and $\eta_x = (\mu_x,\sigma_x)^t$. The primary interest is in estimating β.

To apply the EEE estimator, we first determine the likelihood score functions. It is easy to see that $S_\beta(Y|X) = (1,X)^t\{Y - H(\beta_0+\beta_1 X)\}$ and $S_{\eta_x}(X) = \{X-\mu_X, (X-\mu_X)^2 - \sigma_X^2\}^t$. To estimate η_w, we note that by Jennrich and Schluchter (1986), $S_{\eta_w}(\tilde{W}_i, X) = \{S_{\sigma_u}(\tilde{W}_i, X), S_\rho(\tilde{W}_i, X)\}^t$, where

$$\begin{aligned} S_{\sigma_u}(\tilde{W}_i, X) &= \frac{1}{2}\text{trace}\left[\Sigma_{U_i}^{-1}\{(\tilde{W}_i - X)(\tilde{W}_i - X)^t - \Sigma_{U_i}\}\Sigma_{U_i}^{-1}(\frac{\partial}{\partial\sigma_u}\Sigma_{U_i})\right]; \\ S_\rho(\tilde{W}_i, X) &= \frac{1}{2}\text{trace}\left[\Sigma_{U_i}^{-1}\{(\tilde{W}_i - X)(\tilde{W}_i - X)^t - \Sigma_{U_i}\}\Sigma_{U_i}^{-1}(\frac{\partial}{\partial\rho}\Sigma_{U_i})\right]; \end{aligned}$$

where the (j,j')th element of $(\partial/\partial\rho)\Sigma_{U_i}$ is $\sigma_u^2|t_{ij} - t_{ij'}|\rho^{|t_{ij}-t_{ij'}|-1}$.

As an alternative to likelihood scores for $S_\eta(\tilde{W}_i, X_i)$ we also consider a score function $\Psi(\tilde{W}_i, \eta)$ which does not depend on X_i, and hence derive a pseudo-EEE. We propose a simple method which uses first-order differences of the components of W_{ij}. For $j = 1, \cdots, (k_i - 1)$, let $W_{ij}^* = W_{i(j+1)} - W_{ij}$, which has mean 0 and variance $2\sigma_u^2(1-\rho^{|t_{i(j+1)}-t_{ij}|})$. Further, $\text{cov}(W_{ij}^*, W_{i(j+1)}^*) = \sigma_u^2(\rho^{|t_{i(j+2)}-t_{i(j+1)}|} - 1 - \rho^{|t_{i(j+2)}-t_{ij}|} + \rho^{|t_{i(j+1)}-t_{ij}|})$. Let $\overline{W}_{i\cdot} = \sum_{j=1}^{k_i} W_{ij}/k_i$. Method of moments estimators of σ_u, ρ, μ_x and σ_x can then be based on $\sum \Psi(\tilde{W}_i, \eta) = 0$ where $\sum\ \Psi(\tilde{W}_i, \eta)$ is equal to

$$\left\{\begin{array}{l} \sum_{i=1}^n \left[\sum_{j=1}^{k_i-1}\left\{W_{ij}^{*2} - 2\sigma_u^2(1-\rho^{|t_{i(j+1)}-t_{ij}|})\right\}\right]; \\ \sum_{i=1}^n \left[\sum_{j=1}^{k_i-2}\left\{W_{ij}^* W_{i(j+1)}^* - \sigma_u^2(\rho^{|t_{i(j+2)}-t_{i(j+1)}|} - 1 - \rho^{|t_{i(j+2)}-t_{ij}|} + \rho^{|t_{i(j+1)}-t_{ij}|})\right\}\right]; \\ \sum_{i=1}^n k_i(\overline{W}_{i\cdot} - \mu_x); \\ \sum_{i=1}^n k_i\left\{(\overline{W}_{i\cdot} - \mu_x)^2 - \sigma_x^2 - (1_{k_i}^t\Sigma_{U_i}1_{k_i})/k_i^2\right\}. \end{array}\right. \tag{9}$$

Note that these estimating equations which are the second set of equations in (8) have the advantage of not requiring normality of X_i or of U_{ij}. However, calculation of $E\{S_\beta(Y_i, X_i)|Y_i, \tilde{W}_i\}$ in the first equation of (8) still requires the normality assumption.

4.2 *Regression Calibration*

Wang *et al.* (1996) studied the regression calibration (RC) approximation when errors are correlated and observation times are equally–spaced. Now we consider an extension of the regression calibration method to errors with unequal–spaced observation times. RC is an approximation to likelihood methods, in that X_i is replaced by $E(X_i|\overline{W}_{i\cdot})$ in the complete data estimating equations. Observe that $E(X_i|\overline{W}_{i\cdot}) = \mu_x + \sigma_x^2/\sigma_{\overline{W}_i}^2(\overline{W}_{i\cdot} - \mu_x)$. As in Carroll *et al.* (1995, Chapter 3), the best linear approximation to $E(X_i|\overline{W}_{i\cdot})$ is $\widehat{E}(X_i|\overline{W}_{i\cdot}) = \widehat{\mu}_X + (\widehat{\sigma}_x^2/\widehat{\sigma}_{\overline{W}_{i\cdot}}^2)(\overline{W}_{i\cdot} - \widehat{\mu}_x)$, where $\widehat{\mu}_x = \sum_{i=1}^n k_i\overline{W}_{i\cdot}/\sum_{i=1}^n k_i$, $\widehat{\sigma}_{\overline{W}_{i\cdot}}^2 = \widehat{\sigma}_x^2 + (1_{k_i}^t\Sigma_{U_i}1_{k_i})/k_i^2$. To estimate the nuisance parameters in η, one could solve (9) as was used in the pseudo–EEE estimate of η. Hence, the RC estimator of Θ simultaneously solves $\sum_{i=1}^{n} S_\beta\{Y_i, \widehat{E}(X_i|\overline{W}_{i\cdot})\} = 0$ and $\sum_{i=1}^{n} \Psi(\tilde{W}_i, \eta) = 0$. The asymptotic variance-covariance matrix for the RC estimator may be approximated by a sandwich estimator. The formula is similar to that in Proposition 1 and it takes estimation of η into account.

The RC estimator is an approximation approach. As in Carroll and Stefanski (1990, Section 3) and Carroll *et al.* (1995, page 65), the bias in the approximation is an increasing function of the relative risk parameter $|\beta_1|$ and of the variance of the measurement error. For moderate sample sizes, the bias of the regression calibration estimator may be large if $|\beta_1|$ is large. Further, for any $|\beta_1|$, when the sample size n is sufficiently large, the standard error of the RC estimator will be dominated by the bias. However, when the relative risk is moderate, bias is often not a problem in realistic sample sizes. This will be further investigated in the simulation study which we present next.

4.3 *Simulation Study*

We numerically evaluated the small sample properties of the proposed estimators based on the EEE, the pseudo-EEE and the RC methods through a simulation study. For each subject, the true covariate was generated from a distribution which will be described in each table and the corresponding outcome variable was generated from a logistic model. There were three replicates of W_{ij} and we set $t_{i(j+1)} - t_{ij} = 1, j = 1, 2$. The error process $\tilde{U}_i$ was normal and followed the AR(1) model described in Section 4.1. In addition to the EEE estimators and the RC estimator, we also calculated the naive estimator which simply substituted $\overline{W}_{i\cdot} = \sum_{j=1}^{3} W_{ij}/3$ for X_i in an ordinary logistic regression analysis. Note that estimates of η from RC and pseudo–EEE are the same because they solve the same estimating equation (9).

Tables 1–3 display the results. The true parameters and sample sizes n used in each simulation study are shown in the corresponding tables. In the tables, "bias" was calculated by taking the average of $\widehat{\Theta} - \Theta$ from 500 replicates, "s.d." denotes the sample standard deviation of the estimates, "mean(s.e.)" denotes the average of the estimated standard errors of the estimates. The 95% confidence interval coverage probabilities are included also.

In Table 1, we consider a normally distributed X. Results are displayed for the case when the covariate effect is moderate with $(\beta_0, \beta_1) = (-\ln(2), \ln(2))$, measurement error is large with $\sigma_u/\sigma_x = 1$ and the autocorrelation coefficient is 0.2. Bias in the naive estimator is clearly evident and the three adjusted estimators are all superior. The pseudo–EEE estimates are about as efficient as the EEE estimates when $n = 500$, but are slightly less efficient when $n = 150$. Note that in these simulations the EEE estimator is the ML estimator, as we used $S_\beta = \phi_\beta$ and $S_\eta = \phi_\eta$, and hence the EEE is the most efficient estimator in the class of consistent and asymptotically normal estimators. Interestingly, the RC estimator appears to perform the best in this situation. This phenomenon is similar to some findings for Cox regression (Wang *et al.*, 1997). However, under this setting, when $n = 5,000$ and $\sigma_u/\sigma_x = 2$, the RC estimate for β_1 has bias -0.035 with standard deviation 0.098, and the 95% confidence interval has a coverage probability of 87%. Thus the RC estimator can have poor properties in large samples.

In Table 2, we consider a setting similar to Table 1 but with $(\beta_0, \beta_1) =$ (-ln(5),ln(5)) and with 4 different sample sizes. Inference based on the naive estimator is not valid. When $n = 150$, biases from the EEE and pseudo–EEE methods are relatively large, while those from the RC estimator are smaller. However, EEE is still acceptable in terms of coverage probabilities, *i.e.* confidence interval estimation. It is seen that when n is increased to 300, the bias of EEE or pseudo-EEE is much reduced, while that of the RC estimator is not. When $n = 500$ and 800, EEE and pseudo–EEE are clearly superior and RC has unacceptable biases and coverage probabilities associated with it.

Observe that the EEE estimator requires modelling assumptions about the distribution of X and $\tilde{W}|X$ for estimating Θ. Although the distributions of X and $\tilde{W}|X$ are not required for the pseudo–EEE in estimating η, they are involved in the calculation of the conditional expectations of (5.1) and hence in solving the estimating equation for β. We performed a simulation study in order to determine the sensitivity of the estimation to misspecification of the model for X. One may feel uneasy about making assumptions about X as it is not observable. Table 3 demonstrates sensitivity analyses for incorrectly modeling the distribution of X. We consider the case when the actual X was generated from a uniform $[-\sqrt{3}, \sqrt{3}]$, and a Chi-squared distribution, respectively, but the calculations of EEE used the assumption that X is normal. It is seen that when $|\beta_1|$ is moderate, model misspecification does not cause a serious bias problem. However, the biases may be unacceptable when $|\beta_1|$ is large. While not reported in Table 3, EEE estimation of η is satisfactory when X was generated from a uniform distribution. However, it is not the case when X was generated from a Chi-squared distribution, especially for the estimation of σ_x.

The results of Table 3 demonstrate an important issue in measurement error problems, namely the sensitivity of statistical methods to misspecification of the covariate distribution. To address this problem, Carroll *et al.* (1999) proposed a parametric approach using mixtures of normal distributions as a flexible model for covariates. In general, data analysts should be cautious about this modeling issue in dealing with nonlinear regression with measurement error.

These simulation results suggest that for moderate sample sizes, if $|\beta_1|$ is not too large then the RC estimator has very limited bias and is quite efficient. However, if $|\beta_1|$ is large, sample sizes are reasonably large and if the distributions of X and U can be modeled adequately, then the EEE estimator is better. We also note from Table 2 that in small samples a consistent estimator, such as EEE, may have both large bias and standard error.

5 DATA ANALYSIS

5.1 *Logistic regression analysis*

We illustrate the methods using data from 485 subjects included in a retrospective longitudinal study of childhood growth. Details of the study can be found in Whitaker *et al.* (1997). In brief, the study included all subjects born at Group Health Cooperative, a health maintenance organization in Seattle, during 1/1/65 – 1/1/71, and who had at least one outpatient visit on or after their 21st birthday. Subjects were categorized as obese or non-obese as adults using their average body mass index (BMI) between 21 and 29 years of age. The objective of the analyses displayed here is to determine the extent to which average BMI z–score in early childhood, between ages 2 and 7 years, is predictive of adult obesity. As described by Whitaker *et al.* (1997), the z–score standardizes BMI for age and gender. Thirteen percent of the adults in this study for whom data have been abstracted have been classified as obese as adults. Measurement error analysis in this problem is needed because it is impossible to measure an infinite number of observations in order to obtain the true long–term average BMI z–score between ages 2 and 7 years. We have a finite number of observations and hence only an estimate of the true long–term average.

Instead of observing the child's long–term average BMI z-score between 2 and 7 years of age (denoted by X_i), BMI measurements denoted by W_{ij} are available at times $\{t_{ij}, j = 1, \cdots, k_i\}$ in the interval 2 to 7 years of age. We assume that the data follow the structure defined in Section 5. Thus $P(Y|X)$ follows a logistic regression model which is linear in X. We assume that $\tilde{W}_i$ given X_i is multivariate normal with mean $X_i 1_{k_i}$ and variance Σ_{U_i}, where $1_{k_i} = (1, \ldots, 1)^t$, the (j, j')th element of Σ_{U_i} is equal to $\sigma_u^2 \rho^{|t_{ij} - t_{ij'}|}$ for $j, j' \in \{1, \ldots, k_i\}$ and ρ is the autocorrelation parameter. The number and timing of BMI measurements between ages 2 and 7 years varied amongst subjects and hence the surrogate data $\tilde{W}_i$ are unbalanced.

Each of the methods described previously in Section 4 were applied to the data. Results are displayed in Table 4. As expected, the naive estimator of β_1 is smaller in magnitude than any of the RC, EEE or pseudo–EEE estimators. Although estimates of β_1 from RC, EEE and pseudo–EEE are slightly different, the result indicates an odds ratio (e^{β_1}) of approximately 3.5 to 4.0. We note also that the EEE estimates of (σ_x, σ_u) are close to those from the pseudo–EEE (same as RC) and that the ratio σ_u/σ_x is about 0.5. The de–attenuation ratios from the three adjusted estimates are between 1.12 and 1.18, which can be explained by noting the estimates of σ_x, σ_u and ρ and in particular the fairly small ratio σ_u/σ_x. Perhaps in this example, one should be conservative about the use of RC as it tends to underestimate the relative risk when the effect is a strong one.

5.2 *Clustered data analysis*

Both the simulation studies and the data analyses presented above involved complete data estimating equations which were likelihood score equations, so that EEE was simply an implementation of maximum likelihood. We now consider a setting where the EEE method is not maximum likelihood. Let X_{i1} denote the young child's average BMI z-score between 2 and 7 years of age for subject i. Y_{i1} is an indicator of obesity in the adolescent, between 12 and 17 years of age. Similarly, X_{i2} is the average BMI z–score for the pre-adolescent child, 7 to 12 years old, and Y_{i2} is an indicator of obesity in the young adult between 20 and 29 years. The analysis addresses the questions of how BMI in the young child predicts obesity in the adolescent and how BMI in the pre-adolescent predicts obesity in the adult. Other combinations of predictor and outcome age intervals could also be used but for simplicity we consider only two combinations here. Consider marginal logistic models for Y_{ir} given X_{ir} with parameters β_{r0} and β_{r1}, $r = 1, 2$. Instead of observing $\{X_{ir}, i = 1, \ldots, n, r = 1, 2\}$, we have $\{\tilde{W}_{ir}, i = 1, \cdots, n, r = 1, 2\}$ available, where $\tilde{W}_{ir} = (W_{ir}, \ldots, W_{ir_{ir}})^t$ are repeated BMI measurements of the i–th subject in the r–th age interval. As in the previous illustration, we assume that $\tilde{W}_{ir}$ given X_{ir} is multivariate normal with mean $X_{ir}1_{k_{ir}}$ and variance $\Sigma_{U_{ir}}$. The parameters $(\sigma_u, \rho, \mu_x, \sigma_x)$ are assumed to be common across the two age intervals.

The EEEs described in Section 3.2 were used with $S_{\beta_r}(Y_{ir}, X_{ir}) = (1, X_{ir})\{Y_{ir} - H(\beta_{r0} + \beta_{r1}X_{ir})\}$, $r = 1, 2$, and the S_η being similar to that given in Section 4.1. The pseudo–EEEs for η are also similar to those given in Section 4.1, but we combined age intervals $\tilde{W}_{ir}, r = 1, 2$ and hence all the summations of the form $\sum_{i=1}^{n}$ in (9) become $\sum_{i=1}^{n}\sum_{r=1}^{2}$. The RC estimator of η is the same as that from the pseudo–EEE, and the estimator of β is similar to that in Section 4.2 but with the replacement covariates $\widehat{E}(X_{ir}|\overline{W}_{ir}) = \widehat{\mu}_X + (\widehat{\sigma}_x^2/\widehat{\sigma}^2_{\overline{W}_{ir\cdot}})(\overline{W}_{ir\cdot} - \widehat{\mu}_x), r = 1, 2$. The naive, RC, EEE and pseudo–EEE estimates from this analysis are given in Table 5. As expected, childhood BMI during the younger age interval (2 to 7 years) is predictive of obesity at later ages (12 to 17 years) as is evidenced by an odds ratio $e^{\beta_{11}}$ of about 5.0. Similarly, childhood BMI–z in the age interval 7 to 12 years is highly related to obesity in the young adult ($e^{\beta_{21}} \approx 11$ for EEE or pseudo–EEE; ≈ 9.5 from RC). The parameter estimates for $\sigma_u, \rho, \mu_x, \sigma_x$ are similar to those obtained from the logistic regression analysis of Table 4. Both slope parameters from the naive analyses are attenuated relative to the RC, EEE and pseudo–EEE analyses which adjust for measurement error. Again, there is little difference between the estimates or standard errors for the parameters of interest $(\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$ derived using the EEE or pseudo–EEE methods. That the RC estimate of β_{21} is somewhat smaller than that from EEE is probably due to a very large β_{21} which is the situation where the calibration approximation tends to underestimate the relative risk.

6 DISCUSSION

In this paper we have proposed a general approach to estimation when true covariate data are missing but surrogate or mismeasured data are available instead. The method requires that models be postulated for the covariate and measurement error distributions. Parameter estimates are derived from estimating equations which are expectations of complete data estimating equations conditional on observed data. We note that conditioning need not be on the entire set of observed data but possibly on subsets of the available data. When the complete data scores are score functions derived from the likelihood and conditional expectations are taken with respect to all available data then the method corresponds to maximum likelihood estimation.

The error process considered in this paper is general. In the example considered in detail the errors were correlated and covariates were observed at unequally–spaced times. Such an error structure is likely to be common for longitudinal data. In addition to EEE and pseudo–EEE, RC was studied in this paper. Overall, to apply any of these methods, one should be careful about misspecifying the model for X, especially when the relative risk parameter is large.

Acknowledgments

The research was supported by grants awarded by the National Institutes of Health (AG 15026, CA 53996, HL 57288), USA. We would like to thank Dr. Bob Whitaker for providing us with data

used in the illustrations, and an associate editor and two referees for helpful comments.

REFERENCES

Armstrong, B. K., White, E. and Saracci R. (1992). *Principles of Exposure Measurement in Epidemiology. Monographs in Epidemiology and Biostatistics.* Vol **21**, Oxford: Oxford University Press.

Carroll, R. J., Roeder, L. and Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics*, **55**, 44-54.

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

Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. T. and Abbott, R. D. (1984). On errors–in–variables for binary regression models. *Biometrika*, **71**, 19–26.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, **85**, 652-663.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **39**, 1–38.

Fuller, W. A. (1987). *Measurement Error Models*, John Wiley & Sons, New York.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, **9**, 861–869.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated–measures models with structured covariance matrices. *Biometrics*, **42**, 805–820.

McFadden, D. and Ruud, P. A. (1994). Estimation by simulations. *The review of Economics and Statistics*, **76**, 591-608.

Pepe, M.S. (1992). Inference using surrogate outcome data and a validation sample. *Biometrika*, **79**, 770–778.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics*, **23**, 939–951.

Pepe M. S. and Couper, D. (1997). Modelling partly conditional means with longitudinal data. *Journal of the American Statistical Association*, **92**, 991–998..

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69**, 331-342.

Prentice, R. L. (1996). Dietary fat and breast cancer: measurement error and results from analytic epidemiology. *Journal of the National Cancer Institute*, **88**, 1738–1747.

Reilly, M. and Pepe, M. S. (1995). A mean-score method for missing and auxiliary covariate data in regression models. *Biometrika*, **82**, 299-314.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866.

Rosner, B., Willett, W. C. and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, **8**, 1051-1070.

Rotnitzky, A., Robins, J. M. and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, **93**, 1321-1339.

Wang, C. Y., Hsu, L., Feng, Z. D. and R. L. Prentice (1997). Regression calibration in failure time regression. *Biometrics*, **53**, 131–145.

Wang, N., Carroll, R. J. and Liang, K. Y. (1996). Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics*, **52**, 401–411.

Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, K. D. and Dietz, W. H. (1997). Predicting obesity in young adulthood from childhood and parental obesity. *New England Journal of Medicine*, **337**, 869–873.

Table 1. Simulation study with moderate covariate effect

		n=150				n=500			
		Naive	RC	EEE	Pseudo–EEE	Naive	RC	EEE	Pseudo–EEE
β_0	bias	0.197	-0.067	-0.127	-0.122	0.231	0.017	-0.012	-0.010
	s.d.	0.238	0.353	0.417	0.437	0.123	0.163	0.178	0.178
	mean(s.e.)	0.231	0.366	0.427	0.453	0.125	0.167	0.178	0.183
	95% cov.	0.844	0.952	0.958	0.958	0.542	0.958	0.952	0.962
β_1	bias	-0.195	0.069	0.129	0.124	-0.225	-0.011	0.018	0.016
	s.d.	0.163	0.310	0.381	0.401	0.081	0.135	0.152	0.153
	mean(s.e.)	0.155	0.315	0.378	0.405	0.083	0.138	0.150	0.155
	95% cov.	0.680	0.960	0.970	0.970	0.234	0.962	0.962	0.970
σ_u	bias	—	0.008	0.014	0.008	—	0.002	0.004	0.002
	s.d.	—	0.081	0.079	0.081	—	0.045	0.043	0.045
	mean(s.e.)	—	0.086	0.082	0.086	—	0.045	0.041	0.045
	95% cov.	—	0.926	0.940	0.926	—	0.954	0.938	0.954
ρ	bias	—	0.005	0.011	0.005	—	0.002	0.004	0.002
	s.d.	—	0.128	0.126	0.128	—	0.075	0.070	0.075
	mean(s.e.)	—	0.137	0.122	0.137	—	0.076	0.064	0.076
	95% cov.	—	0.950	0.930	0.950	—	0.952	0.916	0.952
μ_x	bias	—	0.001	0.001	0.001	—	-0.001	-0.001	-0.001
	s.d.	—	0.099	0.099	0.099	—	0.057	0.057	0.057
	mean(s.e.)	—	0.098	0.098	0.098	—	0.053	0.053	0.053
	95% cov.	—	0.940	0.940	0.940	—	0.934	0.932	0.934
σ_x	bias	—	-0.018	-0.024	-0.018	—	-0.007	-0.009	-0.007
	s.d.	—	0.118	0.117	0.118	—	0.060	0.059	0.060
	mean(s.e.)	—	0.122	0.118	0.122	—	0.061	0.055	0.061
	95% cov.	—	0.960	0.960	0.960	—	0.966	0.946	0.966

NOTE: Parameters are $(\beta_0, \beta_1, \sigma_u, \rho, \mu_x, \sigma_x) = (-\log(2), \log(2), 1, 0.2, 1, 1)$ and $P(Y = 1) = 0.5$. The estimating scores were obtained by using a usual Riemann integral from 50 partitions. The result is from 500 replicates.

Table 2. Simulation study with large relative risk

		Naive	RC	EEE	Pseudo–EEE	Naive	RC	EEE	Pseudo–EEE
		n=150				n=300			
β_0	bias	0.600	0.091	-0.345	-0.345	0.622	0.162	-0.101	-0.128
	s.d.	0.270	0.446	1.052	1.030	0.193	0.310	0.506	0.583
	mean(s.e.)	0.272	0.493	1.303	1.312	0.191	0.306	0.486	0.560
	95% cov.	0.372	0.920	0.966	0.968	0.132	0.850	0.942	0.946
β_1	bias	-0.604	-0.093	0.345	0.341	-0.619	-0.158	0.104	0.132
	s.d.	0.203	0.396	1.046	1.080	0.145	0.277	0.476	0.551
	mean(s.e.)	0.198	0.448	1.283	1.275	0.139	0.274	0.456	0.534
	95% cov.	0.182	0.890	0.964	0.960	0.028	0.820	0.944	0.944
		n=500				n=800			
β_0	bias	0.626	0.173	-0.090	-0.083	0.608	0.163	-0.077	-0.081
	s.d.	0.145	0.224	0.401	0.402	0.124	0.188	0.288	0.296
	mean(s.e.)	0.147	0.232	0.369	0.388	0.117	0.179	0.267	0.281
	95% cov.	0.018	0.836	0.972	0.968	0.000	0.806	0.964	0.962
β_1	bias	-0.623	-0.171	0.092	0.086	-0.620	-0.175	0.063	0.067
	s.d.	0.104	0.204	0.383	0.386	0.088	0.163	0.263	0.274
	mean(s.e.)	0.107	0.206	0.346	0.368	0.084	0.158	0.247	0.264
	95% cov.	0.000	0.814	0.956	0.944	0.000	0.730	0.954	0.956

NOTE: Parameters are $(\beta_0, \beta_1, \sigma_u, \rho, \mu_x, \sigma_x) = (-\ln(5), \ln(5), 1, 0.2, 1, 1)$ and $P(Y = 1) = 0.5$. The estimating scores were obtained by using a usual Riemann integral from 50 partitions. The result is from 500 replicates.

Table 3. Sensitivity Analyses of misspecification

		n=150				n=500			
		Naive	RC	EEE	Pseudo–EEE	Naive	RC	EEE	Pseudo–EEE
— Uniform X —									
β_0 = -ln(2)	bias	0.207	-0.027	-0.061	-0.071	0.211	-0.014	-0.041	-0.044
	s.d.	0.253	0.341	0.368	0.402	0.128	0.178	0.192	0.200
	mean(s.e.)	0.230	0.330	0.357	0.385	0.125	0.170	0.183	0.189
	95% cov.	0.816	0.952	0.954	0.962	0.598	0.944	0.948	0.952
β_1 = ln(2)	bias	-0.197	0.038	0.071	0.081	-0.208	0.016	0.044	0.046
	s.d.	0.163	0.269	0.297	0.328	0.086	0.145	0.159	0.167
	mean(s.e.)	0.154	0.278	0.304	0.334	0.083	0.141	0.154	0.161
	95% cov.	0.692	0.962	0.966	0.964	0.290	0.952	0.952	0.962
β_0 = -ln(5)	bias	0.509	-0.005	-0.416	-0.447	0.522	0.036	-0.275	-0.276
	s.d.	0.274	0.403	0.782	0.931	0.155	0.228	0.379	0.385
	mean(s.e.)	0.277	0.487	0.990	1.150	0.150	0.241	0.393	0.418
	95% cov.	0.528	0.956	0.976	0.972	0.096	0.934	0.982	0.990
β_1 = ln(5)	bias	-0.512	0.007	0.420	0.448	-0.525	0.040	0.270	0.271
	s.d.	0.196	0.356	0.760	0.884	0.117	0.201	0.356	0.362
	mean(s.e.)	0.200	0.443	0.956	1.182	0.108	0.216	0.370	0.397
	95% cov.	0.300	0.960	0.982	0.984	0.010	0.932	0.986	0.988
— Chi-squared X —									
β_0 = -ln(2)	bias	0.244	0.020	-0.051	-0.013	0.254	0.058	0.013	0.040
	s.d.	0.236	0.351	0.425	0.417	0.114	0.163	0.183	0.178
	mean(s.e.)	0.228	0.360	0.440	0.425	0.124	0.169	0.190	0.184
	95% cov.	0.780	0.960	0.970	0.970	0.472	0.946	0.954	0.946
β_1 = ln(2)	bias	-0.280	-0.054	0.023	-0.023	-0.283	-0.086	-0.038	-0.069
	s.d.	0.155	0.300	0.384	0.377	0.080	0.136	0.161	0.158
	mean(s.e.)	0.163	0.311	0.406	0.382	0.089	0.140	0.165	0.157
	95% cov.	0.562	0.918	0.950	0.932	0.090	0.870	0.916	0.900
β_0 = -ln(5)	bias	0.675	0.252	-0.042	-0.007	0.680	0.324	0.115	0.183
	s.d.	0.266	0.453	0.770	1.110	0.148	0.226	0.347	0.320
	mean(s.e.)	0.254	0.505	0.884	1.454	0.138	0.217	0.314	0.299
	95% cov.	0.278	0.820	0.912	0.890	0.004	0.614	0.850	0.804
β_1 = ln(5)	bias	-0.839	-0.412	0.122	-0.171	-0.835	-0.477	-0.273	-0.351
	s.d.	0.187	0.414	0.732	1.082	0.103	0.200	0.318	0.293
	mean(s.e.)	0.195	0.462	0.839	1.321	0.106	0.190	0.287	0.270
	95% cov.	0.043	0.617	0.803	0.747	0.000	0.274	0.650	0.566

NOTE: Nuisance parameters are $(\sigma_u, \rho, \mu_x, \sigma_x) = (0.5, 0.2, 1, 1)$. The calculations of EEE and the β part of pseudo–EEE are from normal assumptions of X.

Table 4: Logistic regression analysis of child growth data. Estimated standard errors are shown in parentheses.

	Naive	RC	EEE	pseudo–EEE
β_0	-2.212 (0.168)	-2.214 (0.155)	-2.281 (0.169)	-2.268 (0.166)
β_1	1.151 (0.183)	1.288 (0.216)	1.350 (0.239)	1.316 (0.227)
σ_u	—	0.394 (0.020)	0.411 (0.024)	0.394 (0.020)
ρ	—	0.119 (0.083)	0.203 (0.092)	0.119 (0.083)
μ_x	—	0.021 (0.040)	0.029 (0.037)	0.021 (0.040)
σ_x	—	0.766 (0.033)	0.752 (0.034)	0.766 (0.033)

NOTE: There were 485 subjects used in the analysis, β_0, β_1 are the logistic regression parameters of adulthood obesity on childhood BMI, and ρ is the AR(1) autocorrelation coefficient for the measurement errors.

Table 5: Partly conditional analysis of child growth data.

	Naive	RC	EEE	pseudo–EEE
β_{10}	-2.888 (0.254)	-2.898 (0.237)	-2.974 (0.256)	-2.974 (0.255)
β_{11}	1.386 (0.249)	1.552 (0.253)	1.613 (0.289)	1.595 (0.283)
β_{20}	-2.504 (0.226)	-2.503 (0.207)	-2.659 (0.244)	-2.656 (0.243)
β_{21}	1.974 (0.265)	2.244 (0.340)	2.432 (0.410)	2.403 (0.410)
σ_u	—	0.361 (0.017)	0.363 (0.025)	0.361 (0.017)
ρ	—	0.156 (0.088)	0.198 (0.082)	0.156 (0.088)
μ_x	—	-0.014 (0.040)	-0.029 (0.039)	-0.014 (0.040)
σ_x	—	0.743 (0.033)	0.716 (0.035)	0.743 (0.033)

NOTE: There were 385 subjects used in the analysis. The logistic regression parameters, β_{10} and β_{11}, relate to a model for obesity in the 12 to 17 year old as predicted by BMI in the child at ages 2 to 7 years. Similarly, β_{20} and β_{21} are parameters in the model for adult obesity as predicted by BMI in the 7 to 12 year old. The parameter ρ is the AR(1) autocorrelation coefficient in the measurement error model.