Kastner, Ziegler:

Cross-sectional Analysis of Longitudinal Data with Missing Values in the Dependent Variables: A Comparison of Weighted Estimating Equations with the Complete Case Analysis

Sonderforschungsbereich 386, Paper 64 (1997)

Cross-sectional Analysis of Longitudinal Data with Missing Values in the Dependent Variables: A Comparison of Weighted Estimating Equations with the Complete Case Analysis

C. Kastner * A. Ziegler **

August 27, 1997

Abstract

Inference for cross-sectional models using longitudinal data can be drawn with independence estimating equations (Liang and Zeger, 1986). Many studies suffer from missing data. Robins and coworkers proposed to use weighted estimating equations (WEE) in estimating the mean structure, if missing data are present in dependent variables. In this paper the WEE are compared with complete case analyses for binary responses using simulated data. Our results are in accordance with the theoretical findings of Robins and coworkers. The WEE yield consistent estimates, even if the data are missing at random.

Keywords: Correlated Data Analysis, Generalised Estimating Equations, Horvitz-Thompson Estimation, Marginal Models, Missing Data, Weighted Estimating Equations

1 Introduction

Several approaches for the analysis of cross-sectional models using longitudinal data have been proposed. Application of these marginal models is very popular in the literature. The Generalised Estimating Equations (GEE; Liang and Zeger, 1986) belong to this class of models. The term "generalised" indicates that the association between the responses is modelled in addition to the mean structure which is of primary interest. If the association between the responses is not modelled but taken into account, the corresponding Estimating Equations (EE) are termed Independence Estimating Equations (IEE).

In this paper we restrict our attention to the IEE and do not consider the GEE because Fitzmaurice (1995) and Mancl and Leroux (1996) have shown that the IEE are as efficient as GEE in various situations though the association is

*Institute of Statistics, LMU München, Ludwigstr. 33, 80539 München, Germany, kchris@stat.uni-muenchen.de

**Medical Centre for Methodology and Health Research, Institute of Medical Biometry and Epidemiology, Philipps–University of Marburg, Bunsenstr. 3, 35033 Marburg, Germany, ziegler@mailer.uni-marburg.de

neglected. For example, if only baseline covariates are included so that all explanatory variables are constant within one observational unit, the IEE are as efficient as specific GEE, if the true correlation is equal between any two sample time points.

Furthermore, the GEE underly important implicit assumptions that are required for the validity of this method. First, the association needs to be correctly specified. Otherwise, the association parameters involved might be subject to uncertainty of definition which can result in a breakdown of the asymptotic properties of the estimators (Crowder, 1995). Second, the mean structure needs to be correctly specified as a function of all—probably time point specific—explanatory variables. Otherwise, the resulting estimator might be biased (Pepe and Anderson, 1994). These implicit assumptions do not apply to the IEE.

Standard or advanced statistical methods are designed for complete data sets. However, many studies suffer from missing or incomplete data so that statistical analyses become more complicated. This also applies to models that may be analysed by the IEE. Approaches that ignore systematical differences between complete and incomplete clusters may be biased (Little and Schenker, 1995). Two forms of item non-response are commonly distinguished: non-response in explanatory variables (in the X), and non-response in dependent variables (in the y). In this paper we focus on missing data in dependent variables. Explanatory variables of interest are assumed to be completely observed. For simplicity, we restrict our attention to monotone missing data patterns; that is, once a subject leaves the study, it will never return.

One approach for solving the IEE in presence of missing dependent data received considerable attention (Robins, Rotnitzky and Zhao, 1995; Robins and Rotnitzky, 1995). The basic idea of this approach is to introduce weighted estimating equations (WEE). The weights are inversely proportional to the respective response probabilities. They are estimated by surrogate variables for the response. Thus, the resulting estimators belong to the class of Horvitz-Thompson estimators. The WEE may be applied to data missing at random in Laird's (1988) sense.

The aim of this paper is a comparison of the WEE approach with the "classical" complete case analyses using the IEE for simulated data. Both approaches are implemented in the program MAREG (Kastner, Fieger and Heumann, 1997).

The paper is organised as follows: In the next section we derive the IEE assuming complete observations. In section 3 missing data mechanisms for dependent variables are discussed. In section 4 we consider an approach for estimating the response probabilities. These estimates are used in the WEE of Robins et al. (1995) which are introduced in section 5 assuming a monotone missing data mechanism. The use of the IEE and the WEE in the presence of missing response data is illustrated in section 6. It is especially shown that the WEE yield consistent parameter estimates, if the non-response is ignorable in the sense of Laird (1988).

2 The Independence Estimating Equations

Let y_{it} be a vector of responses from n clusters with T observations for the ith cluster, $i = 1, \dots, n$. For each y_{it} several covariates $\mathbf{x}_{it}$ are available. The data can be summarized to the vector $\mathbf{y}_i$ and the matrix $\mathbf{X}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT})'$. The method can be easily extended to unequal cluster sizes T_i. The pairs $(\mathbf{y}_i, \mathbf{X}_i)$ are assumed to be independently identically distributed. We focus on marginal models so that we do not consider models including state dependence or duration dependence. Furthermore, we consider cross-sectional models with longitudinal data, i.e. we assume $E(y_{it}|\mathbf{x}_{it}) = E(y_{it}|\mathbf{X}_i)$. However, the association within a cluster needs to be taken into account.

For independent observations, the Generalized Linear Model (GLM) allows flexibility in modelling mean and variance structures. In GLM, the mean structure is given by

$$E(y_{it}|\mathbf{x}_{it}) = \mu_{it} = g(\mathbf{x}'_{it}\boldsymbol{\beta}), \tag{1}$$

where g is a non-linear response function and $\boldsymbol{\beta}$ is the $p \times 1$ parameter of interest. g^{-1} is termed link function. In most GLM mean and variance are functionally related: $v_{it} = V(y_{it}|\mathbf{x}_{it}) = h(\mu_{it})$, where h is the variance function. In general, a distributional assumption motivates the link and the variance function of the GLM. If y_{it} is a binary variable, the connection between y_{it} and $\mathbf{x}_{it}$ may be established e.g. via the logit link so that $\mu_{it} = E(y_{it}|\mathbf{x}_{it}) = \text{logit}(\mathbf{x}'_{it}\boldsymbol{\beta})$. The variance function is usually chosen as $v_{it} = \mu_{it}(1 - \mu_{it})$ in this situation. The parameter vector $\boldsymbol{\beta}$ is estimated by solving the IEE

$$\mathbf{u}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{D}'_i \mathbf{V}_i^{-1} \boldsymbol{\epsilon}_i = \frac{1}{n}\mathbf{D}'\mathbf{V}^{-1}\boldsymbol{\epsilon} = \mathbf{0}\,, \tag{2}$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = g(\mathbf{X}_i\boldsymbol{\beta})$ is the vector of the mean structure, $\mathbf{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$ is the diagonal matrix of first derivatives, $\mathbf{V}_i$ is the diagonal matrix of the variances $\mathbf{V}_i = \text{diag}(v_{it})$ and $\boldsymbol{\epsilon}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$. $\mathbf{D}$ and $\boldsymbol{\epsilon}$ are the stacked $\mathbf{D}_i$ matrices and $\boldsymbol{\epsilon}_i$ vectors, respectively. Finally, $\mathbf{V}$ is the (block) diagonal matrix of the $\mathbf{V}_i$.

An analytic solution of (2) exists for the linear model with normal distributed response variables. In general, (2) are solved iteratively by a Fisher-scoring algorithm or iterative weighted least squares (IWLS). If the observations are indeed independent, the estimator $\hat{\boldsymbol{\beta}}_{IEE}$ is consistent and asymptotically normal with variance matrix

$$Cov(\hat{\boldsymbol{\beta}}_{IEE}) = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1} \tag{3}$$

under suitable regularity conditions.

For correlated observations, however, the true variance matrix $Cov(\mathbf{y}_i|\mathbf{X}_i) = \boldsymbol{\Omega}_i \neq \mathbf{V}_i$ is not diagonal. Thus, the use of (3) generally leads to biased estimates. Therefore, Zeger, Liang and Self (1985) proposed to use the sandwich information matrix which yields consistent estimates of $V(\hat{\boldsymbol{\beta}}_{IEE})$, even if the dependent variables within a cluster are correlated

$$\widehat{V(\hat{\boldsymbol{\beta}}_{IEE})} = \left(\sum_{i=1}^{n} \hat{\mathbf{D}}'_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)^{-1} \left(\sum_{i=1}^{n} \hat{\mathbf{D}}'_i \hat{\mathbf{V}}_i^{-1} \hat{\boldsymbol{\Omega}}_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right) \left(\sum_{i=1}^{n} \hat{\mathbf{D}}'_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)^{-1}. \tag{4}$$

$-\sum_{i=1}^{n} \hat{\mathbf{D}}'_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i$ is the estimated Fisher information matrix. The middle matrix on the right hand side of (4) consists of the estimated crossproducts of first order derivatives, also termed estimated outer product gradient (estimated OPG), with $\hat{\mathbf{\Omega}}_i = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)'$. The variance estimator of (4) is also termed robust variance estimator because it is a strongly consistent estimator of $V(\mathbf{y}_i|\mathbf{X}_i)$ regardless of the true variance structure.

3 Missing Data Mechanisms

As noted in section 1, we assume that the explanatory variables of interest $\mathbf{X}_i$ are observed completely; dependent variables y_{it} may be missing. In different applications the missing data mechanism can be explained by additional variables that are not of the investigator's interest for the mean structure. Suppose that surrogate variables $\mathbf{z}_{it}$ are observed in addition to y_{it}, if y_{it} is observed.

Let $\mathbf{w}_{i0} = (\mathrm{vec}(\mathbf{X}_i)', y_{i0}, \mathbf{z}_{i0}')'$ be comprised of covariates $\mathbf{X}_i$ and the observations of y_{i0} and $\mathbf{z}_{i0}$ prior to follow–up. We assume that y_{i0} and $\mathbf{z}_{i0}$ are observed completely. If $\mathbf{X}_i$ are not baseline covariates, they either have to be fixed or their values may not depend on the dependent variables. Furthermore, we set $\mathbf{w}_{it} = (y_{it}, \mathbf{z}_{it}')'$ for $t = 1, \dots, T$. As pointed out by Robins et al. (1995), this notation is redundant, if $\mathbf{z}_{i0}$ is included in the explanatory variables. Bars are used to indicate variables including the whole history except the current observation. For example, $\bar{\mathbf{w}}_{it} = (\mathbf{w}_{i0}', \mathbf{w}_{i1}', \dots, \mathbf{w}_{i(t-1)}')'$.

Let r_{it} denote the missing data indicator, such that $r_{it} = 1$, if the pair $(y_{it}, \mathbf{z}_{it})$ is observed and $r_{it} = 0$, if $(y_{it}, \mathbf{z}_{it})$ is not observed. In the following we assume a monotone missing data pattern so that $r_{i(t+1)} = 0$, if $r_{it} = 0$ for any t. This yields $P(r_{i(t+1)} = 1|r_{it} = 0, \bar{\mathbf{w}}_{i(T+1)}) = 0$. $r_{iT} = 1$ indicates that the data of cluster i are completely observed.

We assume that the data are missing at random (MAR) in the sense of Laird (1988). Then

$$P(r_{it} = 1|r_{i(t-1)} = 1, \bar{\mathbf{w}}_{it}, y_{it}, \dots, y_{iT}) = P(r_{it} = 1|r_{i(t-1)} = 1, \bar{\mathbf{w}}_{it}) \qquad (5)$$

holds for a monotone missing data pattern. Equation (5) implies that the response probability at time t only depends on observations prior to t. Furthermore, $\bar{\mathbf{w}}_{is}, s \geq t$ does not contain any information on the response probability at time t. The idea of the approach proposed by Robins and coworkers is similar to the classical Horvitz-Thompson approach (Horvitz and Thompson, 1952) where the observations are weighted by their inverse inclusion (observation) probability. Therefore, we assume that the response probabilitiy λ_{it} for each study subject remaining in the study is bounded away from 0:

$$\lambda_{it} = P(r_{it} = 1|r_{i(t-1)} = 1, \bar{\mathbf{w}}_{it}) > \delta > 0, \quad t = 1, \dots, T \qquad (6)$$

We do not restrict our attention to data that are missing completely at random (MCAR) in Laird's (1988) sense which is equivalent to

$$P(r_{it} = 1|r_{i(t-1)} = 1, \bar{\mathbf{w}}_{it}, y_{it}, \dots, y_{iT}) = P(r_{it} = 1|r_{i(t-1)} = 1, \mathbf{X}_i) \qquad (7)$$

for monotone missing data patterns.

4 Estimation of Response Probabilities

The response probabilities $\lambda_{it}(\boldsymbol{\gamma}) = \lambda_{it}(\boldsymbol{\gamma}|r_{i(t-1)}, \bar{\mathbf{w}}_{it})$ may depend on an additional parameter $\boldsymbol{\gamma}$ that is modelled as a function of $\bar{\mathbf{w}}_{it}$ and $r_{i(t-1)}$. Note that

$\bar{\mathbf{w}}_{it}$ includes $\mathbf{X}_i$ and $\bar{\mathbf{y}}_{it}$. If the response probabilities $\lambda_{it}(\boldsymbol{\gamma})$ are unknown, $\boldsymbol{\gamma}$ has to be estimated. A natural choice for estimating $\boldsymbol{\gamma}$ is the logit link. For high response probabilities, however, the compound-log-log link should be preferred. If no observation is missing at a specific time point t, λ_{it} need not be estimated. In this situation, $\lambda_{it} = 1$. If at least one obervation is missing and present, respectively, for every time point, an estimate $\hat{\boldsymbol{\gamma}}$ can be obtained by maximizing the normed partial Likelihood function $L(\boldsymbol{\gamma}) = \frac{1}{n}\prod_{i=1}^{n} L_i(\boldsymbol{\gamma})$, where

$$L_i(\boldsymbol{\gamma}) = \prod_{t=1}^{T}\Big(\lambda_{it}(\boldsymbol{\gamma})^{r_{it}}[1-\lambda_{it}(\boldsymbol{\gamma})]^{1-r_{it}}\Big)^{r_{i(t-1)}}. \tag{8}$$

In (8) the contribution of cluster i at time t to the partial Likelihood function is not a constant, if $r_{i(t-1)} = 1$. The maximum is found by solving the EE of a GLM for binary data $\mathbf{u}(\boldsymbol{\gamma}) = \partial L(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}' = \mathbf{0}$. The contribution of cluster i to the score vector for $\boldsymbol{\gamma}$ is given by:

$$\mathbf{u}_i(\boldsymbol{\gamma}) = \frac{\partial \ln L_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \sum_{t=1}^{T}\frac{\partial \lambda_{it}}{\partial \boldsymbol{\gamma}}\Big(\lambda_{it}(1-\lambda_{it})\Big)^{-1}(r_{it} - \lambda_{it}(\boldsymbol{\gamma})r_{i(t-1)}) \tag{9}$$

(9) can be simplified, if the natural link $\lambda_{it}(\boldsymbol{\gamma}) = \text{logit}(\boldsymbol{\gamma}'\mathbf{k}(\bar{\mathbf{w}}_{it}))$ is used for some known vector function $\mathbf{k}$:

$$\mathbf{u}_i(\boldsymbol{\gamma}) = \sum_{t=1}^{T}[r_{it} - \lambda_{it}(\boldsymbol{\gamma})r_{i(t-1)}]\mathbf{k}(\bar{\mathbf{w}}_{it})$$

If the functional relation between λ_{it} and $\boldsymbol{\gamma}$ is correctly specified, the asymptotic variance matrix of $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ is given by the inverse of $V_{\boldsymbol{\gamma}} = V(\mathbf{u}_i(\boldsymbol{\gamma})) = E(\mathbf{u}_i(\boldsymbol{\gamma})\mathbf{u}_i(\boldsymbol{\gamma})')$.

Define $\bar{\pi}_{it} = \bar{\pi}_{it}(\boldsymbol{\gamma}) = \lambda_{i1}(\boldsymbol{\gamma}) \cdot \ldots \cdot \lambda_{it}(\boldsymbol{\gamma})$ which may be interpreted as the conditional probability of observing cluster i at time t given the entirely observed history $\bar{\mathbf{w}}_{it}$, if the data are MAR. These conditional probabilities multiplied with their observational status are collected in a diagonal matrix $\boldsymbol{\Pi}_i = \boldsymbol{\Pi}_i(\boldsymbol{\gamma})$ with elements $r_{it}/\bar{\pi}_{it}$.

5 Weighted Estimating Equations

Equations (5) and (6) are the fundamental assumptions required to identify the marginal mean $E(y_{it}|\mathbf{X}_i)$. They imply

$$E(y_{it}r_{it}/\bar{\pi}_{it}|\mathbf{X}_i) = E(y_{it}|\mathbf{X}_i). \tag{10}$$

(10) is the key result in order to formulate the WEE. The validity of (10) is easily shown, if the data are MCAR. In this situation, r_{it} and y_{it} are independent so that the left hand side of (10) can be factorized to

$$E(y_{it}r_{it}/\bar{\pi}_{it}|\mathbf{X}_i) = E(y_{it}|\mathbf{X}_i)E(r_{it}|\mathbf{X}_i)/\bar{\pi}_{it}. \tag{11}$$

Finally, $E(r_{it}|\mathbf{X}_i) = \bar{\pi}_{it}$ because the missing data pattern is monotone. The more complex case of data that are MAR is discussed in detail by Robins et al. (1995). The WEE are similar to the IEE (2). The WEE rest upon the left hand side of (10), while the IEE are based on the right hand side of (10). The residuals $\boldsymbol{\epsilon}_i$ are weighted by $\boldsymbol{\Pi}_i(\boldsymbol{\gamma})$, the inverse observation probability multiplied by the actual observation status. This weighting implies that unobserved dependent data have no influence in the WEE. Using the notation of section 2, an estimator $\hat{\boldsymbol{\beta}}_{WEE}$ is the solution of the WEE

$$\mathbf{u}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_i(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \boldsymbol{\Pi}_i(\hat{\boldsymbol{\gamma}}) \boldsymbol{\epsilon}_i = \mathbf{0}, \tag{12}$$

Robins et al. (1995, Appendix A) have shown that $\hat{\boldsymbol{\beta}}_{WEE}$ is asymptotically normal under suitable regularity conditions. The mean of $\hat{\boldsymbol{\beta}}_{WEE}$ is $\boldsymbol{\beta}$. Its variance matrix can be consistently estimated by the positive definite variance matrix

$$\widehat{V(\hat{\boldsymbol{\beta}}_{WEE})} = \mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})^{-1} \mathbf{C}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) \mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})^{-1}, \tag{13}$$

where

$$\mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \boldsymbol{\Pi}_i(\hat{\boldsymbol{\gamma}}) \hat{\mathbf{D}}_i$$

and

$$\mathbf{C}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \widehat{\text{resid}}_i \, \widehat{\text{resid}}_i'.$$

$-\mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ is the estimated Fisher information matrix. It is symmetric because $\hat{\mathbf{V}}_i$ and $\boldsymbol{\Pi}_i(\hat{\boldsymbol{\gamma}})$ are (block) diagonal matrices. $\mathbf{C}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ is similar to the estimated OPG. However, it is more complex than the estimated OPG of the IEE (4). The EE (12) depend on an additionally estimated nuisance parameter $\hat{\boldsymbol{\gamma}}$. $\widehat{\text{resid}}_i = \widehat{\text{resid}}[\mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}), \mathbf{u}_i(\hat{\boldsymbol{\gamma}})]$ is the estimated residual of the linear regression of $\mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ on $\mathbf{u}_i(\hat{\boldsymbol{\gamma}})$:

$$\widehat{\text{resid}}_i = \mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) - \Big(\sum_{i=1}^{n} \mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) \mathbf{u}_i(\hat{\boldsymbol{\gamma}})'\Big) \Big(\sum_{i=1}^{n} \mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) \mathbf{u}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})'\Big)^{-1} \mathbf{u}_i(\hat{\boldsymbol{\gamma}}). \tag{14}$$

(14) can be obtained by application of the chain rule together with standard Taylor series expansion yields (Robins et al., 1995).

Robins and Rotnitzky (1995) have shown that prior knowledge concerning the response probabilites does not provide additional information, if (i) the mean structure is correctly specified, (ii) the data are MAR and (iii) the response probabilities are > 0. (iii) ensures that λ_{it} can be estimated from the actual data set. This results also implies that the stronger condition for the non-response process—that is that the data are MCAR—does not provide additional information.

6 Illustration

In a simulation study using binary dependent variables we compare the properties of the WEE estimator with the IEE estimator that uses the information of all observations (available case; AV) and the IEE estimator that uses the information of all clusters without missing data (complete cluster; CC). The IEE estimator of the simulated data before generation of missings (full data estimator; FD) is used as reference.

The simulation proceeds as follows. First, the complete data set without missing observations is generated. Second, observations are deleted from the complete simulated data set using pre-specified missing data mechanisms. Several approaches for simulating correlated binary data have been proposed (Park, Park and Shin, 1996; Gange, Linton, Scott, DeMets and Klein, 1995; Lee, Scott and Soo, 1993; Emrich and Piedmonte, 1991). We do not fix the marginal moments in our approach but base our simulations on the theoretical work of Fitzmaurice and Laird (1993) which requires specification of conditional log-odds ratios. Given the conditional log-odds ratios, the joint multinomial distribution of the response patterns given the explanatory variables can be computed. The joint distribution is used to determine the marginal moments. Random numbers are generated using the DRAND48 generator, which is supplied by SunOS 5.5 as a C-library function (SunOS, 1995, man Pages(3C)). The seed is set using the system time for each experiment. The WEE and IEE are solved by MAREG (Kastner et al., 1997).

We simulate 100 clusters with $t = 0, \dots, 3$ each. $t = 0$ is used as baseline and is assumed to be always observed. We simulated two different marginal models. In both models, the response depends on an intercept and a dummy-coded treatment variable. However, in the first model the intercept and the treatment effect are time-constant, while in the second model the treatment effect is time-varying. Furthermore, we used two different models for the dropout-process. In the first model the response probability at time t depends on a constant so that the data are MCAR. In the second model the response probability at time t depends on a constant and the reponse at time $t-1$. This implies that the data are MAR. In both cases the parameters were chosen so that the dropout rate was about 10%, 20% and 30% at $t = 1, 2$ and 3, respectively. Missing data were generated using the logistic function. Finally, two different degrees of association are used. The first yields a small correlation of about 0.1, while the second results in a high correlation of about 0.5.

Table 1 shows mean and standard deviations (in brackets) of the estimated parameters of the model with time-constant treatment effect. The theoretical parameters are $\beta_{\text{intercpt}} = -0.5$ and $\beta_{\text{treat}} = 1$. The drop-out process was modelled assuming $\gamma_{\text{intercpt}} = 2$, if the data are MCAR. For data that are MAR we used $\gamma_{\text{intercpt}} = 1$ and $\gamma_{y_{(t-1)}} = 3$ as parameters.

Table 2 shows the results in the model with time-varying treatment effects. Here we assumed $\beta_{\text{intercpt}} = -0.5$, and $\beta_{\text{treat}} = 0, 0.5, 1$ and 1 at $t = 0, 1, 2$ and 3, respectively. The same parameters as in the cluster-constant case were chosen for modelling the drop-out process.

Our results are similar to those obtained by Robins et al. (1995) and Robins and Rotnitzky (1995) for continuous dependent variables. The bias is not so pronounced for our simulations that are based on binary response variables compared with the continuous case used for illustration by Robins et al. (1995). This findings can be explained by the naturally lower variation of binary variables compared with continuous data.

As long as the data are MCAR, all approaches yield consistent parameter estimates. However, if the data are MAR, the IEE estimator using the AV or the CC approach yield biased parameter estimates, while the WEE remains consistent. The bias increases with the correlation of the responses. As expected, the FD estimator has the lowest standard errors across all models, while the CC estimator has the largest standard errors. Interestingly, the AV estimator and the CC estimator can lead to quite contrary results. For example, in the model with time-dependent treatment effect and high correlation, the mean parameter estimates of β_{treat_3} are 1.258 and 0.895 for the AV and the CC approach, respectively. Here, the AV approach leads to an overestimation, while the CC approach leads to underestimation of the true treatment effect.

Dropout	ρ	Parameter	FD	WEE	AV	CC
MCAR	0.1	intercpt	$-.509$	$-.509$	$-.509$	$-.506$
			(.162)	(.174)	(.172)	(.193)
		treat	1.017	1.017	1.016	1.011
			(.229)	(.251)	(.250)	(.279)
MCAR	0.5	intercpt	$-.497$	$-.495$	$-.496$	$-.500$
			(.237)	(.252)	(.250)	(.289)
		treat	.999	.998	.997	1.006
			(.329)	(.351)	(.347)	(.403)
MAR	0.1	intercpt	$-.511$	$-.509$	$-.475$	$-.217$
			(.170)	(.199)	(.186)	(.218)
		treat	1.010	1.007	.987	.965
			(.230)	(.262)	(.247)	(.287)
MAR	0.5	intercpt	$-.500$	$-.502$	$-.313$	.059
			(.237)	(.263)	(.250)	(.291)
		treat	1.013	1.014	.975	.936
			(.321)	(.370)	(.337)	(.401)

Table 1: Marginal model with cluster-constant covariates. FD: full data, WEE: weighted estimating equations, AV: available case, CC: complete clusters

Dropout	ρ	Parameter	FD	WEE	AV	CC
MCAR	0.1	intercpt	−.498	−.496	−.494	−.494
			(.165)	(.181)	(.181)	(.205)
		treat_0	−.012	−.017	−.018	−.021
			(.336)	(.345)	(.345)	(.427)
		treat_1	.499	.498	.497	.496
			(.349)	(.374)	(.375)	(.421)
		treat_2	1.005	1.007	1.006	1.009
			(.348)	(.393)	(.393)	(.429)
		treat_3	.998	1.004	1.003	1.002
			(.332)	(.398)	(.398)	(.409)
MCAR	0.5	intercpt	−.499	−.498	−.498	−.496
			(.232)	(.251)	(.249)	(.284)
		treat_0	−.010	−.012	−.011	−.022
			(.385)	(.396)	(.394)	(.467)
		treat_1	.503	.498	.498	.501
			(.376)	(.401)	(.399)	(.458)
		treat_2	1.013	1.024	1.025	1.018
			(.379)	(.433)	(.432)	(.469)
		treat_3	1.001	1.009	1.009	1.007
			(.377)	(.452)	(.451)	(.470)
MAR	0.1	intercpt	−.508	−.505	−.473	−.213
			(.174)	(.206)	(.192)	(.222)
		treat_0	−.007	−.009	−.042	.021
			(.340)	(.362)	(.353)	(.433)
		treat_1	.511	.496	.479	.544
			(.332)	(.378)	(.370)	(.448)
		treat_2	1.007	1.004	1.009	1.065
			(.343)	(.428)	(.417)	(.476)
		treat_3	1.028	1.042	1.075	.816
			(.351)	(.458)	(.440)	(.457)
MAR	0.5	intercpt	−.499	−.496	−.311	.052
			(.237)	(.264)	(.253)	(.298)
		treat_0	−.011	−.013	−.198	−.067
			(.388)	(.402)	(.397)	(.487)
		treat_1	.493	.496	.401	.468
			(.376)	(.422)	(.413)	(.486)
		treat_2	1.011	1.023	1.097	1.037
			(.397)	(.476)	(.465)	(.534)
		treat_3	1.014	1.040	1.258	.895
			(.389)	(.509)	(.486)	(.513)

Table 2: Marginal model with time-varying covariates. FD: full data, WEE: weighted estimating equations, AV: available case, CC: complete clusters

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft.

References

Crowder, M. J. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measurements, *Biometrika* **82**: 407–410.

Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates, *The American Statistician* **45**: 302–304.

Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data., *Biometrics* **51**: 309–317.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika* **80**: 141–151.

Gange, S. J., Linton, K. L. P., Scott, A. J., DeMets, D. L. and Klein, R. (1995). A comparison of methods for correlated ordinal measures with ophtalmic applications, *Statistics in Medicine* **14**: 1961–1974.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from finite universe, *Journal of the American Statistical Association* **47**: 663–685.

Kastner, C., Fieger, A. and Heumann, C. (1997). MAREG and WinMAREG – a tool for marginal regression models, *Computational Statistics and Data Analysis* **24**: 235–241.

Laird, N. M. (1988). Missing data in longitudinal studies, *Statistics in Medicine* **7**: 305–315.

Lee, A. J., Scott, A. J. and Soo, S. (1993). Comparing Liang-Zeger estimates with maximum likelihood in bivariate logistic regression, *Statistical Computation and Simulation* **44**: 133–148.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Little, R. J. A. and Schenker, N. (1995). Missing data, *in* G. Arminger, C. C. Clogg and M. E. Sobel (eds), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Plenum, New York, pp. 39–75.

Mancl, L. A. and Leroux, B. G. (1996). Efficiency of regression estimates for clustered data, *Biometrics* **52**: 500–511.

Park, C. G., Park, T. and Shin, D. W. (1996). A simple method for generating correlated binary variates, *Journal of the American Statistical Association* **50**: 306–310.

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, Part B–Simulation and Computation* **23**: 939–951.

Robins, J. M. and Rotnitzky, A. G. (1995). Semiparametric efficiency in multivariate regression models with missing data, *Journal of the American Statistical Association* **90**: 122–129.

Robins, J. M., Rotnitzky, A. G. and Zhao, L. P. (1995). Analysis of semiparametric regression models repeated outcomes in the presence of missing data, *Journal of the American Statistical Association* **90**: 106–120.

SunOS (1995). *SunOS Reference Manual*, Mountain View.

Zeger, S. L., Liang, K.-Y. and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates, *Biometrika* **72**: 31–38.