Numerical approximations in weighted estimating equations for regression models with missing continuous covariates

By

MICHAEL PARZEN
Emory University, Atlanta, GA, U.S.A.

GARRETT M. FITZMAURICE
Harvard Medical School, Boston, MA, U.S.A.

STUART R. LIPSITZ
Harvard Medical School, Boston, MA, U.S.A.

JOSEPH IBRAHIM
University of North Carolina, U.S.A.

and STEVEN LIPSHULTZ
University of Miami School of Medicine, Miami, FLA, U.S.A.

Submitted to the Journal of Statistical Planning and Inference
Summary

Missing covariate data is a common problem that complicates the fitting of regression models (e.g., generalized linear models) in many applications. A recently developed technique for handling missing covariate data is the weighted estimating equations (WEE) approach (Robins et al., 1994; Lipsitz et al., 1999). In the WEE approach, the contribution to the estimating equations from a complete-case (i.e., an observation with no missing covariate data) is weighted by the inverse ‘probability of being fully observed’. When the missing covariates are continuous, the efficient WEE requires integration over the distribution of the missing data. Lipsitz et al. (1999) proposed a Monte Carlo EM algorithm to obtain the efficient WEE, but, in general, their implementation can be computationally intensive. In this paper, as an alternative to the Monte Carlo EM algorithm, we propose using a simple numerical integration technique, Riemann summation using a midpoint rule, to overcome the computational burden of the efficient WEE. Due to the appealing properties of the WEE, even a relatively crude numerical approximation to the integral produces consistent estimators of the regression parameters. Furthermore, the results of a simulation study indicate that crude integral approximations can yield WEE estimators with relatively high semi-parametric efficiency. When the dimension of the missing covariate vector is large, and Riemann summation becomes computationally demanding, we propose numerical integration with the elements of a low-discrepancy sequence (Niederreiter, 1978) chosen as quadrature points. The proposed methods are illustrated using data from a study of cardiotoxicity among children receiving treatments for cancer.

Key words: Generalized linear models; low-discrepancy sequence, Missing at random; Missing data mechanism; Monte Carlo integration; Riemann summation.
1 Introduction

Missing covariate data is a common problem that complicates the fitting of regression models (e.g., generalized linear models). In this paper, we consider a regression analysis of an outcome $y$ on a vector of covariates, $x = (x_1, \ldots, x_p)'$, that are always observed, and a vector of continuous covariates $z = (z_1, \ldots, z_G)'$, that is not fully observed for many subjects. In general, the pattern of missingness of the components of $z$ can be arbitrary. Our primary interest is in estimating the vector of regression coefficients for the conditional distribution of $y$ given both $x$ and $z$. A complete-case analysis, excluding all subjects with missing covariate data, can yield highly inefficient and potentially biased estimates of the regression parameters (Little and Rubin, 1987).

To increase efficiency and reduce potential bias, one can use a weighted estimating equations (WEE) approach (Robins et al., 1994), which extended previous work proposed by Manski and Lerman (1977), Manski and McFadden (1981) and Cosslett (1981).

When the components of $z$ are continuous, the semi-parametric efficient WEE of Lipsitz et al. (1999) requires integration over the distribution of the missing data. The goal of this WEE is to obtain a semi-parametric efficient estimate of the regression parameters, without having to correctly specify the distribution of the missing data. Lipsitz et al. (1999) proposed a Monte Carlo EM algorithm to obtain the solution to the WEE. However, implementation of the Monte Carlo EM algorithm can be computationally intensive even when $z$ is univariate; in addition, numerical convergence of the Monte Carlo EM algorithm can require an excessively large number of Monte Carlo draws at each iteration (Satten and Datta, 2000). In this paper, we propose alternatives to the Monte Carlo EM; these alternatives do not require any Monte Carlo simulation. Instead, we propose direct numerical integration (quadrature) to approximate the integral in the WEE. Then, given the quadrature approximation to
the integral in the WEE, the Newton-Raphson algorithm is used to obtain the solution to the WEE. In particular, the integral in the original WEE is replaced with a sum (the quadrature approximation), and the solution to this WEE is obtained via the Newton-Raphson algorithm. This proposed technique does not require iteration between an E-step and an M-step like the Monte-Carlo EM algorithm, but instead only requires the specification of a set of quadrature points for which to do the numerical integration in the weighted estimating equations.

When the dimension \((G)\) of \(z\) is relatively small, say \(G \leq 5\), we propose using a simple numerical integration technique, Riemann summation using a midpoint rule, to overcome the computational burden of the efficient WEE. Due to the appealing properties of the WEE, even a relatively crude numerical approximation to the integral produces consistent estimators of the regression parameters. Furthermore, in simulations (reported later), we find that crude integral approximations can yield WEE estimators with relatively high semi-parametric efficiency. When the dimension \((G)\) of \(z\) is relatively large, even a crude numerical approximation, such as Riemann summation, can be computationally demanding. For example, when \(G = 6\), and only 10 quadrature points are used in each dimension, the numerical approximation to the integral requires a sum over \(10^6 = 1,000,000\) quadrature points for each observation.

When \(G \geq 6\), we propose to reduce the number of quadrature points by using the points of a low-discrepancy sequence (Niederreiter, 1978) as the quadrature points. In the unit interval \([0,1]\), a low-discrepancy sequence is a fixed set of \(J\) numbers that more uniformly fills out the interval \([0,1]\) than a set of \(J\) randomly sampled uniform \((0,1)\) random variables. However, in one-dimension, a set of equally spaced points in the \((0,1)\) interval (as one uses in Riemann summation with a midpoint rule), can just as easily be used. In fact, a set of equally spaced numbers is a low-discrepancy sequence in one-dimension. However, the
advantage of using a low-discrepancy sequence as quadrature points occurs in $G$-dimensions. In $G$-dimensions, a low-discrepancy sequence of $J$, $(G \times 1)$ vectors more uniformly fills out the unit hypercube than a set of $J$, $(G \times 1)$ vectors whose elements are independent uniform (0,1) random variables. There have been many proposed low-discrepancy sequences (Halton, 1960; Hammersley, 1960; Sobol, 1967; and Faure, 1982); in this paper, we use the simple low-discrepancy sequence proposed by Halton (1960). Since we have found that crude numerical approximations can lead to WEE with high efficiency, using the Halton sequence as quadrature points will suffice in leading to WEE with high efficiency. For the typical dimension of integrals in a missing data problems (very rarely more than $G = 10$), the low-discrepancy sequence leads to a suitable approximation (Sloan and Wozniakowski, 1998).

The use of WEE requires that the missing data are missing at random (MAR) (Rubin, 1976; Zhao et al., 1996) (MAR), where the probability of $z$ being completely observed can depend on the observed covariates ($x$) and the outcome ($y$), but not on any components of $z$. Robins et al. (1994) show that WEE are applicable to regression analysis when missing covariates are MAR and a model for the missing data process can be correctly specified. In the WEE approach the contribution to the regression estimating equations from a complete observation, $(y, x, z)$, is weighted by the inverse of the probability that $z$ is observed. This requires estimation of a binary regression model for the probability of $z$ being completely observed as a function of $(y, x)$, e.g., a logistic or probit model can be used or parametric assumptions can be relaxed by using generalized additive models (Hastie and Tibshirani, 1990). For ease of exposition, we first consider the case where all components of $z$ are either missing or observed; the generalization to arbitrary patterns of missingness in $z$ is outlined in Section 5.
2 Notation and Model

Consider a regression model involving \( n \) independent subjects, \( i = 1, \ldots, n \). The data collected on the \( i \)th subject are the outcome variable \( y_i \), a vector \( x_i = (x_{i1}, \ldots, x_{ip})' \) of \( p \) covariates that is always observed, and a vector \( z_i = (z_{i1}, \ldots, z_{iG})' \) of \( G \) covariates that has some missing elements for some subjects. Our primary interest is in estimating the vector of regression coefficients for the conditional distribution of \( y_i \) given both \( x_i \) and \( z_i \). We denote this density by

\[
p(y_i|x_i, z_i, \beta),
\]

(1)

where \( \beta \) contains both the regression parameters of primary interest and also nuisance parameters (e.g., dispersion parameters). For example, when (1) belongs to the exponential family, this includes a broad class of widely used regression models, such as standard linear regression, logistic and probit regression for binary data, Poisson regression for count data, and some proportional hazards models (see, for example, McCullagh and Nelder, 1989). However, we note that the proposed methods are not restricted to generalized linear models.

Since \( z_i \) can be missing, we must also consider the density of \( z_i \) given \( x_i \), denoted by \( p(z_i|x_i, \alpha) \), with parameter vector \( \alpha \). We consider both \( y_i \) and \( z_i \) as random, with \( x_i \) fixed. The complete data density of \((y_i, z_i|x_i)\) for the \( i \)th subject is

\[
p(y_i, z_i|x_i, \beta, \alpha) = p(y_i|x_i, z_i, \beta) \ p(z_i|x_i, \alpha) .
\]

(2)

In general, we assume that \( p(y_i|x_i, z_i, \beta) \) and \( p(z_i|x_i, \alpha) \) have continuous first and second derivatives in \( \beta \) and \( \alpha \), respectively.

We first consider the case where all components of \( z_i \) are either missing or observed. As such, we also define the indicator random variable \( r_i \), which equals 1 if \( z_i \) is observed and 0 if \( z_i \) is missing. The distribution of \( r_i \) given \((y_i, x_i, z_i)\) is Bernoulli with probability, \( \pi_i = \text{Pr}(r_i = 1|y_i, x_i, z_i) \), and is referred to as the missing data mechanism. We assume a
missing at random (MAR) mechanism in which \( \pi_i \) does not depend on \( z_i \) (given \( y_i \) and \( x_i \)). A logistic regression model is assumed for the dependence of \( \pi_i \) on both \( y_i \) and \( x_i \),

\[
\pi_i = \pi_i(\omega) = \frac{\exp(-\omega'm_i)}{1 + \exp(-\omega'm_i)},
\]

where \( \omega \) is a vector of unknown parameters and \( m_i \) is some function of \((y_i, x'_i)\); often, we have \( m_i = (y_i, x'_i) \), but \( m_i \) could also include interactions between the elements of \((y_i, x'_i)\). Note, in principle, other suitable link functions relating \( \pi_i \) to \( y_i \) and \( x_i \) could be adopted.

3 Simple Weighted Estimating Equations

The simplest weighted estimating equations are weighted versions of the so-called ‘complete-case’ estimator. The complete-case (CC) estimator, \( \hat{\beta}_{cc} \), is obtained as the solution to the estimating equations \( u_{cc}(\hat{\beta}_{cc}) = 0 \), where

\[
u_{cc}(\beta) = \sum_{i=1}^{n} r_i \left( \frac{\partial}{\partial \beta} \log[p(y_i|x_i, z_i, \beta)] \right). \tag{4}\]

Only complete-cases \((r_i = 1)\) contribute to (4). To obtain a consistent estimator of the regression parameters under a MAR mechanism, we can use the ‘simple’ weighted estimating equations proposed by Robins et al. (1994) and Zhao et al. (1996), where \( r_i \) is replaced in the complete case estimating equation (4) with \( r_i/\pi_i \). In particular, the weighted estimating equations are \( u_{WEE1}(\hat{\beta}_{WEE1}) = 0 \), where

\[
u_{WEE1}(\beta) = \sum_{i=1}^{n} \frac{r_i}{\pi_i} u_i(\beta) = \sum_{i=1}^{n} \frac{r_i}{\pi_i} \left( \frac{\partial}{\partial \beta} \log[p(y_i|x_i, z_i, \beta)] \right). \tag{5}\]

Hereafter, we denote the weighted estimating equations in (5) as WEE1.

Using results from the method of moments, because the estimating equations in (5) are unbiased for 0 and we are solving \( u_{WEE1}(\hat{\beta}_{WEE1}) = 0 \), \( \hat{\beta}_{WEE1} \) defines a consistent estimator for \( \beta \). If \( \pi_i \) is either known or consistently estimated, a consistent estimate of \( \beta \) is obtained. Under a MAR mechanism, \( \pi_i \) can be estimated independently from the model for
In most applications, \(\pi_i(\omega)\) is unknown and needs to be estimated and substituted in (5); this requires that simplifying assumptions, such as MAR, be made for \(\pi_i(\omega)\) to be identifiable and estimable. We propose estimating \(\pi_i\) using logistic regression with outcome \(r_i\) and covariates \(m_i\) given in (3).

### 4 Efficient Weighted Estimating Equations

To motivate the discussion of semi-parametric efficiency and the form of the efficient WEE, we first review maximum likelihood (ML) estimation in this setting. For ML, we must specify the density of \((y_i, z_i, r_i|x_i)\),

\[
p(y_i, z_i, r_i|x_i, \beta, \alpha, \omega) = p(y_i|x_i, z_i, \beta) p(z_i|x_i, \alpha) p(r_i|y_i, x_i, z_i, \omega) \\
= p(y_i|x_i, z_i, \beta) p(z_i|x_i, \alpha) p(r_i|y_i, x_i, \omega),
\]

(6)

where \(p(r_i|y_i, x_i, z_i, \omega)\) does not depend on \(z_i\) because the data are assumed to be MAR. Because \(x_i\) is always observed, we condition on it throughout equation (6).

We assume that (6) is the underlying density generating the data, regardless of whether \(z_i\) is observed or missing. If \(z_i\) is observed, then the observed data are \((r_i = 1, y_i, z_i, x_i)\), with density given by (6). If \(z_i\) is missing \((r_i = 0)\), then the observed data are \((r_i = 0, y_i, x_i)\), with the marginal distribution of \((r_i, y_i|x_i)\) given by

\[
p(r_i, y_i|x_i, \alpha, \beta, \omega) = p(r_i|y_i, x_i, \omega) \int_{z_i} p(y_i|x_i, z_i, \beta) p(z_i|x_i, \alpha) \, dz_i.
\]

(7)

Assuming the components of \(z_i\) are continuous, to obtain the marginal density

\[
p(y_i|x_i, \beta, \alpha) = \int_{z_i} p(y_i|x_i, z_i, \beta) p(z_i|x_i, \alpha) \, dz_i,
\]

(8)

some form of numerical or Monte Carlo integration is usually necessary; in this paper, we propose simple numerical integration. The contribution to the log-likelihood for the \(i^{th}\)
subject is

\[ r_i \log[p(r_i, y_i, z_i|x_i, \alpha, \beta, \omega)] + (1 - r_i) \log[p(r_i, y_i|x_i, \alpha, \beta, \omega)]. \]

Because the missing data are assumed to be MAR, so that \( p(r_i|y_i, x_i, \omega) \) factors outside of the integral in (7), the likelihood factors into a function of \( \omega \) and a function of \((\beta, \alpha)\), and the MLE for \((\beta, \alpha)\) can be obtained from solving \( U(\hat{\beta}, \hat{\alpha}) = 0 \), where

\[
U(\beta, \alpha) = \sum_{i=1}^{n} \left\{ r_i \left( \frac{\partial}{\partial (\beta, \alpha)} \log[p(y_i, z_i|x_i, \beta, \alpha)] \right) + (1 - r_i) \left( \frac{\partial}{\partial (\beta, \alpha)} \log[p(y_i|x_i, \beta, \alpha)] \right) \right\} = \\
\sum_{i=1}^{n} \left\{ r_i \left( \frac{\partial}{\partial (\beta, \alpha)} \log[p(y_i|x_i, z_i, \beta)p(z_i|x_i, \alpha)] \right) + (1 - r_i) \left( \frac{\partial}{\partial (\beta, \alpha)} \log[\int_{z_i} p(y_i|x_i, z_i, \beta)p(z_i|x_i, \alpha)dz_i] \right) \right\}. \tag{9}
\]

The maximum likelihood estimate \((\hat{\beta}, \hat{\alpha})\) is consistent if the distributions of \((y_i|x_i, z_i)\) and \((z_i|x_i)\) are correctly specified. However, if \( p(y_i|x_i, z_i) \) and/or \( p(z_i|x_i) \) are misspecified, then \( \hat{\beta} \) may fail to be consistent. Finally, even when \( p(z_i|x_i) \) is correctly specified, if the numerical approximation to (8) is not sufficiently accurate then the ML estimator may not be consistent. As an alternative to ML estimation, we propose a WEE approach that yields consistent estimators of \( \beta \) even when a crude numerical approximation to the integral is used.

Robins et al. (1994), Zhao et al. (1996) and Lipsitz et al. (1999) describe general forms for weighted estimating equations (WEE). Suppose we replace \( r_i \) with \( r_i/\pi_i \) in (9). Then, assuming \( \omega \) in \( \pi_i = \pi_i(\omega) \) is known, we obtain the ‘weighted estimating equations’ \( S(\hat{\beta}_{WEE}, \hat{\alpha}_{WEE}, \omega) = 0 \), where

\[
S(\beta, \alpha, \omega) = \sum_{i=1}^{n} \left\{ r_i \left( \frac{\partial}{\partial (\beta, \alpha)} \log[p(y_i, z_i|x_i, \beta, \alpha)] \right) + \left(1 - r_i/\pi_i \right) \left( \frac{\partial}{\partial (\beta, \alpha)} \log[p(y_i|x_i, \alpha, \beta)] \right) \right\}. \tag{10}
\]

Since \( \pi_i = \pi_i(\omega) \) is unknown, to obtain \( (\hat{\beta}_{WEE}, \hat{\alpha}_{WEE}) \), we solve \( S(\hat{\beta}_{WEE}, \hat{\alpha}_{WEE}, \hat{\omega}) = 0 \), where we have substituted \( \hat{\pi}_i = \pi_i(\hat{\omega}) \) for \( \pi_i \) in (10). Because of the similarity of (10) to the ML score vector in (9), the same maximization method (e.g., Newton-Raphson, EM algorithm) can be used to solve \( S(\hat{\beta}_{WEE}, \hat{\alpha}_{WEE}, \hat{\omega}) = 0 \). Similar to ML, this requires numerical integration when some components of \( z_i \) are continuous. In particular, Lipsitz et al. (1999)
showed that, similar to usual likelihood theory, (10) can be rewritten as

\[
S(\beta, \alpha, \omega) = \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi_i} \left( \frac{\partial}{\partial(\beta, \alpha)} \log[p(y_i, z_i | x_i, \beta, \alpha)] \right) + \left( 1 - \frac{r_i}{\pi_i} \right) E_{z_i | y_i, x_i} \left( \frac{\partial}{\partial(\beta, \alpha)} \log[p(y_i, z_i | x_i, \beta, \alpha)] \right) \right\},
\]

(11)

where the expectation on the right hand side of (11) is taken over the density of the missing data \((z_i)\) given the observed data \((y_i, x_i)\). The formulation of the score vector in (11) is the exact formulation of the score vector in the EM algorithm, except for the weights \(r_i / \pi_i\).

Nonetheless, Lipsitz et al. (1999) showed that the EM algorithm can still be used. In their Monte Carlo EM algorithm, given an estimate of \((\hat{\beta}, \hat{\alpha})\) from a previous iteration, one must sample \(z_i\) a large number of times from the conditional density

\[
p(z_i | y_i, x_i, \beta, \alpha) = \frac{p(y_i | x_i, z_i, \beta) \ p(z_i | x_i, \alpha)}{\int_{z_i} p(y_i | x_i, z_i, \beta) \ p(z_i | x_i, \alpha)}
\]

using the adaptive rejection Gibbs sampling algorithm of Gilks and Wild (1992). After

\[E_{z_i | y_i, x_i} \left( \frac{\partial}{\partial(\beta, \alpha)} \log[p(y_i, z_i | x_i, \beta, \alpha)] \right)\]

in (11) is approximated by an average over the sampled \(z_i\)’s, an updated estimate of \((\beta, \alpha)\) is then obtained as the solution to \(S(\beta, \alpha, \omega)\) in (11) equal to 0. The Monte Carlo EM algorithm iterates between filling-in the missing \(z_i\)’s and solving for updated estimates of \((\beta, \alpha)\) until convergence. For numerical convergence of the Monte Carlo EM algorithm, an excessively large number of Monte Carlo draws are required at each iteration (Satten and Datta, 2000). Thus, a major motivation for this paper is to find alternatives to the Monte Carlo EM algorithm. We do not explore the Monte Carlo EM algorithm any further in this paper, but instead focus on these alternatives Monte Carlo EM.

In the proposed WEE approach we are interested in obtaining a consistent estimator of \(\beta\) (or, more specifically, the regression parameters that comprise \(\beta\)), with \((\alpha, \omega)\) treated as nuisance parameters. The appeal of the WEE is that, unlike ML, a consistent estimator of \(\beta\) is obtained even if \(p(z_i | x_i)\) is misspecified, provided the model for the probability of \(z_i\) being
complete, \( \pi_i \), is correctly specified. The proof of this result is given in Robins and Ritov (1997) and Lipsitz et al. (1999). When \( p(z_i|x_i) \) is correctly specified, but the approximation to

\[
p(y_i|x_i, \alpha, \beta) = \int_{z_i} p(y_i|x_i, z_i, \beta) p(z_i|x_i, \alpha) \, dz_i
\]

(12)
in (10) is not accurate (e.g., due to a crude numerical approximation), then the WEE estimator remains consistent provided the model for \( \pi_i \) is correctly specified. Thus, when the dimension of \( z_i \) is relatively small, say \( G \leq 5 \), we suggest approximating the integral in (12) using a simple Riemann summation with a midpoint rule. Although not central to this paper, we also note that the WEE estimator has a “double-robustness” property: if \( \pi_i \) is misspecified the estimator is consistent provided \( p(z_i|x_i) \) is correctly specified (Robins and Ritov, 1997; Lipsitz et al., 1999).

First, suppose \( z_i \) is univariate, and we use a Riemann-sum approximation to the integral in (12). Suppose the support of \( z_i \) is a finite interval, and we divide this interval into \( M \) sub-intervals. We label the endpoints of the sub-intervals \( a_0, a_1, a_2, \ldots, a_M \), so that the support of \( z_i \) is \((a_0, a_M)\). In practical applications, we may choose \( a_0 \) to equal \((n-1)/n \) times the observed minimum and \( a_M \) to equal \((n+1)/n \) times the observed maximum (these are the MLE’s for the lower and upper endpoints for the uniform distribution). Alternatively, one can choose the endpoints \( a_0 \) and \( a_M \) to equal the 0.1 and 99.9 percentiles from the estimated marginal distribution of the observed data for \( z_i \). Also, the simplest approach is to let the \( M \) sub-intervals be of equal length \( \Delta \), so that \( \Delta = (a_m - a_{m-1}) = (a_M - a_0)/M \). In simulations, we have found that as few as \( M = 10 \) equally spaced intervals can give estimates with high asymptotic efficiency. We suggest using no more than \( M = 30 \) equally spaced intervals.

The Riemann-sum approximation, using midpoints, to the integral in (12) is

\[
\hat{p}(y_i|x_i, \alpha, \beta) = \Delta \sum_{m=1}^{M} p(y_i|x_i, 0.5a_{m-1} + 0.5a_m, \beta) p(0.5a_{m-1} + 0.5a_m|x_i, \alpha).
\]

(13)
The integral approximation becomes more accurate the larger the number of sub-intervals
However, as discussed above, results from a simulation study to be reported later indicate that a relatively small number of intervals (i.e., small $M$ and large $\Delta$) can lead to a WEE with relatively high semi-parametric efficiency. Furthermore, the choice of small $M$ and large $\Delta$ reduces the computational burden. In principle, one could allow the interval lengths, which we denote $\Delta_m = (a_m - a_{m-1})$, to vary, thus letting the integral approximation be

$$\hat{p}(y_i|x_i, \alpha, \beta) = \sum_{m=1}^{M} p(y_i|x_i, .5a_{m-1} + .5a_m, \beta) p(.5a_{m-1} + .5a_m|x_i, \alpha)\Delta_m.$$  

In practice, varying the interval lengths is usually not necessary in this setting.

Next, suppose $z_i = (z_{i1}, z_{i2})'$ is bivariate, and both $z_{i1}$ and $z_{i2}$ are continuous. We can divide the support of $z_{i1}$ into the sub-intervals $a_0, a_1, a_2, ..., a_M$, with a constant $\Delta_1 = (a_m - a_{m-1}) = (a_M - a_0)/M$; and divide the support of $z_{i2}$ into the sub-intervals $b_0, b_1, b_2, ..., b_Q$, with a constant $\Delta_2 = (b_q - b_{q-1}) = (b_Q - b_0)/Q$. The densities in (12) are $p(y_i|x_i, z_{i1}, z_{i2}, \beta)$ and $p(z_{i1}, z_{i2}|x_i, \alpha)$, and the Riemann-sum approximation to the integral (12) using midpoints is

$$\hat{p}(y_i|x_i, \alpha, \beta) = \Delta_1\Delta_2 \sum_{m=1}^{M} \sum_{q=1}^{Q} p(y_i|x_i, .5a_{m-1}+.5a_m, .5b_{q-1}+.5b_q, \beta) p(.5a_{m-1}+.5a_m, .5b_{q-1}+.5b_q|x_i, \alpha).$$  

(14)

The extension to more than two dimensions is straightforward. Note, using these simple integral approximations, the asymptotic variance of $\hat{\beta}_{WEE}$ can also be obtained using either the so-called “sandwich estimator” described in Zhao et al. (1996) or the jackknife (Lipsitz et al., 1994).

In simulations reported in Section 7, we show that when $z_i$ is one-dimensional, even a crude numerical integral approximation yields WEE estimates with high semi-parametric efficiency. For example, with only $M = 6$ quadrature points, the resulting WEE estimator has efficiency that exceeds 90%. However, when the dimension, $G$, of $z_i = (z_{i1}, \ldots, z_{iG})'$ is relatively large, say $G > 5$, the computational demands of the Riemann-sum technique
quickly become excessive. For example, with \( G = 5 \) and only \( M = 6 \) quadrature points in each dimension, the method requires summing over \( M^G = 6^5 = 7776 \) points; but with \( G = 10 \) and \( M = 6 \) quadrature points, the method requires summing over approximately 6 million points. As a less computationally demanding alternative, we propose the use of numerical integration that uses a low-discrepancy sequence of size \( J \) as the \( J \) quadrature points. In \( G \)-dimensions, a low-discrepancy sequence of \( J, (G \times 1) \) vectors more uniformly fills out the unit hypercube than a set of \( J \) randomly sampled uniform \((0,1)\) random variables. Further, one can choose \( J \) to be much less than \( M^G \). In this paper, we use the simple low-discrepancy sequence proposed by Halton (1960).

First, consider a uni-dimensional Halton sequence. The \( J \) points for a uni-dimensional Halton (1960) low-discrepancy sequence can be constructed by first uniquely writing the \( J \) integers \( j = 1, \ldots, J \) in a chosen base, say \( p \), and then transforming the digits. In particular, we can write \( j \) as

\[
j = s_0 + s_1 p + s_2 p^2 + \ldots + s_r p^r,
\]

where \( p \) is any prime number chosen to be the base, \( s_\ell \) is a non-negative integer less than \( p \), and \( r \) is a positive integer which depends on the base chosen. For example, if \( p = 2 \) is chosen and \( j = 10 \), then \( r = 4 \) and \((s_0, s_1, s_2, s_3) = (0, 1, 0, 1)\). Next, we transform \( j \) to the \( j^{th} \) number in the Halton sequence as

\[
h_p(j) = \frac{s_0}{p} + \frac{s_1}{p^2} + \frac{s_2}{p^3} + \ldots + \frac{s_r}{p^{r+1}}.
\]

(15)

In the above example, if \( p = 2 \) and \( j = 10 \), then \( h_2(10) = 0.3125 \). To construct the \( j^{th} \), \((G \times 1)\) vector in the multi-dimensional Halton sequence, one chooses a set of \( G \) prime numbers, say \( \{p_1, \ldots, p_G\} \), and performs the above transformation \( h_{p_g}(j) \) on the \( g^{th} \) prime number. In particular, the \( j^{th} \), \((G \times 1)\) vector in the multi-dimensional Halton sequence is

\[
h_j = \{h_{p_1}(j), h_{p_2}(j), \ldots, h_{p_G}(j)\},
\]

(16)
for \( j = 1, \ldots, J \). For example, if \( G = 2, p_1 = 5, p_2 = 7, \) and \( J = 10 \), then the 10 vectors of the Halton sequence are

\[
\{(0.20, 0.14), (0.40, 0.29), (0.60, 0.43), (0.80, 0.57), (0.04, 0.71), (0.24, 0.86), (0.44, 0.02), (0.64, 0.16), (0.84, 0.31), (0.08, 0.45)\}.
\]

The elements of the Halton sequence are always between 0 and 1, but can be transformed into the units of a covariate using a linear transformation. For example, in one-dimensional space, where the support of \( z_i \) is \((a_0, a_M)\), we use the transformation

\[
a_0 + (a_M - a_0)h_p(j).
\]

After obtaining the quadrature points from the Halton sequence, we replace \( \hat{p}(y_i|x_i, \alpha, \beta) \) in (13) with

\[
\hat{p}^*(y_i|x_i, \alpha, \beta) = \frac{1}{J} \sum_{j=1}^{J} p(y_i|x_i, h_j, \beta) p(h_j|x_i, \alpha).
\]

where the \( h_j \)'s \((j = 1, \ldots, J)\) are the points from the Halton sequence given in (16). Unlike numerical integration with \( M^G \) quadrature points, for any dimension, the number of quadrature points using the Halton sequence can be fixed in advance at a reasonable number (say \( J = 50 \)) and does not increase as the dimensions increase. Furthermore, this is unlike the Monte Carlo EM algorithm, in which a different Monte Carlo sample must be drawn for each subject at each iteration, and an excessively large number of Monte Carlo draws are required for convergence.

5 Extensions to Arbitrary Missing Data Patterns

In this section we outline the extension of the WEE to arbitrary missing covariate data patterns. As before, we let \( x_i = (x_{i1}, \ldots, x_{ip})' \) be the covariates that are always observed. However, we define \( w_{i,\text{mis}} = (w_{i1}, \ldots, w_{iG_i})' \) to be the set of covariates that are missing for
the \(i^{th}\) subject. Note that \(z_i = (z_{i1}, \ldots, z_{iG})'\) is defined as before, i.e., if a given covariate is missing for at least one subject, then it belongs to \(z_i\). Thus, \(w_{i,mis}\) is a subset of \(z_i\), and \(G_i \leq G\). Next, in a departure from the notation used in previous sections, we define \(r_i\) as the \(G \times 1\) vector of indicators, \(r_i = (r_{i1}, \ldots, r_{iG})'\), whose \(g^{th}\) component, \(r_{ig}\), equals 1 if \(z_{ig}\) is observed for the \(i^{th}\) subject, and 0 if \(z_{ig}\) is missing.

We must specify and estimate a model for the missing data mechanism, i.e., a model for the vector \(r_i\) as a function of \(m_i = (y_i, x_i')'\). For example, the joint multinomial distribution of \(r_i\) can be modelled using standard log-linear models (Agresti, 1990); we denote the parameters of this model by \(\omega\). The parameter vector \(\omega\) can be estimated using a multinomial log-likelihood for the vector \(r_i\). We define the joint probability

\[
\pi_i(r) = \pi_i(r_1, \ldots, r_G) = \Pr(r_{i1} = r_1, \ldots, r_{iG} = r_G | \omega),
\]

and its MLE by \(\hat{\pi}_i(r)\). We also define the indicator variable \(I_i(r)\), which equals 1 if \(r_i = r\) and equals 0 otherwise. Finally, we define \(1\) as a \(G \times 1\) vector of 1’s, so that if \(r_i = 1\), then all covariates are observed, and \(I_i(1) = 1\); if at least one covariate is missing, then \(I_i(1) = 0\). Then, the efficient WEE (Rotnitzky et al., 1998) is the solution to \(S(\hat{\beta}_{WEE}, \hat{\alpha}_{WEE}, \hat{\omega}) = 0\), where

\[
S(\beta, \alpha, \omega) = \sum_{i=1}^{n} \frac{I_i(1)}{\pi_i(1)} \left( \frac{\partial}{\partial (\beta, \alpha)} \log[\log[p(y_i|x_i, z_i, \beta) \ p(z_i|x_i, \alpha)]] \right) \\
+ \sum_{i=1}^{n} \sum_{r \neq 1} \left( I_i(r) - \frac{I_i(1)}{\pi_i(1)} \right) \left( \frac{\partial}{\partial (\beta, \alpha)} \log \left[ \int \ p(y_i|x_i, z_i, \beta) \ p(z_i|x_i, \alpha) \ dw_{i,mis} \right] \right) \tag{19},
\]

where the sum in the second term is over all \(2^G - 1\) vectors \(r\) that do not equal \(1\). Finally, the simple numerical integration techniques discussed in Section 4 can be used to approximate the integral in (19).
6 Application: Study of Cardiotoxicity

The ‘Prospective National Cancer Institute Cardiovascular Status Childhood Cancer Survivor (C3S2) Study’ (Lipshultz et al., 1995) was undertaken to study the cardiotoxicity of radiation to the chest and/or the class of anthracycline drugs in chemotherapy in childhood cancer survivors more than 5 years post-treatment. In this study there are 252 subjects belonging to three treatment groups: Group A consists of cancer patients who received anthracycline exposure and/or radiation to the heart; Group B consists of cancer patients who received neither anthracycline exposure nor radiation to the heart; group C consists of subjects without cancer (controls). The binary outcome of interest is low ($y_i = 1$) versus normal ($y_i = 0$) left ventricular end-diastolic dimension (a measure of the left ventricular function of the heart). The covariates are group ($\text{grp}_{Ai} = 1$ if group A and 0 otherwise; $\text{grp}_{Bi} = 1$ if group B and 0 otherwise), gender ($\text{fem}_i = 1$ if female, 0 if otherwise), $\text{wt}_i =$ standardized weight, and $\text{ht}_i =$ standardized height (height and weight are in standard deviations from the expected height and weight for the subject’s age and gender). We let $x_i = [\text{grp}_{Ai}, \text{grp}_{Bi}, \text{fem}_i]'$. All 252 patients have $y_i$ and $x_i$ observed, but 143 of the subjects (56.7%) are missing $z_i = [\text{wt}_i, \text{ht}_i]'$. Note, in this example, both elements of $z_i$ are either missing or observed. The distribution of $y_i$ given $(x_i, z_i)$ is Bernoulli and we model the probability that $y_i = 1$ given $(x_i, z_i)$ as a logistic regression,

$$\logit[p(y_i = 1|x_i, z_i, \beta)] = \logit[p(x_i, z_i)] = \beta_0 + \beta_1 \text{grp}_{Ai} + \beta_2 \text{grp}_{Bi} + \beta_3 \text{fem}_i + \beta_4 \text{wt}_i + \beta_5 \text{ht}_i.$$  

(20)

We note that one might expect an interaction between height and weight, i.e., that the effect of height to depend on weight in the model for $y_i$. However, in analyses not reported here, we found no statistically discernible interaction ($p > 0.15$).
Given the logistic regression model for \(y_i\), then

\[
p(y_i|x_i, z_i, \beta) = p(x_i, z_i)^{y_i}[1 - p(x_i, z_i)]^{(1-y_i)},
\]

(21)

Conditional on \(x_i = [\text{grp}_{A_i}, \text{grp}_{B_i}, \text{fem}_i]'\), we assume \(z_i = [\text{wt}_i, \text{ht}_i]'\) follows a bivariate normal distribution:

\[
p(\text{wt}_i, \text{ht}_i|x_i, \alpha) = \frac{e^{-\{(\text{wt}_i-x_i'\alpha_1)^2/\sigma_1^2+(\text{ht}_i-x_i'\alpha_2)^2/\sigma_2^2-2\rho(\text{wt}_i-x_i'\alpha_1)(\text{ht}_i-x_i'\alpha_2)/(2\sigma_1\sigma_2(1-\rho^2))\}}}{2\pi\sigma_1^2\sigma_2^2\sqrt{1-\rho^2}}.
\]

(22)

To implement the WEE proposed in Sections 4 and 5, we must use a numerical approximation to (12). We present estimates using both the Riemann-sum approximations and the simple numerical integration with a low-discrepancy sequence. Since we assume that \((z_i|x_i)\) has a bivariate normal distribution, the support of \(\text{wt}_i\) and \(\text{ht}_i\) are both \((-\infty, \infty)\). The minimum and maximum values of both \(\text{wt}_i\) and \(\text{ht}_i\) in the given data set were -5 and +5, and these values were used to guide the choice of \((a_0, a_M)\) and \((b_0, b_Q)\). In particular, for both the Riemann-sum approximations and the numerical integration with a low-discrepancy sequence, we let \(a_0 = b_0 = -5(144/143) = -5.03\), \(a_M = b_Q = 5(144/143) = 5.03\), which are the estimates of the lower and upper support points if the height and weight z-scores were uniformly distributed. For the Riemann-sum approximations, we also varied the number of intervals (of equal length) \(M\) and \(Q\); for simplicity, we report results for \(M = Q = 5\), \(M = Q = 10\) and \(M = Q = 40\). The Riemann-sum approximation to the integral in (12) is given by (14), with densities given by (21) and (22). Numerical integration with a low-discrepancy sequence is given by (18) with the \(J = 10\) quadrature points given in (17); we chose \(J = 10\) to see how a numerical integration technique with a small number of quadrature points compares to the Riemann-sum approximation.

For the WEE, we must also specify and estimate a model \(pr(r_i = 1|y_i, x_i)\). Assuming a
logistic regression, we found the best fitting model for $pr(r_i = 1|y_i, x_i)$ was

$$\logit[pr(r_i = 1|x_i, y_i)] = \omega_0 + \omega_1 y_i + \omega_2 \text{grp}_{A_i} + \omega_3 \text{grp}_{B_i} + \omega_4 \text{fem}_i + \omega_5 y_i * \text{grp}_{A_i} + \omega_6 y_i * \text{grp}_{B_i}.$$  

(23)

This model included an interaction between the outcome (left ventricular function) and group. The estimates for the logistic model in (23) are shown in Table 1. The interaction between outcome and group A is significant ($p < 0.001$) and indicates that subjects in group A (anthracycline exposure and/or radiation to the heart) who had low left ventricular end-diastolic dimension were less likely to have weight and height recorded (when compared to those with normal left ventricular end-diastolic dimension). Furthermore, females were observed to have weight and height recorded less frequently than males, with the evidence approaching conventional levels of significance ($p = .0672$). Although one cannot conclude based on the available data that the missing-data mechanism is MAR; if (23) was the true missing data mechanism, then weight and height would be MAR and we would expect a complete-case analysis to give a biased estimate of $\beta$.

Table 2 presents the estimates of $\beta$ for all of the methods considered; in this table, $\text{WEE}(M)$ denotes the WEE estimate for a given number of intervals $M = Q$ and WEE-LD denotes the WEE estimate using numerical integration with a Halton low-discrepancy sequence for $J = 10$ quadrature points. The estimated standard errors in Table 2 were calculated using the jackknife. We see from Table 2 that the CC estimate of the group A effect appears to be biased, with the CC estimate approximately three times the magnitude of any of the WEE estimates; furthermore, the CC estimate is significant ($p = .020$), whereas the others are not. In general, the WEE estimates of $\beta$ are similar to WEE1 (simple WEE); as the number of intervals $M = Q$ increases, $\text{WEE}(M)$ approaches the semi-parametric efficient estimate (assuming that (22) is the correct specification of $p(z_i|x_i)$). However, the results in Table 2 indicate that any of the integral approximations, even the simplest approx-
imation with WEE-LD or WEE(5), yields very similar estimates of coefficients and standard
errors. In fact for any $M \geq 10$, we obtain practically the same estimate and the same es-
timated standard error. Thus, for these particular data, this suggests that WEE(10), and
perhaps even WEE(5) or WEE-LD, are almost semi-parametrically efficient. In the following
section we examine, via a simulation study, the conjecture that a relatively crude integral
approximations can nonetheless yield almost semi-parametrically efficient estimates.

Finally, the results reported in Table 2 for the WEE($M$) estimates were obtained using
SAS PROC NLMIXED; an example data set, a SAS PROC NLMIXED program, and a
SAS macro to implement the weighted estimating equations is publicly available (at the
first authors website). The computational times varied greatly on a 2.4GHz Pentium IV
computer with 512MB of RAM. For all WEE($M$) estimates, we used a convergence criterion
of 0.00001. The WEE(5) estimate required 1.5 seconds of CPU time, the WEE(10) estimate
required 5 seconds, the WEE(40) estimate required 1 minute and 30 seconds. The WEE-LD
required only .15 seconds. Thus, not only did WEE-LD, WEE(5) and WEE(10) apparently
have relatively high semi-parametric efficiency, they also required substantially less time to
compute than WEE(40).

7 Simulation Study

We present the results from simulations with univariate $z_i$ given $x_i$ following an exponential
distribution. We performed many different simulations with exponential $z_i$, and, in all cases,
WEE with a relatively crude integral approximation led to estimates of $\beta$ with high semi-
parametric efficiency. We compare the estimates obtained from complete cases (CC); simple
weighted estimating equations (WEE1), weighted estimating equations with $(z_i|x_i)$ correctly
specified, and weighted estimating equations with $(z_i|x_i)$ incorrectly specified; for the latter
two, we considered different Riemann-sum midpoint integral approximations.
In the simulation study, there were two fully observed covariates, \((x_{i1}, x_{i2})\) and a possibly missing exponential covariate \((z_i)\). We formulate the true model for the simulations by specifying each term on the right hand side of

\[
p(r_i, z_i, y_i, x_{i1}, x_{i2}, \alpha, \beta, \omega) = p(r_i|y_i, x_{i1}, x_{i2}, z_i, \omega)p(y_i|x_{i1}, x_{i2}, z_i, \beta)p(z_i|x_{i1}, x_{i2}, \alpha_1)p(x_{i2}|x_{i1}, \alpha_2)p(x_{i1}|\alpha_3).
\]

(24)

For the covariate distributions, we let \(x_{i1}\) be Bernoulli with success probability 0.5. We let \(p(x_{i2}|x_{i1}, \alpha_2)\) also be Bernoulli, with the logit of the success probability equal to

\[
\text{logit}[pr(x_{i2} = 1|x_{i1})] = .31 - .35x_{i1}.
\]

(25)

Next, we let \(p(z_i|x_{i1}, x_{i2}, \alpha_1)\) have an exponential distribution,

\[
p(z_i|x_i, \alpha) = e^{-\lambda_i z_i \lambda_i},
\]

(26)

where

\[
\lambda_i = \exp(-1.5 + .5x_{i1} - .5x_{i2}).
\]

(27)

In each simulation, we let \(y_i\) given \((x_i, z_i)\) follow a Bernoulli distribution with the logit of the success probability equal to

\[
\text{logit}[pr(y_i = 1|x_{i1}, x_{i2}, z_i)] = .5 - x_{i1} + x_{i2} - z_i,
\]

(28)

i.e., \(\beta' = (\beta_0, \beta_1, \beta_2, \beta_z) = (.5, -1, 1, -1)\). Here, \(\beta\) is the parameter of primary interest. The true model for \(pr(r_i = 1|y_i, x_{i1}, x_{i2})\) was

\[
\text{logit}\{pr(r_i = 1|y_i, x_{i1}, x_{i2}, \omega)\} = .1 - .75y_i + .5x_{i1} + .5x_{i2} + x_{i1}x_{i2}.
\]

(29)

and was correctly specified when using the WEE approaches. Here, the marginal probability of being observed, i.e., \(pr(r_i = 1)\), is approximately 0.63. We considered \(n = 500\), which corresponds to a moderately large sample size, and performed 1,000 simulation replications, which implies a margin of error for 95% coverage statistics of 1.35%. The results of the simulations are presented in Table 3.
To use the WEE, we must use a numerical approximation to (12). First, we discuss the numerical approximation when \((z_i | x_i)\) is correctly specified as in (26). The support of \(z_i\) for this data set is \((0, \infty)\). For the Riemann-sum approximation using the midpoint rule, we must formulate the support points \((a_0, a_M)\), and the number of intervals (assumed to be of equal length). We obtained the WEE estimate for \(a_0 = 0\) and varying values of \(a_M\), and varying number of intervals \(M\). In each case, we considered an \(M\) that is 5 times larger than \(a_M\). In particular, we let \((a_M = 1.2, M = 6)\), \((a_M = 3, M = 15)\), \((a_M = 6, M = 30)\) and \((a_M = 92, M = 460)\). Here, examining the true marginal distribution of \(z_i\), \(a_M = 1.2\) is the 25th percentile, \(a_M = 3\) is the 50th percentile, and \(a_M = 6\) is the 75th percentile. The upper limit \(a_M = 92\) was derived by noting that the exponential probability of \(z_i\) (given \(x_i\)) being greater than 92 is less than \(10^{-8}\) for all \(n = 500\) subjects in a given simulation. For the correctly specified WEE, we denote the WEE estimate for a given upper limit \(a_M\) and given number of intervals \(M\) by \(\text{WEE}_{\text{exp}}(a_M, M)\).

To explore the properties of WEE with \((z_i | x_i)\) incorrectly specified, we also fit WEE with \((z_i | x_i)\) incorrectly specified as a log-logistic distribution

\[
p(z_i | x_i, \alpha) = \frac{\exp(\log(z_i) - \eta_i)}{[1 + \exp(\log(z_i) - \eta_i)]^2}, \tag{30}
\]

where \(\eta_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2}\). When using the integral approximation for WEE with (30), we used the same support points \((a_0, a_M)\), and number of intervals as when \(p(z_i | x_i, \alpha)\) is correctly specified. We denote the WEE estimate using the log-logistic distribution for \((z_i | x_i)\) by \(\text{WEE}_{\text{ell}}(a_M, M)\).

The results in Table 3 indicate that all WEE estimates feature little bias, whereas CC is biased for \(\beta_0\), \(\beta_1\) and \(\beta_2\). The coverage probabilities for the 95% confidence intervals all methods appear adequate, except for CC for the intercept and \(\beta_2\). For these simulations, we consider \(\text{WEE}_{\text{exp}}(92, 460)\) as the semi-parametric efficient estimate. We see that WEE1 can have very poor efficiency, 55.3% for \(\beta_1\) and 59.5% for \(\beta_2\). However, we see that even the crude
integral approximation used in \( WEE_{\exp}(1.2, 6) \) with correctly specified \((z_i|x_i)\) leads to estimates with over 90% efficiency. Interestingly, the efficiency of WEE with incorrectly specified \((z_i|x_i)\) is higher than WEE1, although this result cannot be expected to hold in general. Of note, the efficiency of WEE with incorrectly specified \((z_i|x_i)\) is not as high as the efficient \( WEE_{\exp}(92, 460) \). In particular, even the best numerical approximation, \( WEE_{\ell\ell}(92, 460) \) did not give quite as high efficiency as \( WEE_{\exp}(92, 460) \); however, a simple numerical approximation with misspecified \( p(z_i|x_i, \alpha) \), say \( WEE_{\ell\ell}(1.2, 6) \) did give very high efficiency relative to \( WEE_{\ell\ell}(92, 460) \). In one-dimension, the equally spaced quadrature points used in a Riemann-sum approximation with the midpoint rule do form a low-discrepancy sequence, so that we can consider these simulation results for \( WEE_{\exp}(a_M, M) \) as corresponding to typical results from a low-discrepancy sequence.

In the simulations, we used SAS PROC NLMIXED to compute \( WEE_{\exp}(a_M, M) \), and \( WEE_{\ell\ell}(a_M, M) \). For all of the WEE estimates, we used a convergence criterion of 0.00001. For a given simulation run, the \( WEE_{\exp}(1.2, 6) \) estimate required approximately 0.3 seconds, the \( WEE_{\exp}(3, 15) \) estimate required 0.5 seconds, the \( WEE_{\exp}(6, 30) \) estimate required 0.9 seconds, and the \( WEE_{\exp}(92, 460) \) estimate required 12.5 seconds. For a given \( a_M \) and \( M \), \( WEE_{\ell\ell}(a_M, M) \) required almost identical CPU time as \( WEE_{\exp}(a_M, M) \). Comparing the WEE estimates, not only did the \( WEE_{\exp}(1.2, 6) \) have relatively high semi-parametric efficiency, it also required much less time to compute. Furthermore, \( WEE_{\exp}(6, 30) \) required substantially less CPU time than \( WEE_{\exp}(92, 460) \) and was almost 100% efficient. The results of the simulations confirm what was suggested in Section 6: a relatively crude integral approximation can nonetheless yield an estimate of \( \beta \) with relatively high semi-parametric efficiency.
Lipsitz et al. (1999) proposed a Monte Carlo EM algorithm to obtain the efficient WEE with missing continuous covariates. For numerical convergence of the Monte Carlo EM algorithm, a large number of Monte Carlo draws are required (Satten and Datta, 2000), which can be computationally intensive and time consuming. Here, when the dimension of the missing covariate vector is small, say less than 6, we have proposed simple numerical integral approximations to use with weighted estimating equations. These estimating equations do not require Monte Carlo approximations; instead, they use simple numerical integration. Via simulations, we have examined the efficiency of the numerical approximations and found that simple numerical integration based on a midpoint rule has very high efficiency (at least 90% compared to the semi-parametric efficient estimator).

When the dimension of the missing covariate vector is large, say greater than 5, we propose simple numerical integration in which the quadrature points are the points of a low-discrepancy sequence (Niederreiter, 1978). Using this type of simple numerical integration, the number of quadrature points $J$ can remain fixed regardless of the dimension of the missing covariate vector, whereas, the number of quadrature points for a simple Riemann-sum approximation increases exponentially as the number of dimensions increase. For the typical dimension of integrals in a missing data problems (very rarely more than $G = 10$), the low-discrepancy sequence will be a suitable approximation (Sloan and Wozniakowski, 1998).

Because of the broad range of possible missing data configurations and underlying probability distributions generating the data, it is difficult to draw definitive conclusions from the simulation studies. Based on the simulation results and due to the simplicity of the estimating equations, we recommend use of the proposed WEE with a simple numerical integral
approximation. Because a simple Riemann-sum numerical approximation yields relatively high efficiency, the choice of the lower and upper bounds, \( a_0 \) and \( a_M \), and the width of the intervals are not so critical. In practical applications, we may choose \( a_0 \) to equal \((n-1)/n\) times the observed minimum and \( a_M \) to equal \((n+1)/n\) times the observed maximum. Furthermore, in simulations, we have found that as few as \( M = 10 \) equally spaced intervals can give estimates with high asymptotic efficiency. We suggest using no more than \( M = 30 \) equally spaced intervals. For simple numerical integration in which the quadrature points are the points of a low-discrepancy sequence, we similarly suggest using at least \( J = 30 \) quadrature points. In summary, the contributions of this paper are showing the computational simplicity of the proposed WEE in addition to its high efficiency relative to the semi-parametric efficient WEE.

Finally, in this paper we have concentrated on the setting where \( \pi_i = \Pr(r_i = 1|y_i, x_i, z_i) \) and \( p(z_i|x_i, \alpha) \) are correctly specified, and one wants to use a simple numerical approximation to gain efficiency. If \( \pi_i \) is misspecified, then the resulting estimate of \( \beta \) will be biased unless \( p(z_i|x_i, \alpha) \) is correctly specified. In this case, one would want to use more intervals in the numerical approximation to better approximate \( p(z_i|x_i, \alpha) \). If \( \pi_i \) is correctly specified, but \( p(z_i|x_i, \alpha) \) is not, then the estimate of \( \beta \) will be consistent. In this case, as seen in the simulations, even the best numerical approximation cannot be expected to yield quite as high efficiency as when \( p(z_i|x_i, \alpha) \) is correctly specified.

Acknowledgments

The authors are grateful for the support provided by the following grants from the US National Institutes of Health: GM 29745, HL 69800, AI 60373, CA 74015, CA 70101, and CA 68484.
References


Table 1. Logistic regression estimates for the model for \( pr(R_i = 1|x_i, y_i) \) from the cardiotoxicity data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3482</td>
<td>0.5092</td>
<td>0.68</td>
<td>0.494</td>
</tr>
<tr>
<td>( y_i )</td>
<td>1.3019</td>
<td>0.6298</td>
<td>2.07</td>
<td>0.039</td>
</tr>
<tr>
<td>( \text{grp}_{Ai} )</td>
<td>1.0372</td>
<td>0.6287</td>
<td>1.65</td>
<td>0.099</td>
</tr>
<tr>
<td>( \text{grp}_{Bi} )</td>
<td>-0.6687</td>
<td>0.7656</td>
<td>-0.87</td>
<td>0.382</td>
</tr>
<tr>
<td>( y_i \times \text{grp}_{Ai} )</td>
<td>-2.5966</td>
<td>0.7528</td>
<td>-3.45</td>
<td>0.001</td>
</tr>
<tr>
<td>( y_i \times \text{grp}_{Bi} )</td>
<td>-1.1855</td>
<td>0.9240</td>
<td>-1.28</td>
<td>0.200</td>
</tr>
<tr>
<td>\text{female}</td>
<td>-0.5090</td>
<td>0.2781</td>
<td>-1.83</td>
<td>0.067</td>
</tr>
</tbody>
</table>
Table 2. Regression parameter ($\beta$) estimates from the cardiotoxicity data.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Approach</th>
<th>$\hat{\beta}$</th>
<th>SE</th>
<th>Z</th>
<th>p—value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>CC</td>
<td>1.0155</td>
<td>0.4679</td>
<td>2.17</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>0.5756</td>
<td>0.3611</td>
<td>1.59</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>WEE(5)$^a$</td>
<td>0.6861</td>
<td>0.3507</td>
<td>1.96</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>0.6860</td>
<td>0.3507</td>
<td>1.96</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>0.6895</td>
<td>0.3462</td>
<td>1.99</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>0.6575</td>
<td>0.3505</td>
<td>1.93</td>
<td>0.054</td>
</tr>
<tr>
<td>$\text{grp}_{A_i}$</td>
<td>CC</td>
<td>-1.2332</td>
<td>0.5286</td>
<td>-2.33</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>-0.3883</td>
<td>0.4256</td>
<td>-0.91</td>
<td>0.362</td>
</tr>
<tr>
<td></td>
<td>WEE(5)</td>
<td>-0.4390</td>
<td>0.4122</td>
<td>-1.06</td>
<td>0.287</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>-0.4390</td>
<td>0.4121</td>
<td>-1.06</td>
<td>0.287</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>-0.4271</td>
<td>0.4033</td>
<td>-1.06</td>
<td>0.290</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>-0.4383</td>
<td>0.4135</td>
<td>-1.06</td>
<td>0.289</td>
</tr>
<tr>
<td>$\text{grp}_{B_i}$</td>
<td>CC</td>
<td>-0.3923</td>
<td>0.7584</td>
<td>-0.52</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>0.0290</td>
<td>0.5013</td>
<td>0.06</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>WEE(5)</td>
<td>-0.2049</td>
<td>0.5010</td>
<td>-0.41</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>-0.2049</td>
<td>0.5010</td>
<td>-0.41</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>-0.1993</td>
<td>0.4963</td>
<td>-0.40</td>
<td>0.689</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>-0.1998</td>
<td>0.4987</td>
<td>-0.40</td>
<td>0.689</td>
</tr>
<tr>
<td>female</td>
<td>CC</td>
<td>0.7655</td>
<td>0.4093</td>
<td>1.87</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>0.9303</td>
<td>0.4066</td>
<td>2.29</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>WEE(5)</td>
<td>0.7714</td>
<td>0.3136</td>
<td>2.46</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>0.7717</td>
<td>0.3135</td>
<td>2.46</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>0.7561</td>
<td>0.3135</td>
<td>2.41</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>0.8047</td>
<td>0.3150</td>
<td>2.55</td>
<td>0.011</td>
</tr>
<tr>
<td>height</td>
<td>CC</td>
<td>-0.3743</td>
<td>0.2217</td>
<td>-1.69</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>-0.3722</td>
<td>0.2278</td>
<td>-1.63</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>WEE(5)</td>
<td>-0.3769</td>
<td>0.2275</td>
<td>-1.66</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>-0.3768</td>
<td>0.2270</td>
<td>-1.65</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>-0.3621</td>
<td>0.2210</td>
<td>-1.65</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>-0.3880</td>
<td>0.2272</td>
<td>-1.71</td>
<td>0.087</td>
</tr>
<tr>
<td>weight</td>
<td>CC</td>
<td>0.5190</td>
<td>0.1968</td>
<td>2.64</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>WEE1</td>
<td>0.4910</td>
<td>0.1960</td>
<td>2.51</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>WEE(5)</td>
<td>0.4920</td>
<td>0.1955</td>
<td>2.52</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>WEE(10)</td>
<td>0.4918</td>
<td>0.1954</td>
<td>2.51</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>WEE(40)</td>
<td>0.4762</td>
<td>0.1900</td>
<td>2.51</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>WEE-LD$^b$</td>
<td>0.5151</td>
<td>0.1919</td>
<td>2.68</td>
<td>0.007</td>
</tr>
</tbody>
</table>

$^a$WEE($M = Q$) denotes WEE method using Riemann-sum approximation with $M = Q$ intervals

$^b$WEE-LD denotes WEE method using numerical integration with a Halton low-discrepancy sequence as the $J = 10$ quadrature points
Table 3. Results of simulation study with correctly and incorrectly specified models for \( p(z_i | x_i, \alpha) \).

<table>
<thead>
<tr>
<th>Approach</th>
<th>( \beta_0 = .5 )</th>
<th>( \beta_1 = -1 )</th>
<th>( \beta_2 = 1 )</th>
<th>( \beta_3 = -1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average CC</td>
<td>0.050</td>
<td>-0.866</td>
<td>1.201</td>
<td>-1.058</td>
</tr>
<tr>
<td>WEE1</td>
<td>0.523</td>
<td>-1.021</td>
<td>1.037</td>
<td>-1.063</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(1,2, 6)^a</td>
<td>0.564</td>
<td>-1.036</td>
<td>1.030</td>
<td>-1.059</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(3, 15)</td>
<td>0.562</td>
<td>-1.034</td>
<td>1.028</td>
<td>-1.058</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(6, 30)</td>
<td>0.567</td>
<td>-1.035</td>
<td>1.030</td>
<td>-1.057</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(92, 460)</td>
<td>0.569</td>
<td>-1.039</td>
<td>1.036</td>
<td>-1.059</td>
</tr>
<tr>
<td>WEE_{exp}(1, 2, 6)^b</td>
<td>0.566</td>
<td>-1.043</td>
<td>1.017</td>
<td>-1.059</td>
</tr>
<tr>
<td>WEE_{exp}(3, 15)</td>
<td>0.561</td>
<td>-1.039</td>
<td>1.014</td>
<td>-1.055</td>
</tr>
<tr>
<td>WEE_{exp}(6, 30)</td>
<td>0.560</td>
<td>-1.038</td>
<td>1.014</td>
<td>-1.056</td>
</tr>
<tr>
<td>WEE_{exp}(92, 460)</td>
<td>0.560</td>
<td>-1.038</td>
<td>1.014</td>
<td>-1.056</td>
</tr>
<tr>
<td>Average Variance of Estimates CC</td>
<td>0.2151</td>
<td>0.1866</td>
<td>0.1938</td>
<td>0.0523</td>
</tr>
<tr>
<td>WEE1</td>
<td>0.1860</td>
<td>0.1966</td>
<td>0.1992</td>
<td>0.0525</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(1,2, 6)</td>
<td>0.1670</td>
<td>0.1219</td>
<td>0.1464</td>
<td>0.0528</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(3, 15)</td>
<td>0.1575</td>
<td>0.1175</td>
<td>0.1422</td>
<td>0.0529</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(6, 30)</td>
<td>0.1557</td>
<td>0.1151</td>
<td>0.1379</td>
<td>0.0529</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(92, 460)</td>
<td>0.1556</td>
<td>0.1151</td>
<td>0.1373</td>
<td>0.0528</td>
</tr>
<tr>
<td>WEE_{exp}(1, 2, 6)</td>
<td>0.1625</td>
<td>0.1156</td>
<td>0.1290</td>
<td>0.0524</td>
</tr>
<tr>
<td>WEE_{exp}(3, 15)</td>
<td>0.1557</td>
<td>0.1109</td>
<td>0.1230</td>
<td>0.0525</td>
</tr>
<tr>
<td>WEE_{exp}(6, 30)</td>
<td>0.1541</td>
<td>0.1090</td>
<td>0.1194</td>
<td>0.0524</td>
</tr>
<tr>
<td>WEE_{exp}(92, 460)</td>
<td>0.1541</td>
<td>0.1088</td>
<td>0.1186</td>
<td>0.0525</td>
</tr>
<tr>
<td>Relative Efficiency (^b) WEE1</td>
<td>82.8</td>
<td>55.3</td>
<td>59.5</td>
<td>100.0</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(1,2, 6)</td>
<td>92.3</td>
<td>89.3</td>
<td>81.0</td>
<td>99.9</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(3, 15)</td>
<td>97.8</td>
<td>92.6</td>
<td>83.4</td>
<td>99.9</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(6, 30)</td>
<td>99.0</td>
<td>94.5</td>
<td>86.0</td>
<td>99.9</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(92, 460)</td>
<td>99.0</td>
<td>94.5</td>
<td>86.4</td>
<td>99.9</td>
</tr>
<tr>
<td>WEE_{exp}(1, 2, 6)</td>
<td>94.8</td>
<td>94.1</td>
<td>91.9</td>
<td>100.0</td>
</tr>
<tr>
<td>WEE_{exp}(3, 15)</td>
<td>99.0</td>
<td>98.1</td>
<td>96.4</td>
<td>100.0</td>
</tr>
<tr>
<td>WEE_{exp}(6, 30)</td>
<td>100.0</td>
<td>99.8</td>
<td>99.3</td>
<td>100.0</td>
</tr>
<tr>
<td>WEE_{exp}(92, 460)</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Coverage Probability (^c) CC</td>
<td>84.8</td>
<td>94.1</td>
<td>92.5</td>
<td>94.1</td>
</tr>
<tr>
<td>WEE1</td>
<td>95.2</td>
<td>95.0</td>
<td>94.8</td>
<td>94.6</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(1,2, 6)</td>
<td>94.7</td>
<td>94.3</td>
<td>94.9</td>
<td>94.0</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(3, 15)</td>
<td>94.5</td>
<td>94.3</td>
<td>94.7</td>
<td>93.8</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(6, 30)</td>
<td>94.6</td>
<td>94.3</td>
<td>94.8</td>
<td>93.8</td>
</tr>
<tr>
<td>WEE_{\ell\ell}(92, 460)</td>
<td>94.5</td>
<td>94.4</td>
<td>94.5</td>
<td>93.6</td>
</tr>
<tr>
<td>WEE_{exp}(1, 2, 6)</td>
<td>94.2</td>
<td>94.2</td>
<td>95.1</td>
<td>94.3</td>
</tr>
<tr>
<td>WEE_{exp}(3, 15)</td>
<td>94.0</td>
<td>94.7</td>
<td>95.3</td>
<td>94.0</td>
</tr>
<tr>
<td>WEE_{exp}(6, 30)</td>
<td>94.3</td>
<td>94.7</td>
<td>95.0</td>
<td>94.0</td>
</tr>
<tr>
<td>WEE_{exp}(92, 460)</td>
<td>94.3</td>
<td>94.9</td>
<td>95.0</td>
<td>94.0</td>
</tr>
</tbody>
</table>

\(^a\) WEE_{\ell\ell}(a_M, M) : WEE with \( p(z_i | x_i, \alpha) \) incorrectly specified based on log-logistic distribution

\(^b\) WEE_{exp}(a_M, M) : WEE with \( p(z_i | x_i, \alpha) \) correctly specified based on exponential distribution

\(^c\) Coverage Probability of 95% Confidence Intervals