A Weighted Estimating Equation for Missing Covariate Data With Properties Similar to Maximum Likelihood

Stuart R. LIPSITZ, Joseph G. IBRAHIM, and Lue Ping ZHAO

In regression analysis, missing covariate data occurs often. A recent approach to analyzing such data is weighted estimating equations. With weighted estimating equations, the contribution to the estimating equation from a complete observation is weighted by the inverse probability of being observed. In this article we propose a weighted estimating equation that is almost identical to the maximum likelihood estimating equations. As such, we propose an EM-type algorithm to solve these weighted estimating equations. Although the weighted estimating equations are a special case of those proposed earlier by Robins et al., our EM-type algorithm to solve them is new. Similar to Robins and Ritov, we give the result that to obtain a consistent estimate of the regression parameters, either the missing-data mechanism or the distribution of the missing data given the observed data must be correctly specified. We compare the weighted estimating equations to maximum likelihood via two examples, a simulation and an asymptotic study.

KEY WORDS: Generalized linear model; Missing at random; Missing completely at random; Missing-data mechanism.

1. INTRODUCTION

Missing covariate data is common in such studies as sample surveys and clinical trials. The most common technique used by data analysts is to naively exclude subjects with missing covariates, then perform a regression analysis with the remaining data. This is called a complete-case analysis. Because subjects with any missing variables are excluded, it is well known (see Little and Rubin 1987) that a complete-case analysis can give highly inefficient estimates. Also, if the complete cases are not a completely random sample of the original data, then the complete-case estimates can be biased. Thus to increase efficiency and reduce the bias, it is important to develop methods that incorporate the partially incomplete data into the analysis. In this article, we consider a regression analysis of an outcome y on a vector $\mathbf{x} = (x_1, \ldots, x_p)'$ of p covariates that are always observed, and a covariate z, which can be missing for some subjects. We consider two examples in Section 6.

First, we analyze a randomized clinical trial in multiple myeloma (Kalish 1992). The outcome variable is the survival time and there are eight covariates, one of which has missing values. The only covariate with missing values is z = bone fractures at diagnosis (1 = yes, 0 = no), which is missing for 84 (37.5%) of the 224 cases in the dataset. This covariate is a very important predictor, as multiple myeloma is a hematologic cancer that attacks the bone marrow, so a patient with bone fractures at diagnosis may have more advanced cancer and thus shorter survival. With such a large fraction of missing data, a complete-case analysis using only the 140 subjects with no missing data could give highly inefficient and/or biased estimates.

Second, we analyze data from a randomized phase II clinical trial to evaluate new treatments in patients with primary liver cancer. This study involved 4 treatments and 140 patients (35 per treatment). The binary outcome is whether or not the patient survived at least 4 months. The covariates of interest are treatment and time from diagnosis until entry onto the study. Unfortunately, 27 subjects (19.3%) are missing the time from diagnosis variable.

We consider two possible missing-data processes. *Missing completely at random* (MCAR) implies that the missing-data process does not depend on the observed outcomes nor on the observed covariates. *Missing at random* (MAR) implies that the missing-data process can depend on the observed covariates ($\mathbf{x}$) and the outcome (y), but not on (z). Developing methods for regression analyses with covariate data that are MAR or MCAR has received much recent attention in the statistical literature. Two recent approaches are weighted estimating equations (WEEs) (Robins, Rotnitzky, and Zhao 1994; Zhao, Lipsitz, and Lew 1996) and maximum likelihood (ML) (Ibrahim 1990; Lipsitz and Ibrahim 1996; Little and Schluchter 1985). With WEE, the contribution to the estimating equation from a complete observation $(y, \mathbf{x}, z)$ is weighted by π, the inverse probability that z was observed. It has been shown that WEE is applicable to regression analysis when missing covariates are either MAR or MCAR (Zhao et al. 1996).

Alternatively, when using ML, a distribution for $(y, z|\mathbf{x})$ must be specified. A common approach to specifying this distribution is to specify a conditional distribution on $(y|\mathbf{x}, z)$ and a marginal distribution on $(z|\mathbf{x})$. In this specification the interest lies in estimating the parameters from the conditional distribution of $(y|\mathbf{x}, z)$ with the parameters from $(z|\mathbf{x})$ treated as nuisance parameters. A popular approach for obtaining the maximum likelihood estimate (MLE) is the EM algorithm if z is discrete (Dempster, Laird, and Rubin 1977; Ibrahim 1990; Lipsitz and Ibrahim 1996), and

Stuart R. Lipsitz is Professor, Medical University of South Carolina, Department of Biometry and Epidemiology. Joseph G. Ibrahim is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana Farber Cancer Institute, Boston, MA 02115 (E-mail: *ibrahim@jimmy.harvard.edu*). Lue Ping Zhao is Associate Professor, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98104. The authors are grateful for the support provided by National Institutes of Health grants CA 55576, CA 55670, CA 70101, and CA 74015.

Journal of the American Statistical Association
December 1999, Vol. 94, No. 448, Theory and Methods

a Monte Carlo EM algorithm if z is continuous (Ibrahim, Chen, and Lipsitz 1999).

In this article we propose a WEE that is almost identical to the ML estimating equations. As such, we propose an EM-type algorithm to solve these WEEs. Although the WEEs are a special case of those proposed by Robins et al. (1994), our EM-type algorithm to solve them is new. Similar to Robins and Ritov (1997), we give the result that to obtain a consistent estimate of the regression parameters, either, but not both, the missing-data mechanism or the distribution of the missing data given the observed data must be correctly specified. Further, we show that both the ML and the WEEs are special cases of a general class of estimating equations that give a consistent estimate of β as long as $p(z|\mathbf{x})$ is correctly specified.

In Section 2 we define the notation. In Section 3, to provide sufficient background for our proposed WEEs, we describe the ML score equations; in Section 4 we describe our proposed WEEs. In Section 5 we describe the EM-type algorithm. In Section 6 we illustrate the methods with the examples. In Section 7 we give results from a simulation study that examines the approximation of the asymptotic results in finite samples, and in Section 8 we present an asymptotic study. We conclude the article with a brief discussion in Section 9.

2. NOTATION AND MODEL

Consider a regression problem involving N independent subjects, $i = 1, \ldots, N$. The data collected on the ith subject are the outcome variable y_i, a vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ of p covariates that are always observed, and a covariate z_i that is missing for some subjects. Because z_i can be missing, 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 $(y_i, \mathbf{x}_i, z_i)$, which is Bernoulli with probability

$$\pi_i = \Pr(r_i = 1|y_i, \mathbf{x}_i, z_i),$$

is called the missing-data mechanism. If missingness is nonignorable, then π_i depends on z_i. In this article, we consider MAR or MCAR missing-data mechanisms in which π_i does not depend on z_i. When the missing data are MAR, π_i can depend on both y_i and $\mathbf{x}_i$. MCAR implies that π_i does not depend on $y_i, \mathbf{x}_i$, or z_i. We pose a logistic regression for the probability of being observed,

$$\pi_i = \pi_i(\boldsymbol{\omega}) = \frac{\exp(-\boldsymbol{\omega}'\mathbf{m}_i)}{1 + \exp(-\boldsymbol{\omega}'\mathbf{m}_i)}. \quad (1)$$

where $\boldsymbol{\omega}$ is a vector of unknown parameters and $\mathbf{m}_i$ is some function of $(y_i, \mathbf{x}_i')$'s. Often we have $\mathbf{m}_i = (y_i, \mathbf{x}_i')'$, but $\mathbf{m}_i$ could also include interactions between the elements of $(y_i, \mathbf{x}_i')'$.

Our interest is in estimating the vector of regression coefficients β of the conditional distribution of y_i given $\mathbf{x}_i$ and z_i. Suppose that $(y_i|\mathbf{x}_i, z_i)$ has a density in the exponential family with the form

$$p(y_i|\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}, \tau) = \exp\{a_i^{-1}(\tau)(y_i\theta_i - b(\theta_i)) + c(y_i, \tau)\}, \quad (2)$$

where $\theta_i = \theta(\beta_0 + \mathbf{x}_i'\beta_1 + z_i\beta_2)$, $\boldsymbol{\beta}' = (\beta_0, \beta_1', \beta_2)$, and τ is the scale parameter. The functions b and c determine a particular family in the class, such as the binomial, normal, or Poisson. The functions $a_i(\tau)$ are commonly of the form $a_i(\tau) = \tau^{-1}s_i^{-1}$, where the s_i's are known weights. The density in (2) includes a large class of regression models, such as normal linear regression, logistic and probit regression, Poisson regression, gamma regression, and some proportional hazards models (see McCullagh and Nelder 1989). Without loss of generality, we assume throughout that $\tau = 1$, as, for example, in logistic and Poisson regression. Because we assume that $\tau = 1$, we write (2) as

$$p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta}, \tau) \equiv p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})$$

throughout.

Since subjects can have z_i missing, we must also consider the density of z_i given $\mathbf{x}_i$, which we denote by $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ with parameter vector $\boldsymbol{\alpha}$. The complete-data density of $(y_i, z_i|\mathbf{x}_i)$ for subject i is now given by

$$p(y_i, z_i|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) = p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha}). \quad (3)$$

The density of $(y_i, z_i, r_i|\mathbf{x}_i)$ for subject i is then given by

$$\begin{aligned} p(y_i, z_i, r_i|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega}) &= p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega}) \\ &= p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})p(r_i|y_i, \mathbf{x}_i, \boldsymbol{\omega}), \end{aligned} \quad (4)$$

because the data are MAR. Because $\mathbf{x}_i$ is always observed, we condition on it throughout (4). We denote the vector of all the parameters as $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \boldsymbol{\omega}')$. Our main interest is in estimation of $\boldsymbol{\beta}$, with $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$ viewed as nuisance parameters.

3. MAXIMUM LIKELIHOOD

To provide sufficient background for our proposed weighted estimating equations, in this section we describe the ML equations when z_i is missing for some subjects. First, if there are no missing data, then the ML estimating equations for $\hat{\gamma}' = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}', \hat{\omega}')$ are

$$\mathbf{u}(\hat{\gamma}) = \begin{bmatrix} \mathbf{u}_1(\hat{\boldsymbol{\beta}}) \\ \mathbf{u}_2(\hat{\boldsymbol{\alpha}}) \\ \mathbf{u}_3(\hat{\boldsymbol{\omega}}) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{u}_{1i}(\hat{\boldsymbol{\beta}}) \\ \mathbf{u}_{2i}(\hat{\boldsymbol{\alpha}}) \\ \mathbf{u}_{3i}(\hat{\boldsymbol{\omega}}) \end{bmatrix} = 0. \quad (5)$$

where

$$\mathbf{u}_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) = \sum_{i=1}^{n} \frac{\partial \log p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \quad (6)$$

$$\mathbf{u}_2(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{u}_{2i}(\boldsymbol{\alpha}; \mathbf{z}_i, x_i) = \sum_{i=1}^{n} \frac{\partial \log p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}. \quad (7)$$

and

$$\mathbf{u}_3(\boldsymbol{\omega}) = \sum_{i=1}^{n} \mathbf{u}_{3i}(\boldsymbol{\omega}; r_i, y_i, \mathbf{x}_i) = \sum_{i=1}^{n} \frac{\partial \log p(r_i|y_i, \mathbf{x}_i, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \sum_{i=1}^{n} \mathbf{m}_i'[r_i - \pi_i(\boldsymbol{\omega})], \quad (8)$$

and π_i is given in (1).

Using the EM algorithm (Dempster et al. 1977) with z_i missing for some subjects, the MLE of γ can be obtained by setting the conditional expectation of the complete-data score vector, denoted by $\mathbf{u}^*(\gamma)$, to 0, and solving for $\hat{\gamma}$. Here the expectation is taken with respect to the conditional distribution of the missing data given the observed data; that is,

$$u^*(\gamma) = \sum_{i=1}^{n} u_i^*(\gamma) = \sum_{i=1}^{n} E\left[\begin{array}{c|c} \mathbf{u}_{1i}(\boldsymbol{\beta}) & \\ \mathbf{u}_{2i}(\boldsymbol{\alpha}) & \text{observed data for subject } i \\ \mathbf{u}_{3i}(\boldsymbol{\omega}) & \end{array}\right]. \quad (9)$$

In particular, if z_i is observed, then the observed data are $(r_i, y_i, \mathbf{x}_i, z_i)$, and subject i's contribution to the expected score is

$$\mathbf{u}_i^*(\gamma) = E\left[\begin{array}{c|c} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) & \\ \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) & z_i, \mathbf{x}_i, y_i, r_i \\ \mathbf{m}_i'(r_i - \pi_i) & \end{array}\right] = \begin{bmatrix} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) \\ \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) \\ \mathbf{m}_i'(r_i - \pi_i) \end{bmatrix}, \quad (10)$$

which is the same as its contribution to the score vector in (5). If z_i is missing, then the observed data are $(r_i, y_i, \mathbf{x}_i)$ and

$$\mathbf{u}_i^*(\gamma) = E\left[\begin{array}{c|c} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) & \\ \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) & r_i, y_i, \mathbf{x}_i \\ \mathbf{m}_i'(r_i - \pi_i) & \end{array}\right] = \begin{bmatrix} E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \\ E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)] \\ \mathbf{m}_i'(r_i - \pi_i) \end{bmatrix}. \quad (11)$$

The conditional expectation on the right side of (11) does not depend on r_i because, with data MAR,

$$\begin{aligned} p(z_i|y_i, \mathbf{x}_i, r_i) &= \frac{p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega})p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})}{\int_{z_i} p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega})p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})} \\ &= \frac{p(r_i|y_i, \mathbf{x}_i, \boldsymbol{\omega})p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})}{p(r_i|y_i, \mathbf{x}_i, \boldsymbol{\omega}) \int_{z_i} p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})} \\ &= p(z_i|y_i, \mathbf{x}_i). \end{aligned} \quad (12)$$

Further, under MAR, $\mathbf{m}_i'(r_i - \pi_i)$ is a function only of the observed data $(r_i, y_i, \mathbf{x}_i)$, so that its conditional expectation given $(r_i, y_i, \mathbf{x}_i)$ is $\mathbf{m}_i'(r_i - \pi_i)$.

Combining (10) and (11), we get

$$\mathbf{u}^*(\gamma) = \sum_{i=1}^{n} \mathbf{u}_i^*(\gamma) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\ \mathbf{u}_{2i}^*(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \mathbf{u}_{3i}^*(\boldsymbol{\omega}) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} r_i\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + (1 - r_i) \\ \times E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \\ r_i\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + (1 - r_i) \\ \times E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)] \\ \mathbf{m}_i'(r_i - \pi_i) \end{bmatrix}. \quad (13)$$

For discrete z_i, Ibrahim (1990) and Lipsitz and Ibrahim (1996) have proposed the EM algorithm by the method of weights to solve $\mathbf{u}^*(\hat{\gamma}) = 0$; for continuous z_i, Ibrahim, Chen, and Lipsitz (1999) have proposed a Monte Carlo EM algorithm to solve $\mathbf{u}^*(\hat{\gamma}) = 0$.

Because $\mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha})$ and $\mathbf{u}_{2i}^*(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are functions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ but not of $\boldsymbol{\omega}$, and $\mathbf{u}_{3i}(\boldsymbol{\omega})$ is a function of $\boldsymbol{\omega}$ but not of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the asymptotic properties of the MLE $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ do not depend on whether the model for π_i is correctly specified. Further, estimation of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ can be done separately from the estimation of $\boldsymbol{\omega}$, which can be done using an ordinary logistic regression package.

If the distributions for $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$ are correctly specified, using method-of-moment ideas, then $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is consistent because $E[\mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha})] = 0$ and $E[\mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha})] = 0$, and we are solving $\mathbf{u}^*(\hat{\gamma}) = 0$ for $\hat{\gamma}$. In particular, consider $E[\mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha})]$:

$$\begin{aligned} E[\mathbf{u}_{1i}^*(\boldsymbol{\beta}, \boldsymbol{\alpha})] &= E\{r_i E_{z_i|y_i,\mathbf{x}_i,r_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\ &\quad + E\{(1 - r_i)E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\ &= E\{r_i E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\ &\quad + E\{(1 - r_i)E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\ &= E\{r_i E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \\ &\qquad + (1 - r_i)E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\ &= E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = 0. \end{aligned} \quad (14)$$

Because of (12), the conditional expectation of z_i given $(y_i, \mathbf{x}_i, r_i)$ does not depend on r_i in the first line of (14). The last line in (14) equals 0, because the usual score vector $\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)$ has mean 0. Note that if $p(y_i|\mathbf{x}_i, z_i)$ and/or $p(z_i|\mathbf{x}_i)$ are misspecified, then $\hat{\boldsymbol{\beta}}$ will not be consistent.

4. WEIGHTED ESTIMATING EQUATIONS

Robins et al. (1994) and Zhao et al. (1996) have described general forms for WEEs, but never made a direct connection between ML methods and WEEs. We now give WEEs whose form is almost identical to the ML score equations in the previous section. Suppose that in the parts of the score vector in (13) corresponding to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ we replace r_i with

r_i/π_i. Then we obtain the WEEs $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$, where

$$\mathbf{S}(\gamma) = \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ \mathbf{S}_2(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ \mathbf{S}_3(\boldsymbol{\omega}) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{S}_{1i}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ \mathbf{S}_{2i}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ \mathbf{S}_{3i}(\boldsymbol{\omega}) \end{bmatrix}$$

$$= \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\pi_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times\, E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \\ \frac{r_i}{\pi_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times\, E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)] \\ \mathbf{m}_i'(r_i - \pi_i) \end{bmatrix}. \tag{15}$$

Because of the similarity of (15) to the ML score vector in (13), we can use an EM type algorithm to solve $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$. In Section 5.1 we describe the EM-type algorithm for solving $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$ when z_i is discrete; in Section 5.2 we describe the Monte Carlo EM-type algorithm for solving $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$ when z_i is continuous.

Again, we are interested in obtaining a consistent estimate of $\boldsymbol{\beta}$, with $(\boldsymbol{\alpha},\boldsymbol{\omega})$ treated as nuisance parameters. Similar to Robins and Ritov (1997), we give the result that to obtain a consistent estimate of $\boldsymbol{\beta}$, either, but not both, π_i or $p(z_i|\mathbf{x}_i)$ must be correctly specified. Robins et al. (1994) and Zhao et al. (1996) only gave the result that π_i must be correctly specified in their WEEs. Further, we show that both the ML and the WEEs are special cases of a general class of estimating equations that give a consistent estimate of $\boldsymbol{\beta}$ as long as $p(z_i|\mathbf{x}_i)$ is correctly specified.

Using a first-order Taylor series expansion, it can be shown that

$$N^{1/2}(\hat{\gamma} - \gamma) \approx N\left\{-\frac{dE[\mathbf{S}(\gamma)]}{d\gamma}\right\}^{-1} N^{-1/2}\mathbf{S}(\gamma).$$

This implies that

$$\begin{aligned} N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx\ & N\mathcal{I}^{11}N^{-1/2}\mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ & + N\mathcal{I}^{12}N^{-1/2}\mathbf{S}_2(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}) \\ & + N\mathcal{I}^{13}N^{-1/2}\mathbf{S}_3(\boldsymbol{\omega}), \end{aligned} \tag{16}$$

where $\mathcal{I}^{11}, \mathcal{I}^{12}$, and $\mathcal{I}^{13}$ are the appropriate submatrices of $\{-dE[\mathbf{S}(\gamma)]/d\gamma\}^{-1}$. If one can show that all three terms on the right side of (16) have expectation 0, then $\hat{\boldsymbol{\beta}}$ will be asymptotically unbiased. In the next two sections we show that this is true when either π_i or $p(z_i|\mathbf{x}_i)$ is correctly specified.

4.1 π_i Correctly Specified

Suppose that the model for π_i and the distribution of $(y_i|\mathbf{x}_i, z_i)$ are correctly specified, but the distribution for $(z_i|\mathbf{x}_i)$ is not correctly specified. Then, in the WEEs in (15), we rewrite

$$E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$$

and

$$E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$$

as

$$E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$$

and

$$E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)],$$

where the superscript "*" represents the expectation taken over the wrongly specified distribution for $(z_i|\mathbf{x}_i)$. Here we show that as long as the model for π_i and the distribution of $(y_i|\mathbf{x}_i, z_i)$ are correctly specified, even if $E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$ and $E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$ are not the true expectations, $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased. In particular, we show that $E[\mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega})] = 0, E[\mathbf{S}_3(\boldsymbol{\omega})] = 0$, and $\mathcal{I}^{12} = 0$, implying that each term on the right side of (16) has 0 expectation and that $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased.

When the model for $\pi_i = E[r_i|z_i, y_i, \mathbf{x}_i]$ is correctly specified, we have

$$E[\mathbf{S}_{3i}(\boldsymbol{\omega})] = E[\mathbf{m}_i'(r_i - \pi_i)] = 0,$$

so that the third term on the right side of (16) has expectation equal to 0.

If the model for π_i and the distribution of $(y_i|\mathbf{x}_i, z_i)$ are correctly specified, then the first term of $\mathbf{S}_{1i}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega})$ in (15) has expectation equal to 0; that is,

$$\begin{aligned} & E\left[\frac{r_i}{\pi_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)\right] \\ & = E\left[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) E_{r_i|z_i,y_i,\mathbf{x}_i}\left(\frac{r_i}{\pi_i}\right)\right] \\ & = E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) \cdot 1] \\ & = 0, \end{aligned}$$

because if π_i is correctly specified, then

$$E_{r_i|z_i,y_i,\mathbf{x}_i}\left(\frac{r_i}{\pi_i}\right) = 1,$$

and if the distribution of $(y_i|\mathbf{x}_i, z_i)$ is correctly specified, then the usual score has mean 0; that is,

$$E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = 0. \tag{17}$$

Now the second term of $S_{1i}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega}))$ has expectation equal to

$$\begin{aligned} & E\left[\left(1 - \frac{r_i}{\pi_i}\right) E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right] \\ & = E\left[E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] E_{r_i|z_i,y_i,\mathbf{x}_i}\left(1 - \frac{r_i}{\pi_i}\right)\right] \\ & = E[E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \cdot 0] \\ & = 0, \end{aligned}$$

regardless of whether $E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$ equals the true expectation $E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$. Then $E[\mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\omega})] = 0$, and the first term on the right side of (16) has expectation equal to 0.

Now if the distribution of $(z_i|\mathbf{x}_i)$ is incorrectly specified as $p^*(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$, then

$$E[\mathbf{u}_{2i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = E\left[\frac{\partial \log p^*(z_i|\mathbf{x}_i, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right] \neq 0,$$

which implies that

$$E[\mathbf{S}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] \neq 0.$$

However, the second term on the right side of (16) still has expectation equal to 0. Using the theory of partitioned matrices, it can be shown that $\mathcal{I}^{12} = 0$ if

$$E\left[\frac{d\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})}{d\boldsymbol{\alpha}}\right] = 0.$$

Now

$$\begin{aligned}
E&\left[\frac{d\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})}{d\boldsymbol{\alpha}}\right] \\
&= E\left[\left(\frac{r_i}{\pi_i}\right)\frac{d\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)}{d\boldsymbol{\alpha}} + \left(1 - \frac{r_i}{\pi_i}\right) \right. \\
&\qquad \left. \times \frac{dE^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]}{d\boldsymbol{\alpha}}\right] \\
&= E\left[\left(\frac{r_i}{\pi_i}\right)\cdot 0 + \left(1 - \frac{r_i}{\pi_i}\right) \right. \\
&\qquad \left. \times \frac{dE^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]}{d\boldsymbol{\alpha}}\right] \\
&= E\left[\left(\frac{dE^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]}{d\boldsymbol{\alpha}}\right) \right. \\
&\qquad \left. \times E_{r_i|z_i,y_i,\mathbf{x}_i}\left(1 - \frac{r_i}{\pi_i}\right)\right] \\
&= E\left[\frac{dE^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]}{d\boldsymbol{\alpha}}\cdot 0\right] \\
&= 0. \qquad (18)
\end{aligned}$$

Thus the second term on the right side of (16) has expectation equal to 0, so that all three terms on the right side of (16) have expectation equal to 0 and $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased. Note that if π_i is correctly specified, as long as $\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)$ in (15) has mean 0, even if it is not the first derivative of the log-likelihood of $(y_i|\mathbf{x}_i, z_i)$ (it could, for example, be a quasi-likelihood estimating equation), then the WEEs $\mathbf{S}(\hat{\boldsymbol{\gamma}}_{\text{WEE}}) = 0$ will give a consistent estimate of $\boldsymbol{\beta}$.

4.2 $p(z_i|\mathbf{x}_i)$ Correctly Specified

Suppose that the distributions of $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$ are correctly specified but π_i is not correctly specified. In particular, suppose that

$$E^*_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$$

but π_i is misspecified as π_i^*, which is still only a function of $\mathbf{x}_i$ and y_i, say $\pi_i^* = \pi_i^*(\mathbf{x}_i, y_i)$. For example, in modeling π_i, one might leave out interaction terms between y_i and some of the elements of $\mathbf{x}_i$. We show that $E[\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] = 0, E[\mathbf{S}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] = 0$, and $\mathcal{I}^{13} = 0$, implying that each term on the right side of (16) has 0 expectation and $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased.

If the distributions of $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$ are correctly specified, then

$$\begin{aligned}
E&[\mathbf{S}_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] \\
&= E\left\{\frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)} E_{z_i|y_i,\mathbf{x}_i,r_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right\} \\
&\quad + E\left\{\left(1 - \frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)}\right) E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right\} \\
&= E\left\{\frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)} E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right\} \\
&\quad + E\left\{\left(1 - \frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)}\right) E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right\} \\
&= E\left\{\frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)} E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \right. \\
&\qquad \left. + \left(1 - \frac{r_i}{\pi_i^*(y_i, \mathbf{x}_i)}\right) E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\right\} \\
&= E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = 0. \qquad (19)
\end{aligned}$$

Just as in (14), because we showed that $p(z_i|y_i, \mathbf{x}_i, r_i) = p(z_i|y_i, \mathbf{x}_i)$ in (12), the conditional expectation of z_i given $(y_i, \mathbf{x}_i, r_i)$ does not depend on r_i in the first line of (19). Thus, because $E[\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] = 0$, the first term on the right side of (16) has expectation equal to 0. It can be similarly shown that $E[\mathbf{S}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})] = 0$, and thus the second term on the right side of (16) has expectation equal to 0.

Now if π_i is incorrectly specified as π_i^*, then $E[\mathbf{S}_3(\boldsymbol{\omega})] = E[\mathbf{m}_i'(r_i - \pi_i^*)] \neq 0$. However, the third term on the right side still has expectation equal to 0. Using the theory of partitioned matrices, it can be shown that $\mathcal{I}^{13} = 0$ if

$$E\left[\frac{d\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})}{d\boldsymbol{\omega}}\right] = 0.$$

Suppose that we let

$$\Delta_i = \frac{d}{d\boldsymbol{\omega}}\left(\frac{r_i}{\pi_i^*}\right).$$

Then

$$\begin{aligned}
E&\left[\frac{d\mathbf{S}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\omega})}{d\boldsymbol{\omega}}\right] \\
&= E\{\Delta_i E_{z_i|y_i,\mathbf{x}_i,r_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\
&\quad - E\{\Delta_i E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\
&= E\{\Delta_i E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\
&\quad - E\{\Delta_i E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]\} \\
&= 0. \qquad (20)
\end{aligned}$$

Then the third term on the right side of (16) has expectation equal to 0, as do the first two terms, and $\hat{\beta}$ is asymptotically unbiased. Here, if π_i is misspecified, then the distributions of both $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$ must be correctly specified for $\hat{\beta}$ to be asymptotically unbiased.

In general, if we let $h(r_i, y_i, \mathbf{x}_i)$ be any function of $(r_i, y_i, \mathbf{x}_i)$ that is finite with probability 1, then we can define the estimating equations

$$Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \begin{bmatrix} h(r_i, y_i, \mathbf{x}_i)\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) \\ + [1 - h(r_i, y_i, \mathbf{x}_i)] \\ \times E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] \\ h(r_i, y_i, \mathbf{x}_i)\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) \\ + [1 - h(r_i, y_i, \mathbf{x}_i)] \\ \times E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)] \end{bmatrix}. \quad (21)$$

Using similar logic as for (19), if the distributions of $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$ are correctly specified, then $E[Q(\boldsymbol{\alpha}, \boldsymbol{\beta})] = 0$, and using method-of-moments ideas, the solution to $Q(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$ will produce a consistent estimate of $\boldsymbol{\beta}$. The MLE is a special case of (21) in which $h(r_i, y_i, \mathbf{x}_i) = r_i$, and our WEE is a special case of (21) in which $h(r_i, y_i, \mathbf{x}_i) = r_i/\pi_i$. In general, if $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ is correctly specified, then the asymptotically efficient estimate is the MLE, with $h(r_i, y_i, \mathbf{x}_i) = r_i$. One loses some efficiency by setting $h(r_i, y_i, \mathbf{x}_i) = r_i/\pi_i$, but when $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ is misspecified, this loss of efficiency can be offset by the reduction in bias obtained by correctly specifying π_i. In the asymptotic efficiency calculations in Section 7, the WEE appears to be at least 90% efficient when both π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ are correctly specified.

5. EM-TYPE ALGORITHM

5.1 Categorical z_i

In the score equations (15), when z_i is missing, we see that we need to calculate $E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)]$ and $E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)]$. If z_i is discrete with J levels, then we can write these two conditional expectations as

$$E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = \sum_{z} \mathrm{pr}[z_i = z|y_i, \mathbf{x}_i]\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z)$$

and

$$E_{z_i|y_i,\mathbf{x}_i}[\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)] = \sum_{z} \mathrm{pr}[z_i = z|y_i, \mathbf{x}_i]\mathbf{u}_{2i}(\boldsymbol{\alpha}; z, \mathbf{x}_i),$$

where the sum extends over all J possible values of z_i. Thus for discrete z_i, the score vector in (15) becomes

$$\mathbf{S}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\pi_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times \sum_z w_{iz}\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z) \\ \frac{r_i}{\pi_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times \sum_z w_{iz}\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) \\ m_i'(r_i - \pi_i) \end{bmatrix}, \quad (22)$$

with "weights"

$$w_{iz} = \mathrm{pr}[z_i = z|y_i, \mathbf{x}_i] = \frac{p(y_i|\mathbf{x}_i, z, \boldsymbol{\gamma})p(z|\mathbf{x}_i, \boldsymbol{\gamma})}{\sum_z p(y_i|\mathbf{x}_i, z, \boldsymbol{\gamma})p(z|\mathbf{x}_i, \boldsymbol{\gamma})} \quad (23)$$

equal to the conditional probability $z_i = z$, given the observed data $(y_i, \mathbf{x}_i)$. We can think of (22) as a weighted score vector. In particular, the components of the score vector corresponding to $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ in (22) take the form of weighted complete-data score vectors based on $N = \sum_{i=1}^{n}[r_i + J]$ observations. If z_i is missing ($r_i = 0$), then subject i contributes the sum of J weighted complete-data score vectors to (22), with weights $(1 - r_i/\pi_i)w_{iz} = w_{iz}$. If z_i is observed, then subject i contributes the usual complete-data score vector weighted by π_i^{-1}, plus J weighted complete-data score vectors, with weights $(1 - \pi_i^{-1})w_{iz}$. The WEEs are $\mathbf{S}(\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}) = 0$. Because the estimating equation for π_i does not involve $\boldsymbol{\beta}$ or $\boldsymbol{\alpha}$, just as when using ML, $\boldsymbol{\omega}$ can be obtained separately from $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. In fact, $\hat{\boldsymbol{\omega}}_{\mathrm{WEE}} = \hat{\boldsymbol{\omega}}$ is the MLE obtained via ordinary logistic regression discussed earlier.

Then, given the MLE of π_i, say $\hat{\pi}_i$, we can use an EM-type algorithm to obtain the estimates of $\hat{\boldsymbol{\beta}}_{\mathrm{WEE}}$ and $\hat{\boldsymbol{\alpha}}_{\mathrm{WEE}}$. Suppose that we define the function

$$\mathbf{S}(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)}) = \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right) \\ \times \sum_z w_{iz}^{(t)}\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z) \\ \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right) \\ \times \sum_z w_{iz}^{(t)}\mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) \\ \mathbf{m}_i'(r_i - \hat{\pi}_i) \end{bmatrix},$$

where $w_{iz}^{(t)}$ is the conditional probability in (23) evaluated at the given value $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(t)}$. The EM-type iterative algorithm to solve $\mathbf{S}(\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}) = 0$ is as follows:

1. Obtain an initial estimate $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(1)}$, say, by complete cases. At the tth step, we have $\boldsymbol{\gamma}^{(t)}$.
2. Using $\boldsymbol{\gamma}^{(t)}$, calculate $w_{iz}^{(t)} = w_{iz}(\boldsymbol{\gamma}^{(t)})$.
3. Treating $w_{iz}^{(t)}$ as fixed, solve $\mathbf{S}(\boldsymbol{\gamma}^{(t+1)}|\boldsymbol{\gamma}^{(t)}) = 0$ for $\boldsymbol{\gamma}^{(t+1)}$, using any generalized linear model program that allows for weights. The weights will be $\hat{\pi}_i^{-1}$, $(1 - \hat{\pi}_i^{-1})w_{iz}^{(t)}$, or $w_{iz}^{(t)}$.
4. Iterate until convergence, $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} = \hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}$ which gives the solution to $\mathbf{S}(\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}) = 0$.

The asymptotic variance of $\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}$ has a "sandwich" form (White 1982),

$$\mathrm{Var}(\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}) = \left\{\sum_{i=1}^{n} E[\dot{\mathbf{S}}_i(\boldsymbol{\gamma})]\right\}^{-1} \sum_{i=1}^{n} E[\mathbf{S}_i(\boldsymbol{\gamma})\mathbf{S}_i(\boldsymbol{\gamma})'] \times \left\{\sum_{i=1}^{n} E[\dot{\mathbf{S}}_i(\boldsymbol{\gamma})]'\right\}^{-1}, \quad (24)$$

where

$$\dot{\mathbf{S}}_i(\boldsymbol{\gamma}) = \frac{\partial \mathbf{S}_i(\boldsymbol{\gamma})'}{\partial \boldsymbol{\gamma}}.$$

We consistently estimate (24) by replacing $\boldsymbol{\gamma}$ with $\hat{\boldsymbol{\gamma}}_{\mathrm{WEE}}$. The upper $p \times p$ block of this estimate is consistent for the asymptotic variance of $\hat{\beta}_{\text{WEE}}$. When z_i is continuous, all of the theory and methods presented in this section can still be used. We use a Monte Carlo modification of the foregoing EM-type algorithm, as discussed in the following section.

5.2 Continuous z_i

When z_i is continuous, the WEEs are

$$\mathbf{S}(\gamma) = \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\pi_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times \int_z w_{iz} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z)\, dz \\ \frac{r_i}{\pi_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\pi_i}\right) \\ \times \int_z w_{iz} \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i)\, dz \\ \mathbf{m}_i'(r_i - \pi_i) \end{bmatrix},$$

with conditional probability

$$w_{iz} = p(z|y_i, \mathbf{x}_i, \gamma) = \frac{p(y_i|\mathbf{x}_i, z, \gamma)p(z|\mathbf{x}_i, \gamma)}{\int_z p(y_i|\mathbf{x}_i, z, \gamma)p(z|\mathbf{x}_i, \gamma)}.$$

Rather than use numerical integration to solve $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$, we use a Monte Carlo method. To obtain the solution $\mathbf{S}(\hat{\gamma}_{\text{WEE}}) = 0$, we again define the function

$$\mathbf{S}(\gamma|\gamma^{(t)}) = \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right) \\ \times \int_z w_{iz}^{(t)} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z)\, dz \\ \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right) \\ \times \int_z w_{iz}^{(t)} \mathbf{u}_{2i}(\boldsymbol{\alpha}; z, \mathbf{x}_i)\, dz \\ \mathbf{m}_i'(r_i - \hat{\pi}_i) \end{bmatrix}, \quad (25)$$

where

$$w_{iz}^{(t)} = p(z|y_i, \mathbf{x}_i, \gamma^{(t)}).$$

In iterations of our proposed Monte Carlo algorithm, rather than solve $\mathbf{S}(\hat{\gamma}^{(t+1)}|\gamma^{(t)}) = 0$, we now solve a Monte Carlo approximation $\mathbf{S}^*(\gamma^{(t+1)}|\gamma^{(t)}) = 0$. The Monte Carlo algorithm for this problem can be summarized as follows:

1. Obtain an initial estimate of $\gamma = \gamma^{(1)}$, say, by complete cases. At the tth step, we have $\gamma^{(t)}$.
2. Sample z from the conditional density $w_{iz}^{(t)} = p(z|y_i, \mathbf{x}_i, \gamma^{(t)})$ using the adaptive rejection algorithm of Gilks and Wild (1992). Repeat this L times, with the lth draw of z_i denoted by $z_i^{l(t)}$.
3. Solve

$$\mathbf{S}^*(\gamma^{(t+1)}|\gamma^{(t)}) = \sum_{i=1}^{n} \begin{bmatrix} \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right)\frac{1}{L} \\ \times \sum_{l=1}^{L} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i^{l(t)}) \\ \frac{r_i}{\hat{\pi}_i}\, \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i, \mathbf{x}_i) + \left(1 - \frac{r_i}{\hat{\pi}_i}\right)\frac{1}{L} \\ \times \sum_{l=1}^{L} \mathbf{u}_{2i}(\boldsymbol{\alpha}; z_i^{l(t)}, \mathbf{x}_i) \\ \mathbf{m}_i'(r_i - \hat{\pi}_i) \end{bmatrix} = 0 \quad (26)$$

for $\gamma^{(t+1)}$, using any generalized linear model program that allows weights. The weights will be $\hat{\pi}_i^{-1}, (1 - \hat{\pi}_i^{-1})L^{-1}$, or L^{-1}.

4. Iterate until convergence to obtain $\gamma^{(t+1)} = \gamma^{(t)} = \hat{\gamma}_{\text{WEE}}$.

Note that in (26) we have approximated $\int_z w_{iz}^{(t)} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z)\, dz$ with

$$\frac{1}{L}\sum_{l=1}^{L} \mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i^{l(t)}).$$

Given the $z_i^{l(T)}$'s from the last (Tth) iteration of the Monte Carlo algorithm, computation of the asymptotic covariance matrix is straightforward using (24) evaluated at $\hat{\gamma}$.

6. EXAMPLES

6.1 Multiple Myeloma

To illustrate our proposed approach, we consider data on a subset of $n = 224$ patients from Eastern Cooperative Oncology Group (ECOG) clinical trial E2479 (Kalish 1992). The main purpose of the trial was to evaluate whether vincristine, BCNU, melphalan, cyclophosphamide, and prednosone (VBMCP) should replace Melphalan plus prednosone (MP) as a standard therapy for patients with previously untreated multiple myeloma. We are primarily interested in how treatment affects survival time, the time of entry into the study until death; we are also interested in how survival is predicted by seven other baseline characteristics. The other seven covariates of interest are bone fractures at diagnosis, 1 = yes, 0 = no (FRAC); logarithm of the blood urea nitrogen (LOGBUN); hemoglobin (HGB); platelet count (PLATELET); logarithm of the white blood cell count (LOGWBC); logarithm of plasma cells in bone marrow (LOGPBM); and serum calcium (SCALC). Patients with high values of LOGBUN, HGB, LOGPBM, and SCALC and low values of PLATELET and LOGWBC are expected to have shorter survival. The FRAC covariate (z) is missing for 84 (37.5%) of the 224 patients; all other covariates are completely observed. We wish to demonstrate strategies that can be used for modeling $(z_i|\mathbf{x}_i)$ and $(r_i|y_i, \mathbf{x}_i)$. Recall when using ML, we do not need to estimate π_i to obtain consistent estimates of $\boldsymbol{\beta}$. We do, however, need to correctly model the densities $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$. For the WEEs, to obtain consistent estimates of $\boldsymbol{\beta}$, we need to correctly model π_i and also specify $\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)$ such that $E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = 0$, or we can misspecify π_i but need to correctly model the densities $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$.

For illustration, we assume that the survival time is exponentially distributed. Let T_i be the true failure time for subject i. We note here that 10 of the 224 (2.4%) cases have a censored survival time. If we let U_i instead of T_i be the censoring time, then we observe $Y_i = \min(T_i, U_i)$ and the censoring indicator $\delta_i = I\{T_i \le U_i\}$. Under noninformative censoring, the density of $(y_i, \delta_i|\mathbf{x}_i, z_i)$ is

$$p(y_i, \delta_i|\mathbf{x}_i, z_i, \boldsymbol{\beta}) \propto e^{-\lambda_i y_i}\lambda_i^{\delta_i},$$

where

$$\lambda_i = \exp(\beta_0 + \boldsymbol{\beta}_1' \mathbf{x}_i + \beta_2 z_i) \tag{27}$$

and $\mathbf{x}_i$ contains LOGBUN, HGB, PLATELET, LOGWBC, LOGPBM, and SCAL. Because z_i (FRAC) can be missing, to use ML we need to specify the density $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$. Because z_i is binary, given $\mathbf{x}_i$, it follows a Bernoulli distribution with probability of success that we specify with a logistic regression

$$\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha}) = \frac{\exp(\alpha_0 + \boldsymbol{\alpha}_1' g(\mathbf{x}_i))}{1 + \exp(\alpha_0 + \boldsymbol{\alpha}_1' g(\mathbf{x}_i))},$$

where $\mathrm{g}(\mathbf{x}_i)$ is some vector function of the elements of $\mathrm{g}(\mathbf{x}_i)$ (which could include interactions between the elements of $\mathbf{x}_i$). Note that the estimate of $\boldsymbol{\alpha}$ is not of interest in itself, but we need to fit the density for $(z_i|\mathbf{x}_i)$ close to the truth to minimize the bias in our estimate of the β. As such, in finding the best-fitting model for $\Pr(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$, we do not worry about a high type I error rate; we would rather overspecify than underspecify, as underspecification could bias the estimates of β. Thus we suggest keeping any parameter in the model that is significant at the .20 level of significance. To determine a suitable model for $(z_i|\mathbf{x}_i)$, one could fix the model of interest for $[y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta}]$ and do variable selection on $(z_i|\mathbf{x}_i)$ using the full likelihood for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. One could then use the likelihood ratio or Akaike information criterion (AIC) to evaluate the fit of each model. We considered the main effects model in x_i as the baseline model for $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$ and used a step-up logistic regression approach from the main effects (keeping all effects that were significant at the .20 level). The best-fitting model for $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$ contained pairwise interactions between LOGBUN and TREATMENT, LOGBUN and LOGPBM, and SCALC and TREATMENT. The results are shown in Table 1.

To use the WEEs, we must also specify a model for $\pi_i = \Pr(r_i = 1|y_i, \mathbf{x}_i)$. We again propose a logistic regression, as in (1). To protect ourselves from bias in β caused from misspecifying $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$, when using WEEs, we need to fit π_i well. As such, in finding the best-fitting model for π_i again, we do not worry about a high type I error rate; we would rather overspecify than underspecify, as underspecification could bias the estimates of β. Thus we suggest keeping any parameter in the model for π_i that is significant at .20 level of significance. However, we can find the best fitting model for π_i using ordinary logistic regression, without having to worry about the specification of the model for $[y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta}]$. We considered the main effects model in $\mathbf{x}_i$ and log censoring time as the baseline model for π_i, and used a step-up logistic regression approach from the main effects (keeping all effects in that were significant at the .20 level). The best-fitting model for π_i contained pairwise interactions between log(survival) and SCALC, LOGBUN and HGB, and HGB and SCALC. The results are shown in Table 2.

Table 3 gives estimates and standard errors for the regression coefficients based on ML, WEE, and complete-case estimation. From Table 3, we see that the ML and WEE estimates are similar, although HGB is significant at the 5% level using WEE but significant at the 10% level using

Table 1. EM Parameter Estimates for the Model for $Pr(z_i = 1|\mathbf{x}_i)$ From the Myeloma Data

Parameter	Estimate	Standard error	Z value	p value
INTERCEPT	−14.072	7.734	−1.82	.069
LOGBUN	8.446	3.365	2.51	.012
HGB	−.025	.059	−.42	.674
PLATELET	−.387	1.164	−.33	.739
LOGWBC	−.309	.280	−1.10	.270
LOGPBM	4.103	1.957	2.10	.036
SCALC	−.203	.117	−1.73	.083
TREATMENT	.546	2.556	.21	.831
LOGBUN*TREATM	−1.540	1.061	−1.45	.147
LOGBUN*LOGPBM	−1.911	.859	−2.23	.026
SCALC*TREATM	.285	.162	1.76	.078

Table 2. MLEs for the Model for $Pr(r_i = 1|\mathbf{x}_i, y_i)$ From the Myeloma Data

Parameter	Estimate	Standard error	Z value	p value
INTERCEPT	2.615	7.050	.37	.711
LOGSURV	−1.776	.657	−2.70	.007
LOGBUN	−3.772	2.518	−1.50	.134
HGB	.260	.677	.38	.701
PLATELET	−.550	.680	−.81	.417
LOGWBC	.295	.318	.93	.353
LOGPBM	−.089	.188	−.48	.634
SCALC	.651	.460	1.41	.158
TRT	−.107	.252	−.42	.673
LOGSURV*SCALC	.165	.064	2.59	.010
LOGBUN*HGB	.367	.259	1.42	.156
HGB*SCALC	−.103	.047	−2.17	.030

Table 3. Regression Parameter (β) Estimates for the Myeloma Data

Effect	Approach	$\hat{\beta}$	Standard error	Z statistic	p value
INTERCEPT	EM	−6.101	.838	−7.28	0
	WEE	−6.104	.859	−7.10	0
	CC	−5.952	.947	−6.28	0
FRAC	EM	−.027	.112	−.24	.810
	WEE	−.037	.111	−.33	.742
	CC	−.024	.116	−.21	.834
LOGBUN	EM	.347	.202	1.71	.087
	WEE	.346	.214	1.62	.105
	CC	.273	.232	1.18	.238
HGB	EM	−.039	.024	−1.68	.094
	WEE	−.040	.020	−1.98	.047
	CC	−.036	.026	−1.40	.160
PLATELET	EM	−.153	.140	−1.09	.274
	WEE	−.152	.196	−.78	.435
	CC	−1.400	.600	−2.33	.024
LOGWBC	EM	.075	.135	.56	.578
	WEE	.077	.172	.45	.656
	CC	.208	.156	1.33	.183
LOGPBM	EM	.272	.075	3.60	0
	WEE	.272	.073	3.70	0
	CC	.279	.085	3.30	.001
SCALC	EM	.079	.038	2.06	.039
	WEE	.079	.034	2.33	.020
	CC	.050	.043	1.16	.244
TREATMENT	EM	−.059	.099	−.59	.553
	WEE	−.058	.095	−.61	.542
	CC	−.045	.112	−.41	.685

NOTE: CC denotes complete-case estimation.

ML, and LOGBUN is significant at the 10% level using ML but not using WEE. For a few covariates, we see striking differences when comparing the complete case to ML and WEE. The effect of SCALC is significant at the 5% level using both ML and WEE but is not even significant at the .20 level using complete-case estimation. Further, the PLATELET effect is significant at the 5% level when using complete-case estimation but is not even significant at the .25 level using ML or WEE. We see here that the complete-case analysis can be misleading and can give results that are completely contradictory to ML or WEE analysis. Further, the standard deviations using ML and WEE appear similar, whereas we also seem to loose efficiency when using complete case estimation. ML and WEE give very similar results in this example, but in Section 7 we look at a simulation comparing the approaches and in Section 8 compare the asymptotic bias and asymptotic relative efficiency of ML and WEE.

In this example, when using WEEs, we had to correctly specify the logistic regression model for either π_i or $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$. When using ML, we had to correctly specify a logistic regression model for $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$. Thus, in this example, besides correctly specifying the regression model for y_i, both approaches necessitate correctly specifying a logistic regression model, although the WEEs are a little more robust in that we can misspecify $\Pr(z_i = 1|\mathbf{x}_i, \boldsymbol{\alpha})$ and still get consistent estimates.

We note here that the convergence criterion used for the EM-type algorithm was that the distance between the tth iteration and the $(t+1)$th iteration in each parameter be less than 10^{-6}. The number of iterations required for convergence was 28, with the CC estimates used as starting values. We programmed the algorithm in SAS, and our SAS macro took 3 minutes and 40 seconds in real time to calculate the weighted estimate on a SPARC20 Workstation.

6.2 Liver Data

Although there are some difficulties in determining the precise number of deaths, each year approximately 2500 persons in the United States die of liver cancer. ECOG undertook a randomized phase II clinical trial (Falkson et al. 1990) to evaluate four new treatments in patients with primary liver cancer. The binary outcome (y_i) equals 1 if subject i lived 4 months or longer after entering the study and 0 otherwise. The covariates are treatment and time from diagnosis (in weeks) until entering the study. Because there are four treatments, we form three binary covariates corresponding to treatment, $\mathbf{x}_i = [x_{i1}, x_{i2}, x_{i3}]'$, where x_{ij} equals 1 if subject i was randomized to treatment j and 0 otherwise. All 140 patients have (y_i) and treatment $(\mathbf{x}_i)$ observed, but 27 of the subjects (19.3%) are missing time from diagnosis (z_i). The distribution of y given $(\mathbf{x}_i, z_i)$ is Bernoulli; we model the probability $y_i = 1$ given $(\mathbf{x}_i, z_i)$ as a logistic regression,

$$\operatorname{logit}[\Pr(y_i = 1|\mathbf{x}_i, z_i, \boldsymbol{\beta})] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 z_i.$$

The covariate z_i, the time since diagnosis, is continuous. Given treatment $(\mathbf{x}_i)$, we assume that z_i follows an exponential distribution

$$p(z_i|\mathbf{x}_i, \boldsymbol{\alpha}) = \lambda_i e^{-\lambda_i z_i},$$

where

$$\lambda_i = \exp(\alpha_0 + \boldsymbol{\alpha}_1' \mathbf{x}_i). \tag{28}$$

To use the WEEs, we specify $\Pr(r_i = 1|y_i, \mathbf{x}_i)$ as a logistic regression model,

$$\operatorname{logit}[\Pr(r_i = 1|\mathbf{x}_i, y_i, \boldsymbol{\omega})] = \omega_0 + \omega_1 x_{i1} + \omega_2 x_{i2} + \omega_3 x_{i3} + \omega_4 y_i. \tag{29}$$

We added each pairwise interaction term separately to (29), and none were significant at the 5% level. The estimates for the logistic model in (29) are shown in Table 4. Time from diagnosis is apparently missing completely at random; that is, missingness appears independent of the outcome and treatment. As such, we expect a complete-case analysis to give unbiased but possibly inefficient estimates of $\boldsymbol{\beta}$. This is confirmed in Table 5. The estimates of $\boldsymbol{\beta}$ from the three approaches are very similar. Also, the estimated standard errors are very similar for the WEEs and ML; however, the estimated standard errors from the complete-case analysis are much larger than those from the other two approaches.

The number of Monte Carlo iterations within each iteration of the EM-type algorithm was $L = 10{,}000$. Our convergence criterion was .0001. Thirty-five iterations were required so that the distance between the tth iteration and the $(t+1)$th iteration in each parameter was less than .0001, with the complete-case estimates used as starting values. We ran the algorithm for 100 iterations, and after the 35th iteration, the estimates did bounce around some in the fifth decimal place, but never more than .00002. For any iteration after the 35th, the results were identical to Table 5. To reduce the convergence criterion to .00001, we postulate that we should increase L by a factor of 10; that is, set $L = 100{,}000$. But this was infeasible using our SAS macro, because the SAS datasets created in the macros were too large for our computer, a SPARC20 workstation. We are currently devising a more efficient computer program. The SAS macro took 2 hours in real time to do 35 iterations on the SPARC20 workstation.

Table 4. MLEs for the Model for $\Pr(r_i = 1|\mathbf{x}_i, y_i)$ From the Liver Cancer Data

Parameter	Estimate	Standard error	Z value	p value
INTERCEPT	.852	.509	1.67	.094
Y	.059	.455	.13	.897
TREATMENT 1	.966	.624	1.55	.121
TREATMENT 2	1.166	.669	1.74	.081
TREATMENT 3	.246	.552	.45	.656

7. A SIMULATION STUDY

We performed a simulation study based on the myeloma example discussed in Section 6.1, with response survival

Table 5. Regression Parameter (β) Estimates for the Liver Cancer Data

Effect	Approach	$\hat{\beta}$	Standard error	Z statistic	p value
INTERCEPT	EM	.9632	.4018	2.40	.016
	WEE	.9803	.4015	2.44	.015
	CC	1.3104	.5122	2.56	.011
TIME DIAG	EM	−.0220	.0128	−1.72	.085
	WEE	−.0227	.0131	−1.72	.085
	CC	−.0215	.0145	−1.48	.138
TREATMENT 1	EM	−.1409	.5414	−.26	.795
	WEE	−.1430	.5440	−.26	.793
	CC	−.7093	.6325	−1.12	.262
TREATMENT 2	EM	−.7466	.5265	−1.42	.156
	WEE	−.7528	.5259	−1.43	.152
	CC	−1.0749	.6275	−1.71	.087
TREATMENT 3	EM	−.3185	.5272	−.60	.549
	WEE	−.3057	.5259	−.58	.562
	CC	−.5425	.6497	−.84	.404

NOTE: CC denotes complete-case estimation.

time and covariates TRT (x_{i1}), SCALC (x_{i2}), and FRAC (z_i). For the models described here, we sampled covariate data from the observed complete cases in the myeloma study and generated the responses and the missingness indicators r_i for whether the covariate z_i (FRAC) is observed.

We used the covariate data from the 140 subjects with z_i observed and randomly drew 84 other values for $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, z_i)$ from the 140 complete cases, to increase the total sample size in a simulation to 224 as in the example. These covariate data remained fixed over all simulations. The simulation consisted of 500 replications and compared the WEE approach, ML via the EM algorithm, the simple complete-case analysis, and ML if we observed z_i for all 224 subjects. (Even if $r_i = 0$ in a simulation, we still have the value of z_i.) We ran 500 simulations.

In each simulation the exponential model for survival (y_i) had the hazard

$$\lambda_i = \exp(-3 - x_{i1} + x_{i2} + z_i); \tag{30}$$

that is, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_z) = (-3.8, -1, 1, 1)'$. We then sampled y_i accordingly, without censoring any values. In all simulations the true model for $\Pr(z_i = 1|x_{i1}, x_{i2})$ was

$$\operatorname{logit}\{\Pr(z_i = 1|x_{i1}, x_{i2}, \boldsymbol{\alpha})\} = .5 - .5x_{i1} - .5x_{i2} + .5x_{i1}x_{i2} \tag{31}$$

and the true model for $\Pr(r_i = 1|y_i, x_{i1}, x_{i2})$ was

$$\operatorname{logit}\{\Pr(r_i = 1|y_i, x_{i1}, x_{i2}, \boldsymbol{\omega})\} = .5 - .5\log(y_i) + .5x_{i1} + .5x_{i2} - x_{i1}x_{i2}. \tag{32}$$

On average, about 50% of the z_i values will be missing in any given simulation. When using the EM algorithm and WEEs in the simulations, we specified various models for $\Pr(z_i = 1|x_{i1}, x_{i2})$ and π_i. First, we specified $\Pr(z_i = 1|x_{i1}, x_{i2})$ and π_i correctly when using the EM algorithm and WEEs, denoted by EM(z^+) and WEE(z^+, r^+), where z^+ and r^+ means that both the models for $\Pr(z_i = 1|x_{i1}, x_{i2})$ and π_i are correctly specified. Leaving out an interaction term is the most likely misspecification of $\Pr(z_i = 1|x_{i1}, x_{i2})$ and/or π_i, so we looked at misspecifications of this type. Then, for WEEs we correctly specified $\Pr(z_i = 1|x_{i1}, x_{i2})$ but left out the interaction term in the model π_i, denoted by WEE(z^+, r^-), and left out the interaction term in $\Pr(z_i = 1|x_{i1}, x_{i2})$ and correctly modeled π_i, denoted by WEE(z^-, r^+). Both of these should provide consistent estimates of $\boldsymbol{\beta}$. For WEEs, we also left out the interaction terms in both $\Pr(z_i = 1|x_{i1}, x_{i2})$ and π_i, denoted by WEE(z^-, r^-); the resulting estimate of $\boldsymbol{\beta}$ could be biased. Finally, for the EM algorithm, we left out the interaction term in $\Pr(z_i = 1|x_{i1}, x_{i2})$, denoted by EM($z^-$). In this case the EM algorithm could also give a biased estimate of $\boldsymbol{\beta}$. The results are summarized in Table 6.

Despite the fact that $\Pr(z_i = 1|x_{i1}, x_{i2})$ and/or $\Pr(r_i = 1|y_i, x_{i1}, x_{i2})$ are misspecified in some of the simulations, both WEE and EM appear to give approximately unbiased estimates of $\boldsymbol{\beta}$ for all simulations. Even when $p(z_i|\mathbf{x}_i)$ is misspecified for EM, EM(z^-), and when both $p(z_i|\mathbf{x}_i)$ and π_i are misspecified for WEE, WEE(z^-, r^-), there is only a very small bias in the estimate for β_2. However, the complete-case estimate appears to be heavily biased. In addition, the coverage probabilities are highly accurate when using WEE and EM, but not when using the complete case approach.

Table 7 gives the efficiency for the consistent estimates WEE(z^+, r^+), WEE(z^+, r^-), and WEE(z^-, r^+) versus the asymptotically efficient estimator EM(z^+). We see that correctly specifying $p(z_i|\mathbf{x}_i)$ and π_i to obtain WEE(z^+, r^+) leads to the greatest efficiency. For WEE, correctly specifying $p(z_i|\mathbf{x}_i)$ appears to lead to more efficiency in estimating β_2 than correctly specifying π_i, in that WEE(z^+, r^-) is more efficient than WEE(z^-, r^+). For β_z, it does not appear that there is much difference in specifying one of $p(z_i|\mathbf{x}_i)$ or π_i correctly, in that the efficiency of WEE(z^+, r^-) and WEE(z^-, r^+) is about the same; specifying both correctly does appear to increase the efficiency.

Even though the bias is mostly negligible in these simulations, this is not always the case, as we show in the next section, which looks at asymptotic bias and efficiency.

8. STUDY OF ASYMPTOTIC BIAS AND EFFICIENCY

8.1 Asymptotic Bias

In this section we study the asymptotic bias in estimating $\boldsymbol{\beta}$ using ML and WEE. We formulate the true model by specifying each term on the right side of

$$p(r_i, z_i, y_i, x_{i1}, x_{i2}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \delta) = p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega})p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})p(\mathbf{x}_i|\delta). \tag{33}$$

For simplicity, we let the covariate that is always observed, x_i, be a Bernoulli random variable with $\Pr(x_i = 1) = .5$. We let $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ be a Bernoulli distribution with the logit of the probability of success equal to

$$\operatorname{logit}[\Pr(z_i = 1|\mathbf{x}_i)] = 1 + \alpha_1 x_i. \tag{34}$$

Table 6. Simulation Based on Myeloma Example

	Approach[a]	$\beta_0 = -3$	$\beta_1 = 1$	$\beta_2 = -1$	$\beta_z = 1$
Estimate	All	−3.002	.996	−.995	1.002
	EM(z^+)	−3.004	1.007	−.999	1.000
	WEE(z^+, r^+)	−3.004	.996	−.993	1.006
	WEE(z^+, r^-)	−3.008	1.007	−.997	1.009
	WEE(z^-, r^+)	−2.993	.994	−1.001	1.006
	EM(z^-)	−2.988	1.000	−1.016	1.007
	WEE(z^-, r^-)	−2.997	1.002	−1.012	1.004
	CC	−2.619	.826	−1.059	.945
Bias	All	−.002	−.004	.005	.002
	EM(z^+)	−.004	.007	.001	.000
	WEE(z^+, r^+)	−.004	−.004	.007	.006
	WEE(z^+, r^-)	−.008	.007	.003	.009
	WEE(z^-, r^+)	.007	−.006	−.001	.006
	EM(z^-)	.012[b]	.000	−.016[b]	.007
	WEE(z^-, r^-)	.003	.002	−.012[b]	.004
	CC	.381[b]	−.174[b]	−.059[b]	−.055[b]
Simulation variance	All	.0080	.0086	.0073	.0082
	EM(z^+)	.0110	.0115	.0111	.0124
	WEE(z^+, r^+)	.0110	.0119	.0117	.0154
	WEE(z^+, r^-)	.0110	.0120	.0121	.0166
	WEE(z^-, r^+)	.0110	.0136	.0134	.0164
	EM(z^-)	.0102	.0109	.0096	.0104
	WEE(z^-, r^-)	.0097	.0123	.0127	.0155
	CC	.0265	.0227	.0178	.0188
Average of variances	All	.0083	.0087	.0086	.0081
	EM(z^+)	.0101	.0113	.0106	.0116
	WEE(z^+, r^+)	.0121	.0128	.0121	.0143
	WEE(z^+, r^-)	.0118	.0139	.0129	.0140
	WEE(z^-, r^+)	.0115	.0131	.0124	.0141
	EM(z^-)	.0105	.0116	.0107	.0115
	WEE(z^-, r^-)	.0100	.0133	.0130	.0140
	CC	.0199	.0180	.0156	.0148
Coverage probability	All	94.8	94.4	95.2	94.4
	EM(z^+)	94.2	95.0	96.0	94.8
	WEE(z^+, r^+)	95.0	94.6	93.4	94.2
	WEE(z^+, r^-)	93.2	97.2	95.4	93.2
	WEE(z^-, r^+)	94.4	94.2	94.0	93.8
	EM(z^-)	94.4	95.2	95.0	97.0
	WEE(z^-, r^-)	95.4	95.8	94.2	93.0
	CC	25.2	71.4	89.8	88.6

[a] EM(z^-), WEE(z^-, r^-), and complete-case (CC) are not consistent; all others are.
[b] $p < .05$ when testing if bias = 0 using $z = \sqrt{500}$ bias/SE$_{sim}$.

where we let α_1 vary from −3 to 3. We let $p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})$ be a Bernoulli distribution with the logit of the probability of success equal to

$$\text{logit}[\Pr(y_i = 1|\mathbf{x}_i, z_i)] = \beta_0 + \beta_x x_i + \beta_z z_i = 1 - x_i + z_i. \quad (35)$$

Finally, we let $p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega})$ be a Bernoulli distribution with the logit of the probability of success equal to

$$\begin{aligned}\text{logit}[\pi_i] &= \omega_0 + \omega_x x_i + \omega_y y_i + \omega_{xy} x_i y_i \\ &= 1 + x_i - y_i + \omega_{xy} x_i y_i, \quad (36)\end{aligned}$$

where we let ω_{xy} vary from −3 to 3.

Table 7. Simulation Efficiency (%)

Approach	β_0	β_1	β_2	β_z
WEE(z^+, r^+)	100.0	96.6	94.9	80.5
WEE(z^+, r^-)	100.0	95.8	91.7	74.7
WEE(z^-, r^+)	100.0	84.6	82.8	75.6

Letting $\hat{\boldsymbol{\beta}}$ denote an estimate of $\boldsymbol{\beta}$ from a given approach, then $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^*$ is not necessarily equal to $\boldsymbol{\beta}$. Here we are primarily interested in assessing the asymptotic bias, $(\boldsymbol{\beta}^* - \boldsymbol{\beta})$. Following Rotnitzky and Wypij (1994), because all random variables are discrete, the asymptotic bias can be ascertained by simply considering an artificial sample comprised of one suitably weighted observation for each possible realization of $(r_i, z_i, y_i, z_i, \mathbf{x}_i)$. Then we can solve for $\boldsymbol{\beta}^*$ in the usual way, except that each individual's contribution to the ML estimating equations or WEEs is weighted by its respective probability. Because the WEE is asymptotically biased only if both π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ are misspecified, we look at the bias of WEE when we wrongly set $\omega_{xy} = 0$ in estimating (36) and wrongly set $\alpha_1 = 0$ in estimating (34); we denote this estimate by WEE(z^-, r^-). Similarly, because EM will give a biased estimate when we misspecify $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$, we look at the bias of EM when we set $\alpha_1 = 0$ in estimating (34); we denote this estimate by EM(z^-).

Letting $\omega_{xy} = -1.5$ in π_i, Figure 1 plots the relative bias in estimating $\beta_x = -1$ and $\beta_z = 1$ as a function of α_1 as α_1

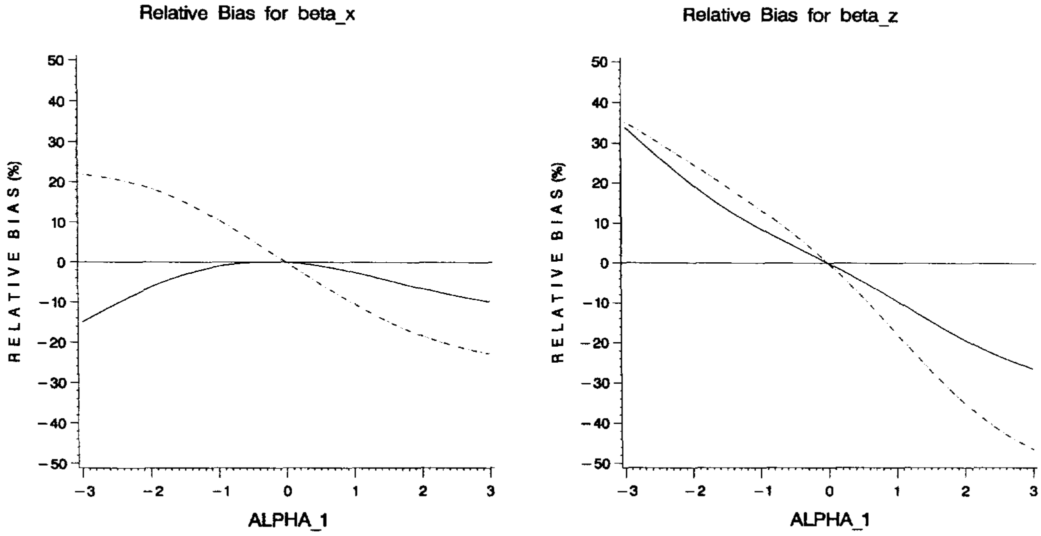

Figure 1. Plot of the Relative Asymptotic Bias in Estimating $\beta_x = -1$ (a) and $\beta_z = 1$ (b) as a Function of α_1 as α_1 Varies from −3 to 3. ———, WEE(z^-, r^-); − · − · − ·, EM(z^-).

varies from −3 to 3. As a function of α_1, we see that the absolute value of the relative bias is uniformly smaller for WEE(z^-, r^-) than for EM(z^-) for estimating both β_x and β_z. As α_1 gets further from 0 (i.e., the strength of association between z_i and x_i gets stronger), the relative bias gets larger; this makes sense because both WEE(z^-, r^-) and EM(z^-) wrongly set $\alpha_1 = 0$. The relative bias is greater for estimating β_z, the coefficient of the possibly missing variable, for both ML and WEE.

Letting $\alpha_1 = -1.5$ in (34), Figure 2 plots the relative bias in estimating β_x and β_z as a function of ω_{xy} as ω_{xy} varies from −3 to 3. As a function of ω_{xy}, we see that the absolute value of the relative bias is much smaller for WEE(z^-, r^-) than for EM(z^-) for estimating β_x. For estimating β_z, the relative bias is smaller using WEE(z^-, r^-), except when ω_{xy} is between −1 and −.4.

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 this asymptotic study. We can make only general suggestions. In our limited asymptotic study, we have seen that the relative bias for an estimate of $\boldsymbol{\beta}$ tends to be smaller using WEE when both π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ are misspecified than with ML with the same misspecification of $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$. In other words, if $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ has the same misspecification for both ML and WEE, then there is a very good chance that WEE will reduce the bias of ML if π_i is modeled approximately correctly.

8.2 Asymptotic Efficiency

In this section we study the asymptotic efficiency in estimating $\boldsymbol{\beta}$ using ML and WEE. The efficiency is calculated as the ratio of asymptotic variances of ML to WEE, where the asymptotic variance of ML as calculated from the inverse of the negative of the expected value of the second

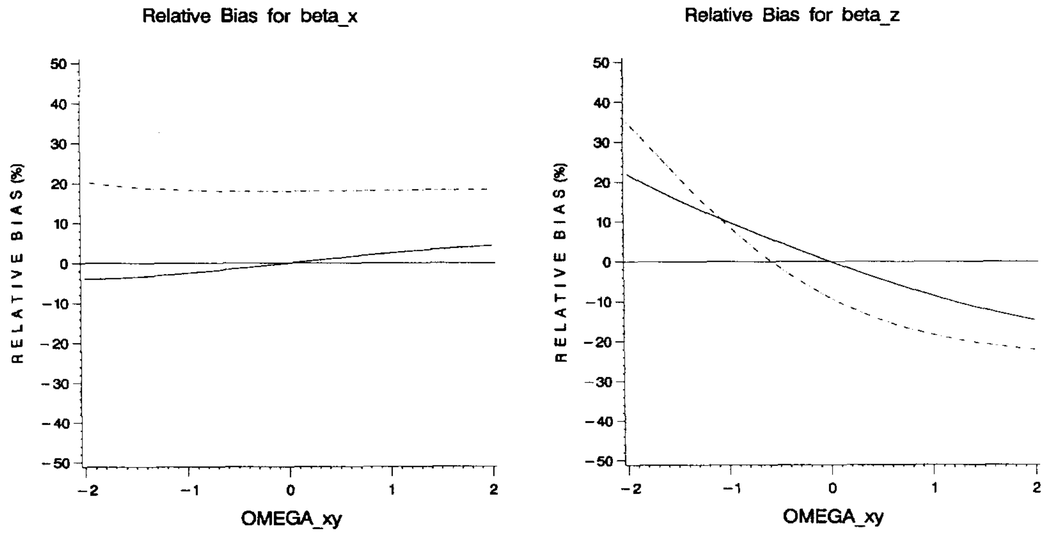

Figure 2. Plot of the Relative Bias in Estimating β_x (a) and β_z (b) as a Function of ω_{xy} as ω_{12} Varies from −3 to 3. ——, WEE(z^-, r^-); − · − · − ·, EM(z^-).

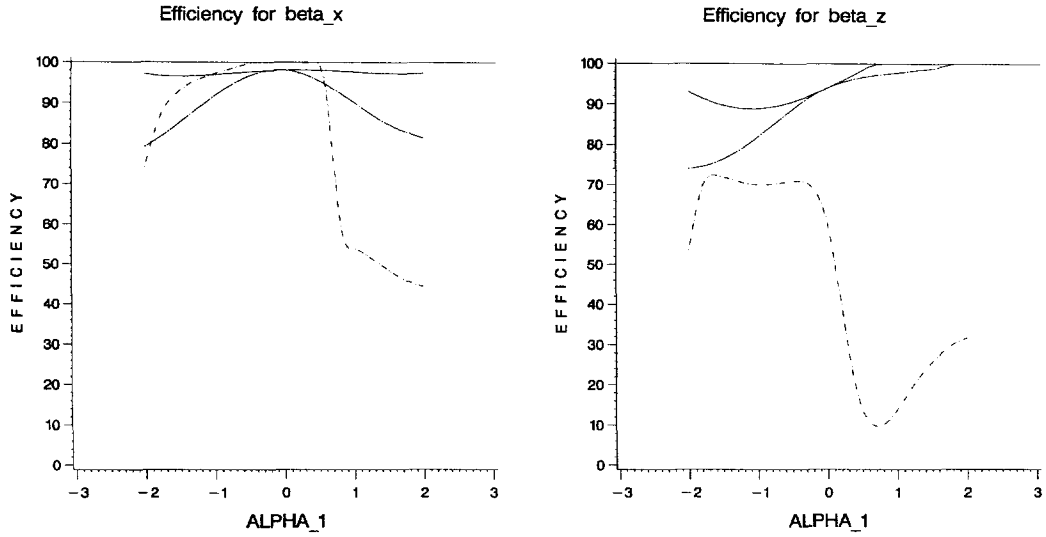

Figure 3. Plot of the Asymptotic Relative Efficiency in Estimating β_x (a) and β_z (b) as a Function of α_1 as α_1 Varies from −2 to 2. ——, WEE(z^+, r^+); — · — · —·, WEE(z^+, r^-); – · – · – ·, WEE(z^-, r^+).

derivative of the log-likelihood and the asymptotic variance of WEE is calculated using (24).

We again formulate the true model by specifying each term on the right side of

$$p(r_i, z_i, y_i, x_{i1}, x_{i2}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \delta)$$
$$= p(r_i|y_i, \mathbf{x}_i, z_i, \boldsymbol{\omega})p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})p(\mathbf{x}_i|\delta). \quad (37)$$

In (37), $p(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta}), p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$, and $p(\mathbf{x}_i|\delta)$ are specified as in the previous section. However, unlike in the previous section, the model for π_i is

$$\text{logit}[\pi_i] = \omega_0 + \omega_x x_i + \omega_y y_i = 1 + x_i + \omega_y y_i, \quad (38)$$

where we let ω_y vary from −2 to 2.

Using ML, we correctly specify $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$, so that ML is the asymptotically efficient estimator. We let WEE(z^+, r^+) denote the WEE with both π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ correctly specified, let WEE(z^-, r^+) denote the WEE with π_i correctly specified and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ misspecified by setting $\alpha_1 = 0$, and let WEE(z^+, r^-) denote the WEE with $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ correctly specified and π_i misspecified by setting $\omega_y = 0$.

Letting $\omega_y = -1.5$ in π_i, Figure 3 plots the asymptotic relative efficiency in estimating β_x and β_z as a function of α_1 as α_1 varies from −2 to 2. We see that on average, WEE(z^+, r^+) has the highest efficiency of the WEEs for estimating β_x, with about 96% efficiency for all values of α_1. On average, WEE(z^+, r^-) has the second-highest efficiency of the WEEs for estimating β_x, with at least 80% efficiency for all values of α_1. When the true α_1 is less than .7, WEE(z^-, r^+) actually has slightly higher efficiency than WEE(z^+, r^-), but the efficiency drops off remarkably when α_1 is greater than .7. For estimating β_z, WEE(z^+, r^+) has the highest efficiency (greater than 90% throughout), followed by WEE(z^+, r^-), for which the efficiency tails off when $\alpha_1 < 0$. Unfortunately, the efficiency of WEE(z^-, r^+) is poor. Thus, as a function of α_1 for this given model

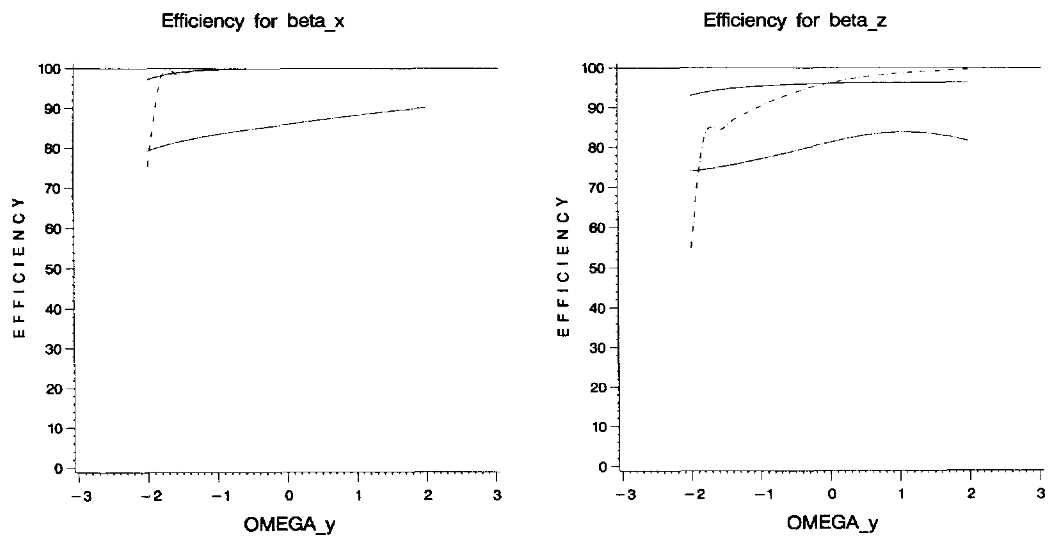

Figure 4. Plot of the Asymptotic Relative Efficiency in Estimating β_x (a) and β_z (b) as a Function of ω_y as ω_y Varies From −2 to 2. ——, WEE(z^+, r^+); — ·— · —·, WEE(z^+, r^-); – ·– ·– ·, WEE(z^-, r^+).

for π_i, more efficiency is gained when correctly modeling $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ than when correctly modeling π_i.

Letting $\alpha_1 = -1.5$, Figure 4 plots the asymptotic relative efficiency in estimating β_x and β_z as a function of ω_y as ω_y varies from -2 to 2. We see again that on average, WEE(z^+, r^+) has the highest efficiency of the WEEs for estimating both β_x and β_z, with about 100% efficiency for β_x and 95% for efficiency for β_z. On average, WEE(z^-, r^+) has the second-highest efficiency of the WEEs for estimating β_x and β_z, except for ω_y less than -1.5. On average, WEE(z^+, r^-) tends to have the lowest efficiency for this configuration, even though it is still at least 85% efficient for estimating β_x and at least 75% efficient for estimating β_z.

Because of the broad range of possible missing-data configurations and underlying probability distributions generating the data, it is also difficult to draw definitive conclusions about efficiency from this asymptotic study. We can make only general suggestions. In our limited asymptotic study, we have seen that the relative efficiency for an estimate of $\boldsymbol{\beta}$ tends to be largest using WEE when both π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ are correctly specified. As for which misspecification can lead to the most inefficient estimate, it appears to depend on the underlying true models for π_i and $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$. We can say that over all of the configurations we have looked at, WEE(z^-, r^+) can have very low efficiency, as shown in Figure 3.

9. DISCUSSION

We have developed WEEs whose form is almost identical to the EM algorithm estimating equations. The WEEs and EM algorithm estimating equations both have strengths and weaknesses. Using ML, the distributions of $(z_i|\mathbf{x}_i)$ and $(y_i|\mathbf{x}_i, z_i)$ must be correctly specified, but we can misspecify (or do not even have to estimate) the distribution of $(r_i|y_i, \mathbf{x}_i)$. For the WEEs, to obtain a consistent estimate of $\boldsymbol{\beta}$ we need to correctly model π_i and also specify $\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)$ such that $E[\mathbf{u}_{1i}(\boldsymbol{\beta}; y_i, \mathbf{x}_i, z_i)] = 0$, or we can misspecify π_i but need to correctly model the densities $(y_i|\mathbf{x}_i, z_i)$ and $(z_i|\mathbf{x}_i)$. Thus WEE is more robust than ML in that we can obtain consistent estimates using WEE under the same conditions as ML, or if π_i is correctly specified. However, unlike ML, WEE requires a sufficient amount of missing data so π_i can be estimated with some precision (Robins et al. 1994). With missing covariates that are MAR, ML easily generalizes to more than one missing covariate with any pattern of missingness, and mixed categorical and continuous covariates (Ibrahim, Chen, and Lipsitz, 1999). The WEEs become more difficult, because a model for the missing-data mechanism must be specified and estimated, which is easy to do only for monotone missing-data patterns.

For a possibly missing continuous covariate, we have demonstrated how to extend the Monte Carlo EM algorithm to a Monte Carlo WEE. These approaches are flexible and can be applied to very large classes of models, including generalized linear models, proportional hazards models, and nonlinear models. We have shown that our procedure is computationally feasible. The sensitivity of the estimates to the choice of covariate distribution is an important issue in the likelihood-based approaches proposed here and is currently being investigated, as are theoretical convergence properties of the proposed Monte Carlo method. Our asymptotic study suggests that if π_i is not modeled correctly but is approximately correct, then there is a very good chance that WEE will reduce the bias of ML with $p(z_i|\mathbf{x}_i, \boldsymbol{\alpha})$ misspecified.

[Received October 1997. Revised March 1999.]

REFERENCES

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.

Falkson, G., Cnaan, A., and Simson, I. W. (1990), A randomized Phase II study of acivicin and 4′Deoxydoxorubicin in patients with hepatocellular carcinoma in an Eastern Cooperative Oncology Group Study. *American Journal Of Clinical Oncology*, 13, 510–515.

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Ibrahim, J. G. (1990), "Incomplete Data in Generalized Linear Models," *Journal of the American Statistical Association*, 85, 765–769.

Ibrahim, J. G., Chen, M. H., and Lipstiz, S. R. (1999), "Monte Carlo EM for Missing Covariates in Parametric Regression Models," *Biometrics*, 55, 591–596.

Kalish, L. A. (1992), "Phase III Multiple Myeloma: Evaluation of Combination Chemotherapy in Previously Untreated Patients," Technical Report 726E, Dana-Farber Cancer Institute, Dept. of Biostatistics.

Lipsitz, S. R., and Ibrahim, J. G. (1996), "A Conditional Model for Incomplete Covariates in Parametric Regression Models," *Biometrika*, 83, 916–922.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley.

Little, R. J. A., and Schluchter, M. (1985), "Maximum Likelihood Estimation for Mixed Continuous and Categorical Data With Missing Values," *Biometrika*, 72, 497–512.

Louis, T. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

McCullagh, R., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Robins, J. M., and Ritov, Y. (1997), "Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models," *Statistics in Medicine*, 16, 285–319.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors are not Always Observed," *Journal of the Royal Statistical Society*, 89, 846–866.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

Wei, G. C., and Tanner, M. A. (1990), "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms," *Journal of the American Statistical Association*, 85, 699–704.

White, H. (1982), "Maximum Likelihood Estimation Under Mis-Specified Models," *Econometrica*, 50, 1–26.

Zhao, L. P., Lipsitz, S. R., and Lew, D. (1996), "Regression Analysis With Missing Covariate Data Using Estimating Equations," *Biometrics*, 52, 1165–1182.