Simple adjustments for randomized trials with nonrandomly missing or censored outcomes arising from informative covariates

STUART G. BAKER
Biometry Research Group, Division of Cancer Prevention, National Cancer Institute,
Bethesda, MD, USA
sb16i@nih.gov

GARRETT M. FITZMAURICE
Division of General Medicine, Brigham and Women's Hospital and Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

LAURENCE S. FREEDMAN
Department of Mathematics, Statistics, and Computer Science, Bar Ilan University,
Ramat Gan 52900, Israel.

BARNETT S. KRAMER
Office of Disease Prevention, National Institutes of Health, Bethesda, MD, USA

Stuart G. Baker is the corresponding author. Other authors are listed in alphabetical order.
SUMMARY

In randomized trials with missing or censored outcomes, standard maximum likelihood estimates of the effect of intervention on outcome are based on the assumption that the missing-data mechanism is ignorable. This assumption is violated if there is an unobserved baseline covariate that is informative, namely a baseline covariate associated with both outcome and the probability the outcome is missing or censored. Incorporating informative covariates in the analysis has the desirable result of ameliorating the violation of this assumption. Although this idea of including informative covariates is recognized in the statistics literature, it is not appreciated in the literature on randomized trials. Moreover, to our knowledge, there has been no discussion on how to incorporate informative covariates into a general likelihood-based analysis with partially missing outcomes to estimate the quantities of interest. Our contribution is a simple likelihood-based approach for using informative covariates to estimate the effect of intervention on a partially missing outcome in a randomized trial. The first step is to create a propensity-to-be-missing score for each randomization group and divide the scores into a small number of strata based on quantiles. The second step is to compute stratum-specific estimates of outcome derived from a likelihood-analysis conditional on the informative covariates, so that the missing-data mechanism is ignorable. The third step is to average the stratum-specific estimates and compute the estimated effect of intervention on outcome. We discuss the computations for univariate, survival, and longitudinal outcomes, and present an application involving a randomized study of dual versus triple combinations of HIV-1 reverse transcriptase inhibitors.

Keywords: Binary; Longitudinal; Maximum likelihood; Propensity score; Survival
1. INTRODUCTION

In many randomized trials outcome data are partially missing due to loss-to-follow-up or subjects not returning for scheduled outcome determination. Examples include a univariate outcome missing in some subjects, longitudinal outcomes missing due to dropout, and censoring of times of death. In analyzing these data the usual goal is to estimate the effect of intervention on the outcome that would be observed if the outcome data were not missing. The appropriate method of analysis depends on the missing-data mechanism (Rubin, 1976). A missing-data mechanism is ignorable if it factors from the likelihood allowing separate likelihood based inference for the parameters of interest; otherwise it is non-ignorable. Two conditions are required for a missing-data mechanism to be ignorable. First, the probability of missing depends on only observed variables, in which case the missing variables are said to be missing at random. Second the parameters modeling the missing-data mechanism are distinct from those modeling the parameters of interest. For a non-ignorable missing-data mechanism with distinct parameters, the probability an outcome variable is missing can depend on one or more of the following: (i) the outcome variable itself, (ii) another partially observed variable, or (iii) an indirectly observed variable such as a potential outcome for noncompliance (Baker, 2000).

When estimating the effect of intervention on outcome in a randomized trial with a partially missing outcome, a standard and often implicit assumption is that the probability the outcome is missing depends only on the randomization group and, in the case of longitudinal data, previously observed outcomes, so that the missing-data mechanism is ignorable. This assumption underlies maximum likelihood estimation in the guise of complete-case analyses with missing univariate outcomes (Baker and Laird, 1988), survival models that ignore the censoring mechanism and, linear mixed effects longitudinal models that ignore the missing-data mechanism (Molenberghs et al, 2004).
We define an informative covariate as a baseline covariate associated with both outcome and the probability that the outcome is missing. We say a set of informative covariates is completely informative if (i) the probability the outcome is missing depends on the set of informative covariates, as well as other variables that are unrelated to outcome, but not the partially observed outcome and (ii) if the distribution of outcome depends on the set of informative covariates. As discussed later, this definition implies that outcome and the probability the outcome is missing are independent conditional on the set of informative covariates.

If all informative covariates are not included in the model for the effect of intervention on outcome, the missing-data mechanism would be nonignorable because the probability of missing outcome would be associated with the outcome through the unobserved informative covariate. In this case standard analyses that assume an ignorable missing-data mechanism would incorrectly estimate the effect of intervention on outcome. Although it is sometimes possible to estimate the effect of intervention on outcome using a model for the nonignorable missing-data mechanism, such models require many covariates and strong unverifiable assumptions for identifiability or sufficient precision (Baker et al, 2003).

Of course the simplest way to avoid the problems of an unobserved informative covariate is to include informative covariates in the analysis. If the completely informative set of covariates is included in the model for the missing-data mechanism, the missing-data mechanism will be ignorable. The idea of creating an ignorable missing-data mechanism from a non-ignorable missing-data mechanism by including additional covariates is well-recognized. (Collins et al, 2001 and references therein). However its use is not well-appreciated in randomized trials. In fact, a major review of the use of baseline covariates in randomized trials (Pocock et al, 2002) did not mention the role of baseline covariates in adjusting for the missing outcome data. In addition, to our knowledge, there has been no discussion of how to use informative covariates to estimate
quantities of interest in a likelihood-based analysis with missing outcomes. Collins et al (2001) note that currently existing statistical software programs that incorporate likelihood-based approaches for handling missing data do not facilitate the use of informative covariates.

Our contribution is a simple method for incorporating informative covariates into a likelihood-based analysis for estimating the effect of intervention on a partially missing outcome in a randomized trial. The method capitalizes on existing software and methodology for maximum likelihood estimation with missing outcomes.

2. MOTIVATING EXAMPLE

To motivate our approach, we consider data from a randomized, double-blind, study of AIDS patients with advanced immune suppression (CD4 count of ≤ 50 cells/mm³) (Henry et al., 1998). Patients were randomized to one of four daily regimens containing 600 mg of zidovudine: zidovudine alternating monthly with 400 mg didanosine; zidovudine plus 2.25 mg of zalcitabine; zidovudine plus 400 mg of didanosine; or zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine. A total of 1313 patients were evaluated at baseline and followed up to 2 years. For the analyses presented here, we focus on the comparison of the first three treatment regimens (dual therapy) with the fourth (triple therapy). The primary outcome of interest is survival (i.e., time to death) and the goal of our analyses is to compare the dual and triple therapy groups in terms of survival rates at 18 months. The Kaplan-Meier estimate of the probability of survival at 18 months is 0.6439 (SE = 0.0176) and 0.7189 (SE = 0.0287) in the dual and triple therapy groups respectively. The estimated survival difference at 18 months is 0.0750 (SE = 0.0337, p < 0.03), indicating a survival benefit of 7.5% for triple therapy over dual therapy. A comparison of the entire survival curves for the two treatment groups, via both log rank and Wilcoxon tests, produces chi-square statistics of 6.34 (p < 0.02) and 6.46 (p < 0.02) respectively. Overall, the results of these standard
analyses comparing the two treatment groups suggest a beneficial effect of triple therapy over dual therapy in prolonging survival among individuals with advanced HIV disease.

The concern is that these results may be incorrect because there are two candidate unobserved informative covariates, namely baseline age and CD4 count, that are likely associated with both censoring and survival. Our goal is to estimate the survival difference between the two randomization groups after adjusting for the informative covariates of age and CD4 count. The underlying assumption is that a missing-data mechanism based on age, CD4 count, and randomization group is ignorable. We will return to this example after presenting the methodology.

3. METHODOLOGY

Let \( i \) index subject in each randomization group \( z \). Let \( S_{\text{obs}}(z) \) and \( S_{\text{miss}}(z) \) denote subjects with observed and missing outcomes, respectively, in randomization group \( z \).

Let \( x^*_z \) denote a vector of informative covariates for subject \( i \) in group \( z \). Let \( y_{zi} \) denote one or more outcomes (univariate, survival, or longitudinal) for subject \( i \) in group \( z \) with missing outcomes denoted \( y_{\text{miss}}(zi) \). Let \( w_{zi} \) denote variables that predict the probability the outcome is missing but not outcome and which need to be included when we cannot distinguish them from \( x^*_z \). Let \( m_{zi} \) indicate which outcomes, if any, are missing for subject \( i \) in group \( z \).

We assume that \( x^*_z \) are completely informative, namely that \( pr(m_{zi} \mid y_{\text{miss}}(zi), x^*_z, w_{zi}) = pr(m_{zi} \mid x^*_z, w_{zi}) \) and \( pr(y_{zi} \mid x^*_z, w_{zi}) = pr(y_{zi} \mid x^*_z) \). This implies that \( pr(m_{zi}, y_{zi} \mid x^*_z) = pr(m_{zi} \mid x^*_z) pr(y_{zi} \mid x^*_z) \), namely that outcome and the indicator of missingness are independent conditional on the set of informative covariates (Appendix A).

To avoid sparse-data problems when the set of informative covariates is of high dimensions we invoke the following procedure. First we model the missing data
mechanism separately for each randomization group. Because one may not know if a

covariate is informative, one should include many covariates likely related to outcome in
the model for the missing-data mechanism and test if their coefficients statistically differ
from zero. This approach is related to the inclusive strategy of Collins et al (2001) with
the added advantage of removing covariates that do not predict missingness in outcome.

Second, we define the propensity-to-be-missing score for each randomization group \( z \) by

\[
e_z(\mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}) = pr(\mathbf{m}_{z,i} \mid \mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}).
\]

By definition, if \( e_z(\mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}) = e \), then \( pr(\mathbf{m}_{z,i} \mid e) = pr(\mathbf{m}_{z,i} \mid \mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}) \) for all \( \mathbf{x}_{z,i}^* \) and \( \mathbf{w}_{z,i} \) such that \( e_z(\mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}) = e \). In other words, if

the model for the propensity-to-be-missing score is correctly specified, the propensity-to-

be-missing score has the same information for predicting missing outcomes as the vector

of covariates. As a further step, in the case of survival and longitudinal data (as will be
discussed), we write \( e_z(\mathbf{x}_{z,i}^*, \mathbf{w}_{z,i}) = h_z(\mathbf{x}_{z,i}^* \Theta_{z1} + \mathbf{w}_{z,i} \Theta_{z2}) \) where \( h_z(\cdot) \) is a

monotonically increasing function. We form \( k \) quantiles of \( \mathbf{x}_{z,i}^* \Theta_{z1} + \mathbf{w}_{z,i} \Theta_{z2} \) and let \( x_z \)

index each of the \( k \) strata. The variable \( x_z \) is a single categorical informative covariate

that should capture most of the information in \( \mathbf{x}_{z,i}^* \).

Let \( \theta_{z} = pr(X_z = x \mid z), \ \Theta_z = (\theta_{z1}, \theta_{z2}, \ldots \theta_{z(k-1)}) \), and let \( \alpha_z \) denote a vector of

parameters characterizing the distribution of outcome conditional on the informative
covariate and randomization group \( z \). If data on the informative covariate are collected,
the missing-data mechanism is ignorable, and the likelihood kernel is \( L = \prod_z L_z \),

where

\[
L_z = \prod_{i \in S_{\text{obs}(z)}} pr(Y_{z_i} = y_{z_i} \mid x_{z_i}, \alpha_z) \prod_{i \in S_{\text{miss}(z)}} \left\{ \int_{y \in y_{\text{miss}(z)}} pr(Y_{z_i} = y_{z_i} \mid x_{z_i}, \alpha_z) dy_{\text{miss}(z_i)} \right\}.
\]  \hspace{1cm} (1)
The parameters $\alpha_x$ can be estimated using simple formulas or standard software. By construction of $k$ quantiles in each randomization group, $\hat{\theta}_{zx} = 1/k$. The estimated difference in outcomes between randomization groups is

$$
\hat{\delta} = \sum_{x=1}^{k} \hat{f}_{1x} \hat{\theta}_{1x} - \sum_{x=1}^{k} \hat{f}_{0x} \hat{\theta}_{0x} = \sum_{x=1}^{k} (\hat{f}_{1x} - \hat{f}_{0x}) / k,
$$

where $\hat{f}_{zx} = E(y | Z = z, x; \alpha_x)$ for univariate or longitudinal data and $pr(Y \geq y | Z = z, x; \alpha_x)$ for survival data. Let $d_z = \{ (\hat{f}_{z1} - \hat{f}_{zk}), (\hat{f}_{z2} - \hat{f}_{zk}), \ldots, (\hat{f}_{z(k-1)} - \hat{f}_{zk}) \}$, and let $N_z$ denote the total number of subjects in group $z$. The estimated asymptotic variance is

$$
\hat{v}\hat{a}\hat{r}(\hat{\delta}) = \sum_{z=0}^{1} \sum_{x=1}^{k} \hat{v}\hat{a}\hat{r}(\hat{f}_{zx}) \hat{\theta}_{zx}^2 + d_z \cdot \hat{v}\hat{a}\hat{r}(\hat{\theta}_z; N_z) \cdot d_z^T,
$$

where $\hat{v}\hat{a}\hat{r}(\hat{f}_z(x))$ is typically obtained from standard software packages and $\hat{v}\hat{a}\hat{r}(\hat{\theta}_z; N_z)$ is a matrix with diagonal elements of $(1/k) \{1-(1/k)\} / N_z$ and off-diagonal elements of $(-1/k^2) / N_z$. In the remainder of this section we discuss special formulations for univariate, longitudinal, and survival outcomes.

### 3.1 Continuous univariate outcome

Let $S_{obs(z)}$ denote subjects in group $z$ with observed outcomes and let $y_{zi}$ denote a univariate outcome for subject $i$ in group $z$. Let $M_{zi} = 1$ if the outcome for subject $i$ in group $z$ is missing and 0 otherwise. The propensity-to-be-missing score can be created by fitting $logit \{ pr(M_{zi} = 1 \mid x_{zi}^*, \mathbf{w}_{zi} ) \} = \beta_{z0} + x_{zi}^* \beta_{z1} + \mathbf{w}_{zi} \beta_{z2}$ in the following likelihood, $\prod_{i \in S_{obs(z)}} pr(M_{zi} = 0 \mid x_{zi}^*) \prod_{i \in S_{miss(z)}} pr(M_{zi} = 1 \mid x_{zi}^*)$. The strata $x_z$ are formed by quantiles of $x_{zi}^* \beta_{z1} + \mathbf{w}_{zi} \beta_{z2}$.
Because the integral in (1) equals one in this situation, the likelihood for \( \alpha_z \) is based only on subjects with observed outcomes, \( \prod_{i \in S_{\text{obs}}} pr(Y_{zi} = y_{zi} | x_{zi}; \alpha_z) \). Let \( I_{zzi} \) equal 1 if the informative covariate for individual \( i \) in group \( z \) equals \( x \), and 0 otherwise.

As an example, one can fit the following model to observed data from the trial: \( y_{zi} = \alpha_{z0} + \sum_{x=1}^{k-1} \alpha_{z1x} I_{zxi} + \epsilon_{zi} \) where \( \epsilon_{zi} \sim N(0, \sigma_z^2) \). The estimated mean difference in outcome between randomization groups is \( \hat{\delta}_{CU} = \sum_{x=1}^{k} \left\{ E(y_{zi} | x; \hat{\alpha}_1) - E(y_{0i} | x; \hat{\alpha}_0) \right\} / k = (\hat{\alpha}_{10} + \sum_{x=1}^{k-1} \hat{\alpha}_{11x} / k) - (\hat{\alpha}_{00} + \sum_{x=1}^{k-1} \hat{\alpha}_{01x} / k). \)

### 3.2 Binary univariate outcome

The propensity-to-be-missing score can be created by fitting a logistic regression for the probability of missing an outcome, similar to the situation with a continuous univariate outcome. Let \( n_{zxy} \) denote the number of subjects in randomization group \( z \) with informative covariate \( x \) and outcome \( y \). Let \( \alpha_{y|x} = \text{pr}(Y = y | z, x) \). The likelihood conditional on the informative covariate is \( \prod_x \prod_y \alpha_{y|x}^{n_{zxy}} \). The maximum likelihood estimate is \( \hat{\alpha}_{1|x} = n_{z1}/\sum_y n_{zxy} \). The estimated mean difference in the probability of outcome between the randomization groups is \( \hat{\delta}_{BU} = \text{pr}(Y_1 = 1 | \hat{\alpha}_1) - \text{pr}(Y_0 = 1 | \hat{\alpha}_0) = \sum_{x=1}^{k} \left\{ \hat{\alpha}_{1|x} - \hat{\alpha}_{0|x} \right\} / k. \)

### 3.3 Survival outcomes

For survival data, we make the analogy between censoring mechanisms and missing-data mechanisms as has been done for discrete-time survival analyses (Baker et al, 1993; Baker, 1994) and more generally (Heitjan and Rubin, 1991). Let \( S_{\text{obs}(z)} \) and \( S_{\text{cens}(z)} \) denote the subjects with observed and censored outcomes, respectively. Let \( Y_{zi} \) denote the time after randomization when the outcome occurs, \( C_{zi} \) denote the time after randomization when censoring occurs, and let \( t_{zi} \) denote the time of failure or censoring,
whichever comes first. If either \( Y_{zi} \) or \( C_{zi} \) equals time \( t_{zi} \), the other time is greater than or equal to \( t_{zi} \).

The propensity-to-be-missing scores can be created by fitting a proportional hazards model to the hazard for censoring, \( \log \{ pr(C_{zi} = t_{zi} \mid x_{zi}, w_{zi}) / \, pr(C_{zi} \geq t_{zi} \mid x_{zi}, w_{zi}) \} = x_{zi}^{*} \beta_{1z} + w_{zi} \beta_{2z} + h_{z}(t_{zi}) \) where \( h_{z}(t_{zi}) \) is a baseline hazard function. The likelihood for censoring in group \( z \) is

\[
L_{cens}(z) = \prod_{i \in S_{obs}(z)} pr(C_{zi} \geq t_{zi} \mid x_{zi}^{*}, w_{zi}) \quad \prod_{i \in S_{cens}(z)} pr(C_{zi} = t_{zi} \mid x_{zi}^{*}, w_{zi})
\]

where censoring is a "failure" event and failure means the time of censoring is "censored". The \( x_{z} \) are computed by taking \( k \) quantiles of \( x_{zi}^{*} \beta_{1z} + w_{zi} \beta_{2z} \).

The component of the likelihood involving the parameters of interest is

\[
L_{surv}(z) = \prod_{i \in S_{obs}(z)} pr(Y_{zi} = t_{zi} \mid x_{zi}; \alpha_{z}) \quad \prod_{i \in S_{cens}(z)} pr(Y_{zi} \geq t_{zi} \mid x_{zi}; \alpha_{z}).
\]

Let \( S_{xz}(t; \alpha_{z}) \) denote the probability of surviving to time \( t \) for subjects in group \( z \) and stratum \( x_{z} \) of the informative covariate, which can be estimated in a variety of ways including Kaplan-Meier, proportional hazards or Weibull regression. The estimated mean difference in the probability of survival to time \( t \) is

\[
\hat{\delta}_{SURV} = \Sigma_{z=1}^{k} \{(S_{1z}(t; \hat{\alpha}_{1}) - S_{0z}(t; \hat{\alpha}_{0}) \} / k.
\]

### 3.4 Continuous longitudinal outcomes

We consider only the situation in which subjects dropout over time with no intermittent missing outcomes. Let \( y_{zij} \) denote the outcome a time \( j \) for subject \( i \) in group \( z \). Let \( M_{zij} = 1 \) if subject \( i \) in group \( z \) is missing at time \( j \), and 0 otherwise, with realization \( m_{zij} \). Let \( p_{zij} = pr(M_{zij} = 1 \mid x_{zi}^{*}, w_{zi}, y_{zi(j-1)} \}. \) Let \( S_{core}(z) \) denote the set
of subjects with at least one outcome. Subjects without any outcomes are dropped from
the analysis because they make no contribution, as in the case of univariate data. The
likelihood for the missing-data mechanism is

$$L_{CLmiss(z)} = \prod_{i \in S_{core(z)}} \prod_{t=1}^{\Sigma_i(1-m_{zij})} (1-p_{zi})^{(1-m_{zij})} m_{zij} p_{zi}^t,$$

which only includes contributions up to dropout time $j$. The propensity-to-be-missing
score is created by modeling $\text{logit}(p_{zi}) = x^*_{zi} \beta_{z1} + w_{zi} \beta_{z2} + y_{zi(j-1)} \beta_{z3}$.

Importantly the model contains a time-invariant component $x^*_{zi} \beta_{z1} + w_{zi} \beta_{z2}$ and a
time-varying component that is the outcome prior to dropout. The direct inclusion of
additional history of prior outcomes in the model can be problematic because the set of
parameters would vary by dropout time, affecting interpretation of parameters.

Alternatively, a suitable function of the history of prior outcomes, e.g., arithmetic mean,
could be included without affecting interpretation. In general, the outcome prior to
dropout likely captures most of the information for predicting dropout. The $x_{zi}$ are
computed by taking quantiles of $x^*_{zi} \beta_{z1} + w_{zi} \beta_{z2}$.

We partition the set of subjects $S_{core(z)}$ into $S_{obs(z)}$, the subjects in group $z$ not missing
any outcome, and $S_{miss(z)}$ the subjects in group $z$ missing at least one outcome. Let $y_{zi}$
$= (y_{zi1}, y_{zi2}, \ldots, y_{zik})$ denote all outcomes for subject $i$ and let $y_{miss(zi)} =$
$(y_{zi1}, y_{zi(j+1)}, \ldots)$ denote a vector of missing outcomes for subject $i$ in group $z$, who
drops out at time $j$. The component of likelihood involving the parameters of interest is

$$L_{CL}(z) = \prod_{i \in S_{obs(z)}} \text{pr}(X_{zi} = y_{zi} \mid x_{zi}; \alpha_z) \prod_{i \in S_{miss(z)}} \int_{y \in y_{miss(z,i)}} \text{pr}(Y_{zi} = y_{zi} \mid x_{zi}; \alpha_z) dy.$$

Although the likelihood involves a complicated integral, it is relatively easy to maximize
if $y_{zi}$ follows a multivariate normal distribution. The reason is that standard models for
multivariate normal distributions involve a reproducible parametrization. We term a parametrization reproducible if a subset of the same parameters can be used for any likelihood contribution regardless of the pattern of missing outcomes. This definition of a reproducible parametrization is more general than that in Liang et al (1992). The multivariate normal distribution is reproducible because any marginal distribution involves a subset of parameters from the full multivariate normal distribution.

In the context of longitudinal data, it is convenient to model the multivariate normal distribution using a regression for the outcomes and a structured covariance matrix (Molenberghs et al, 2003). One model is

\[ y_{zi} = \alpha_{z0} + \sum_{x=1}^{k-1} \alpha_{1x} I_{xzi} + t_{zi} \alpha_{z2} + t_{zi}^2 \alpha_{z3} + \epsilon_{zi}, \]

where \( t_{zi} \) is a vector of times, \( \epsilon_{zi} \sim N(0, W_z) \), and \( W_z \) is a structured covariance matrix. Alternatively, the covariance can be modeled via the introduction of random effects, i.e. \( \epsilon_{zi} \) can be decomposed into between- and within-subject sources of variability. The model can be fit using standard software packages with maximum likelihood (or restricted maximum likelihood) estimation of continuous longitudinal data with missing outcomes. The estimated mean difference between randomization groups in the average outcome over \( J \) time periods is

\[ \hat{\delta}_{CL} = \sum_{x=1}^{k} \sum_{j=1}^{J} \left\{ E(y_{1ij} | x, \tilde{\alpha}_1) \right\} - E(y_{0ij} | x; \tilde{\alpha}_0) \} / (Jk) \]

3.5 Binary longitudinal outcomes

As with continuous longitudinal data, it is preferable to model binary longitudinal data using a reproducible parametrization. Various approaches have been proposed to model the marginal distribution of longitudinal binary outcomes using a reproducible parametrization (Baker, 1995; Diggle et al; 2002; Farmer and Tutz, 2001; Heagerty 2002, and Molenberghs and Lesaffre, 1994). However these methods are complicated and require specialized software. A simpler alternative is to model distributions of binary outcomes conditional on past history and combine estimates from this model to estimate
the marginal distributions of interest. If the only type of missing outcomes are dropouts, it is easy to fit the following reproducible parametrization based on a first-order Markov chain. Without loss of generality, consider three time periods indexed by \( j \). Let \( Y_j \) denote the binary outcome at time \( j \) where \( Y_j = 1 \) if a success occurs and \( 0 \) if a failure occurs at time \( j \). The propensity-to-be-missing score is computed in the same way as for continuous longitudinal data, and the resulting strata from \( k \) quantiles are denoted \( x_z \).

Let \( \alpha_{y_1,z}^{(1)} = pr(Y_{z1} = y_{z1} \mid x_z) \) and \( \alpha_{y(j-1),y;zz}^{(j)} = pr(Y_{zj} = y_j \mid Y_{z(j-1)} = y_{z(j-1)}, x_z) \). The likelihood and the maximum likelihood estimates are given in Appendix. B. The estimated difference between randomization groups in the average number of successes over three time intervals is 

\[
\delta_{CL} = \frac{\sum_{j=1}^{3} (\hat{f}_{1x} - \hat{f}_{0x})}{k},
\]

where 

\[
\hat{f}_{xx} = \prod_{y_1} \Pi_{y_2} \Pi_{y_3} \alpha_{y_1,z}^{(1)} \alpha_{y_1,y;zz}^{(2)} \alpha_{y_2,y;zz}^{(3)} (y_1 + y_2 + y_3)/3.
\]

4. APPLICATION

Using the aforementioned methodology we reanalyzed the data from the AIDS trial discussed in Section 2 to adjusted for the informative covariates of baseline age and log transformed CD4 count. Because 14 subjects were missing baseline CD4 count the analysis involves data from 1299 subjects without missing informative covariates.

In ACTG Study 193A approximately 20% of subjects prematurely discontinued from the study (for reasons other than death or protocol completion), primarily because of subject refusal or the clinical unit's inability to contact the subject. We fit a proportional hazards model for the hazard for censoring in which the logarithm of the hazard was proportional to the logarithm of the CD4 count plus 1. The inclusion of a term for age did not significantly improve the fit, so was excluded. The estimated parameters and standard errors are reported in Table 1. For each randomization group, we formed 5 strata based on quintiles of the proportionality factor (Table 2). For each quintile we also computed the median probability of censoring in the quintile at 18 months. Within each stratum of each group, we estimated the probability of survival at 18 months, \( \hat{S}_{xx} \), via the
Kaplan-Meier estimator (see Table 2). Finally, we averaged the treatment differences in survival probability over the 5 strata and obtained an adjusted estimate of the survival difference of 0.0732 (SE = 0.0336, p < 0.03). In this case, the adjusted and unadjusted estimates and standard errors are similar.

We conclude that even if censoring is related to baseline CD4 count, the triple therapy results in significantly better survival than the dual therapy. The fact that the adjustment for an informative covariate yields a similar estimate adds to the robustness of the conclusion and clinical implications for patients similar to those who fulfilled eligibility criteria in the trial. The reason adjustment may have made little difference here is that the range of CD4 count was small, which weakens the association of CD4 count with outcome and probability the outcome is missing.

To investigate how the informative covariate could have a larger impact on results, we used the estimates in Table 2 to construct a data set with a binary endpoint. The unadjusted and adjusted estimates of .0768 (SE=.0451) and .0733 (SE=.0454) were similar to those from the survival data, although the standard errors are larger. We then switched the estimated survival for triple therapy in stratum 1 and stratum 5. We think this example is realistic because it uses the original survival probabilities but with one simple switch. The adjusted estimate remained the same (which is mathematically required because the weights are the same over strata) with a similar standard error, .0733 (SE=.0448). However the unadjusted estimate differed somewhat, .0615 (SE=.046), because the weights differed over strata.

5. RELATION TO INVERSE WEIGHTING METHODS

The propensity-to-be-missing score is similar to the response propensity for adjusting for nonresponse in sample surveys (Little,1986) and closely related to the propensity scores for analyzing data from observational studies (Rosenbaum and Rubin, 1984). Two general strategies for using propensity scores are stratification (stratifying subjects by
quantiles of the score) and inverse weighting (weighting each subject by the reciprocal of the propensity score). When the propensity score is used to adjust for selection bias in observational studies, both stratification and inverse weighting are popular (Rosenbaum, 1987; Lunceford and Davidian, 2004). In contrast, when using propensity scores to adjust for missing outcomes, the predominant approach in the statistics literature is inverse weighting (Rosenbaum, 1987, Robins et al 1995, Rotnitzky and Robins 1995) although Little (1986) and Lavori et al (1995) discussed stratification (but differently from our approach).

The following is a simple comparison of inverse weighting and stratification. Let denote the strata obtained by cross-classifying the informative covariates. If there are many informative covariates, the strata will be very sparse making estimation problematic. Stratification and inverse weighting are ways to circumvent the problem of sparse data. We compare the two approaches in the simple case of estimating a mean outcome (in one group) when some of the outcomes are missing, there is cross-classification of informative covariates into categories indexed by and there are zero or few subjects in many of the categories.

With the proposed approach for stratification, a propensity-to-be-missing score is computed and, based on this score, individuals are divided into a small number of strata indexed by . Let denote the mean outcome in stratum , the number of subjects in stratum with observed outcome, the total number of subjects in stratum , and the total number of subjects. With our stratification approach the estimated mean outcome is .

For the inverse weighting approach, the following derivation follows that in Rosenbaum et al (1987). Let index individuals within stratum and let if the outcome for individual is observed and otherwise. Let denote the total number of subjects in stratum , and let denote the number of subjects with observed outcome in stratum . Let denote the response for subject in stratum . The
population mean is $\bar{y}_{\text{pop}} = \sum_{i=1}^{N_s} y_{si} / N = \sum_{s=1}^{S} \sum_{i=1}^{N_s} \bar{y}_{\text{pop}(s)} N_s / N$, where

$\bar{y}_{\text{pop}(s)} = \sum_{i=1}^{N_s} y_{si} / N_s$ is the population mean for stratum $s$. Let $\bar{y}_s = \sum_{i=1}^{N_s} y_{si} / n_s$ denote the mean for the observed outcomes in stratum $s$. Substituting $\bar{y}_s$ for $\bar{y}_{\text{pop}(s)}$ in the equation for $\bar{y}_{\text{pop}}$ gives an estimated mean of

$$
\sum_{s=1}^{S} \bar{y}_s / N = \sum_{s=1}^{S} \bar{y}_s (n_s / N) \quad n_s / N
$$

Because $n_s$ is very small or zero, a direct estimate of $n_s / N$ is impossible or unstable. The key to the inverse weighting approach is to substitute the propensity-to-be-missing score evaluated at $s$ for $n_s / N$ giving an estimate of $\sum_{s=1}^{S} \bar{y}_s e(s)^{-1} (n_s / N)$.

In essence, the stratification approach avoids the problem of sparse data up front by stratifying the propensity-to-be-missing score into categories. In contrast the inverse weighting approach keeps the original sparse categories but deftly substitutes the inverse of propensity-to-be-missing score to create a weight that augments the observed sample. For a comparison of the two approaches in a more complex setting see Czajka et al (1992) and Rotnitzky and Robins (1995).

In many settings an advantage of our approach over inverse weighting approaches (e.g. Rotnitzky and Robins, 1985) is that the final phase of estimating the effect of intervention on outcome does not require a model for the missing-data mechanism. Although the preliminary phase of computing the propensity-to-be-missing score requires the appropriate covariates for modeling the missing-data mechanism, the exact functional form is not critical because the categorical informative covariate is created by stratifying the propensity-to-be-missing score into quantiles. The reason is that within each quantile, the probability of missing the outcome is similar for all subjects regardless of the model. The tradeoff for using the stratified propensity-to-be-missing score instead of inverse weighting is that it requires correct specification of the effect of the intervention and the informative covariate (created from the propensity-to-be-missing score) on outcome. Therefore we recommend saturating the model for the effect of informative
covariate and intervention on outcome, e.g., by including an interaction term between the informative covariate and outcome, as was illustrated in the examples.

6. DISCUSSION

The use of informative covariates transforms a non-ignorable missing-data mechanism into an ignorable missing-data mechanism, which avoids the need for strong assumptions about the missing data mechanism. The novel aspect presented here involves using informative covariates to simply estimate the effect of intervention on a partially missing outcome in a likelihood based analysis.

Investigators should consider many possible covariates in the model for the missing-data mechanism and eliminate those that do not predict the probability of missing outcome. Intuitively, the inclusion of covariates that do not predict the probability an outcome is missing will add noise that will blur the strata and reduce the differences among stratum-specific probabilities of missing outcomes, making the adjustment for missing outcomes less effective.

If the investigator thinks some informative covariates were not identified or the nonignorable missing-data mechanism arises in another way, a sensitivity analysis would be needed after adjusting for informative covariates. For a missing binary outcome, Baker and Freedman (2003) proposed a sensitivity analysis that first adjusts for informative covariates and then uses the randomization to reduce the user input.

Other examples of informative covariates may be instructive. One is health-seeking behavior that indicates if a subject has frequently sought health care prior to the trial. Subjects with health-seeking behavior may be more likely to drop out of the study prior to a scheduled visit and have a more favorable outcome than subjects without health-seeking behavior. Health seeking behavior is a very common characteristic of human nature, and its manifestations may be increasing as patients have increased access to Internet-based health information and direct-to-consumer advertising. Another second
example of an informative covariate is glucose intolerance (diabetes) in a randomized trial of two anti-cancer regimens, one of which contains prednisone. Subjects with diabetes may have worse overall survival and also be more likely to go off study due to unstable blood sugars caused by prednisone than subjects without glucose intolerance. A third example of an informative covariate is date of entry into a randomized trial with staggered entry and a time-to-event outcome. If there is a shift in prognosis during the early phase of the trial, then those who enter earlier will also have different risk of outcome than those who enter later. In addition subjects who enter the study earlier, versus those who enter later, are less likely to be censored.

If the probability of missing the outcome depends only on a single categorical informative covariate, a propensity-to-be-missing score is not needed. In addition there would be a gain in efficiency because \( \theta_{zx} = \theta_x \) by virtue of the randomization. In this case \( \theta_x \) would be estimated by the fraction of subjects in both groups in category \( x \).

In some situations, the utility of the proposed approach depends on the question. Suppose that some subjects randomized to intervention \( A \) switch from intervention \( A \) to \( B \) during the trial but the outcomes are reported for all subjects. In this case one could perform an intent-to-treat analysis with no missing data. However this will dilute the contrast between the effect of \( A \) and \( B \). Alternatively one could define outcome by outcome given assigned treatment and view subjects who switch from \( A \) to \( B \) as having a missing outcome for \( A \). In this case the proposed methodology would be applicable, but only if one identifies all informative covariates and assumes no other nonignorable missing-data mechanism. (This is a very strong assumption, however.) In the case of all-or-none compliance, the preferred approach involves a potential outcomes model that makes use of the randomization (Baker and Kramer, 2004).

A topic for future research is whether this approach is appropriate when \( x \) includes variables that are observed after randomization and are related to both outcome and the probability of missing outcome. Because intervention would affect outcome both
directly and through \( x \), there may be greater sensitivity to model misspecification. In this case one might consider an alternative model involving the distribution of \( y \) conditional on \( z \), and the distribution of \( x \) conditional on \( y \) and \( z \) (e.g. Baker, 2000).

The proposed approach has an important implication for trial design, namely that investigators should collect data on variables that they think may be informative and incorporate them into the analysis of differences in outcomes between intervention groups. Collecting data on informative covariates should be easy because investigators routinely collect data on medical conditions, medical history and demographics (Pocock et al, 2002). If these data are collected, the approach suggested here, as opposed to an approach without informative covariates, can yield better estimates of the effect of intervention on outcome in randomized clinical trials with missing outcomes.

**APPENDIX A**

Based on the definition of a completely informative covariate, \( \text{pr}(m_{zi}, y_{\text{miss}(zi)} \mid x^*_z, w_{zi}) = \text{pr}(m_{zi} \mid y_{\text{miss}(zi)}, x^*_z, w_{zi}) \times \text{pr}(y_{\text{miss}(zi)} \mid x^*_z, w_{zi}) = \text{pr}(m_{zi} \mid x^*_z, w_{zi}) \text{pr}(y_{\text{miss}(zi)} \mid x^*_z) \). Multiplying both sides of the aforementioned equation by \( \text{pr}(w_{zi} \mid x^*_z) \) and integrating over \( w_{zi} \) gives \( \text{pr}(m_{zi}, y_{\text{miss}(zi)} \mid x^*_z) = \text{pr}(m_{zi} \mid x^*_z) \text{pr}(y_{\text{miss}(zi)} \mid x^*_z) \).

**APPENDIX B**

Let \( n^{(j)}_{y_{z1} \ldots y_{z(j-1)} y_{zj} x} \) denote the number of subjects in group \( z \) and covariate level \( x \) with a history of outcomes to time \( j \) of \( \{y_{z1}, \ldots, y_{z(j-1)}, y_{zj}\} \). The likelihood kernel (for \( j = 1, 2, 3 \)) for the parameters of interest is

\[
\prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} \prod_{yi=0} \prod_{yi=0} \{ \alpha_{yi \mid x} \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right] \\
= \prod_x \left[ \prod_{yi=0} \{ \alpha_{yi \mid x} \}^{n^{(1)}_{yi \mid x}} + n^{(2)}_{yi \mid y_{2}y_{g}z} \prod_{yi=0} \{ \alpha_{yi \mid y_{2}y_{g}z} \}^{(2)} \right]
\]
\[ \prod_{y_2=0}^{1} \prod_{y_3=0}^{1} \{ \alpha_{y_2 y_3 x}^{(3)} \} \]

where "+" in the subscript indicates summation over the indicated index. The maximum likelihood estimates are therefore

\[ \alpha_{y_1 x}^{(1)} = \frac{n_{y_1 x - 2}^{(1)} + n_{y_1 x - 2}^{(2)} + n_{y_1 x - 2}^{(3)}}{n_{y_1 x - 2}^{(3)} + n_{y_1 x - 2}^{(3)} + n_{y_1 x - 2}^{(3)}} , \quad \alpha_{y_1 y_2 x}^{(2)} = \frac{n_{y_1 y_2 x}^{(2)} + n_{y_1 y_2 x}^{(3)} + n_{y_1 y_2 x}^{(3)}}{n_{y_1 y_2 x}^{(3)} + n_{y_1 y_2 x}^{(3)} + n_{y_1 y_2 x}^{(3)}} , \quad \alpha_{y_2 y_3 x}^{(3)} = \frac{n_{y_2 y_3 x}^{(3)}}{n_{y_2 y_3 x}^{(3)}} . \]

ACKNOWLEDGMENT

The authors are grateful to Dr. Keith Henry for sharing data from AIDS Clinical Trial Group (ACTG) Study 193A. Dr. Fitzmaurice is grateful for support provided by grant GM29745 from the U. S. National Institutes of Health.

REFERENCES


four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 19, 339-349.


Table 1. Proportional hazards regression for censoring on baseline CD4 count, for each therapy group in ACTG study 193A

<table>
<thead>
<tr>
<th>group</th>
<th>parameter</th>
<th>estimate</th>
<th>standard error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dual</td>
<td>log(CD4 + 1)</td>
<td>0.07158</td>
<td>0.04948</td>
<td>0.1480</td>
</tr>
<tr>
<td>triple</td>
<td>log(CD4 + 1)</td>
<td>0.17919</td>
<td>0.08513</td>
<td>0.0353</td>
</tr>
</tbody>
</table>

Table 2. Stratum specific estimates of the probability of survival at 18 months in the dual and triple therapy groups of the ACTG study 193A.

<table>
<thead>
<tr>
<th>strat-um*</th>
<th>dual therapy</th>
<th></th>
<th></th>
<th>triple therapy</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>percent censored†</td>
<td>estimated survival</td>
<td>standard error</td>
<td>percent censored†</td>
<td>estimated survival</td>
<td>standard error</td>
</tr>
<tr>
<td>1</td>
<td>60.7</td>
<td>0.5392</td>
<td>0.0418</td>
<td>1</td>
<td>66.0</td>
<td>0.5839</td>
</tr>
<tr>
<td>2</td>
<td>59.2</td>
<td>0.6119</td>
<td>0.0399</td>
<td>2</td>
<td>62.0</td>
<td>0.7357</td>
</tr>
<tr>
<td>3</td>
<td>57.9</td>
<td>0.5915</td>
<td>0.0407</td>
<td>3</td>
<td>58.7</td>
<td>0.7915</td>
</tr>
<tr>
<td>4</td>
<td>56.9</td>
<td>0.6949</td>
<td>0.0380</td>
<td>4</td>
<td>56.3</td>
<td>0.6584</td>
</tr>
<tr>
<td>5</td>
<td>55.9</td>
<td>0.7932</td>
<td>0.0339</td>
<td>5</td>
<td>53.2</td>
<td>0.8277</td>
</tr>
</tbody>
</table>

* based on quintiles of proportional hazards model for censoring, 0.07158 log (CD4+1)
** based on quintiles of proportional hazards model for censoring, 0.17919 log (CD4+1)
† median percent censored at 18 months within stratum