HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2020 September 30.

Published in final edited form as:

Stat Med. 2019 September 30; 38(22): 4453–4474. doi:10.1002/sim.8319.

Maximum likelihood estimation with missing outcomes: From simplicity to complexity

Stuart G. Baker

Division of Cancer Prevention, National Cancer Institute, Bethesda MD, USA

Abstract

Many clinical or prevention studies involve missing or censored outcomes. Maximum likelihood (ML) methods provide a conceptually straightforward approach to estimation when the outcome is partially missing. Methods of implementing ML methods range from the simple to the complex, depending on the type of data and the missing-data mechanism. Simple ML methods for ignorable missing-data mechanisms (when data are missing at random) include complete-case analysis, complete-case analysis with covariate adjustment, survival analysis with covariate adjustment, and analysis via propensity-to-be-missing scores. More complex ML methods for ignorable missing-data mechanisms include the analysis of longitudinal dropouts via a marginal model for continuous data or a conditional model for categorical data. A moderately complex ML method for categorical data with a saturated model and either ignorable or nonignorable missing-data mechanisms is a perfect fit analysis, an algebraic method involving closed-form estimates and variances. A complex and flexible ML method with categorical data and either ignorable or nonignorable missing-data mechanisms is the method of composite linear models, a matrix method requiring specialized software. Except for the method of composite linear models, which can involve challenging matrix specifications, the implementation of these ML methods ranges in difficulty from easy to moderate.

Keywords

composite linear model; double sampling; latent class instrumental variable; missing-data mechanism; perfect fit analysis; randomized trial

1. INTRODUCTION

In many clinical or prevention studies the outcome is missing or censored. Maximum likelihood (ML) methods are a conceptually simple approach for estimation in this setting. The landmark 1976 paper by Rubin[1] made several key innovations for ML estimation with missing data: a missing-data indicator as a random variable, a comprehensive likelihood framework, and the concept of ignorable and nonignorable missing-data mechanisms. Wu

Correspondence Stuart G. Baker, Divisions of Cancer Prevention, National Cancer Institute, 9609 Medical Center Dr 5E638 Bethesda MD 20892-9789., sb16i@nih.gov.

DATA AVAILABILITY STATMENT

The data used in the analyses are available in the tables of the paper.

and Carroll,[2] Heitjan and Rubin,[3] and Little and Rubin,[4] extended this approach to censoring mechanisms.

The basic set-up follows. The goal of the analysis is to estimate parameters in an outcome model, a model for the effect of treatment or covariates on outcome. Coupled with the outcome model is a missing-data mechanism, a model for the probability that the outcome is missing or censored. The sets of parameters for the outcome model and the missing-data mechanism do not overlap and do not constrain each other.

In the context of likelihood-based inference, an ignorable missing-data mechanism is a missing-data mechanism whose parameters factor from the likelihood and hence do not contribute to likelihood-based inference for the outcome model. Rubin[1] showed that an ignorable missing-data mechanism depends only on completely observed variables, in which case the data are said to be Missing at Random (MAR). A special case of MAR is Missing Completely at Random (MCAR), corresponding to a constant probability the data are missing.

A nonignorable missing-data mechanism is simply a missing-data mechanism that is not ignorable. This tutorial introduces the terminology of directly and indirectly nonignorable missing-data mechanisms. A directly nonignorable missing-data mechanism is a nonignorable missing-data mechanism in which the probability of missing a variable depends on that variable and possibly on other variables. An indirectly ignorable missing-data mechanism is a missing-data mechanism in which the probability of missing a variable does *not* depend on that variable but depends on at least one other variable that is partially missing. Table 1 summarizes this missing-data taxonomy in the context of missing outcomes.

Implementation of ML methods with missing outcomes can range from simple computations to complex modeling with specialized software. Because ML methods are often tailored to specific missing-data scenarios and there are numerous missing-data scenarios, it is not possible to cover all ML methods here. Table 2 lists the ML methods discussed in this tutorial.

2. COMPLETE-CASE ANALYSIS

Consider a randomized trial in which missing in univariate outcome Y depends on randomization group Z. As an example, missing in outcome depends on side effects of the experimental treatment. Complete cases are participants who are not missing the outcome. For this scenario, the ML method is a complete case analysis, an analysis involving only complete cases. Separate derivations involve continuous and binary outcomes.

2.1 Continuous outcomes

Let subscript i index trial participant. Let Y_i denote the outcome with realization y_i. Let $MissY_i$ denote the missing-data indicator, where $MissY_i = 1$ if y_i is missing and 0 otherwise. Let $\{MissY\}$ and $\{ObsY\}$ denote the set of persons with missing and observed outcomes, respectively. Let Z_i denote the randomly assigned group with realization z_i. The outcome

model, $pr(y_i|z_i;\ \theta)$, is the distribution of outcome y_i given randomization to group z_i, which is modeled by parameter set θ. The missing-data mechanism, $pr(MissY_i = 1|z_i;\ \beta)$, is the probability of missing outcome Y_i given randomization to group z_i, which is modeled by parameter set β. By definition, $pr(MissY_i = 0|z_i;\ \beta) = 1 - pr(MissY_i = 1|z_i;\ \beta)$. An example of this missing-data mechanism is $pr(MissY_i = 1|Z_i = 0;\ \beta) = \beta_0 = 1/2$ for participants randomized to group 0, and $pr(MissY_i = 1|Z_i = 1;\ \beta) = \beta_1 = 1/3$ for participants randomized to group 1, where $\beta = \{\beta_0\ \beta_1\}$. The parameter sets θ and β do not overlap and do not constrain one another.

The likelihood is the product of a factor for participants missing outcome, L_{MissY}, and a factor for participants with observed outcome, L_{ObsY},

$$\begin{aligned} Lik_{CC}(\theta,\ \beta) &= L_{MissY} \times L_{ObsY}, \text{ where} \\ L_{MissY} &= \prod_{i \in \{MissY\}} \int pr(MissY_i = 1|z_i;\ \beta) \times pr(y_i|z_i;\ \theta) dy_i \\ &= \prod_{i \in \{MissY\}} pr(MissY_i = 1|z_i;\ \beta), \\ L_{ObsY} &= \prod_{i \in \{ObsY\}} pr(MissY_i = 0|z_i;\ \beta) \times pr(y_i|z_i;\ \theta) \end{aligned} \tag{1}$$

The factor L_{MissY} integrates over the missing continuous outcome. Rewriting the likelihood in equation (1) by defining $f_{CC}(\beta)$ as a function of parameters involving only β and defining $Lik_{CC:Ign}(\theta)$ as a function of parameters involving only θ yields

$$\begin{aligned} &Lik_{CC}(\theta,\ \beta) = f_{CC}(\beta) Lik_{CC:Ign}(\theta), \text{ where} \\ &f_{CC}(\beta) = \prod_{i \in \{MissY\}} pr(MissY_i = 1|z_i;\ \beta) \times \prod_{i \in \{ObsY\}} pr(MissY_i = 0|z_i;\ \beta), \\ &Lik_{CC:Ign}(\theta) = \prod_{i \in \{ObsY\}} pr(y_i|z_i;\ \theta). \end{aligned} \tag{2}$$

Because $f_{CC}(\beta)$ factors from the likelihood in equation (2), the missing-data mechanism is ignorable, so ML estimation for θ involves only $Lik_{CC:Ign}(\theta)$. Moreover, because $Lik_{CC:Ign}(\theta)$ involves only observed values of outcome, ML estimation for θ involves only complete cases.

2.2 Example 1

A hypothetical trial randomizes participants to dietary supplement or placebo. The outcome is a continuous biomarker. Missing in outcome depends only on randomization group. If the biomarker is normally distributed with a different mean for each randomization group, a simple ML estimate for the effect of treatment on outcome is the difference in mean biomarkers levels between randomization groups among the complete cases.

2.3 Binary outcomes

A similar derivation applies to binary outcomes. Let n_{zy} denote the number of persons randomized to group z =0, 1, with observed outcome y=0, 1. Let w_z denote the number of persons randomized to group z =0, 1, with a missing outcome. See Table 3. The outcome model, $pr(Y = 1|z;\ \theta) = \theta_z$, is the probability of outcome 1 given randomization to group z. The missing-data mechanism, $pr(MissY = 1|z;\ \beta) = \beta_z$, is the probability of missing outcome y given randomization group z. The likelihood with $\beta = \{\beta_0, \beta_1\}$ and $\theta = \{\theta_0, \theta_1\}$ is

$$\begin{aligned} Lik_{CC}(\theta,\ \beta) &= L_{MissY} \times L_{ObsY}, \text{ where} \\ L_{MissY} &= \prod_z \left\{ \beta_z (1 - \theta_z) + \beta_z\ \theta_z \right\}^w \\ &= \prod_z \beta_z^{\ w_z} \\ L_{ObsY} &= \prod_z \left\{(1 - \beta_z)(1 - \theta_z)\right\}^{n_{z0}} \times \left\{(1 - \beta_z)\,\theta_z\right\}^{n_{z1}}. \end{aligned} \tag{3}$$

The factor L_{MissY} sums over the missing binary outcomes. Let "+" in a subscript denote summation over the index in the subscript, so $n_{z+} = n_{z0} + n_{z1}$. Rewriting the likelihood in equation (3) yields

$$\begin{aligned} Lik_{CC}(\theta,\ \beta) &= f_{CC}(\beta) Lik_{CC:Ign}(\theta), \text{ where} \\ f_{CC}(\beta) &= \prod_z \beta_z^{\ w_z} \times (1 - \beta_z)^{n_{z+}}, \\ Lik_{CC:Ign}(\theta) &= \prod_z (1 - \theta_z)^{n_{z0}} \times \theta_z^{\ n_{z1}}. \end{aligned} \tag{4}$$

ML estimation for θ comes from $Lik_{CC:Ign}(\theta)$, which involves only the complete cases $\{n_{zy}\}$.

2.4 Example 2

A hypothetical trial randomizes participants to dietary supplement or placebo. The outcome is a binary biomarker. Missing in outcome depends only on randomization group. For this scenario, a simple ML estimate of treatment effect is $d = \theta_{(EST)1} - \theta_{(EST)0}$, where $\theta_{(EST)z} = n_{z0}/n_{z+}$. The estimated standard error is $se = \sqrt{v}$, where $v = \sum_z \theta_{(EST)z}(1 - \theta_{(EST)z})/n_{z+}$. For the hypothetical counts in Table 3, d= 0.150 with standard error 0.022.

3. COMPLETE-CASE ANALYSIS WITH COVARIATE ADJUSTMENT

Consider a randomized trial in which missing in outcome Y depends on randomization group Z and baseline covariate X. If the covariate X is not included in the outcome model, the missing-data mechanism is nonignorable leading to challenging ML estimation. The simple expedient of conditioning on the baseline covariate X in the outcome model yields an

ignorable likelihood and simple ML estimation based on complete cases with covariate adjustment. Separate derivations involve continuous and binary outcomes.

3.1 Continuous outcomes

Let X_i with realization x_i denote the covariate for person i. The outcome model, $pr(y_i|z_i, x_i;\ \theta)$, is the distribution of outcome y_i given randomization to group z_i, and covariate x_i. The missing-data mechanism, $pr(MissY = 1|z_i, x_i;\ \beta)$, is the probability of missing outcome y_i given covariate x_i and randomization to group z_i. For example, suppose the probability of missing outcome due to a side effect of treatment is highest among participants in randomization group 1 who are age 60 or older at randomization. Let X_i =0 if age at randomization is less 60, and 0 otherwise. An example of this missing-data mechanism is
$pr(MissY_i = 1|Z_i = 0, X_i = 0,\ \beta) = \beta_{00} = 1/5, pr(MissY_i = 1|Z_i = 0, X_i = 1,\ \beta) = \beta_{01} = 1/5, pr(MissY_i = 1|Z_i = 1, X_i = 0,\ \beta) = \beta_{10} = 1/5, pr(MissY_i = 1|Z_i = 1, X_i = 1,\ \beta) = \beta_{11} = 1/2$
, where $\beta = \{\beta_{00}\ \beta_{01}\ \beta_{10}\ \beta_{11}\}$. The parameter sets θ and β do not overlap and do not constrain one another. The likelihood is

$$\begin{aligned} &Lik_{CCX}(\theta,\ \beta) = L_{MissY} \times L_{ObsY}, \text{ where} \\ &L_{MissY} = \prod_{i \in \{MissY\}} \int pr(MissY_i = 1|z_i, x_i;\ \beta) \times pr(y_i|z_i, x_i;\ \theta) dy_i \\ &\qquad = \prod_{i \in \{MissY\}} pr(MissY_i = 1|z_i, x_i;\ \beta), \\ &L_{ObsY} = \prod_{i \in \{ObsY\}} pr(MissY_i = 0|z_i, x_i;\ \beta) \times pr(y_i|z_i, x_i;\ \theta). \end{aligned} \tag{5}$$

Rewriting the likelihood in equation (5) by defining $f_{CCX}(\beta)$ as a function of parameters involving only β and defining $Lik_{CCX:Ign}(\theta)$ as a function of parameters involving only θ yields

$$\begin{aligned} &Lik_{CCX}(\theta,\ \beta) = f_{CCX}(\beta) \times Lik_{CCX:Ign}(\theta), \text{ where} \\ &f_{CCX}(\beta) = \prod_{i \in \{MissY\}} pr(MissY_i = 1|z_i, x_i;\ \beta) \prod_{i \in \{ObsY\}} pr(MissY_i = 0|z_i, x_i;\ \beta) \\ &Lik_{CCX:Ign}(\theta) = \prod_{i \in ObsY} pr(y_i|x_i, z_i;\ \theta). \end{aligned} \tag{6}$$

Because $f_{CCX}(\beta)$ factors from the likelihood, the missing-data mechanism is ignorable. Moreover, because $Lik_{CCX:Ign}(\theta)$ involves only observed values of outcome, ML estimation of θ involves only complete case with covariates.

If X is partially MCAR, the likelihood based on all the data is indirectly nonignorable, leading to challenging ML estimation. However, the simple of expedient of considering only the random subset of the data with the observed covariate X yields a likelihood factor involving only θ, a result related to the formulation of Little et al.[5]. Moreover, this likelihood factor involves only complete cases with observed values of covariate X. See Appendix A.

3.2 Example 1

A hypothetical trial randomizes participants to dietary supplement or placebo. The outcome is a continuous biomarker. Missing in outcome depends only on randomization group and age. Under this scenario, ML estimation can involve fitting to the complete cases a linear regression for the biomarker as a function of randomization group and age. The estimated treatment effect is the estimated coefficient for randomization group in the linear regression.

3.3 Binary outcomes

Consider a binary outcome and categorical baseline covariate. Suppose that missing in outcome depends only on treatment group and covariate. Let n_{zxy} denote the number of persons randomized to group z =0, 1 with baseline covariate x = 0, 1 and observed outcome y=0, 1. Let w_{zx} denote the number of persons with randomized to group z =0, 1 with baseline covariate x=0, 1 who had a missing outcome. See Table 4. . Let $pr(Y = 1|z, \quad x; \theta) = \theta_{zx}$ denote the probability of outcome 1 given randomization to group z. Let $pr(MissY = 1|z, x; \beta) = \beta_{zx}$ denote the probability of missing outcome given randomization group z. The likelihood with $\beta = \{\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}\}$ and $\theta = \{\theta_{00}, \theta_{00}, \theta_{10}, \theta_{11}\}$ is

$$\begin{aligned} Lik_{CCX}(\theta, \beta) &= L_{MissY} \times L_{ObsY}, \text{ where} \\ L_{MissY} &= \prod_z \prod_x \left\{ \beta_{zx}(1 - \theta_{zx}) \quad + \beta_{zx} \ \theta_{zx} \right\}^{w_{zx}} \\ &= \prod_z \prod_x \beta_{zx}{}^{w_{zx}}, \\ L_{ObsY} &= \prod_z \prod_x \left\{(1 - \beta_{zx})(1 - \theta_{zx})\right\}^{n_{zx0}} \times \left\{(1 - \beta_{zx})\,\theta_{zx}\right\}^{n_{zx1}}. \end{aligned} \tag{7}$$

Rewriting the likelihood in equation (7) yields

$$\begin{aligned} Lik_{CCX}(\theta, \beta) &= f_{CCX}(\beta) Lik_{CCX:Ign}(\theta), \text{ where} \\ f_{CCX}(\beta) &= \prod_z \prod_x (1 - \beta_{zx})^{w_{zx}} \times \beta_{zx}{}^{n_{zx+}}, \\ Lik_{CCX:Ign}(\theta) &= \prod_z \prod_x (1 - \theta_{zx})^{n_{zx0}} \times \theta_{zx}{}^{n_{zx1}}. \end{aligned} \tag{8}$$

Because $Lik_{CCX:Ign}(\theta)$ involves only observed values of Y, ML estimation of θ involves only complete case with covariates.

3.4 Example 2

A hypothetical trial randomizes participants to dietary supplement or placebo. The outcome is a continuous biomarker. Missing in outcome depends only on randomization group and a categorical covariate. Let π_x denote the known probability the covariate takes value x in a target population. An ML estimate of treatment effect in the target population is $d = \sum_x (\theta_{(EST)1x} - \theta_{(EST)0x})\,\pi_x$, where $\theta_{zx(EST)} = n_{zx0}/n_{zx+}$. The estimated standard error is

$se = \sqrt{v}$, where $v = \sum_z \sum_x \left\{ \theta_{(EST)zx}(1 - \theta_{(EST)zx})/n_{zx+} \right\} \pi_x^2$ For the counts in Table 4 with $\pi_X = 0.5$, $d = 0.114$ and standard error 0.023.

4. SURVIVAL ANALYSIS WITH COVARIATE ADJUSTMENT

Consider a randomized trial where outcomes are survival times and censoring depends on randomization group Z and baseline covariate X. Let F denote the failure time in the absence of censoring, and let C denote censoring time in the absence of failure. Censoring at time c implies F occurs at time c or later, and failure at time f implies C occurs after time f. Let $pr(F_i = f_i | z_i, x_i; \theta)$ denote the probability of failure (in the absence of censoring) at time f_i, given randomization group z_i and covariate x_i. Let $pr(C_i = c_i | z_i, x_i; \beta)$ denote the probability of censoring (in the absence of failure) at time t_i, given randomization group z_i and covariate x_i. The parameter sets θ and β do not overlap and do not constrain one another.

If the covariate X is not included in the outcome model, the censoring mechanism is nonignorable. The simple expedient of including X in the outcome model leads to an ignorable censoring mechanism. The likelihood is

$$
\begin{aligned}
Lik_{SurvX}(\theta, \beta) &= L_{Cens} \times L_{Fail}, \quad \text{where} \\
L_{Cens} &= \prod_{i \in \{Cens\}} \int_{c_i}^{\infty} pr(C_i = c_i | z_i, x_i; \beta) \times pr(F_i = f_i | z_i, x_i; \theta) df_i \\
&= \prod_{i \in \{Cens\}} pr(C_i = c_i | z_i, x_i; \beta) \times pr(F \geq c_i | z_i, x_i; \theta), \\
L_{Fail} &= \prod_{i \in \{Fail\}} \int_{f_i}^{\infty} pr(C_i = c_i | z_i, x_i; \beta) \times pr(F_i = f_i | z_i, x_i; \theta) dc_i \\
&= \prod_{i \in \{Fail\}} pr(C_i > f_i | z_i, x_i; \beta) \quad \times pr(F_i = f_i | z_i, x_i; \theta).
\end{aligned}
\tag{9}
$$

The factor L_{Cens} integrates over the unobserved failure times. The factor L_{Fail} integrates over the unobserved censoring times. Rewriting the likelihood in equation (9) by defining $f_{SurvX}(\beta)$ as a function of parameters involving only β and defining $Lik_{SurvX:Ign}(\theta)$ as a function of parameters involving only θ yields

$$
\begin{aligned}
Lik_{SurvX}(\theta, \beta) &= f_{SurvX}(\beta) \times Lik_{SurvX:Ign}(\theta), \quad \text{where} \\
f_{SurvX}(\beta) &= \prod_{i \in \{Cens\}} pr(C_i = c_i | z_i, x_i; \beta) \\
&\quad \times \prod_{i \in \{Fail\}} pr(C_i > f_i | z_i, x_i; \beta), \\
Lik_{SurvX:Ign}(\theta) &= \prod_{i \in \{Cens\}} pr(F \geq c_i | z_i, x_i; \theta) \\
&\quad \times \prod_{i \in \{Fail\}} pr(F = f_i | z_i, x_i; \theta).
\end{aligned}
\tag{10}
$$

Because $f_{SurvX}(\beta)$ factors from the likelihood in equation (10), the censoring mechanism is ignorable, and ML estimation of θ involves only $Lik_{SurvX:Ign}(\theta)$.

If X is partially MCAR, the likelihood based on all the data is indirectly nonignorable, making ML estimation difficult. However, the simple of expedient of considering only a random subset of the data with the observed covariate X leads to a likelihood with the covariate that involves only θ. See Appendix B.

4.2 Example

A hypothetical trial randomizes participants negative on a biomarker to dietary supplement or placebo. The outcome is time until the biomarker is positive. Loss-to-follow-up depends only on randomization group and age. ML estimation can involve fitting a proportional hazards model in which the hazard for failure depends on randomization group and age. The estimated treatment effect is the estimated coefficient for randomization group in the model.

5. PROPENSITY-TO-BE-MISSING SCORES

The method of propensity-to-be-missing scores[6] simplifies a complete-case analysis or a survival analysis when adjusting for multiple baseline covariates. It also avoids having to specify a function for incorporating multiple covariates into the outcome model and yields an easily interpretable difference estimate. The method of propensity-to-be-missing scores involves the following three steps.

Step 1. *Fit a separate model to the missing-data mechanism in each randomization group.* For a univariate outcome with randomization group z, fit a model for the missing-data mechanism, $pr(MissY_i = 1|Z_i = z, x_i; \beta_z)$. For a survival outcome with randomization group z, fit a model for the censoring mechanism, $pr(C_i = c_i|Z_i = z, x_i; \beta_z)$. For a proportional hazards model for the censoring mechanism in randomization group z, let $c^*(z, x_i; \beta_z)$ denote the proportionality component of the model, where the other component is the baseline hazard for censoring. Let $\beta_{(EST)z}$ denotes the estimate of β_z

Step 2. Compute propensity-to-be-missing scores. For a univariate outcome, let $score_{zi} = pr(MissY_i = 1|z_i, x_i; \beta_{(EST)z})$. For a survival outcome with a proportional hazards model for censoring, let $score_{zi} = c^*(z_i, x_i; \beta_{(EST)z})$.

Step 3. Compute estimated treatment effect and its standard error based on estimates in each quintile of scores. Divide the set of scores for each randomization group z, $\{score_{zi}\}$, into quintiles. For randomization group z and quintile j, let f_{zj} denote the estimated probability of outcome or the probability of survival to a pre-specified time. Let se_{zj} denote the estimated standard error of f_{zj}. Let N_z denote the number in randomization group z. The estimated treatment effect is the treatment effect averaged over the quintiles,

$$d = \sum_j (f_{1j} - f_{0j})/5 = (\sum_j f_{1j} - \sum_j f_{0j})/5 . \quad (11)$$

The estimated standard error of d is $se = \sqrt{v}$, where

$$v = \sum_z \left\{ se_{zj}^2/25 + \sum_j (f_{zj} - f_{z5})^2 4/(25N_z) - \sum_{j>k} 2(f_{zj} - f_{z5})/(25N_z) \right\}. \quad (12)$$

5.1 Example

The AIDS Clinical Trials Group randomized patients to dual therapy (z=0) versus triple therapy (z=1) into groups of equal size of N_z= 328.[7] Let d denote the estimated difference in survival to 18 months with triple instead of dual therapy. Approximately 20% of subjects were missing outcomes due to refusal to continue the study or loss-to-follow-up. Two baseline covariates, age and CD4 count, are likely related to both survival and dropout. Following Baker et al.,[6] let f_{zj} denote the Kaplan-Meier estimate of the probability of surviving 18 months among participants in quintile j of the scores in randomization group z. Let se_{zj} denote the estimated standard error of f_{zj}. Substituting the values f_{zj} and se_{zj} from Table 5 into the equations (11) and (12) gives d = 0.72 with standard error 0.34.

6. LONGITUDINAL DROPOUTS

Consider a randomized trial involving longitudinal outcomes in which dropout depends on previously observed outcomes and possibly randomization group and covariates. For example, participants with an unfavorable outcome at a previous time may be more likely to drop out than those with a favorable outcome at a previous time. ML estimation involving this ignorable missing-data mechanism is discussed separately for continuous and binary outcomes

6.1 Continuous outcome

Without loss of generalizability, consider outcomes at three times, denoted Y_1, Y_2, and Y_3, with Y_1 always observed. The outcome model, $pr(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta)$, is the joint distribution of outcomes y_{1i}, y_{2i}, and y_{3i} given randomization to group z_i with covariate x_i. The covariate x_i could be a baseline covariate or covariate that varies over time in a predetermined manner, such as time of observation. The missing-data mechanism $pr(MissY_{2i} = 1 | y_{1i}, z_i, x_i; \beta)$, is the probability of missing outcome Y_{2i} given outcome y_{1i}, randomization to group z_i, and covariate x_i. The missing-data mechanism, $pr(MissY_{3i} = 1 | MissY_{2i} = 0, y_{1i}, y_{2i}; \beta)$, is the probability of missing outcome Y_{3i} given not missing outcome Y_{2i}, outcome y_{2i}, outcome y_{1i}, randomization to group z_i, and covariate x_i. The parameter sets θ and β do not overlap and do not constrain one another. The probabilities of dropout at time 2, dropout at time 3, and no dropout are, respectively,

$$\begin{aligned} &f_{drop2}(y_{1i}, z_i, x_i; \beta) = pr(MissY_{2i} = 1 | y_{1i}, z_i, x_i; \beta), \\ &f_{drop3}(y_{1i}, y_{2i}, z_i, x_i; \beta) = pr(MissY_{3i} = 1 | MissY_{2i} = 0, y_{1i}, y_{2i}, z_i, x_i; \beta) \times pr \\ &(MissY_{2i} = 0 | y_{1i}, z_i, x_i; \beta), \\ &f_{nodrop}(y_{1i}, y_{2i}, z_i, x_i; \beta) = pr(MissY_{3i} = 0 | MissY_{2i} = 0, y_{1i}, y_{2i}, z_i, x_i; \beta) \\ &\qquad \times pr(MissY_{2i} = 0 | y_{1i}, z_i, x_i; \beta). \end{aligned} \quad (13)$$

The likelihood is the product of three factors corresponding to the three subsets of participants defined by dropouts, $\{DropOutTime_2\}$, $\{DropOutTime3\}$, and $\{NoDropOut\}$,

$$
\begin{aligned}
&L_{LD}(\theta, \beta) = L_{DropOutTime2} \times L_{DropoutTime3} \times L_{NoDropOut}, \quad \text{where} \\
&\quad L_{DropOutTime2} = \prod\nolimits_{i \in \{DropOutTime2\}} \int\int f_{drop2}(y_{1i}, z_i, x_i; \beta) \times pr \\
&(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta) dy_{3i} dy_{2i}, \\
&\quad L_{DropOutTime3} = \prod\nolimits_{i \in \{DropOutTime3\}} \int\int f_{drop3}(y_{1i}, y_{2i}, z_i, x_i; \beta) \\
&\times pr(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta) dy_{3i}, \\
&\quad L_{NoDropOut} = \prod\nolimits_{i \in \{NoDropOut\}} f_{nodrop}(y_{1i}, y_{2i}, z_i, x_i; \beta) \times pr(y_{1i}, y_{2i}, \\
&y_{3i} | z_i, x_i; \theta).
\end{aligned}
\tag{14}
$$

Rewriting the likelihood in equation (14) by defining $f_{LD}(\beta)$ as a function of parameters involving only β and defining $L_{LD:Ign}(\theta)$ as a function of parameters involving only θ yields

$$
\begin{aligned}
&L_{LD}(\theta, \beta) = f_{LD}(\beta) L_{LD:Ign}(\theta), \quad \text{where} \\
&\quad f_{LD}(\beta) = \prod\nolimits_{i \in \{DropOutTime2\}} f_{drop2}(y_{1i} | z_i, x_i; \beta) \\
&\times \prod\nolimits_{i \in \{DropOutTime3\}} f_{drop3}(y_{1i}, y_{2i} | z_i, x_i; \beta) \\
&\quad \times \prod\nolimits_{i \in \{NoDropOut\}} f_{nodrop}(y_{1i}, y_{2i} | z_i, x_i; \beta), \\
&L_{LD:Ign}(\theta) = \prod\nolimits_{i \in \{DropOutTime2\}} \int\int pr(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta) dy_{3i} dy_{2i} \\
&\quad \times \prod\nolimits_{i \in \{DropOutTime3\}} \int pr(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta) dy_{3i} \\
&\quad \times \prod\nolimits_{i \in \{NoDropOut\}} pr(y_{1i}, y_{2i}, y_{3i} | z_i, x_i; \theta).
\end{aligned}
\tag{15}
$$

ML estimation of θ in $L_{LD:Ign}(\theta)$ in equation (15) typically involves a marginal outcome model in which outcome at each time is a function of time, treatment, and covariates, but not previous outcomes.

6.2 Example 1

A standard marginal outcome model assumes a multivariate normal distribution with a model for the mean outcome at each time and a structured variance covariance matrix arising from random effects or temporal correlations.[8] Using the commercial software SAS Proc Mixed,[9] Allison[10] fit a multivariate normal model marginal model to continuous longitudinal outcomes with dropout. In Allison's model, the longitudinal outcome was the logarithm of hourly wage and covariates were sex and year.

6.3 Binary outcomes

With longitudinal binary outcomes, a conditional model is often easier to implement than a marginal model. Without loss of generality consider 3 times. For simplicity of notation, covariates are implicit. Let y_j denote the binary outcome at time j. Let $n_1(y_1)$ denote the number of participants who dropped out after outcome y_1. Let $n_2(y_1, y_2)$ denote the number of participants who dropped out after outcomes y_1 and y_2. Let $n_3(y_1, y_2, y_3)$ denote the number of participants with observed outcomes y_1, y_2, and y_3. The conditional model factors the joint distribution of outcomes as $pr(y_1, y_2, y_3; \theta) = pr(y_1; \theta) \times pr(y_2 | y_1; \theta) \times pr(y_3 | y_2, y_1; \theta)$. With obvious extension of notation from the continuous outcome scenario, the likelihood is

$$
\begin{aligned}
&L_{LD}(\theta, \beta) = L_{DropOutTime2} \times L_{DropoutTime3} \times L_{NoDropOut}, \quad \text{where} \\
&L_{DropOutTime2} = \prod_{y1} \{ f_{drop2}(y_1; \beta) \times pr(y_1; \theta) \}^{n_1(y_1)}, \\
&L_{DropoutTime3} = \prod_{y1} \prod_{y2} \{ f_{drop3}(y_1, y_2; \beta) \times pr(y_1; \theta) \times pr(y_2 | y_1; \theta) \}^{n_2(y_1, y_2)}, \\
&L_{NoDropOut} = \prod_{y1} \prod_{y2} \prod_{y3} \{ f_{nodrop}(y_1, y_2; \beta) \times pr(y_1; \theta) \times pr(y_2 | y_1; \theta) \\
&\qquad \times pr(y_3 | y_2, y_1; \theta) \}^{n_3(y_1, y_2, y_3)}.
\end{aligned}
\tag{16}
$$

Rewriting the likelihood in equation (16) gives

$$
\begin{aligned}
&L_{LD}(\theta, \beta) = f_{LD}(\beta) L_{LD:Ign}(\theta), \quad \text{where} \\
&f_{LD}(\beta) = \prod_{y1} \prod_{y2} \prod_{y3} f_{drop2}(y_1; \beta)^{n_1(y_1)} \times f_{drop3}(y_1, y_2; \beta)^{n_2(y_1, y_2)} \\
&\times f_{nodrop}(y_1, y_2; \beta)^{n_3(y_1, y_2, +)}, \\
&L_{LD:Ign}(\theta) = \prod_{y1} pr(y_1; \theta)^{n_1(y_1) + n_2(y_1, +) + n_3(y_1, +, +)} \\
&\qquad \times \prod_{y1} \prod_{y2} pr(y_2 | y_1; \theta)^{n_2(y_1, y_2) + n_3(y_1, y_2, +)} \\
&\qquad \times \prod_{y1} \prod_{y2} \prod_{y3} pr(y_3 | y_2, y_1; \theta)^{n_3(y_1, y_2, y_3)}.
\end{aligned}
\tag{17}
$$

ML estimation can involve fitting a logit model to the various factors and then combining estimates. A more parsimonious outcome model that conditions only on the previous outcome for times after the first would have a likelihood, with $\theta = \{\theta_1, \theta_2\}$,

$$
\begin{aligned}
L_{LD:Ign*}(\theta) = &\prod_{y1} pr(y_1; \theta_1)^{n_1(y_1) + n_2(y_1, +) + n_3(y_1, +, +)} \\
&\times \prod_{y1} \prod_{\mathbf{y2}} pr(y_2 | y_1; \theta_2)^{n_2(y_1, y_2) + n_3(y_1, y_2, +)} \\
&\times \prod_{\mathbf{y2}} \prod_{\mathbf{y3}} pr(y_3 | y_2; \theta_2)^{n_3(+, y_2, y_3)}.
\end{aligned}
\tag{18}
$$

6.4 Example 2

A false positive (FP) on cancer screening is a positive screening test followed by a negative work-up or biopsy. The goal is to estimate the probability of least one FP in a program of screens when the number of screens received varies among participants. A convenient simplification uses screen number instead of time as the longitudinal metric, so all missing FP's are dropouts. For example, receiving screens at times 1 and 3 and missing the screen at time 2 corresponds to receiving screens 1 and 2 and then dropping out. Following Baker et al.,[11] let outcome Y_j denote FP status (0 for no FP or 1 for FP) if screen j were received, which is missing when screen j is missing. Missing a screen likely depends on the FP status of the previous screens and possibly observed covariates, so the missing-data mechanism is ignorable.

Consider the count data in Table 6 from Baker et al.[11], which corresponds to ages 50 to 54 at first screen. As will become apparent, for the goal of estimating the probability of at least one FP in a screening program, it is only necessary to consider participants with no FP on the previous screen. One covariate is time interval x_j since last screen, with $x_j = 1$ (9–12 months), 2 (13–15 months), or 3 (16–18 months). A second covariate is screen number j. Let m_y denote the number of participants with outcome y on screen 1. For $j>1$, let n_{jxy} denote the number of participants with outcome y on screen j among participants with outcome $Y=0$= no FP on screen j–1 and for whom screen j occurred at time interval x_j since screen j–1. See Table 6. Based on an extension of equation (18), the likelihood factor involving $\theta = \{\theta_1, \theta_2\}$is

$$L_{LD:Ign}(\theta) = \prod_y pr(Y_1 = y;\ \theta_1)^{m_y} \times \prod_{j>1} \prod_y pr(Y_j = y \mid Y_{(j-1)} = 0, x, j;\ \theta_2)^{n_{jxy}}. \tag{19}$$

ML estimation can include fitting a logit model, $pr(Y_j = 1 \mid Y_{(j-1)} = 0, x, j;\ \theta_2) = expit(\theta_{20} + \theta_{21} j + \theta_{22} x_j)$. The resulting estimates, $\theta_{(EST)21} = 0.23$ with standard error 0.15 and $\theta_{(EST)22} = 0.017$ with standard error of 0.15, suggest a more parsimonious model, $pr(Y_j = 1 \mid Y_{(j-1)} = 0;\ \theta_2) = \theta_2$. Let $pr(Y_1 = 1;\ \theta_1) = \theta_1$. The ML estimates are $\theta_{(EST)1} = m_1/m_+$ and $\theta_{(EST)2} = n_{++1}/n_{+++}$. The estimated probability of at least one false positive in 5 screens is $r = 1 - (1 - \theta_{(EST)1})(1 - \theta_{(EST)2})^4$ with standard error $se = \sqrt{v}$, where $v = (\partial r/\theta_{(EST)1})^2 \theta_{(EST)1}(1 - \theta_{(EST)1})/m_+ + (\partial r/\theta_{(EST)2})^2 \theta_{(EST)2}(1 - \theta_{(EST)2})/n_{+++}$. Based on the counts in Table 6. $\theta_{(EST)1} = 0.0179$, $\theta_{(EST)2} = 0.0069$, and $r = 0.045$ with standard error 0.004.

7. PERFECT FIT ANALYSIS

A perfect fit analysis is an algebraic method of ML estimation with partially observed categorical data and a saturated model. In a saturated model the number of independent parameters equals number of independent cell counts. The advantage of using a saturated

model is that it makes as few assumptions as possible. A perfect fit analysis involves the following steps:

1. Set observed counts equal to expect counts and solve for closed-form parameter estimates.
2. Compute the statistic of interest from the parameter estimates.
3. Compute the standard error using the Multinomial-Poisson (MP) transformation.

The MP transformation[12] changes a complicated multinomial likelihood into a simpler likelihood for Poisson random variables with the same the ML estimates and variances. Let$\{n_u\}$ denote the set of observed counts, indexed by u. Let d denote the statistic from the perfect fit analysis, which is closed-form function of $\{n_u\}$. With a saturated model, the MP transformation treats n_u as a Poisson random variable with mean and variance equal to n_u. Applying the delta method, the estimated variance of d is

$$varMP(d) = \sum_u (\partial d / \partial n_u)^2 n_u, \tag{20}$$

a quantity easily calculated using symbolic computing. A caveat of the perfect fit analysis is that the parameter estimates are ML only if the parameter estimates lie in the interior of the parameter space.

7.1 Example 1

The Prostate Cancer Prevention Trial randomized participants to placebo (z=0) or finasteride (z=1). [13] One outcome of interest was prostate cancer status determined on biopsy (y = 0 = n0, y = 1 = yes), which is missing if there is no biopsy. An auxiliary variable is a variable observed after randomization that is related to outcome. Biopsy recommendation based on a test for prostate specific antigen (a= 0 = no or a = 1 = yes) is an auxiliary variable which is strongly related to the probability of missing the outcome. Incorporating this auxiliary variable into the model improves the adjustment for missing outcomes. Let n_{zay} denote the number of participants in randomization group z with auxiliary variable a and observed outcome y. Let w_{za} denote the number of participants in randomization group z with observed auxiliary variable a and missing outcome. See Table 7.

The outcome model is $pr(Y = y|z) = \theta_{y|z}$. The auxiliary variable model, $pr(A = a|y, z) = \lambda_{a|zy}$, is the probability of auxiliary variable a given outcome y and randomization group z. The missing-data mechanism is $pr(MissY = 1|a, z) = \beta_{za}$. The model is saturated because there are 10 independent parameters (2 for $\theta_{1|z}$, 4 for $\lambda_{1|zy}$, and 4 for β_{za}) and 10 independent cell counts (8 for $\{n_{zay}\}$and 4 for $\{w_{za}\}$ minus 2 because $n_{z++} + w_{z+}$ is fixed). The likelihood is

$$Lik(\theta, \beta) = L_{MissY} \times L_{ObsY}, \quad \text{where}$$
$$L_{MissY} = \prod_z \prod_a \left(\sum_y \theta_{y|z}\, \lambda_{a|zy}\, \beta_{za}\right)^{w_{za}}, \tag{21}$$
$$L_{ObsY} = \prod_z \prod_a \prod_y \left\{\theta_{y|z}\, \lambda_{a|zy}(1 - \beta_{za})\right\}^{n_{zay}}.$$

The perfect fit analysis yields ML estimates without numerical maximization of the likelihood in equation (21). It involves the following steps.

Step 1. *Set observed counts equal to expect counts and solve for closed-form parameter estimates.* Let $N_z = n_{z++} + w_{z+}$. The relevant equations are

$$N_z \sum_y \theta_{y|z}\, \lambda_{a|zy}\, \beta_{za} = w_{za}. \tag{22}$$

$$N_z\, \theta_{y|z}\, \lambda_{a|zy}(1 - \beta_{za}) = n_{zay}, \tag{23}$$

Summing equation (22) over y and adding it to equation (23) yields

$$N_z \sum_y \theta_{y|z}\, \lambda_{a|zy} = n_{za+} + w_{za}. \tag{24}$$

Substituting equation (24) into equation (22) and solving for β_{za} gives $\beta_{(EST)za} = w_{za}/(n_{za+} + w_{za})$. Substituting $\beta_{(EST)za}$ for β_{za} in equation (23) and simplifying gives

$$\theta_{y|z}\, \lambda_{a|zy} = (n_{zay} + w_{za} n_{zay}/n_{za+})/N_z \tag{25}$$

Summing both sides of equation (25) over a and solving for $\theta_{y|z}$ gives

$$\theta_{(EST)y|z} = m_{zy}/m_{z+}, \text{ where } m_{zy} = n_{z+y} + \sum_a w_{za} n_{zay}/n_{za+}. \tag{26}$$

Step 2. Compute the statistic of interest from the parameter estimates.

The statistic of interest is $d = \theta_{(EST)1|1} - \theta_{(EST)1|0}$.

Step 3. Compute the standard error using the MP transformation. The estimated standard error is $se = \sqrt{v}$, where $v = varMP(d) = \sum_z \sum_a \sum_y (\partial d/\partial n_{zay})^2 n_{zay} + \sum_z \sum_a (\partial d/\partial w_{za})^2 w_{za}$. Based on the counts in Table 7, which come from Baker[14], $d = -0.10$ with standard error 0.007, indicating that finasteride decreases prostate cancer on biopsy.

Baker[15] extended this perfect fit analysis to include l_{za} participants randomized to group z with auxiliary variable a who are missing outcome, yielding $\theta_{(EST)y|z} = m_{zy}/m_{z+}$, where $m_{zy} = n_{z+y} + \sum_a w_{za} n_{zay}/n_{za} + l_{za}$. For the Prostate Cancer Prevention Trial, Baker et al.[14]

implemented a more complicated perfect fit analysis involving biopsy recommendation, biopsy result, and surgery result.

7.2 Example 2

A hypothetical trial randomizes smokers to a behavioral intervention or no intervention to stop smoking. The binary outcome is self-report of smoking cessation. Missing in outcome depends only on the unobserved outcome. See Appendix C for a perfect fit analysis based on Baker and Laird.[16] Under this scenario, for the hypothetical data in Table 3, the estimated risk difference between the two randomization groups is d=0.677 with standard error 0.10.

7.5 Example 3

A hypothetical diagnostic testing study involves two samples cross-classifying binary test results: (i) a reference test versus a new test and (ii) a gold standard versus a reference test. The goal is to estimate the sensitivity and specificity of the new test versus the gold standard. The assumptions are an ignorable missing mechanism, the same sensitivity and specificity of both new and reference test (relative to the gold standard) in both samples, and conditional independence of reference and new test results given the gold standard. See Appendix D for a perfect fit analysis based on Baker.[17] For the hypothetical data in Table 8, the estimated sensitivity of the new test relative to the gold standard is 0.80 with standard error 0.17. The estimated specificity of the new test relative to the gold standard is 0.60 with standard error 0.10.

7.4. Example 4

Some randomized trials involve all-or-none compliance, the switching of treatments immediately at randomization. For all-or-none compliance (or related all-or-none availability in before-and-after studies), Baker and Lindeman[18] and Imbens and Angrist[19] (followed by Angrist, Imbens, and Rubin[20]) independently developed a method, later called latent class instrumental variables,[21] that uses potential outcomes with reasonable assumptions to estimate the causal effect of treatment among the latent class of compliers (participants who would receive the assigned treatment regardless of randomization group to which they are actually assigned). Baker[22] used a perfect fit analysis for ML estimation involving a randomized cancer screening trial, discrete-time survival data, all-or-none compliance, and latent class instrumental variables. Baker and Kramer[23] formulated a perfect fit analysis for ML estimation involving a randomized trial, a partially observed binary endpoint, all-or-none compliance, and latent class instrumental variables. See Appendix E. For the hypothetical data in Table 9 under the latter scenario, the estimated treatment effect based on the perfect fit analysis is 0.4 with standard error 0.095.

8. COMPOSITE LINEAR MODELS

Composite linear models[24] provide a flexible approach to ML estimation with complex missing data patterns involving categorical data and either ignorable or nonignorable missing-data mechanisms. Let U_{obs} denote a vector of expected values for observed counts. Let U denote a vector of expected counts if there were no missing data. Let C denote a

matrix of 0's and 1's that indicates the unobserved expected counts summed to yield observed expected counts. A composite linear model has the form,

$$\begin{aligned} &U_{obs} = C\,U, \text{ where} \\ &U = N\exp\Big\{\sum\nolimits_k Q^{(k)}\Big\}, \\ &Q^{(k)} = q^{(k)}(W^{(k)}, G^{(k)}H^{(k)}), \\ &H^{(k)} = h^{(k)}(Z^{(k)}, X^{(k)}\phi^{(k)}), \end{aligned} \tag{27}$$

$W^{(k)}$, $G^{(k)}$, $H^{(k)}$, $Z^{(k)}$, and $X^{(k)}$ are matrices, $h^{(k)}(\,)$ and $q^{(k)}(\,)$ are functions, k indexes model components, and the parameter vector $\phi^{(k)}$ involves either the outcome model parameters θ or the missing-data mechanism parameters β. The parameter sets θ and β do not overlap and do not constrain one another. See Appendix F for an illustration of the matrices involved with discrete-time hazard models.

Once the matrices and function are specified, maximization is automatic, beginning with an EM algorithm, which is insensitive to poor starting values, and then switching to a Newton-Raphson algorithm, which converges faster and yields standard errors. Examples include two-phase surveys[25], regression analysis of grouped survival data with missing covariate[26], and misclassification.[27] Software for fitting composite linear models, written in Mathematica[28], is available at https://prevention.cancer.gov/about-dcp/staff-search/stuart-g-baker-scd/composite-linear-models. The user needs to specify the matrices and functions, a task simplified by using a previous example as a template, but nevertheless challenging.

8.1 Example 1

Investigators were interested in the effect of drain, a tube for removing fluid from a wound, on wound infection following surgery. Because of the expense and difficulty of following all patients after hospital discharge to determine wound infection status, investigators implemented the following double sampling design. For a random partial follow-up sample, investigators followed 1,236 patients after surgery until wound infection, hospital release, or the end of the study at 30 days after surgery, whichever occurred first. For a random full follow-up sample, investigators followed 194 patients after surgery until wound infection (either in the hospital or after release from the hospital) or the end of the study at 30 days after surgery, whichever occurred first. Time since surgery involves 4 intervals: 1 (0–4 days), 2 (5–7) days, 3 (8–30 days), and 4 (more than 30 days). See Table 10.

Baker et al.[29] formulated the following model to analyze these data. Let $h_{f|x}$ denote the hazard function for wound infection (in the absence of censoring) in time interval f =1, 2, 3 given drain status x = 0 = no drain or x = 1=drain. The outcome model is

$$logit(h_{f|x}) = \theta_0 + \theta_2\,(\text{if } f = 2) + \theta_3\,(\text{if } f = 3) + \theta_X\,(\text{if } x = 1)\,. \tag{28}$$

Let $c_{t|yd}$ denote the hazard function for hospital discharge at the start of time interval t given wound infection occurs in time interval f (for $t \geq f$). The missing-data mechanism is

$$logit(c_{t|fx}) = \beta_0 + \beta_2\,(\text{if } t = 2) + \beta_3\,(\text{if } t = 3) + \beta_X\,(\text{if } x = 1) + \beta_{RL}\,(\text{if } t = f) \quad (29)$$

.

The parameter β_{RL}, where the subscript denote response-linked, makes the missing-data mechanism directly nonignorable. Using the method of composite linear models, Baker et al. [29] estimated β_{RL} as $\beta_{(EST)RL} = -7.12$ with standard error of 1.44. The estimated effect of drain on wound infection was $\theta_{(EST)X} = 1.40$ with standard error 0.24.

8.2 Example 2

The Muscatine Coronary Risk Factor Study collected data on obesity outcome (yes or no) in girls and boys at three times (initially and 2 and 4 year later).[30] Missing outcomes occurred at one or more times, yielding 7 patterns of missing data. Baker[31] used composite linear models to fit a marginal outcome model in which the probability of obesity at each time depends on age at that time and gender. The outcome model coupled with a nonignorable missing-data mechanism fit substantially better than the outcome model coupled with an ignorable missing-data mechanism nested within the nonignorable missing-data mechanism. The estimated coefficient for sex in the logistic outcome model was 0.15 with standard error 0.08, indicating higher obesity levels for girls than boys.

9. DISCUSSION

An often-overlooked consideration with missing-data analyses is the need for missing-data adjustments to make sense. One criterion for sensible missing-data adjustment is that the unobserved data exist or could be ascertained. For example, if a biopsy result is missing because an eligible person did not arrive at the clinic, there exists an unobserved biopsy result that could have been ascertained if the person had arrived at the clinic. However, if the biopsy result is missing because of death, there does not exist an unobserved biopsy result that could have been ascertained. A less stringent criterion for sensible missing-data adjustment is that the unobserved result could be observed in a relevant target population where the missingness could be prevented, as might apply if missing in biopsy was due to accidental death and the target population specified no accidental deaths.

An important component of many missing-data analyses is a sensitivity analysis to determine how assumptions about the missing-data mechanisms affect estimates of treatment effect in the outcome model. *A model-based* sensitivity analysis computes estimated treatment effect under multiple missing-data mechanisms, as when fitting composite linear models. If an outcome model coupled with non-ignorable missing-data mechanism fits substantially better than the same outcome model coupled with a nested ignorable missing-data mechanism, reported estimates should be based on the former model. A *parameter-based* sensitivity analysis computes estimated treatment effect when varying a parameter measuring the association between missing outcome and outcome.[36] A *data-based* sensitivity analysis computes estimated treatment effect when imputing values for the missing outcome.[37, 38] A *randomization-based* sensitivity analysis using the randomization distribution to bound the estimated treatment effect if missing in outcome depends on an

unobserved binary variable.[39] When implementing a sensitivity analyses, prior knowledge helps to limit the range of possible values.

In summary, the ML methods discussed here range from the simple to the complex. The simplest methods are complete-case analysis and complete-case analysis adjusted for covariates Survival analyses adjusted for covariates are easy to implement using standard software. For missing in univariate or survival outcome based on multiple covariates, the propensity-to-be-missing score is preferable and easy to implement. More complicated ML methods are needed for fitting models with longitudinal dropouts. Commercial software is available with continuous outcomes. With binary longitudinal outcomes, a conditional model is easy to fit but extra work is needed to combine parameter estimates to estimate the quantity of interest. The perfect fit analysis is an underappreciated approach to obtain to obtain closed form ML estimates and variances for complicated likelihoods involving saturated models fit to categorical data. Some work is needed in the algebraic derivation, but it is generally simpler to implement than iterative numerical fitting. The most complex method discussed is the method of composite linear models, which is a flexible approach involving categorical data. Except for composite linear models, which awaits the development of more user-friendly software, all the above methods can contribute to the toolkit of statisticians for analyzing clinical and prevention studies with missing outcomes.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health.

APPENDIX A

This Appendix discusses ML estimation for a randomized trial in which missing in outcome Y depends on randomization group Z and baseline covariate X in which the baseline covariate is MCAR among some participants. Four subsets of participants defined by the pattern of missing data are missing both Y and X, {*MissY:MissX*}; missing Y with observed X, {*MissY:ObsX*}; observed Y with missing X,{*ObsY:MissX*}; and observed Y and observed X, {*ObsY:ObsX*}. Let $\beta = (\beta_1, \beta_2)$. Let $pr(x_i; \lambda)$ denote the distribution of x_i, which is modeled by parameter set λ. Let *MissX* denote the missing-data indicator for X. The probability of missing X is constant, as denoted by $pr(MissX_i = 1; \beta_2)$. The likelihood is

$$
\begin{aligned}
&Lik_{CCX} = L_{MissY:MissX} \times L_{MissY:ObsX} \times L_{ObsY:MissX} \times L_{ObsY:ObsX}, \quad \text{where} \\
&L_{MissY:MissX} = \prod\nolimits_{i \in \{MissY:MissX\}} \int\int pr(MissY_i = 1 | MissX_i = 1, z_i, x_i; \beta_1) \\
&\times pr(MissX_i = 1; \beta_2) \\
&\qquad \times pr(y_i | x_i, z_i; \ \theta) \times pr(x_i; \ \lambda) dx_i dy_i, \\
&L_{MissY:ObsX} = \prod\nolimits_{i \in \{MissY:ObsX\}} \int pr(MissY_i = 1 | MissX_i = 0, z_i, x_i; \beta_1) \times pr \\
&(MissX_i = 0; \beta_2) \\
&\qquad \times pr(y_i | x_i, z_i; \ \theta) \times pr(x_i; \ \lambda) dx_i dy_i, \\
&L_{ObsY:MissX} = \prod\nolimits_{i \in \{ObsY:MissX\}} \int pr(MissY_i = 0 | MissX_i = 1, z_i, x_i; \ \beta_1) \\
&\times pr(MissX_i = 1; \ \beta_2) \\
&\qquad \times pr(y_i | x_i, z_i; \ \theta) \times pr(x_i; \ \lambda) dx_i, \\
&L_{ObsY:MissX} = \prod\nolimits_{i \in \{ObsY:ObsX\}} pr(MissY_i = 0 | MissX_i = 0, z_i, x_i; \ \beta_1) \times pr \\
&(MissX_i = 0; \ \beta_2) \\
&\qquad \times pr(y_i | x_i, z_i; \ \theta) pr(x_i; \ \lambda).
\end{aligned}
\tag{A.1}
$$

The likelihood in equation (A.1) is indirectly nonignorable. There is no factor of the likelihood involving θ without β because β linked to λ in $L_{MissY:MissX}$ and λ is linked to θ in $L_{ObsY:MissX}$. To simplify ML estimation, a simple expedient is to consider the likelihood for the random sample of participants with observed values of covariate,

$$
\begin{aligned}
&Lik_{CCX:ObsX} = L_{MissY:ObsX} \times L_{ObsY:ObsX} \\
&= f_{CCX}(\beta, \lambda) \times Lik_{CCX:ObsX:Ign}(\theta), \text{ where} \\
&f_{CCX}(\beta, \lambda) = \prod\nolimits_{i\{MissY:ObsX\}} pr(MissY_i = 1 | MissX_i = 0, z_i, x_i; \beta_1) \times pr \\
&(MissX_i = 0; \beta_2) \times pr(x_i; \ \lambda) \\
&\qquad \times \prod\nolimits_{i\{ObsY:ObsX\}} pr(MissY_i = 0 | MissX_i = 0, z_i, x_i; \ \beta_1) \times pr \\
&(MissX_i = 0; \ \beta_2) \times pr(x_i; \ \lambda), \\
&Lik_{CCX:ObsX:Ign}(\theta) = \prod\nolimits_{i \in \{ObsY:Ob\}} pr(y_i | x_i, z_i; \ \theta).
\end{aligned}
\tag{A.2}
$$

ML estimation for θ in equation (A.2) involves only $Lik_{CCX:ObsX:Ign}(\theta)$, which translates into complete case analysis adjusted for observed values of covariate *X*.

APPENDIX B

This Appendix discusses ML estimation in a randomized trial with survival times in which the probability of censoring depends on randomization group *Z* and a partially observed

baseline covariate X. Four subsets of participants defined by the pattern of missing data are censored with missing X,{$CensMissX$}, censored with observed X, {$CensObsX$}, failure with missing X, {$FailMissX$}, and failure with observed X, {$FailObsX$}. Let $\beta = (\beta_1, \beta_2)$. The probability of missing X is constant, as denoted by $pr(MissX_i = 1; \beta_2)$. The missing data patterns give rise to the following likelihood,

$$Lik_{SurvX}(\theta, \beta) = L_{CensMissX} \times L_{CensObsX} \times L_{FailMissX} \times L_{FailObsX}, \quad \text{where}$$

$$L_{Cens;MissX} = \prod_{i \in \{CensMissX\}} \int_{c_i}^{\infty} \int pr(C_i = c_i | MissX_i = 1, z_i, x_i; \beta_1) \times pr(MissX_i = 1; \beta_2) \times pr(F_i = f_i | z_i, x_i; \theta) \times pr(x_i; \lambda)) dx_i df_i,$$

$$L_{Cens:ObsX} = \prod_{i \in \{CensObsX\}} \int_{c_i}^{\infty} pr(C_i = c_i | MissX_i = 0, z_i, x_i; \beta_1) \times pr(MissX_i = 0; \beta_2) \times pr(F_i = f_i | z_i, x_i; \theta) \times pr(x_i; \lambda) df_i, \quad \text{(B.1)}$$

$$L_{Fail:MissX} = \prod_{i \in \{FailMisX\}} \int_{f_i}^{\infty} \int pr(C_i = c_i | MissX_i = 1, z_i, x_i; \beta_1) \times pr(MissX_i = 1; \beta_2) \times pr(F_i = f_i | z_i, x_i; \theta) \quad pr(x_i; \lambda)) dx_i dc_i,$$

$$L_{Fail:ObsX} = \prod_{i \in \{FailObsX\}} \int_{f_i}^{\infty} pr(C_i = c_i | MissX_i = 0, z_i, x_i; \beta_1) \times pr(MissX_i = 0; \beta_2) \times pr(F_i = f_i | z_i, x_i; \theta) \times pr(x_i; \lambda) dc_i.$$

The likelihood in equation (B.1) is indirectly nonignorable. There is no factor of the likelihood involving θ without β because β is linked to λ in $L_{Cens:MissX}$ and λ is linked to θ in $L_{Fail:MissX}$. To simplify ML estimation, a simple expedient is to consider the likelihood for the random sample of participants with observed values of covariate,

$$Lik_{SurvX:ObsX}(\theta, \beta) = L_{CensObsX} \times L_{FailObsX}$$

$$= f_{SurvX}(\beta, \lambda) \times Lik_{SurvX:ObsX:Ign}(\theta), \quad \text{where}$$

$$f_{SurvX}(\beta, \lambda) = \prod_{i \in \{CensObsX\}} pr(C_i = c_i | MissX_i = 0, z_i, x_i; \beta_1) \times pr$$

$$(MissX_i = 0; \beta_2) \times pr(x_i; \lambda)$$

$$\prod_{i \in \{FailObsX\}} \int_{f_i}^{\infty} pr(C_i = c_i | MissX_i = \quad 0, z_i, x_i; \beta_1) \times pr(MissX_i \tag{B.2}$$

$$= 0; \beta_2) \times pr(x_i; \lambda) dc_i .$$

$$Lik_{SurvX:ObsX:Ign}(\theta) = \prod_{i \in \{CensObsX\}} pr(F_i \geq c_i | z_i, x_i; \theta)$$

$$\times \prod_{i \in \{FailObsX\}} pr(F_i = f_i | z_i, x_i; \theta) .$$

ML estimation of θ equation (B.2) involves only $Lik_{SurvX:ObsX:Ign}(\theta)$, which translates into a survival analysis for the random sample of cases with observed values of covariate X.

APPENDIX C

This Appendix derives the perfect fit estimates for a randomized trial with a binary outcome Y in which missing in outcome depends on the outcome Y but not on the randomization group Z Let n_{zy} denote the number of participants randomized to group z =0, 1, with outcome y=0, 1. Let w_z denote the number of participants randomized to group z =0, 1 who are missing the outcome. See Table 3. The outcome model, $pr(Y = 1 | z; \theta) = \theta_z$, is the probability of outcome 1 given randomization to group z. The directly nonignorable missing-data mechanism, $pr(MissY = 1 | y; \beta) = \beta_y$, is the probability of missing outcome y given outcome y. The model is saturated because there are 4 independent parameters (2 for $\theta_{1|z}$ and 2 for β_y) and 4 independent cell counts (4 for $\{n_{zy}\}$ and 2 for $\{w_z\}$ minus 2 because $n_{z+} + w_z$ is fixed). The perfect fit analysis follows.

Step 1. *Set expected counts equal to observed counts and solve for parameter estimates*. Let $\phi_y = \beta_y / (1 - \beta_y)$ and $\mu_{zy} = n_z + \theta_{y|z} (1 - \beta_y)$. The relevant equations are

$$\mu_{zy} = n_{zy} \tag{C.1}$$

$$\sum_y \mu_{zy} \phi_y . = w_z . \tag{C.2}$$

Simultaneously solving equations (C.1) and (C.2) yields

$$\phi_{(EST)0} = (n_{11}w_0 - n_{01}w_1)/(n_{11}n_{00} - n_{01}n_{10}), \quad \text{(C.3)}$$

$$\phi_{(EST)1} = (n_{00}w_1 - n_{10}w_0)/(n_{11}n_{00} - n_{01}n_{10}). \quad \text{(C.4)}$$

If $\phi_{(EST)y} \geq 0$, $\phi_{(EST)y}$ is the ML estimate. If $\phi_{(EST)0}$ or $\phi_{(EST)1}$ is negative, the ML estimates are on the boundary of the parameter space.

S*tep 2. Compute the statistic of interest.* The estimated risk difference is $d = p_1 - p_0$, where $p_z = \left\{n_{z1}(1 + \phi_{(EST)1})\right\} / \sum_y n_{zy}(\phi_{(EST)y})$.

Step 3. Compute the standard error using the MP transformation. The standard error is $se = \sqrt{v}$, where $v = varMP(d)$.

APPENDIX D

This Appendix presents a perfect fit analysis for the diagnostic testing data in Table 8. Data set 1 involves $\{z_{ij}\}$, the number of persons with reference test result $i = 0, 1$ and new test result $j = 0, 1$. Data set 2 involves $\{w_{jk}\}$, the number of persons with reference test result $i=0, 1$ and gold standard result $k = 0, 1$. The model assumes independence of the test results given the gold standard result and an ignorable missing-data mechanism. Let $\psi_{i|k}$ denote the probability of reference test result i given gold standard result k. Let $\theta_{j|k}$ denote the probability of new test result j given gold standard result k. Let π_k denote the probability of gold standard result k in data set 1 and let ρ_k denote the probability of gold standard k in data set 2. The outcome model is saturated with 6 independent parameters (2 for $\psi_{1|k}$, 2 for $\theta_{j|k}$, 1 for π_k, and 1 for ρ_k) and 6 independent cell counts (3 for $\{z_{ik}\}$ and 3 for $\{w_{ik}\}$).

Step 1. *Set observed counts equal to expect counts and solve for closed-form parameter estimates.* The relevant equations ignore the missing-data mechanism,

$$z_{++}\sum_k \psi_{i|k}\theta_{j|k}\pi_k = z_{ij}, \quad \text{(D.1)}$$

$$w_{++}y_{i|k}\rho_k = w_{ik}. \quad \text{(D.2)}$$

Summing both sides of equation (D.2) over i and solving for ρ_k gives $\rho_{(EST)k} = w_{+k}/w_{++}$. Substituting $\rho_{(EST)k}$ into equation (D.2) and solving for $\psi_{i|k}$ gives $\psi_{(EST)i|k} = w_{ik}/w_{+k}$. Substituting $\psi_{(EST)i|k}$ into equation (D.1) gives

$$z_{++}\sum_k (w_{ik}/w_{+k})\theta_{j|k}\pi_k = z_{ij}. \quad \text{(D.3)}$$

Rewriting equation (D.3) as separate equations for i=0 and i=1gives

$$(w_{00}/w_{+0})\,(\theta_{j|0}\pi_0) + (w_{01}/w_{+1})\,(\theta_{j|1}\pi_1) = z_{0j}/z_{++}, \tag{D.4}$$

$$(w_{10}/w_{+0})\,(\theta_{j|0}\pi_0) + (w_{11}/w_{+1})\,(\theta_{j|1}\pi_1) = z_{1j}/z_{++}. \tag{D.5}$$

Simultaneously solving equations (D.4) and (D.5) gives

$$\theta_{j|0}\pi_0 = w_{+0}(w_{00}\,z_{1j} - w_{10}\,z_{0j})/\left\{(w_{00}\,w_{11} - w_{10}\,w_{01})z_{++}\right\}, \tag{D.6}$$

$$\theta_{j|1}\pi_1 = w_{+1}(w_{01}\,z_{1j} - w_{11}\,z_{0j})/\left\{(w_{00}\,w_{11} - w_{10}\,w_{01})z_{++}\right\}, \tag{D.7}$$

Summing both sides of equation (D.6) and (D.7) over j and solving for π_0 and π_1 yields

$$\pi_{(EST)0} = w_{+0}(w_{00}\,z_{1+} - w_{10}\,z_{0+})/\left\{(w_{00}\,w_{11} - w_{10}w_{01})z_{++}\right\}. \tag{D.8}$$

$$\pi_{(EST)1} = w_{+1}(w_{01}\,z_{1+} - w_{11}\,z_{0+})/\left\{(w_{00}\,w_{11} - w_{10}w_{01})z_{++}\right\}. \tag{D.9}$$

Substituting $\pi_{(EST)0}$ into equation (D.6) and $\pi_{(EST)1}$ into equation (D.7) and solving for $\theta_{j|0}$ and $\theta_{j|1}$ yields

$$\theta_{(EST)(j|0)} = (w_{00}\,z_{1j} - w_{10}\,z_{0j})/(w_{00}\,z_{1+} - w_{10}\,z_{0+}), \tag{D.10}$$

$$\theta_{(EST)(j|1)} = (w_{11}\,z_{0j} - w_{01}\,z_{1j})/(w_{11}\,z_{0+} - z_{01}\,z_{1+}). \tag{D.11}$$

Step 2. Compute the statistic of interest. Specificity is $\theta_{(EST)0|0}$. Sensitivity is $\theta_{(EST)1|1}$.

Step 3. Compute the standard error using the MP transformation. The standard error is $se = \sqrt{v}$, where $v = varMP(d)$ and $d = \theta_{(EST)0|0}$ or $\theta_{(EST)1|1}$.

APPENDIX E

This Appendix derives perfect fit analysis for estimating treatment effect in a randomized trial with all-or-none compliance and binary outcome in which missing depends on outcome and randomization group. Let n_{zby} denote the number of participants randomized to treatment z who receive treatment b immediately after randomization and experience outcome y. Let w_{zb} denote the number of persons randomized to treatment z who receive treatment b immediately after randomization and are missing the outcome. See Table 9.

Let s index latent classes defined by potential outcomes of treatment received. Under the monotonicity assumption s takes three possible values: A = always-takers, who would receive the new treatment regardless of which randomization group to which they might be assigned, N = never-takers who would receive the old treatment regardless of which

randomization to which they might be assigned, and C = compliers who receive the assigned treatment regardless of which randomization group to which they might be assigned.

The outcome model, $\text{pr}(Y = y|z, s;\ \theta) = \theta_{szy}$, is the probability of outcome y given randomization group z and latent class s. The missing-data mechanism, $\text{pr}(MissY = 1|z, s; \beta) = \beta_{zs}$, is the probability of missing outcome given randomization group z and latent class s. Let $\text{pr}(S = s) = \pi_s$ denote the probability of being in latent class s. Under the compound exclusion restriction assumption, the probabilities of outcome and missing in outcome do not depend on randomization group for always-takers and never-takers, namely $\theta_{t|zs} = \theta_{t|z}$ and $\beta_{zs} = \beta_z$ for $s = A$ and N. The model is saturated because there 10 independent parameters ($\theta_{1|A}$ $\theta_{1|0C}$, $\theta_{1|1C}$, $\theta_{1|N}$, β_A, β_{0C}, β_{1C}, β_N, π_A, and π_C) and 10 independent cells counts (8 for $\{n_{zby}\}$ and 4 for $\{w_{zb}\}$ minus 2 because $n_{z++} + w_{z++}$ is fixed). The perfect fit analysis follows.

Step 1. *Set expected counts equal to observed counts and solve for parameter estimates*.

Let $N_z = n_{z++} + w_{z++}$. The relevant equations based on the definitions of the latent classes A, C, and N, are

$$N_0\left\{\theta_{y|N}(1-\beta_N)\pi_N + \theta_{y|0C}(1-\beta_{0C})\pi_C\right\} = n_{00y}, \tag{E.1}$$

$$N_0\left\{\theta_{y|A}(1-\beta_A)\pi_A\right\} = n_{01y}, \tag{E2}$$

$$N_1\left\{\theta_{y|N}(1-\beta_N)\pi_N\right\} = n_{10y}, \tag{E.3}$$

$$N_1\left\{\theta_{y|1C}(1-\beta_{1C})\pi_C + \theta_{y|A}(1-\beta_A)\pi_A\right\} = n_{11y}, \tag{E.4}$$

$$N_0(\boldsymbol{\beta}_N\pi_N + \boldsymbol{\beta}_{0C}\pi_C) = w_{00}, \tag{E.5}$$

$$N_0\boldsymbol{\beta}_A\pi_A = w_{01}, \tag{E.6}$$

$$N_1\boldsymbol{\beta}_N\pi_N = w_{10}, \tag{E.7}$$

$$N_1(\beta_{1C}\pi_{1C} + \beta_A\pi_A) = w_{11}. \tag{E.8}$$

Summing equation (E.2) over y and adding to equation (E.6) yields

$$N_0\pi_A = n_{01} + w_{01}. \tag{E.9}$$

Summing equation (E.4) over y and adding to equation (E.8) yields

$$N_1(\pi_{1C} + \pi_A) = n_{11} + w_{11}. \tag{E.10}$$

Subtracting equation (E.9) from equation (E.10) and solving for π_{1C} gives

$$\begin{aligned} \pi_{(EST)C} &= p_1 - p_0, \text{ where} \\ p_1 &= (n_{11+} + w_{11})/N_1 \text{and } p_0 = (n_{01+} + w_{01})/N_0. \end{aligned} \tag{E.11}$$

Subtracting equation (E.6) from equation (E.8) and solving gives

$$\begin{aligned} \beta_{(EST)1C}\pi_{(EST)C} &= q_{11} - q_{01}, \text{ where} \\ q_{11} &= w_{11}/N_1 \text{ and } q_{01} = w_{01}/N_0. \end{aligned} \tag{E.12}$$

Subtracting equation (E.7) from equation (E.5) and solving gives

$$\begin{aligned} \beta_{(EST)0C}\pi_{(EST)C} &= q_{00} - q_{10}, \text{ where} \\ q_{00} &= w_{00}/N_0 \text{ and } q_{10} = w_{10}/N_1. \end{aligned} \tag{E.13}$$

Subtracting equation (E.2) from equation (E.4) and solving for $\theta_{y|1C}$ based on equations (E. 11) and (E.12) gives

$$\begin{aligned} \theta_{(EST)y|1C} &= (n_{11y}/N_1 - n_{01y}/N_0)/[\{(1 - \beta_{(EST)1C})\}\pi_{(EST)C}] \\ &= (n_{11y}/N_1 - n_{01y}/N_0)/\{(p_1 - p_0) - (q_{11} - q_{01})\}. \end{aligned} \tag{E.14}$$

Subtracting equation (E.3) from (E.1) and solving for $\theta_{y|0C}$ based on equations (E.11) and (E.13) gives

$$\begin{aligned} \theta_{(EST)y|0C} &= (n_{00y}/N_0 - n_{10y}/N_1)/[\{(1 - \beta_{(EST)0C})\}\pi_{(EST)C}]. \\ &= (n_{00y}/N_0 - n_{10y}/N_1)/\{(p_1 - p_0) - (q_{00} - q_{10})\}. \end{aligned} \tag{E.15}$$

Step 2. Compute the statistic of interest. The perfect fit ML estimate of the treatment effect in compliers is $d = \theta_{(EST)1|1C} - \theta_{(EST)1|0C}$ ..

Step 3. Compute the standard error using the MP transformation. The standard error is $se = \sqrt{v}$, where $v = varMP(d)$.

Appendix F

This Appendix presents some of the matrix components in a composite linear model for discrete-time survival. Let $h_{f|x}(\theta)$ denote the hazard for failure (in the absence of censoring) at time f=1, 2 for covariate x =0, 1. Let $c_{t|x}(\beta)$ denote the hazard for censoring (in the absence of failure) at time t=1 for covariate x = 0, 1, where censoring in an interval implies failure is not observed in the interval. Consider two simple models: $logit\{h_{f|x}(\theta)\} = \theta_{f0} + \theta_{f1}\, x$ and $logit\{c_{t|x}(\beta)\} = \beta_{t0} + \beta_{t1}\, x$. Let N denote the sample size. Let u_{Ftx} denote the expected number of persons who fail in interval t with covariate at

level x. Let u_{Ctx} denote the expected number of persons censored in interval t with covariate at level x. The 8×1 column vector of expected counts with no missing data are $U_{8\times1} = (u_{F10}, u_{F20}, u_{C10}, u_{C20}, u_{F11}, u_{F21}, u_{C11}, u_{C21})^{\mathrm{T}}$, where, for covariate x,

$$u_{F1x} = N \quad log\{h_{1|x}(\theta)\}, \tag{F.1}$$

$$u_{F2x} = N\{1 - h_{1|x}(\theta)\} \times h_{2|x}(\theta) \times \{1 - c_{1|x}(\beta)\}, \tag{F.2}$$

$$u_{C1x} = N\{1 - h_{1|x}(\theta)\} \times c_{1|x}(\beta), \tag{F.3}$$

$$u_{C2x} = N\{1 - h_{1|x}(\theta)\} \times \{1 - h_{2|x}(\theta)\} \times \{1 - c_{1|x}(\beta)\}. \tag{F.4}$$

In matrix notation for composite linear models, the expected counts with no missing data are

$$U = N\exp\left\{\sum_k Q^{(k)}\right\}, \quad \text{where}$$

$$Q^{(k)} = q^{(k)}(W^{(k)}, G^{(k)}H^{(k)}),$$

$$h^{(k)}(Z^{(k)}, X^{(k)}\,\theta^{(k)}) \text{ for the outcome model,}$$

$$H^{(\mathrm{k})} = \{$$

$$h^{(k)}(Z^{(k)}, X^{(k)}\,\beta^{(k)}) \text{ for } k = 2, \text{ for the missing} - \text{data mechanism}.$$

Outcome model.

The H-component for the outcome model expresses in matrix form $log\{h_{t|x}(\theta)\} = (\theta_{t0} + \theta_{t1}\,x) - log\{1 + (\exp\theta_{t0} + \theta_{t1}\,x)\}$ and $log\{1 - h_{t|x}(\theta)\} = -log\{1 + (\exp\theta_{t0} + \theta_{t1}\,x)\}$ The H-component is $H_{8\times1} = log(h_{1|0}, 1 - h_{1|0}, h_{2|0}, 1 - h_{2|0}, h_{1|1}, 1 - h_{1|1}, h_{2|1}, 1 - h_{2|1})^{\mathrm{T}}$. In matrix notation, $H_{8\times1}^{(1)} = Z_{8\times1}^{(1)} \circ (X_{8\times4}^{(1)}\,\theta_{4\times1}) - log\{1 + \exp(X_{8\times4}^{(1)}\,\theta_{4\times1})\}$, where $Z_{8\times1}^{(1)} = (1, 0, 1, 0, 1, 0, 1, 0)^{\mathrm{T}}$, $X_{8\times4}^{(1)} = ((X_{4\times2}{}^{*}, 0_{4\times2}), (X_{4\times2}{}^{*}, X_{4\times2}{}^{*}))$, $X_{4\times2}{}^{*} = ((1, 0), (1, 0), (0, 1), (0, 1))$, and $0_{4\times2}$ is a 4 × 2 matrix of 0's , $\theta_{4\times1} = (\theta_{10}, \theta_{20}, \theta_{11}, \theta_{21})^{\mathrm{T}}$, and the symbol "∘" denotes element-by-element multiplication instead of matrix multiplication. The top half of $X_{8\times2}^{(1)}$ corresponds to x=0 and the bottom half to x=1.

The Q-component of the outcome model is $Q_{8\times1}^{(1)} = log(h_{1/0}, (1-h_{1/0})\ h_{2/0}, (1-h_{1/0}), (1-h_{1/0})(1-h_{2/0}), h_{1/1}, (1-h_{1/1})\ h_{2/1}, (1-h_{1/1}), (1-h_{1/1})\ (1-h_{2/1}))^{T}$. In matrix notation, $Q_{8\times1}^{(1)} = W_{8\times1}^{(1)} + G_{8\times8}^{(1)}\ H_{8\times1}^{(1)}$, where $W_{8\times1}^{(1)} = 0_{8\times1}$, $G_{8\times8}^{(1)} = ((G_{4\times4}*, 0_{4\times4}), (0_{4\times4}, G_{4\times4}*))$, and $G_{4\times4}* = ((1, 0, 0, 0), (0, 1, 1, 0), (0, 1, 0, 0), (0, 1, 0, 1))$. The top half of $G_{8\times8}^{(1)}$ corresponds to x=0 and the bottom half to x=1.

Missing-data mechanism.

The H-component for the censoring model is $H_{4\times1}^{(2)} = log(c_{1/0}, 1-c_{1/0}, c_{1/1}, 1-c_{1/1})^{T}$. In matrix notation, $H_{4\times1}^{(2)} = Z_{4\times1}^{(2)} \circ (X_{4\times2}^{(2)}\ \beta_{2\times1}) - log\{1+ \exp(X_{4\times2}^{(2)}\ \beta)\}$, where $Z_{4\times1}^{(2)} = (0, 1, 0, 1)^{T}$, $X_{4\times2}^{(2)} = ((1, 0), (1, 0), (1, 1), (1, 1))$ and $\beta_{2\times1} = (\beta_{10}, \beta_{11})^{T}$.

The Q-component for the censoring model is $Q_{8\times1}^{(2)} = log(1, (1-c_{1/0}), c_{1/0}, (1-c_{1/0}), 1, (1-c_{1/1}), c_{1/1}, (1-c_{1/1}))^{T}$. In matrix notation, $Q_{8\times1}^{(2)} = W_{8\times1}^{(2)} + G_{8\times4}^{(2)}\ H_{4\times1}^{(2)}$, where $W_{8\times1}^{(2)} = 0_{8\times1}$ and $G_{8\times4}^{(2)} = ((G_{4\times2}**, 0_{4\times2}), (0_{4\times2}, G_{4\times2}**))$, where $G_{4\times2}** = ((0, 0), (0, 1),(1, 0), (0, 1))$. The top half of $G_{8\times4}^{(2)}$ corresponds to x=0 and the bottom half to x=1.

REFERENCES

1. Rubin DB. Inference and missing data. Biometrika 1976; 63(3):581–592.
2. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics 1988; 44(1):175–188.
3. Heitjan DF, Rubin DB. Ignorability and coarse data. Ann. Statist 1991; 19(4):2244–2253.
4. Little RJ, Rubin DB. Statistical Analysis with Missing Data Hoboken, NJ: John Wiley & Sons; 2002.
5. Little RJ, Rubin DB and Zangeneh SZ. Conditions for ignoring the missing-data mechanism in likelihood inferences for parameter subsets. J Am Stat Assoc 2017; 112(517):314–320.
6. Baker SG, Fitzmaurice G. Freedman LS, Kramer BS. Simple adjustments for randomized trials with nonrandomly missing or censored outcomes arising from informative baseline covariates. Biostatistics 2006; 7(1):29–40. [PubMed: 15923407]
7. Henry K, Erice A, Tierney C, Balfour HH Jr, et al. A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. AIDS Clinical Trial Group 193A Study Team. J Acquir Immune Defic Syndr Hum Retrovirol 1998 19(4):339–349. [PubMed: 9833742]
8. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. Biometrics1986; 42(4):805–820. [PubMed: 3814725]
9. SAS/STAT(R) 9.22 User's Guide. SAS https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mixed_sect034.htm Accessed March 27, 2019.
10. Allison PD. Handling missing data by maximum likelihood SAS Global Forum 2012 Paper 313–2012 http://www.statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf. Accessed March 27, 2019
11. Baker SG, Erwin D. and Kramer BS. Estimating the cumulative risk of false positive cancer screenings. BMC Med Res Method. 2003; 3:11.
12. Baker SG. The Multinomial-Poisson transformation. The Statistician 1994; 43(4):495–504.
13. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003; 349(3):215–224. [PubMed: 12824459]
14. Baker SG, Darke AK, Pinsky P, Parnes HL, Kramer BS. Transparency and reproducibility in data analysis: The Prostate Cancer Prevention Trial. Biostatistics 2010;11(3): 413–418. [PubMed: 20173101]

15. Baker SG. Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. J Am Stat Assoc 2000; 95(449):43–50.
16. Baker SG, Laird NM. Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. J Am Stat Assoc 1988; 83(401):62–69.
17. Baker SG. Evaluating a new test using a reference test with estimated sensitivity and specificity. Commun Stat 1991; 20(9):2739–2752.
18. Baker SG. and Lindeman KS. The paired availability design, a proposal for evaluating epidural analgesia during labor. Stat Med 1994; 13(21):2269–2278. [PubMed: 7846425]
19. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. Econometrica. 1994; 62(2): 467–475.
20. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996; 91(434):444–455.
21. Baker SG, Kramer BS, Lindeman KL. Latent class instrumental variables. A clinical and biostatistical perspective. Stat Med 2016; 35:147–160. [PubMed: 26239275]
22. Baker SG. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. J. Am Stat Assoc 1998; 93(443): 929–934.
23. Baker SG and Kramer BS. Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. Stat Methods Med Res 2005; 14(4):349–367. [PubMed: 16178137]
24. Baker SG. Composite linear models for incomplete multinomial data. Stat Med 1994; 13(5–7): 609–622. [PubMed: 8023037]
25. Baker SG, Ko C, and Graubard B. A sensitivity analysis for nonrandomly Missing categorical data arising from a national health disability survey. Biostatistics 2003; 4(1):41–56. [PubMed: 12925329]
26. Baker SG. Regression analysis of grouped survival data with incomplete covariates: nonignorable missing-data and censoring mechanisms. Biometrics 1994; 50(3):821–826. [PubMed: 7981401]
27. Baker SG. A simple method for computing the observed information matrix when using the EM algorithm with categorical data. J Comp Graph Stat 1992; 1(1):63–76
28. Wolfram Research, Inc., Mathematica, Version 11.3, Champaign, IL; 2018.
29. Baker SG, Wax Y, Patterson BH. Regression analysis of grouped survival data: informative censoring and double sampling. Biometrics 1993; 49(2):379–389. [PubMed: 8369374]
30. Lauer RM, Connor WE, Leaverton PE, et al. Coronary heart disease risk factors in school children: the Muscatine study. J Pediatr 1975; 86(5):697–706. [PubMed: 1133650]
31. Baker SG. Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. Biometrics 1995; 51(3):1042–1052. [PubMed: 7548689]
32. Vach W, Blettner M. Logistic regression with incompletely observed categorical covariates--investigating the sensitivity against violation of the missing at random assumption. Stat Med 1995; 14(12):1315–1329. [PubMed: 7569490]
33. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Stat Med 2014; 33(24):4170–4185. [PubMed: 24845086]
34. Hollis S A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. Stat Med 2002; 21(24):3823–3834. [PubMed: 12483769]
35. Baker SG, Freedman LS. A simple method for analyzing data from a randomized trial with a missing binary outcome. BMC Med Res Method 2003; 3:8.

Table 1.

Missing-data taxonomy applied to missing outcomes

	Missing-data mechanism	
	Ignorable	**Non-ignorable**
Definition	Likelihood-based inference for the outcome model does not involve parameters modeling the missing-data mechanism *	Not ignorable
Implication	Missing in outcome depends only observed variables	Missing in outcome depends on at least one unobserved variable
Outcome is said to be	Missing at random (MAR)	Missing not at random (MNAR)
Special cases	Missing completely at random (MCAR) Missing in outcome occurs with constant probability	Directly non-ignorable Missing in outcome depends only on outcome Indirectly non-ignorable Missing in outcome does not depend on outcome and depends on other partially missing variables

* The sets of parameters for the outcome model and missing-data mechanism do not overlap and do not constrain one other

Table 2.

Overview of ML estimation methods

Method	Indications for Use	Missing-Data Mechanism	Implementation
Complete case analysis	Missing in outcome depends on randomization group	Ignorable	Compute simple statistics for complete cases (participants not missing the outcome).
Complete case analysis with covariate adjustment	Missing in outcome depends on randomization group and covariate	Ignorable after covariate adjustment	Fit the outcome model (as a function of randomization group and covariates) to complete cases with the covariate.
Survival analysis with covariate adjustment	Censoring depends on randomization group and covariate	Ignorable after covariate adjustment	Fit the outcome model (as a function of randomization group and covariates) to survival data.
Analysis via propensity-to-be-missing scores	Missing in outcome or censoring depends on randomization group and many covariates	Ignorable after covariate adjustment	(1) Fit a model for the missing-data mechanism. (2) Use the fitted model to compute scores. (3) Compute overall estimate based on quintiles of scores.
Longitudinal dropout analysis	Dropout depends on previous observed outcome and possibly randomization group and covariate	Ignorable	For a continuous longitudinal outcome, fit a marginal model using commercial software. For a longitudinal binary outcome, fit a conditional model.
Perfect fit analysis	Saturated models with categorical data	Ignorable or nonignorable	(1) Set expected counts equal to observed counts and solve for parameter estimates. (2) Compute statistic from parameter estimates. (3) Compute estimated variance using MP transformation.
Composite linear models	Flexible models with categorical data	Ignorable or nonignorable	Fit using specialized software.

Table 3.

Hypothetical counts for complete-case analysis

Randomization group	**Outcome**		
	Y=0	Y =1	Missing
Z=0	n_{00} (400)	n_{01} (600)	w_0 (200)
Z=1	n_{10} (200)	n_{11} (600)	w_1 (400)

Table 4.

Hypothetical counts for complete-case analysis with covariate adjustment

Randomization group	**Covariate**	**Outcome**		
		Y=0	Y=1	Missing
Z=0	X=0	n_{000} (100)	n_{001} (200)	w_{00} (100)
	X=1	n_{010} (300)	n_{011} (400)	w_{01} (100)
Z=1	X=0	n_{100} (100)	n_{101} (200)	w_{10} (100)
	X=1	n_{110} (100)	n_{111} (400)	w_{11} (300)

Table 5.

Estimate and standard errors with propensity-to-be-missing score

Randomization group Z=0 (Dual therapy)			Randomization group Z=1 (Triple Therapy)		
Quintile	**Estimate***	**Standard Error**	**Quintile**	**Estimate***	**Standard Error**
1	0.539	0.041	1	0.584	0.070
2	0.619	0.040	2	0.736	0.062
3	0.592	0.041	3	0.792	0.057
4	0.695	0.038	4	0.658	0.071
5	0.793	0.034	5	0.828	0.054

*Estimated probability of surviving to 18 months in each quntile.

Table 6.

Counts for false positive status given screen and time since late screen

Screen	Time interval since last screen	No false positive	False positive
1	Not applicable	n_{10} (4509)	n_{11} (82)
2	1	n_{210} (1662)	n_{211} (7)
	2	n_{220} (1634)	n_{221} (13)
	3	n_{230} (291)	n_{231} (1)
	4	n_{240} (406)	n_{241} (2)
3	1	n_{310} (1589)	n_{311} (9)
	2	n_{320} (1488)	n_{321} (10)
	3	n_{330} (218)	n_{331} (2)
	4	n_{340} (204)	n_{341} (2)
4	1	n_{410} (1087)	n_{411} (13)
	2	n_{420} (1467)	n_{421} (10)
	3	n_{430} (193)	n_{431} (2)
	4	n_{440} (48)	n_{441} (0)

Table 7.

Counts for Prostate Cancer Prevention Trial

Randomization group	Auxiliary variable (biopsy recommendation)	Outcome (prostate cancer on biopsy)		
		Y=0 (No)	Y=1 (Yes)	Missing
Z=0 (Placebo)	A=0 (No)	n_{000} (618)	n_{001} (3675)	w_{00} (3955)
	A=1 (Yes)	n_{010} (524)	n_{011} (479)	w_{01} (215)
Z=1 (Finasteride)	A=0 (No)	n_{100} (381)	n_{101} (3791)	w_{10} (4169)
	A=1 (Yes)	n_{110} (409)	n_{111} (458)	w_{11} (214)

Table 8.

Hypothetical counts for diagnostic testing example

Data Set		**New test**	
1	Reference test	0 (negative)	1 (positive)
	0 (negative)	z_{00} (84)	z_{01} (46)
	1 (positive)	z_{10} (26)	z_{11} (44)
		Gold standard	
2	Reference test	0 (negative)	1 (positive)
	0 (negative)	w_{00} (18)	w_{01} (4)
	1 (positive)	w_{10} (2)	w_{11} (6)

Table 9.

Hypothetical counts for latent class instrumental variables with missing outcome

Randomization group	**Treatment received**	**Outcome**		
		Y=0	Y=1	Missing
Z=0	T=0	n_{000} (100)	n_{001} (200)	w_{00} (100)
	T=1	n_{010} (400)	n_{011} (300)	w_{01} (100)
Z=1	T=0	n_{100} (300)	n_{101} (200)	w_{10} (200)
	T=1	n_{110} (100)	n_{111} (100)	w_{11} (300)

Table 10.

Counts for double sampling study of wound infection

Sample	Drain	Interval of hospital discharge with no prior infection	Interval of infection			No infection by day 30	Hospital discharge without follow-up
			0–4 days	5–7 days	8–30 days		
Partial follow-up sample	No	None	6	10	7	1	
		0–4 days	--	--	--	--	180
		5–7 days	0	--	--	--	544
		8–30 days	0	0	--	--	232
	Yes	None	7	15	11	0	
		0–4 days	--	--	--	--	8
		5–7 days	0	--	--	--	87
		8–30 days	0	0	--	--	128
Full follow-up sample	No	None	2	1	1	0	
		0–4 days	0	0	3	39	
		5–7 days	0	0	3	89	
		8–30 days	0	0	0	30	
	Yes	None	3	3	0	0	
		0–4 days	0	0	0	2	
		5–7 days	0	0	0	10	
		8–30 days	0	0	1	7	