Handbook of Statistics, Vol. 27
ISSN: 0169-7161

DOI: 10.1016/S0169-7161(07)27019-1

https://doi.org/10.20378/irbo-55153

Incomplete Data in Epidemiology and Medical Statistics

Susanne Rässler, Donald B. Rubin and Elizabeth R. Zell[☆]

Abstract

Missing data are a common problem in most epidemiological and medical studies, including surveys and clinical trials. Imputation, or filling in the missing values, is an intuitive and flexible way to handle the incomplete data sets that arise because of such missing data. Here, in addition to imputation, including multiple imputation (MI), we discuss several other strategies and their theoretical background, as well as present some examples and advice on computation. Our focus is on MI, which is a statistically valid strategy for handling missing data, although we review other less sound methods, as well as direct maximum likelihood and Bayesian methods for estimating parameters, which are also valid approaches. The analysis of a multiply-imputed data set is now relatively standard using readily available statistical software. The creation of multiply-imputed data sets is more challenging than their analysis but still straightforward relative to other valid methods of handling missing data, and we discuss available software for doing so. Ad hoc methods, including using singly-imputed data sets, almost always lead to invalid inferences and should be eschewed, especially when the focus is on valid interval estimation or testing hypotheses.

1. Introduction

Missing data are a common problem with large databases in general and with epidemiological, medical, and health-care databases in particular. Missing data also occur in clinical trials when subjects fail to provide data at one or more time points or drop out, for reasons including lack of interest or untoward side effects. Data may also be "missing" due to death, although the methods described here

[☆] The findings and conclusions in this chapter are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention.

are generally not appropriate for such situations because such values are not really missing (see Little and Rubin, 2002, Example 1.7; Zhang and Rubin, 2003; Rubin, 2006).

Epidemiological and medical databases nearly always have missing data. Unit nonresponse occurs when a selected unit (e.g., patient, doctor, hospital) does not provide any of the information being sought. Item nonresponse occurs when a unit responds to some items but not to others. Discussions of many issues related to missing data are contained in the three volumes produced by the Panel on Incomplete Data of the Committee on National Statistics in 1983 (Madow et al., 1983a, 1983b; Madow and Olkin, 1983), as well as in the volume stimulated by the 1999 International Conference on Survey Nonresponse (Groves et al., 2002).

A classical textbook on analysis with missing data (Little and Rubin, 1987, 2002) categorizes methods for analyzing incomplete data into four main groups. The first group comprises simple procedures such as complete-case analysis (also known as "listwise deletion") and available-case analysis, which discards the units with incomplete data in different ways. Although these simple methods are relatively easy to implement, they can often lead to inefficient and biased estimates. The second group of methods comprises weighting procedures, which deals with unit nonresponse by increasing the survey weights for responding units in an attempt to account for the nonrespondents, who are dropped from further analysis. The third group comprises imputation-based procedures, a standard approach for handling item nonresponse, especially in databases that are to be shared by many users. Imputation methods fill in values that are missing, and the resultant completed data are then analyzed as if there never were any missing values.

Of particular interest, multiple imputation (MI) is a method for reflecting the added uncertainty due to the fact that imputed values are not actual values, and yet still allows using complete-data methods to analyze each data set completed by imputation. The final group of methods comprises direct analyses using model-based procedures, in which models are specified for the observed data, and inferences are based on likelihood or Bayesian analyses. In general, only MI and direct analysis can lead to valid inferences. By valid inferences we mean ones that satisfy three criteria:

(a) approximately unbiased estimates of population estimates (e.g., means, correlation coefficients),
(b) interval estimates with at least their nominal coverage (e.g., 95% intervals for a population mean should cover the true population mean at least 95% of the time), and
(c) tests of significance should reject at their nominal level or less frequently when the null hypothesis is true (e.g., a 5% test of a zero population correlation should reject at most 5% of the time when the population correlation is zero).

Resampling methods, such as the bootstrap and jackknife, can satisfy criteria (b) and (c) asymptotically, but give no guidance on how to satisfy criterion (a) in the presence of missing data, but rather implicitly assume that estimates satisfying (a) have already been obtained (see Efron, 1994 and the discussion by Rubin,

1994). Such methods are only briefly discussed in Section 4.2 because they do not represent a complete approach to the problem of missing data.

This chapter reviews these four classes of approaches to handling missing data, with a focus on MI, which we believe is the most generally useful approach for medical and epidemiological databases. Before presenting our review of approaches, we start with a basic discussion of missing-data mechanisms, i.e., the processes that govern why certain values are missing and others are observed.

2. Missing-data mechanisms and ignorability

When data are missing, it is important to distinguish various missing-data mechanisms, which describe to what extent missingness depends on the observed and/or unobserved data values. Many simple methods for dealing with missing data are based, either implicitly or explicitly, on the assumption of a particularly simple missing-data mechanism, and these methods' behavior can be influenced strongly by differences between the assumed and the true mechanisms. More formally, let Y represent the $N \times P$ matrix of complete data, and let R represent the $N \times P$ matrix of indicator values for observed and missing values in Y. Then, the missing-data mechanism gives the probability of the matrix of indicator variables, R, given Y and possible parameters governing this process, $\xi : p(R|Y, \xi)$.

Key concepts about missing-data mechanisms were formalized by Rubin (1976), and following this work, subsequent statistical literature (e.g., Little and Rubin, 2002, p. 12) distinguishes three cases: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). This language was chosen to be consistent with much older terminology in classical experimental design for completely randomized, randomized, and not randomized studies.

MCAR refers to missing data for which missingness does not depend on any of the data values, missing or observed. Thus, the probability that units provide data on a particular variable does not depend on the value of that variable or the value of any other variable: $p(R|Y, \xi) = p(R|\xi)$. The MCAR assumption can be unrealistically restrictive and can be contradicted by the observed data, for example, when men are observed to have a higher rate of missing data on postoperative blood pressure than women.

Often, it is plausible to assume that missingness can be explained by the observed values in the data set. For example, in an epidemiological survey, the missingness for certain medical variables might depend on completely observed variables such as gender, age group, health conditions, social status, etc. If the probability of units responding to items depends only on such observed values but not on any missing values, then the missing data are MAR, but not necessarily MCAR because of the following possible dependence: $p(R|Y, \xi) = p(R|Y_{\text{obs}}, \xi)$, where Y_{obs} are observed values in Y, $Y = (Y_{\text{obs}}, Y_{\text{mis}})$, Y_{mis} being the missing values in Y. Thus, if the value of blood pressure at the end of a clinical trial is more likely to be missing when some previously observed values of blood pressure are high, and given these, the probability of missingness is

independent of the value of blood pressure at the end of the trial, the missingness mechanism is MAR.

If, even given the observed values, missingness still depends on data values that are missing, the missing data are NMAR. This could be the case, for example, with final blood pressure, if people with higher final blood pressure tend to be less likely to provide their blood pressure than people with lower final blood pressure, even though they have the exact same observed values of race, education, all previous blood pressure measurements, etc. Obviously, the richer the data set in terms of observed variables, the more plausible the MAR assumption becomes.

In addition to defining formally the concepts underlying MCAR, MAR, and NMAR, Rubin (1976) defined the concept of ignorability. Suppose that, in a situation with missing data, parametric models have been specified for: (1) the distribution of the data that would occur in the absence of missing value, $p(Y|\psi)$, and (2) the missing-data mechanism, $p(R|Y,\xi)$. Rubin (1976) showed that if the missing data are MAR and the parameters of the data distribution, ψ, and the missing-data mechanism, ξ, are distinct (which means, in disjoint parameter spaces and, if Bayesian models are used, a priori independent), then valid inferences about the distribution of the data can be obtained using a likelihood function that does not contain a factor for the missing-data mechanism and is simply proportional to $p(Y_{\text{obs}}|\psi) = \int p(Y|\psi)dY_{\text{mis}}$. In this situation, the missing-data mechanism may be "ignored" for likelihood or Bayesian inferences.

In many cases, it is reasonable to assume that the parameters of the data distribution and the missing-data mechanism are distinct, so that the practical question of whether the missing-data mechanism is ignorable often reduces to a question of whether the missing data are MAR. This argument requires some care, however, when using random parameter models, where there can exist ambiguity between unknown parameters and missing data (see Shih, 1992). Also, even when the parameters are not distinct, if the missing data are MAR, then inferences based on the likelihood ignoring the missing-data mechanism are still potentially valid in the sense of satisfying criteria (a)–(c) of Section 1, but may not be fully efficient. Thus, the MAR condition is typically regarded as the more important one in considerations of ignorability. Little and Rubin (2002, Section 6.2) include further discussion of these ideas, as does Rubin (1978b) in a very simple but instructive artificial example.

It is common to make the ignorability assumption in analyses of incomplete data even when it is not known to be correct, and it can be advantageous to do so for a variety of reasons. First, it can simplify analyses greatly. Second, the MAR assumption is often reasonable, especially when there are fully observed covariates available in the analysis to "explain" the reasons for the missingness; further, MAR cannot be contradicted by the observed data without the incorporation of external assumptions such as exact normality of variables. Third, even when the missing data are NMAR, an analysis based on the assumption of MAR can be helpful in reducing bias by effectively imputing missing data using relationships that are observed. Finally, even if the missing data are NMAR, it is usually not at all easy to specify a correct nonignorable model, for the simple reason that any

evidence concerning the relationship of missingness to the missing values is absent because the missing values are, by definition, not observed (for example, see Rubin et al., 1995).

3. Simple approaches to handling missing data

3.1. Complete-case analysis

The simplest analysis with incomplete data is to delete all units (cases) with at least one missing variable, i.e., to use "complete-case" analysis (sometimes called listwise deletion). This approach is generally biased unless the missing data are MCAR; the degree of bias depends on (a) the amount of the missing data, (b) the degree to which the assumption of MCAR is violated, and (c) the particular analysis being implemented. Even when complete-case analysis is unbiased, it can be highly inefficient, especially with multivariate data sets. For example, consider a data set with 20 variables, each of which has probability of being missing of .05, and suppose that missingness on each variable is independent of missingness on the other variables. Then, the probability of a unit having complete data is $(.95)^{20} = .36$, so that complete-case analysis would be expected to include only 36% of the units, and many of the discarded units have a large fraction of their values observed.

3.2. Available-case analysis

A simple alternative to the complete-case method is to include all units that have complete data on the variables that are needed for the analysis being considered. This approach, "available-case" analysis, can be regarded as "complete-case analysis restricted to the variables of interest." Available-case analysis retains at least as many of the data values as does complete-case analysis. However, it can be problematic when more than one quantity is estimated and the different estimates are compared or combined, because the sample base generally changes from one estimated quantity to the next. For example, if summaries of different variables are to be compared, the set of units for which each variable is summarized can differ across variables, and the summaries can be incomparable if the missing data are not MCAR; an extreme artificial illustration of incomparable estimation using available-case analyses would occur if last year's mean cholesterol were based on males because it was not collected for females, and this year's were based on females because it was not collected for males. As an extreme example in the context of combining estimates, if the covariance of two variables and their individual standard deviations have been estimated using available-case analyses, when these estimates are combined to estimate the correlation between the two variables, the resulting estimated correlation can lie outside the range $[-1, 1]$.

Complete-case analysis and available-case analysis were often the default treatments of missing data in older software packages, and they are simple to implement, which is undeniably attractive. However, as just discussed, they can

have serious deficiencies, which can be avoided when using more modern and more appropriate methods.

3.3. *Weighting adjustments*

For the case of unit nonresponse in surveys, a modification of complete-case analysis that can help to remove bias when the missing data are not MCAR is to weight the complete cases (i.e., the respondents) based on background information that is available for all of the units in the survey. For example, when a nonrespondent matches a respondent with respect to background variables that are observed for both, the nonrespondent's weight is simply added to the matching respondent's weight, and the nonrespondent is discarded. Because the match is defined by observed variables, such adjustments implicitly assume MAR.

A weighting procedure was used, for example, in the National Health Interview Survey (Botman et al., 2000). Typically, even if there were no adjustments for unit nonresponse in a survey, each sampled unit would already be weighted by the inverse of its probability of its selection, so that unbiased estimates of certain population quantities, such as totals, under repeated sampling could be calculated using those weights. The basic idea underlying a weighting adjustment for unit nonresponse is to treat unit nonresponse as an extra layer of sampling, which is accurate assuming ignorability, and then to weight each responding unit by the inverse of its estimated probability of both selection *and* response. For dealing with item nonresponse, the use of weighting adjustments is nearly always problematic, in large part because discarding the incomplete cases discards additional observed data that are not used in creating the weighting adjustment. Therefore, the standard method for handling item nonresponse in surveys is imputation, discussed in the next two sections. For further discussion of weighting procedures for nonresponse in general, see Bethlehem (2002), Gelman and Carlin (2002), and Little and Rubin (2002, Section 3.3).

4. Single imputation

Single imputation refers to imputing one value for each missing datum. Singly imputed data sets are straightforward to analyze using standard complete-data methods, which is again an undeniably attractive feature. Little and Rubin (2002, p. 72) offer the following guidelines for creating imputations. They should be: (1) conditional on observed variables; (2) multivariate, to reflect associations among missing variables; and (3) randomly drawn from predictive distributions rather than set equal to means, to ensure that correct variability is reflected. Methods for single imputation typically assume ignorability, and for simplicity, we concentrate discussion on the ignorable case.

4.1. *Simple imputation methods*

Unconditional mean imputation, which replaces each missing value with the mean of the observed values of that variable, meets none of the three guidelines

listed above. Regression imputation can satisfy the first two guidelines by replacing the missing values for each variable with the values predicted from a regression (e.g., least squares, logistic) of that variable on other variables. Replacing missing values of each variable with the mean of that variable calculated within cells defined by categorical variables is a special case of regression imputation. Stochastic regression imputation adds random noise to the value predicted by the regression model, and when done properly can meet all three guidelines for single imputation.

Hot-deck imputation replaces each missing value with a random draw from a "donor pool" consisting of values of that variable observed on units similar to the unit with the missing value. Donor pools are selected, for example, by choosing units with complete data who have "similar" observed values to the unit with missing data, e.g., by exact matching on their observed values or using a distance measure (metric) on observed variables to define "similar." When the distance is defined as the difference between units on the predicted value of the variable to be imputed (Rubin, 1986), the imputation procedure is termed "predictive mean matching imputation" (Little, 1988). Hot-deck imputation, when done properly, can also satisfy all three of the guidelines listed above for single imputation.

Suppose that single imputations have been created following the three guidelines of Little and Rubin (2002) mentioned above. Then, analyzing such a singly imputed data set with standard complete-data techniques is straightforward and can lead to approximately unbiased point estimates under ignorability. This approach then satisfies criterion (a) of Section 1. However, the resulting analyses will nearly always result in estimated standard errors that are too small, confidence intervals that are too narrow, and p-values for hypothesis tests that are too significant, regardless of how the imputations were created, thus failing to satisfy criteria (b) and (c). The reason is that imputed data are treated by standard complete-data analyses as if they were known with no uncertainty. Thus, single imputation followed by a complete-data analysis that does not distinguish between real and imputed values is almost always statistically invalid.

4.2. Interval estimation after single imputation

Special methods for variance estimation following single imputation have been developed for specific imputation procedures and estimation problems; see, for example, Schafer and Schenker (2000) and Lee et al. (2002). However, such techniques need to be customized to the imputation method used and to the analysis methods at hand, and they often require the user to have information from the imputation model that is not typically available in shared data sets. A more broadly applicable but computationally intensive approach with singly imputed data is to use a replication technique such as balanced repeated replication, the jackknife, or the bootstrap for variance estimation, with the imputation procedure repeated separately for each replicate; see, for example, Efron (1994) and Shao (2002). But, again, such replication methods assume criterion (a) has been satisfied by the single imputation method.

Multiple imputation (MI), described in Section 5, is a generally valid approach (i.e., satisfying criteria (a)–(c)), that is broadly applicable but less computationally intensive than the replication approach just mentioned, and it is thus particularly useful in the context of creating data sets to be shared by many users. MI simply involves repeating the drawing of single imputations several times, but its exact validity requires that the imputations are "proper" (Rubin, 1987), or more generally "confidence proper" (Rubin, 1996), both of which satisfy the three criteria of Little and Rubin (2002) for imputation.

4.3. Properly drawn single imputations

For notational simplicity, assume ignorability of the missing-data mechanism, even though the ignorability assumption is not necessary for MI to be appropriate. A proper imputation is often most easily obtained as a random draw from the "posterior predictive distribution" of the missing data given the observed data, which formally can be written as: $p(Y_{\text{mis}}|Y_{\text{obs}}) = \int p(Y_{\text{mis}}, \psi|Y_{\text{obs}})d\psi = \int p(Y_{\text{mis}}|Y_{\text{obs}}, \psi)p(\psi|Y_{\text{obs}})d\psi$. This expression effectively gives the distribution of the missing values, Y_{mis}, given the observed values, Y_{obs}, under a model for Y governed by ψ, $p(Y|\psi)p(\psi)$, where $p(\psi)$ is the prior distribution on ψ. The distribution $p(Y_{\text{mis}}|Y_{\text{obs}})$ is called "posterior" because it is conditional on the observed Y_{obs}, and it is called "predictive" because it predicts the missing Y_{mis}. It can be proper" because it reflects all uncertainty, including in parameter estimation, by taking draws of ψ from its posterior distribution, $p(\psi|Y_{\text{obs}})$, before using ψ to impute the missing data, Y_{mis}, from $p(Y_{\text{mis}}|Y_{\text{obs}}, \psi)$. More details are given in Sections 4.4 and 4.5.

Rubin (1987, Chapter 4) labeled imputation methods that do not account for all sources of variability as "improper." Thus, for example, fixing ψ at a point estimate $\hat{\psi}$, and then drawing m imputations for Y_{mis} independently with density $p(Y_{\text{mis}}|Y_{\text{obs}}, \hat{\psi})$, would constitute an improper MI procedure.

For simple patterns of missing data, such as with only one variable subject to missingness, the two-step paradigm of drawing ψ from $p(\psi|Y_{\text{obs}})$ and then drawing Y_{mis} from $p(Y_{\text{mis}}|Y_{\text{obs}}, \psi)$ is relatively straightforward to implement. For a simple example, Rubin and Schenker (1987) described its use in the context of fully parametric imputation involving logistic regression models. These steps can also incorporate more nonparametric analogs. The simple hot-deck procedure that randomly draws imputations for incomplete cases from matching complete cases is not proper because it ignores the sampling variability due to the fact that the population distribution of complete cases is not known, but rather it is estimated from the complete cases in the sample. Rubin and Schenker (1986, 1991) described a two-step procedure, termed "approximate Bayesian bootstrap imputation," which draws a bootstrap sample from the complete cases and then draws imputations randomly from the bootstrap sample. The initial bootstrap step is a nonparametric analog to the process of drawing a value ψ^* with density $p(\psi|Y_{\text{obs}})$, and the subsequent hot-deck step is a nonparametric analog to the process of drawing a value of Y_{mis} with density $p(Y_{\text{mis}}|Y_{\text{obs}}, \psi^*)$. Dorey et al. (1993) combined an initial bootstrap step with a fully parametric second step,

whereas Schenker and Taylor (1996) combined a fully parametric first step with predictive mean matching imputation at the second step. Finally, Heitjan and Little (1991) combined an initial bootstrap step with bivariate predictive mean matching imputation at the second step.

4.4. Properly drawing imputations with monotone missingness

If the missing data follow a monotone pattern, it is straightforward to draw random samples from $p(Y_{\text{mis}}|Y_{\text{obs}})$. When the missing data are not monotone, iterative computational methods are generally necessary, as described in Section 4.5. A missing-data pattern is monotone if the rows and columns of the data matrix can be sorted so that an irregular staircase separates Y_{obs} and Y_{mis}. Figure 1 illustrate monotone missing-data patterns. Missing data in clinical trials are often monotone or nearly monotone when data are missing due to patient dropout, where once a patient drops out, the patient never returns. Similarly some longitudinal surveys have monotone or nearly monotone missingness patterns when people who drop out never return.

Let Y_0 represent fully observed variables, Y_1 the incompletely observed variable with the fewest missing values, Y_2 the variable with the second fewest missing values, and so on, and assume a monotone pattern of missingness. Proper imputation with a monotone missing-data pattern begins by fitting an appropriate model to predict Y_1 from Y_0 and then using this model to impute the missing values in Y_1. For example, fit a least squares regression of Y_1 on Y_0 using the units with Y_1 observed, draw the regression parameters of this model from their posterior distribution, and then draw the missing values of Y_1 given these drawn parameters and the observed values of Y_0. Next impute the missing values for Y_2 using Y_0 and the observed and imputed values of Y_1; for example, if Y_2 is dichotomous, use a logistic regression model for Y_2 given (Y_0, Y_1). Continue to impute the next most complete variable until all missing values have been imputed. The collection of imputed values is a proper imputation of the missing data, Y_{mis}, under this model, and the collection of univariate prediction models defines the implied full imputation model, $p(Y_{\text{mis}}|Y_{\text{obs}})$. When missing data are not monotone, this method of imputation as described cannot be used directly to define $p(Y_{\text{mis}}|Y_{\text{obs}})$.

4.5. Properly drawing imputations with nonmonotone missingness

Creating imputations when the missing-data pattern is nonmonotone generally involves iteration because the distribution $p(Y_{\text{mis}}|Y_{\text{obs}})$ is often difficult to draw from directly. However, the Data-Augmentation algorithm (DA; Tanner and Wong, 1987), a stochastic version of the Expectation-Maximization algorithm (EM; Dempster et al., 1977), is often straightforward to implement. Briefly, DA involves iterating between randomly sampling missing data given a current draw of the model parameters and randomly sampling model parameters given a current draw of the missing data. The draws of Y_{mis} form a Markov Chain whose stationary distribution is $p(Y_{\text{mis}}|Y_{\text{obs}})$.

Thus, once the Markov Chain has reached effective convergence, a draw of Y_{mis} obtained by DA is effectively a single proper imputation of the missing data from the correct target distribution $p(Y_{mis}|Y_{obs})$, the posterior predictive distribution of Y_{mis}. Many of the programs discussed in Section 5.3 use DA or variants of DA to impute missing values. Other algorithms that use Markov Chain Monte Carlo methods for imputing missing values include the Gibbs sampler (Geman and Geman, 1984) and the Metropolis–Hastings algorithm (Metropolis and Ulam, 1949; Hastings, 1970). See, e.g., Gelman et al. (2003) for more details for these algorithms in general, and Schafer (1997) for the application of DA for imputation. This general approach is also discussed in Section 6.

An alternative to doing imputation under one specified model is to do imputation under potentially incompatible models, e.g., a potentially incompatible Gibbs sampler. These iterative simulation methods run a regression (e.g., least squares, logistic) on each variable with some missing data on all other variables using previously imputed values for these other variables, and then cycle through each variable with missing data. In fact, such regression imputation methods that are not necessarily derived from a joint distribution for all of the data have been more extensively developed recently, and they provide very flexible tools for creating imputations. As we will see in Section 5, such methods have gained prominence for the creation of MIs in recent years, although they have a relatively long history of application (e.g., Kennickell, 1991; Van Buuren and Oudshoorn, 2000; Raghunathan et al., 2001; Münnich and Rässler, 2005; Van Buuren et al., 2006). Further research should lead to greater understanding of the theoretical properties of such methods as well as to refinements of the methods in practice.

5. Multiple imputation

Multiple imputation (MI) was introduced by Rubin (1978a) and discussed in detail in Rubin (1987, 2004a, 2004b); it is an approach that retains the advantages of single imputation while allowing the uncertainty due to the process of imputation to be assessed directly and included to create valid inferences in many situations. MI is a simulation technique that replaces the missing values Y_{mis} with $m>1$ plausible values, and therefore reveals and quantifies uncertainty in the imputed values. Each set of imputations (i.e., each single imputation Y_{mis}) thus creates a completed data set, thereby creating m "completed" data sets: $Y^{(1)}$, ..., $Y^{(l)}$, ..., $Y^{(m)}$, where $Y^{(l)} = (Y_{obs}, Y^{(l)}_{mis})$. Typically m is fairly small; $m = 5$ is a standard number of imputations to use. Each of the m completed data sets is then analyzed as if there were no missing data, just as with single imputation, and the results of the m analyses are combined using simple rules described shortly.

Obtaining proper multiple-imputations is no more difficult than obtaining a single proper imputation because the process for obtaining a proper single imputation is simply repeated independently m times. Schafer (1997) is an excellent source for computational guidance on creating multiple-imputations

under a variety of models for the data Y. Multiple-imputations can be created under both ignorable and nonignorable models for missingness, although the use of ignorable models has been the norm, in part based on considerations of the type discussed at the conclusion of Section 2.

5.1. Combining rules for proper multiple imputation – scalar point estimates

Let θ represent the scalar estimand of interest (e.g., the mean of a variable, a relative risk, the intention-to-treat effect, etc.), let $\hat{\theta}$ represent the standard complete-data estimator of θ (i.e., the quantity calculated treating all imputed values of Y_{mis} as observed data), and let $\hat{V}(\hat{\theta})$ represent the standard complete-data estimated variance of $\hat{\theta}$.

Suppose MI has been used to create m completed data sets. A standard complete-data analysis of each will produce m completed data sets, each associated with completed-data statistics, say $\hat{\theta}_l$ and $\hat{V}_l = \hat{V}(\hat{\theta})_l$, $l = 1, \ldots, m$. The m sets of statistics are combined to produce the final point estimate $\hat{\theta}_{\text{MI}} = m^{-1}\Sigma_{l=1}^{m}\hat{\theta}_l$ and its estimated variance $T = W + (1 + m^{-1})B$, where $W = m^{-1}\Sigma_{l=1}^{m}\hat{V}_l$ is the "within-imputation" variance, $B = (m-1)^{-1}\Sigma_{l=1}^{m}(\hat{\theta}_l - \hat{\theta}_{\text{MI}})^2$ is the "between-imputation" variance, and the factor $(1 + m^{-1})$ reflects the fact that only a finite number of completed-data estimates $\hat{\theta}_l$, $l = 1, \ldots, m$, are averaged together to obtain the final point estimate. The quantity $\hat{\gamma} = (1 + m^{-1})B/T$ estimates the fraction of information about θ that is missing due to the missing data.

Inferences from multiply imputed data are based on $\hat{\theta}_{\text{MI}}$, T, and a Student's t reference distribution. Thus, for example, interval estimates for θ have the form $\hat{\theta}_{\text{MI}} \pm t(1-\alpha/2)\sqrt{T}$, where $t(1-\alpha/2)$ is the $(1-\alpha/2)$ quantile of the t distribution. Rubin and Schenker (1986) provided the approximate value $v_{\text{RS}} = (m-1)\hat{\gamma}^{-2}$ for the degrees of freedom of the t distribution, under the assumption that with complete data, a normal reference distribution would have been appropriate (i.e., the complete data would have had large degrees of freedom). Barnard and Rubin (1999) relaxed the assumption of Rubin and Schenker (1986) to allow for a t reference distribution with complete data, and proposed the value $v_{\text{BR}} = (v_{\text{RS}}^{-1} + \hat{v}_{\text{obs}}^{-1})^{-1}$ for the degrees of freedom in the MI analysis, where $\hat{v}_{\text{obs}} = (1-\hat{\gamma})(v_{\text{com}})(v_{\text{com}}+1)(v_{\text{com}}+3)$, and v_{com} is the complete-data degrees of freedom.

See Rubin and Schenker (1991) for additional methods for combining vector-valued estimates, significance levels, and likelihood ratio statistics; also see Little and Rubin (2002, Section 10.2). These sources summarize work done in Meng and Rubin (1992) and Li et al. (1991).

5.2. Discussion of MI in practice

A feature of imputation, either single or multiple, that gives such procedures great inherent flexibility and is especially attractive in the context of data sets that are shared by many users, is that the implicit or explicit model used for imputation, i.e., that leads to $p(Y_{\text{mis}}|Y_{\text{obs}})$, need not be the same as the explicit or implicit model used in subsequent analyses of the completed data. Thus, for example, an organization distributing public-use data can do its best job at imputing missing

data, and then secondary analysts are free to explore a variety of models for analyzing the completed data. The formal derivation of procedures for analyzing multiply imputed data, however, is based on the assumption that the imputer's and analyst's models are compatible, in the sense that the imputation model is proper or confidence proper. Formally, the imputer's and analyst's models must be "congenial" (Meng, 1994) for the resulting analyses to be fully valid. Such congeniality can be enforced more easily when the imputer and analyst are the same entity or communicate with each other. In the context of shared data sets, however, to promote near-congeniality of the imputer's and user's implicit models, so that analyses based on multiply imputed data will be at least approximately valid, the imputer should include as rich a set of variables in the imputation model as possible in order to accommodate the variety of analyses that might be carried out by secondary analysts. For example, when the data come from a complex sample survey, variables reflecting features of the sample design should be included as well (e.g., variables used to determine sampling weights, these weights themselves, stratification indicators); this was done, for instance, when NHANES III was multiply imputed (Ezzati-Rice et al., 1993) as well as when NMES was multiply imputed (Rubin, 2003).

This advice to include as many variables as possible in an MI model was present from the beginning (e.g., Rubin, 1987). Especially important is to include variables used in the design of the data collection, such as variables used to derive sampling weights, or the sampling weights themselves. Also critical is to include domain indicators when domain estimates are to be obtained by subsequent users. There are some criticisms of MI's sampling variance estimation equations in situations when such critical variables are excluded from the MI model (e.g., Kim et al., 2006). Obviously, if a statistical method is implemented in a way that does not even approximate its correct use, resulting answers cannot be valid in general. Although the focus in these criticisms has been on sampling variance estimation, even the point estimates based on an imputation model that excludes weights or domain indicators will be biased in general, so the issue of biased sampling variance estimation becomes secondary.

5.3. Software for multiple imputation

Many standard statistical software packages now have built-in or add-on functions for creating and analyzing multiply-imputed data sets. Routines for creating such data sets include, for example, the S-plus libraries NORM, CAT, MIX, and PAN, for multiply imputing normal, categorical, mixed, and panel data, respectively, which are freely available (see http://www.stat.psu.edu/~jls/misoftwa.html). NORM is also available as a stand-alone version, as is MICE–MI by chained equations (see http://web.inter.nl.net/users/S.van.Buuren/mi/hmtl/mice.htm). In addition, IVEware is very flexible and freely available; it can be called using SAS or can be run as a stand-alone version (http://www.isr.umich.edu/src/smp/ive/). SAS now has procedures PROC MI and PROC MIANALYZE making the analysis of multiply imputed data sets easy. Other software packages have been

developed specifically for creating multiply-imputed data sets, for example, the commercially available SOLAS (http://www.statsol.ie/solas/solas.htm), which has been available for years, is most appropriate for data sets with a monotone or nearly monotone pattern of missing data. Additionally, STATA provides MI routines based on the chained equation approach and supports analyses of multiply-imputed data sets. For more information, see www.multiple-imputation.com or for some historical perspective, see Horton and Lipsitz (2001).

6. Direct analysis using model-based procedures

Direct analyses of the incomplete data can be implemented by specifying a model for the complete data and then basing inferences on the likelihood or posterior distribution under that model. In its full generality, modeling the incomplete data is accomplished by simultaneously modeling both Y and R, as explicitly introduced in Rubin (1976). Selection models (e.g., Heckman, 1976) specify the marginal distribution of Y as well as how the distribution of R depends on Y, as follows:

$$p(Y, R|\psi, \xi) = p(Y|\psi)p(R|Y, \xi), \tag{1}$$

where ψ and ξ are unknown parameters. In contrast, pattern-mixture models (e.g., Rubin, 1977, 1978a; Little, 1993) specify the distribution of Y for each pattern of missing data (implied by R) as well as the probability of the various patterns occurring, as follows:

$$p(Y, R|\phi, \pi) = p(Y|R, \phi)p(R|\pi),$$

where ϕ and π are unknown parameters. When R is independent of Y, the missing data are MCAR, and the selection and pattern-mixture specifications are equivalent when $\psi = \phi$ and $\xi = \pi$, i.e., the implied models are the same. When the missing data are not MCAR, the two specifications generally differ.

Little and Rubin (2002, Chapter 15) discuss the use of selection and pattern-mixture approaches in the context of nonignorable missingness for a variety of types of data. As discussed earlier, the correct specification of nonignorable models is usually difficult due to lack of information in the data about the relationship between the missing-data mechanism and the missing values themselves. For this reason, selection models and pattern-mixture models for nonignorable missing data tend to depend strongly on assumptions about specific distributions. Thus, although they offer different and interesting approaches to modeling nonignorable missing data, it is suggested that they be used primarily for sensitivity analyses; as in Rubin (1977) and Little (1993), with a baseline analysis under ignorability being used as a primary point of comparison.

Consider now the situation of ignorable missing data. The observed data are Y_{obs} and R, and under the selection model specification given by expression (1),

the likelihood function based on the observed data is

$$L(\psi, \xi | Y_{obs}, R) \propto \int p(Y_{obs}, Y_{mis}|\psi)p(R|Y_{obs}, Y_{mis}, \xi)dY_{mis}. \quad (2)$$

As shown by Rubin (1976) and discussed previously, if the missing data are MAR (i.e., $p(R|Y_{obs}, Y_{mis}, \xi) = p(R|Y_{obs}, \xi)$), and if ψ and ξ are distinct, then inferences for ψ based on expression (2) are equivalent to inferences for ψ based on the likelihood for ψ ignoring the missing-data mechanism

$$L(\psi|Y_{obs}) \propto \int p(Y_{obs}, Y_{mis}|\psi)dY_{mis}, \quad (3)$$

because (2) factors into (3) and a factor that is free of ψ. Articles have appeared in the literature describing analyses of incomplete data under the assumption of ignorable missingness for a vast number of different analytic problems. Little and Rubin (2002, Chapters 11–14) review several such examples.

The remainder of this section describes two general techniques: (1) the EM algorithm (Dempster et al., 1977) and its extensions for maximum likelihood estimation of ψ, and (2) DA (Tanner and Wong, 1987) and its extensions for Bayesian posterior simulation. These techniques can be applied in the context of nonignorable missing data as well as that of ignorable missing data, but the presentation here is in the latter context for simplicity.

In many missing-data problems, even the observed-data likelihood (3) is complicated, and explicit expressions for maximum likelihood estimation of ψ are difficult to derive. The EM algorithm, a technique for computing maximum likelihood estimates iteratively, takes advantage of the facts that: (1) if ψ were known, it would be relatively easy to estimate many functions of Y_{mis}, and (2) if the data were complete, computation of maximum likelihood estimates would be relatively simple. Starting with an initial estimate of ψ, the EM algorithm iterates between two steps, an E-step (E for expectation) and an M-step (M for maximization), until convergence. Given the estimate of ψ at iteration t, $\psi^{(t)}$, the E-step computes the expected value of the complete-data log-likelihood given Y_{obs} and $\psi = \psi^{(t)}$, $Q(\psi|\psi^{(t)}) = \int \log L(\psi|Y)p(Y_{mis}|Y_{obs}, \psi = \psi^{(t)})dY_{mis}$; this step often involves computing the expected values of the complete-data sufficient statistics, which are linear in the data for exponential family distributions. Then, the M-step determines $\psi^{(t+1)}$ by maximizing the expected complete-data log-likelihood $Q(\psi|\psi^{(t)})$. For discussions of the theoretical properties of the EM algorithm, examples of its use, methods for obtaining standard errors based on the algorithm, and extensions, see Dempster et al. (1977), McLachlan and Krishnan (1997), Schafer (1997), and Little and Rubin (2002, Chapters 8, 9, and 11–15). Extensions of EM include the ECM (Meng and Rubin, 1993), ECME (Liu and Rubin, 1994), AECM (Meng and van Dyk, 1997), and PXEM (Liu et al., 1998) algorithms.

Bayesian inferences for ψ are based on the observed-data posterior distribution with density $p(\psi|Y_{obs}) \propto p(\psi)L(\psi|Y_{obs})$, where $p(\psi)$ is the prior density for ψ.

As is the case with maximum likelihood estimation, working explicitly with the observed-data posterior distribution can be difficult. DA, introduced in Section 4.5, facilitates the creation of draws of ψ from density $p(\psi|Y_{obs})$ using steps that are analogous to those of the EM algorithm but that involve simulation. In a simple version, DA begins with an initial approximation to $p(\psi|Y_{obs})$ and then iterates between two steps, an I-step, which imputes an updated value for Y_{mis}, and a P-step, which draws a value from an updated conditional posterior distribution for ψ, until convergence of the distribution of draws of Y_{mis} and ψ. Specifically, given the drawn value of ψ at iteration t, $\psi^{(t)}$, the I-step draws a value $Y_{mis}^{(t+1)}$ from density $p(Y_{mis}|Y_{obs},\psi^{(t)})$, and then the P-step draws a value $\psi^{(t+1)}$ from density $p(\psi|Y_{obs}, Y_{mis}^{(t+1)})$. As t increases, the draws $(Y_{mis}^{(t)},\psi^{(t)})$ converge in distribution to draws from joint density $p(Y_{mis},\psi|Y_{obs})$, and thus the draws $\psi^{(t)}$ converge in distribution to draws from density $p(\psi|Y_{obs})$. The empirical distribution of such multiple draws of ψ can be used to approximate the observed-data posterior distribution of ψ. The draws at successive iterations are serially correlated, however. Therefore, to obtain multiple independent draws from the observed-data posterior distribution of ψ, it is standard practice either to independently repeat the entire iterative procedure until convergence multiple times to generate multiple draws or to implement the iterative procedure once until convergence and then take every kth draw thereafter, with k chosen large enough to achieve approximate independence. For discussions of theoretical properties, extensions of DA, and examples of the use of Bayesian iterative simulation methods, see Tanner and Wong (1987), Gelfand and Smith (1990), Schafer (1997), and Little and Rubin (2002, Chapters 10–14). Gelman et al. (2003) is a good reference for related MCMC methods such as the Gibbs sampler and the Metropolis–Hastings algorithm.

For a specific problem, if the sample is large, likelihood-based analyses and Bayesian analyses under diffuse prior distributions are expected to give similar results, because the likelihood would be expected to dominate the prior distribution. For small samples, however, Bayesian analyses have the advantage of avoiding the assumption of asymptotic normality of the likelihood that is typically made. Moreover, results under various prior assumptions can be compared.

7. Examples

The examples we present here are from a randomized clinical trial and epidemiological databases. All use MI to address missing data rather than any of the ad hoc methods described at the start of this chapter or methods of direct analysis just described. We believe this emphasis is generally appropriate in epidemiology and medical statistics. In special cases, of course, methods other than MI can also be appropriate or even more appropriate.

7.1. *Missing data in Genzyme's Randomized Trial of Fabrazyme®*

Fabrazyme® is a synthetic enzyme developed by Genzyme Corporation to treat Fabry's disease, a rare and serious X-linked recessive genetic disorder that occurs due to an inability to metabolize creatinine. Preliminary results from a randomized trial of Fabrazyme® versus placebo revealed that the Fabrazyme® appeared to work well in patients in their 30s, who were not yet severely ill, in the sense that it lowered their serum creatinine substantially. A similar randomized clinical trial involved older patients who were more seriously ill. Since there was no other fully competitive treatment, it was desired to make Fabrazyme® commercially available earlier than initially planned, a decision that would allow patients randomized to placebo to begin taking Fabrazyme®, but would create missing $Y(0)$ outcome data among placebo patients once they began taking Fabrazyme®. The study had staggered enrollment because of the rareness of the condition, so that the number of monthly observations of serum creatinine for each placebo patient depended on the time of entry into the study. Figure 1 illustrates the general pattern of monotone missing data with the same length follow-up intended for each patient. Again, X represents baseline covariates, $Y(0)$ represents the repeated measures of serum creatinine for placebo patients, and $Y(1)$ represents the repeated measures of serum creatinine for Fabrazyme® patients.

In order to impute the missing outcomes under placebo, a complex hierarchical Bayesian model was developed for the progression of serum creatinine in untreated Fabry patients. In this model, inverse serum creatinine varies linearly and quadratically in time, and the prior distribution for the quadratic trend in placebo patients is obtained from the posterior distribution of the quadratic trend in an analogous model fit to a historical database of untreated Fabry patients. Thus, the historical patients' data only influence the imputations of the placebo patients' data rather subtly – via the prior distribution on the quadratic trend parameters.

Although the model fitting algorithm is complex, it is straightforward to use the algorithm to draw ψ from $p(\psi|Y_{obs})$ for the placebo patients, and then draw Y_{mis} in the placebo group conditional on the drawn value of ψ, where, as earlier, ψ represents all model parameters. Drawing the missing values in this way creates a sample from $p(Y_{mis}|Y_{obs})$ and thus an imputation for the missing values in the placebo group.

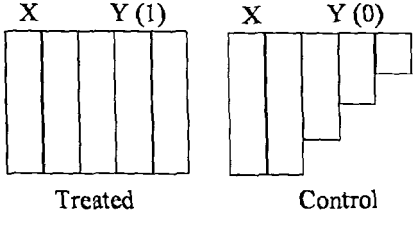

Fig. 1. Pattern of missing data for Genzyme trial.

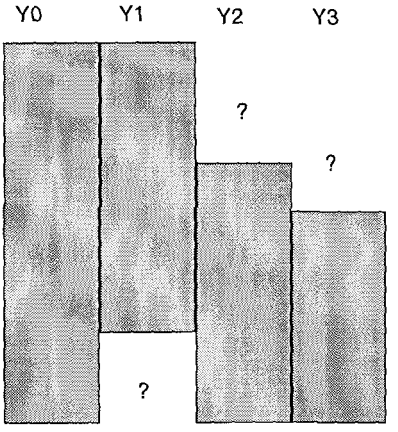

Fig. 2. Illustrative display for type of pattern of missing data in NMES.

7.2. Missing Data in NMES

The NMES collects data on a random sample of approximately 30,000 members of the US population, including hundreds of measurements of medical expenditures, background information, and demographic information. MI for NMES was more complicated than in the previous two examples because the missing-data pattern was not monotone. Figure 2 depicts a tremendous simplification of the missing-data pattern for NMES, where, if $Y1$ were fully observed, the missing-data pattern would be monotone.

Rubin (2003) imputed the missing data in NMES by capitalizing on the simplicity of imputation for monotone missing data by first imputing the missing values that destroyed the monotone pattern (the "nonmonotone missing values") and then proceeding as if the missing-data pattern were in fact monotone, and then iterating this process. More specifically, after choosing starting values for the missing data, iterate between the following two steps. (1) Regress each variable with any nonmonotone missing values (i.e., $Y1$), on all the other variables (i.e., $Y0$, $Y2$, $Y3$), treating the current imputations as true values, but use this regression to impute only the nonmonotone missing values. (2) Impute the remaining missing values in the monotone pattern; first impute the variable with the fewest missing values ($Y2$ in Fig. 2), then the variable with the second fewest missing values ($Y3$ in Fig. 2), and so on, treating the nonmonotone missing values inputed in Step 1 as known. This process was repeated five times to create five sets of imputations in the NMES example.

7.3. Missing data in the ABCs, a disease surveillance system

The Active Bacterial Core surveillance (ABCs) system is population-based and laboratory-based surveillance network. Five bacterial pathogens are monitored through the ABCs. These pathogens are: group A streptococcus, group B streptococcus, *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria*

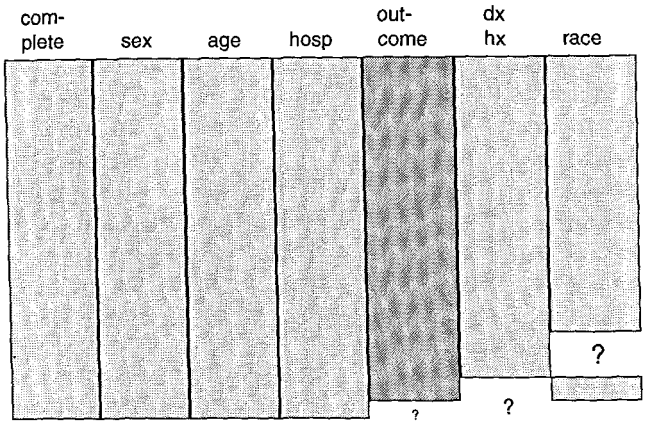

Fig. 3. Pattern of missing data in ABCs 2002.

meningitidis (Schuchat et al., 2001). A case of invasive disease is identified through bacteria isolated from a normally sterile site from an individual residing within the defined surveillance areas for each pathogen. Cases are identified through active contact with clinical laboratories – hence the label "active surveillance." Chart reviews are conducted to obtain demographic, clinical, and risk factor information. Additional susceptibility testing and serotyping is completed for ABCs pathogens at reference laboratories. Chart reviews generally suffer from missing data because not all information, particularly demographic data, are recorded in the medical record.

Multiple imputation (MI) ($m = 5$) was used to complete the data using a sequential regression multivariate approach (Raghunathan et al., 2001) implemented with IVEware (Raghunathan et al., 2002). This approach allows a model that accounts for the categorical and continuous nature of the variables in the database. It also allows for skip patterns, which are important because surveillance systems evolve over time with the case report forms adding or modifying variables. Different sites have implemented these changes to the case report form at different times. Also, for example, one site by law cannot report on a specific underlying disease. The missingness in this data set is close to monotone, as shown in Fig. 3.

8. Literature review for epidemiology and medical studies

Over the last decade, there has been increasing use of MI for databases in many areas of public health, clinical research, and epidemiology. Examples include cross-sectional survey data, longitudinal studies, clinical trials, surveillance systems, case–control studies, etc. These databases address quality of care, descriptive health statistics, AIDS-related studies, cancer mortality and survival rates, comparison of health-related costs and outcomes across countries, prognostic factors for cancer survival, and many other epidemiological, medically and clinically important questions. An intensive literature search and review brought up at least 40 articles using MI, as summarized in Appendix A.

There also have been many articles comparing, evaluating, and reviewing approaches to deal with missing data in many disciplines. For example, articles dealing with longitudinal studies have often compared "last observation carried forward" to MI techniques with the primary recommendation that MI is preferable to last observation carried forward. Others studies have compared complete-case analysis with MI and have found clear advantages when using MI to retain all observations for data analysis. There have been many review articles across different disciplines in health care (e.g., nursing) on the importance of addressing the missing-data problem correctly. A common theme when MI is used is the ease of data analysis using complete-data methods of analysis on multiply-imputed data sets, and the ease of creating multiply-imputed data sets using readily available statistical software packages. More than a hundred articles can be found very easily regarding the comparison and evaluation of missing-data techniques, as summarized in Appendix B.

9. Summary and discussion

Missing values are a common problem in medical and epidemiological databases. This entry has discussed concepts regarding mechanisms that create missing data, as well as strengths and weaknesses of commonly used approaches. Simple approaches, such as complete-case analysis and available-case analysis, are generally valid only when the missing data are MCAR. Even then, such approaches can be problematic.

Multiple imputation (MI) is especially useful in the context of data sets to be shared by many users, because of its general applicability and flexibility, as well as the fact that it allows the data producer to create one "adjustment" for missing data that can be used by all secondary data analysts. MI is also a useful technique in the context of designed missing data, such as when split questionnaire designs (also known as matrix sampling designs) are used to reduce costs and respondent burden (e.g., Raghunathan and Grizzle, 1995). Moreover, it offers potential for new analyses, e.g., in the context of censored data (see Gartner and Rässler, 2005 or Jensen et al., 2006).

For specific analyses problems in the presence of missing data, especially when the data producer and data analyst are the same entity, direct analyses of the incomplete data can be conducted. Techniques such as the EM and DA algorithms and their extensions are useful for handling the complexities created by missing data. MI has the advantage of flexibility over direct analyses, in the sense that the imputer can use one model to fill in the missing data, whereas the analyst can use a different model to draw inferences from the completed data. However, incompatibility of the two models can degrade the approximations underlying MI methods somewhat, although many evaluations in practice suggest that this degradation is often quite limited.

Because of uncertainties about correct models in the presence of missing data, it is useful to conduct sensitivity analyses under different modeling assumptions. In fact, this was one of the original motivations for MI. Rubin (1978a, 1987,

Chapter 1) recommended the creation of imputations under multiple models for purposes of sensitivity analysis, in addition to the creation of repeated imputations under a single model for assessments of variability due to missing data under that model. For examples of such sensitivity analyses, see Rubin (1977, 1986) and Rässler (2002).

Many of the approaches discussed herein can be applied under the assumption of either ignorable or nonignorable missing data. The assumption of ignorability cannot be contradicted directly by the observed data, and procedures that assume ignorability typically lead to at least partial corrections for bias due to missing data. Nonignorable models can be very difficult to specify, and their performance can be quite sensitive to modeling assumptions. Therefore, a sensible approach is to use ignorability as a "baseline" assumption, and to conduct additional sensitivity analyses using nonignorable models. For comparisons of the performance of ignorable and nonignorable models, see Glynn et al. (1986), Rubin et al. (1995), and Baker et al. (2003).

For interested readers, some recent books containing further discussion of topics covered in this chapter, as well as related topics, include Robert and Casella (1999), Groves et al. (2002), Little and Rubin (2002), and Gelman and Meng (2004).

Acknowledgements

This chapter borrows in places from two previous reviews written with other combinations of coauthors, and we thank them for their generosity. These two other articles are presented in the publications by Cook and Rubin (2005) and Rässler et al. (2007).

Appendix A

Boshuizen, H.C., Viet, A.L., Picavet, H.S.J., Botterweck, A., van Loon, A.J.M. (2006). Non-response in a survey of cardiovascular risk factors in the Dutch population: Determinants and resulting biases. *Public Health* **120**(4), 297–308.

Brancato, G., Pezzotti, P., Rapiti, E., Perucci, C.A., Abeni, D., Babbalacchio, A., Rezza, G. (1997). Multiple imputation method for estimating incidence of HIV infection. The Multicenter Prospective HIV Study. *International Journal of Epidemiology* **26**(5), 1107–1114.

Burd, R.S., Jang, T.S., Nair, S.S. (2006). Predicting hospital mortality among injured children using a national trauma database. *Journal of Trauma-Injury Infection & Critical Care* **60**(4), 792–801.

Choe, J.H., Koepsell, T.D., Heagerty, P.J., Taylor, V.M. (2005). Colorectal cancer among Asians and Pacific Islanders in the U.S.: Survival disadvantage for the foreign-born. *Cancer Detection & Prevention* **29**(4), 361–368.

Clark, T.G., Stewart, M.E., Altman, D.G., Gabra, H., Smyth, J.F. (2001). A prognostic model for ovarian cancer. *British Journal of Cancer* **85**(7), 944–952.

Clarke, A.E., Petri, M., Manzi, S., Isenberg, D.A., Gordon, C., Senecal, J.L., Penrod, J., Joseph, L., St. Pierre, Y., Fortin, P.R., Sutcliffe, N., Richard Goulet, J., Choquette, D., Grodzicky, T., Esdaile, J.M., Tri-Nation Study Group. (2004). The systemic lupus erythematosus Tri-Nation Study: Absence of a link between health resource use and health outcome. *Rheumatology* **43**(8), 1016–1024.

De Roos, A.J., Hartge, P., Lubin, J.H., Colt, J.S., Davis, S., Cerhan, J.R., Severson, R.K., Cozen, W., Patterson Jr., D.G., Needham, L.L., Rothman, N. (2005). Persistent organochlorine chemicals in plasma and risk of non-Hodgkin's lymphoma. *Cancer Research* **65**(23), 11214–11226.

European Collaborative Study. (1994). Caesarean section and risk of vertical transmission of HIV-1 infection. *Lancet* **343**, 1464–1467.

Fitten, L.J., Perryman, K.M., Wilkinson, C.J., Little, R.J., Burns, M.M., Pachana, N., Mervis, J.R., Malmgren, R., Siembieda, D.W., Ganzell, S. (1995). Alzheimer and vascular dementias and driving – A prospective road and laboratory study. *Journal of the American Medical Association* **273**, 1360–1365.

Fitzgerald, A.P., DeGruttola, V.G., Vaida, F. (2002). Modelling HIV viral rebound using non-linear mixed effects models. *Statistics in Medicine* **21**(14), 2093–2108.

Foshee, V.A., Bauman, K.E., Ennett, S.T., Suchindran, C., Benefield, T., Linder, G.F. (2005). Assessing the effects of the dating violence prevention program "safe dates" using random coefficient regression modeling. *Prevention Science* **6**(3), 245–258.

Freedman, V.A., Wolf, D.A. (1995). A case study on the use of multiple imputation. *Demography* **32**(3), 459–470.

Garfield, R., Leu, C.S. (2000). A multivariate method for estimating mortality rates among children under 5 years from health and social indicators in Iraq. *International Journal of Epidemiology* **29**(3), 510–515.

Hediger, M.L., Overpeck, M.D., McGlynn, A., Kuczmarski, R.J., Maurer, K.R., Davis, W.W. (1999). Growth and fatness at three to six years of age of children born small- or large-for-gestational age. *Pediatrics* **104**(3), e33.

Hennekens, C.H., Lee, I.-M., Cook, N.R., Hebert, P.R., Karlson, E.W., LaMotte, F., Mason, J.E., Buring, J.E. (1996). Self-reported breast implants and connective-tissue diseases in female health professionals. *Journal of the American Medical Association* **275**(8), 616.

Heo, M., Leibel, R.L., Boyer, B.B., Chung, W.K., Koulu, M., Karvonen, M.K., Pesonen, U., Rissanen, A., Laakso, M., Uusitupa, M.I., Chagnon, Y., Bouchard, C., Donohoue, P.A., Burns, T.L., Shuldiner, A.R., Silver, K., Andersen, R.E., Pedersen, O., Echwald, S., Sorensen, T.I., Behn, P., Permutt, M.A., Jacobs, K.B., Elston, R.C., Hoffman, D.J., Allison, D.B. (2001). Pooling analysis of genetic data: The association of leptin receptor (LEPR) polymorphisms with variables related to human adiposity. *Genetics* **159**(3), 1163–1178.

Hill, J.L., Waldfogel, J., Brooks-Gunn, J., Han, W.J. (2005). Maternal employment and child development: A fresh look using newer methods. *Developmental Psychology* **41**(6), 833–850.

Hopman, W.M., Berger, C., Joseph, L., Towheed, T., Van den Kerkhof, E., Anastassiades, T., Adachi, J.D., Ioannidis, G., Brown, J.P., Hanley, D.A.,

Papadimitropoulos, E.A., CaMos Research Group. (2006). The natural progression of health-related quality of life: Results of a five-year prospective study of SF-36 scores in a normative population. *Quality of Life Research* **15**(3), 527–536.

Hopman, W.M., Berger, C., Joseph, L., Towheed, T., Van den Kerkhof, E., Anastassiades, T., Cranney, A., Adachi, J.D., Loannidis, G., Poliquin, S., Brown, J.P., Murray, T.M., Hanley, D.A., Papadimitropoulos, E.A., Tenenhouse, A., CaMos Research Group. (2004). Stability of normative data for the SF-36: Results of a three-year prospective study in middle-aged Canadians. *Canadian Journal of Public Health-Revue Canadienne de Sante Publique* **95**(5), 387–391.

Huo, D., Lauderdale, D.S., Li, L. (2003). Influence of reproductive factors on hip fracture risk in Chinese women. *Osteoporosis International* **14**(8), 694–700.

Jackson, L.W., Schisterman, E.F., Dey-Rao, R., Browne, R., Armstrong, D. (2005). Oxidative stress and endometriosis. *Human Reproduction* **20**(7), 2014–2020.

Jiang, Y., Hesser, J.E. (2006). Associations between health-related quality of life and demographics and health risks. Results from Rhode Island's 2002 behavioral risk factor survey. *Health & Quality of Life Outcomes* **4**, 14.

Kagan, R.S., Joseph, L., Dufresne, C., Gray-Donald, K., Turnbull, E., St. Pierre, Y., Clarke, A.E. (2003). Prevalence of peanut allergy in primary-school children in Montreal, Canada. *Journal of Allergy & Clinical Immunology* **112**(6), 1223–1228.

Kessler, R.C., Adler, L., Barkley, R., Biederman, J., Conners, C.K., Demler, O., Faraone, S.V., Greenhill, L.L., Howes, M.J., Secnik, K., Spencer, T., Ustun, T.B., Walters, E.E., Zaslavsky, A.M. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey Replication. *American Journal of Psychiatry* **163**(4), 716–723.

Kessler, R.C., Adler, L.A., Barkley, R., Biederman, J., Conners, C.K., Faraone, S.V., Greenhill, L.L., Jaeger, S., Secnik, K., Spencer, T., Ustun, T.B., Zaslavsky, A.M. (2005). Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: Results from the National Comorbidity Survey Replication. Biological *Psychiatry* **57**(11), 1442–1451.

Kessler, R.C., Birnbaum, H., Demler, O., Falloon, I.R.H., Gagnon, E., Guyer, M., Howes, M.J., Kendler, K.S., Shi, L., Walters, E., Wu, E.Q. (2005). The prevalence and correlates of nonaffective psychosis in the National Comorbidity Survey Replication (NCS-R). *Biological Psychiatry* **58**(8), 668–676.

Lento, J., Glynn, S., Shetty, V., Asarnow, J., Wang, J., Belin, T.R. (2004). Psychologic functioning and needs of indigent patients with facial injury: A prospective controlled study. *Journal of Oral & Maxillofacial Surgery* **62**(8), 925–932.

Lu, M., Ma, C. (2002). Consistency in performance evaluation reports and medical records. *The Journal of Mental Health Policy & Economics* **5**(4), 141–152.

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research* **51**(5), 339–343.

Powers, J.R., Mishra, G., Young, A.F. (2005). Differences in mail and telephone responses to self-rated health: Use of multiple imputation in correcting for response bias. *Australian & New Zealand Journal of Public Health* **29**(2), 149–154.

Salvan, A., Thomaseth, K., Bortot, P., Sartori, N. (2001). Use of a toxicokinetic model in the analysis of cancer mortality in relation to the estimated absorbed dose of dioxin (2,3,7,8-tetrachlorodibenzo-*p*-dioxin, TCDD). *Science of the Total Environment* **274**(1–3), 21–35.

Sbarra, D.A., Emery, R.E. (2005). Coparenting conflict, nonacceptance, and depression among divorced adults: Results from a 12-year follow-up study of child custody mediation using multiple imputation. *American Journal of Orthopsychiatry* **75**(1), 63–75.

Serrat, C., Gomez, G., Garcia de Olalla, P., Cayla, J.A. (1998). CD4+ lymphocytes and tuberculin skin test as survival predictors in pulmonary tuberculosis HIV-infected patients. *International Journal of Epidemiology* **27**(4), 703–712.

Sheppard, L., Levy, D., Norris, G., Larson, T.V., Koenig, J.Q. (1999). Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987–1994. *Epidemiology* **10**(1), 23–30 (see Comment in *Epidemiology*, 1999, **10**(1), 1–4, PMID: 9888271; Comment in *Epidemiology*, 2000, **11**(3), 367–368, PMID: 10784265).

Stadler, W.M., Huo, D., George, C., Yang, X., Ryan, C.W., Karrison, T., Zimmerman, T.M., Vogelzang, N.J. (2003). Prognostic factors for survival with gemcitabine plus 5-fluorouracil based regimens for metastatic renal cancer. *Journal of Urology* **170**(4 Pt 1), 1141–1145.

Stefan, V.H. (2004). Assessing intrasample variation: Analysis of Rapa Nui (Easter Island) museum cranial collections example. *American Journal of Physical Anthropology* **124**(1), 45–58.

Taylor, J.M.G., Cooper, K.L., Wei, J.T., Sarma, A.V., Raghunathan, T.E., Heeringa, S.G. (2002). Use of multiple imputation to correct for nonresponse bias in a survey of urologic symptoms among African-American men. *American Journal of Epidemiology* **156**(8), 774–782.

Taylor, J.M., Munoz, A., Bass, S.M., Saah, A.J., Chmiel, J.S., Kingsley, L.A. (1990). Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation. Multicentre AIDS Cohort Study. *Statistics in Medicine* **9**(5), 505–514.

van Dijk, M.R., Steyerberg, E.W., Stenning, S.P., Habbema, J., Dik, F. (2006). Survival estimates of a prognostic classification depended more on year of treatment than on imputation of missing values. *Journal of Clinical Epidemiology* **59**(3), 246–253 (Review, 25 refs.).

Weinfurt, K.P., Castel, L.D., Li, Y., Sulmasy, D.P., Balshem, A.M., Benson 3rd, A.B., Burnett, C.B., Gaskin, D.J., Marshall, J.L., Slater, E.F., Schulman, K.A., Meropol, N.J. (2003). The correlation between patient characteristics and expectations of benefit from Phase I clinical trials. *Cancer* **98**(1), 166–175.

Weisfelt, M., van de Beek, D., Spanjaard, L., Reitsma, J.B., de Gans, J. (2006). Clinical features, complications, and outcome in adults with pneumococcal

meningitis: A prospective case series. *Lancet Neurology* **5**(2), 123–129 (see Comment in *Lancet Neurology*, 2006, **5**(2), 104–105, PMID: 16426981).

Appendix B

Ali, M.W., Siddiqui, O. (2000). Multiple imputation compared with some informative dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. *Journal of Biopharmaceutical Statistics* **10**(2), 165–181.

Allison, P.D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology* **112**(4), 545–557.

Arnold, A.M., Kronmal, R.A. (2003). Multiple imputation of baseline data in the cardiovascular health study. *American Journal of Epidemiology* **157**(1), 74–84.

Baccarelli, A., Pfeiffer, R., Consonni, D., Pesatori, A.C., Bonzini, M., Patterson Jr., D.G, Bertazzi, P.A., Landi, M.T. (2005). Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. *Chemosphere* **60**(7), 898–906.

Barnard, J., Meng, X.L. (1999). Applications of multiple imputation in medical studies: From AIDS to NHANES. *Statistical Methods in Medical Research* **8**(1), 17–36 (Review, 44 refs.).

Barnes, S.A., Lindborg, S.R., Seaman Jr., J.W. (2006). Multiple imputation techniques in small sample clinical trials. *Statistics in Medicine* **25**(2), 233–245.

Barzi, F., Woodward, M. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology* **160**(1), 34–45.

Bebchuk, J.D., Betensky, R.A. (2000). Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in Medicine* **19**(3), 405–419.

Bechger, T.M., Boomsma, D.I., Koning, H. (2002). A limited dependent variable model for heritability estimation with non-random ascertained samples. *Behavior Genetics* **32**(2), 145–151.

Beunckens, C., Molenberghs, G., Kenward, M.G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials* **2**(5), 379–386.

Carabin, H., Gyorkos, T.W., Joseph, L., Payment, P., Soto, J.C. (2001). Comparison of methods to analyse imprecise faecal coliform count data from environmental samples. *Epidemiology & Infection* **126**(2), 181–190.

Catellier, D.J., Hannan, P.J., Murray, D.M., Addy, C.L., Conway, T.L., Yang, S., Rice, J.C. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine & Science in Sports & Exercise* **37**(11 Suppl.), S555–S562.

Cheung, Y.B. (2002). Early origins and adult correlates of psychosomatic distress. *Social Science & Medicine* **55**(6), 937–948.

Choi, Y.J., Nam, C.M., Kwak, M.J. (2004). Multiple imputation technique applied to appropriateness ratings in cataract surgery. *Yonsei Medical Journal* **45**(5), 829–837.

Clark, T.G., Altman, D.G. (2003). Developing a prognostic model in the presence of missing data: An ovarian cancer case study. *Journal of Clinical Epidemiology* **56**(1), 28–37.

Collins, L.M., Schafer, J.L., Kam, C.M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* **6**(4), 330–351.

Cook, N.R. (1997). An imputation method for non-ignorable missing data in studies of blood pressure. *Statistics in Medicine* **16**(23), 2713–2728.

Cordell, H.J. (2006). Estimation and testing of genotype and haplotype effects in case–control studies: Comparison of weighted regression and multiple imputation procedures. *Genetic Epidemiology* **30**(3), 259–275.

Croy, C.D., Novins, D.K. (2005). Methods for addressing missing data in psychiatric and developmental research. *Journal of the American Academy of Child & Adolescent Psychiatry* **44**(12), 1230–1240.

Demirtas, H. (2005). Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* **24**(15), 2345–2363.

Dorey, F.J., Little, R.J., Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in Medicine* **12**(17), 1589–1603.

Faris, P.D., Ghali, W.A., Brant, R., Norris, C.M., Galbraith, P.D., Knudtson, M.L. (2002). Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *Journal of Clinical Epidemiology* **55**(2), 184–191.

Faucett, C.L., Schenker, N., Taylor, J.M.G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58**(1), 37–47.

Foster, E.M., Fang, G.Y., Conduct Problems Research Group. (2004). Alternative methods for handling attrition: An illustration using data from the fast track evaluation. *Evaluation Review* **28**(5), 434–464.

Fox-Wasylyshyn, S.M., El-Masri, M.M. (2005). Handling missing data in self-report measures. *Research in Nursing & Health* **28**(6), 488–495 (Review, 29 refs.).

Fridley, B., Rabe, K., de Andrade, M. (2003). Imputation methods for missing data for polygenic models. *BMC Genetics* **4**(Suppl. 1), S42.

Gauderman, W.J., Thomas, D.C. (2001). The role of interacting determinants in the localization of genes. *Advances in Genetics* **42**, 393–412 (Review, 63 refs.).

Geskus, R B. (2001). Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. *Statistics in Medicine* **20**(5), 795–812.

Goetghebeur, E., Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics* **56**(4), 1139–1144.

Graham, J.W., Hofer, S.M., Piccinin, A.M. (1994). Analysis with missing data in drug prevention research. *NIDA Research Monograph* **142**, 13–63 (Review, 27 refs.).

Greenland, S., Finkle, W.D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology* **142**(12), 1255–1264 (Review, 26 refs.).

Herrchen, B., Gould, J.B., Nesbitt, T.S. (1997). Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Computers & Biomedical Research* **30**(4), 290–305.

Hopke, P.K., Liu, C., Rubin, D.B. (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: Time-series concentrations of pollutants in the Arctic. *Biometrics* **57**(1), 22–33.

Houck, P.R., Mazumdar, S., Koru-Sengul, T., Tang, G.M, Benoit H., Pollock, B.G., Reynolds 3rd, C.F. (2004). Estimating treatment effects from longitudinal clinical trial data with missing values: Comparative analyses using different methods. *Psychiatry Research* **129**(2), 209–215.

Hunsberger, S., Murray, D., Davis, C.E., Fabsitz, R.R. (2001). Imputation strategies for missing data in a school-based multi-centre study: The Pathways study. *Statistics in Medicine* **20**(2), 305–316.

Joseph, L., Belisle, P., Tamim, H., Sampalis, J.S. (2004). Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. *Journal of Clinical Epidemiology* **57**(2), 147–153.

Kang, T., Kraft, P., Gauderman, W.J., Thomas, D., Framingham Heart Study. (2003). Multiple imputation methods for longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genetics* **4**(Suppl. 1), S43.

Kim, K.Y., Kim, B.J., Yi, G.S. (2004). Reuse of imputed data in microarray analysis increases imputation efficiency. *BMC Bioinformatics* **5**, 160.

Kistner, E.O., Weinberg, C.R. (2005). A method for identifying genes related to a quantitative trait, incorporating multiple siblings and missing parents. *Genetic Epidemiology* **29**(2), 155–165.

Kmetic, A., Joseph, L., Berger, C., Tenenhouse, A. (2002). Multiple imputation to account for missing data in a survey: Estimating the prevalence of osteoporosis. *Epidemiology* **13**(4), 437–444.

Kneipp, S.M., McIntosh, M. (2001). Handling missing data in nursing research with multiple imputation. *Nursing Research* **50**(6), 384–389 (Review, 12 refs.).

Kosorok, M.R., Wei, W.H., Farrell, P.M. (1996). The incidence of cystic fibrosis. *Statistics in Medicine* **15**(5), 449–462.

Kraft, P., Cox, D.G., Paynter, R.A., Hunter, D., De Vivo, I. (2005). Accounting for haplotype uncertainty in matched association studies: A comparison of simple and flexible techniques. *Genetic Epidemiology* **28**(3), 261–272.

Kristman, V.L., Manno, M., Cote, P. (2005). Methods to account for attrition in longitudinal data: Do they work? A simulation study. *European Journal of Epidemiology* **20**(8), 657–662.

Lam, K.F., Fong, D.Y.T., Tang, O.Y. (2005). Estimating the proportion of cured patients in a censored sample. *Statistics in Medicine* **24**(12), 1865–1879.

Lavori, P.W., Dawson, R., Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine* **14**(17), 1913–1925.

Lipsitz, S.R., Molenberghs, G., Fitzmaurice, G.M., Ibrahim, J. (2000). GEE with Gaussian estimation of the correlations when data are incomplete. *Biometrics* **56**(2), 528–536.

Little, R., Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics* **52**(4), 1324–1333.

Liu, G., Gould, A.L. (2002). Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of Biopharmaceutical Statistics* **12**(2), 207–226.

Liu, M., Taylor, J.M., Belin, T.R. (2000). Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics* **56**(4), 1157–1163.

Longford, N.T. (2001). Multilevel analysis with messy data. *Statistical Methods in Medical Research* **10**(6), 429–444.

Lu, K., Tsiatis, A.A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* **57**(4), 1191–1197.

Lubin, J.H., Colt, J.S., Camann, D., Davis, S., Cerhan, J.R., Severson, R.K., Bernstein, L., Hartge, P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives* **112**(17), 1691–1696.

Lyles, R.H., Fan, D., Chuachoowong, R. (2001). Correlation coefficient estimation involving a left censored laboratory assay variable. *Statistics in Medicine* **20**(19), 2921–2933.

Lynn, H.S. (2001). Maximum likelihood inference for left-censored HIV RNA data. *Statistics in Medicine* **20**(1), 33–45.

Manca, A., Palmer, S. (2005). Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. *Applied Health Economics & Health Policy* **4**(2), 65–75.

Mishra, G.D., Dobson, A.J. (2004). Multiple imputation for body mass index: Lessons from the Australian Longitudinal Study on Women's Health. *Statistics in Medicine* **23**(19), 3077–3087.

Molenberghs, G., Williams, P.L., Lipsitz, S. (2002). Prediction of survival and opportunistic infections in HIV-infected patients: A comparison of imputation methods of incomplete CD4 counts. *Statistics in Medicine* **21**(10), 1387–1408.

Moore, L., Lavoie, A., LeSage, N., Liberman, M., Sampalis, J.S., Bergeron, E., Abdous, B. (2005). Multiple imputation of the Glasgow Coma Score. *Journal of Trauma-Injury Infection & Critical Care* **59**(3), 698–704.

Morita, S., Kobayashi, K., Eguchi, K., Matsumoto, T., Shibuya, M., Yamaji, Y., Ohashi, Y. (2005). Analysis of incomplete quality of life data in advanced stage cancer: A practical application of multiple imputation. *Quality of Life Research* **14**(6), 1533–1544.

Munoz, A., Carey, V., Taylor, J.M., Chmiel, J.S., Kingsley, L., Van Raden, M., Hoover, D.R. (1992). Estimation of time since exposure for a prevalent cohort. *Statistics in Medicine* **11**(7), 939–952.

Naeim, A., Keeler, E.B., Mangione, C.M. (2005). Options for handling missing data in the Health Utilities Index Mark 3. *Medical Decision Making* **25**(2), 186–198.

Newgard, C.D. (2006). The validity of using multiple imputation for missing out-of-hospital data in a state trauma registry. *Academic Emergency Medicine* **13**(3), 314–324.

Nixon, R.M., Duffy, S.W., Fender, G.R.K. (2003). Imputation of a true endpoint from a surrogate: Application to a cluster randomized controlled trial with partial information on the true endpoint. *BMC Medical Research Methodology* **3**, 17.

Oostenbrink, J.B., Al, M.J. (2005). The analysis of incomplete cost data due to dropout. *Health Economics* **14**(8), 763–776.

Oostenbrink, J.B., Al, M.J., Rutten-van Molken, M.P.M.H. (2003). Methods to analyse cost data of patients who withdraw in a clinical trial setting. *Pharmacoeconomics* **21**(15), 1103–1112.

Paik, M.C. (1997). Multiple imputation for the Cox proportional hazards model with missing covariates. *Lifetime Data Analysis* **3**(3), 289–298.

Pan, W. (2000a). A multiple imputation approach to Cox regression with interval-censored data. *Biometrics* **56**(1), 199–203.

Pan, W. (2000b). A two-sample test with interval censored data via multiple imputation. *Statistics in Medicine* **19**(1), 1–11.

Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies. *Biometrics* **57**(4), 1245–1250.

Pan, W., Kooperberg, C. (1999). Linear regression for bivariate censored data via multiple imputation. *Statistics in Medicine* **18**(22), 3111–3121.

Patrician, P.A. (2002). Multiple imputation for missing data. *Research in Nursing & Health* **25**(1), 76–84 (Review, 24 refs.).

Penny, K.I., Jolliffe, I.T. (1999). Multivariate outlier detection applied to multiply imputed laboratory data. *Statistics in Medicine* **18**(14), 1879–1895 (discussion 1897).

Perez, A., Dennis, R.J., Gil, J.F.A., Rondon, M.A., Lopez, A. (2002). Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. *Statistics in Medicine* **21**(24), 3885–3896.

Pinnaduwage, D., Beyene, J., Fallah, S. (2003). Genome-wide linkage analysis of systolic blood pressure slope using the Genetic Analysis Workshop 13 data sets. *BMC Genetics* **4**(Suppl. 1), S86.

Raghunathan, T.E. (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health* **25**, 99–117.

Ratcliffe, J., Young, T., Longworth, L., Buxton, M. (2005). An assessment of the impact of informative dropout and nonresponse in measuring health-related quality of life using the EuroQol (EQ-5D) descriptive system. *Value in Health* **8**(1), 53–58.

Richardson, B.A., Flack, V.F. (1996). The analysis of incomplete data in the three-period two-treatment cross-over design for clinical trials. *Statistics in Medicine* **15**(2), 127–143.

Rubin, D.B., Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* **10**(4), 585–598 (Review, 26 refs.).

Schafer, J.L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research* **8**(1), 3–15 (Review, 29 refs.).

Schafer, J.L., Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods* **7**(2), 147–177.

Schiattino, I., Villegas, R., Cruzat, A., Cuenca, J., Salazar, L., Aravena, O., Pesce, B., Catalan, D., Llanos, C., Cuchacovich, M., Aguillon, J.C. (2005). Multiple imputation procedures allow the rescue of missing data: An application to determine serum tumor necrosis factor (TNF) concentration values during the treatment of rheumatoid arthritis patients with anti-TNF therapy. *Biological Research* **38**(1), 7–12.

Sinharay, S., Stern, H.S., Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods* **6**(4), 317–329 (Review, 31 refs.).

Souverein, O.W., Zwinderman, A.H., Tanck, M.W.T. (2006). Multiple imputation of missing genotype data for unrelated individuals. *Annals of Human Genetics* **70**(Pt 3), 372–381.

Stadler, W.M., Huo, D., George, C., Yang, X., Ryan, C.W., Karrison, T., Zimmerman, T.M., Vogelzang, N.J. (2003). Prognostic factors for survival with gemcitabine plus 5-fluorouracil based regimens for metastatic renal cancer. *Journal of Urology* **170**(4 Pt 1), 1141–1145.

Streiner, D.L. (2002). The case of the missing data: Methods of dealing with dropouts and other research vagaries. *Canadian Journal of Psychiatry-Revue Canadienne de Psychiatrie* **47**(1), 68–75.

Stuck, A.E., Aronow, H.U., Steiner, A., Alessi, C.A., Büla, C.J., Gold, M.N., Yuhas, K.E., Nisenbaum, R., Rubenstein, L.Z., Beck, J.C. (1995). A trial of annual in-home comprehensive geriatric assessments for elderly people living in the community. *The New England Journal of Medicine* **333**, 1184–1189.

Tang, L., Song, J., Belin, T.R., Unutzer, J. (2005). A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine* **24**(14), 2111–2128.

Touloumi, G., Babiker, A.G., Kenward, M.G., Pocock, S.J., Darbyshire, J.H. (2003). A comparison of two methods for the estimation of precision with incomplete longitudinal data, jointly modelled with a time-to-event outcome. *Statistics in Medicine* **22**(20), 3161–3175.

Twisk, J., de Vente, W. (2002). Attrition in longitudinal studies. How to deal with missing data. *Journal of Clinical Epidemiology* **55**(4), 329–337.

Van Beijsterveldt, C.E.M., van Boxtel, M.P.J., Bosma, H., Houx, P.J., Buntinx, F., Jolles, J. (2002). Predictors of attrition in a longitudinal cognitive

aging study: The Maastricht Aging Study (MAAS). *Journal of Clinical Epidemiology* **55**(3), 216–223.

Van Buuren, S., Boshuizen, H.C., Knook, D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* **18**(6), 681–694.

Wang, C.Y., Anderson, G.L., Prentice, R.L. (1999). Estimation of the correlation between nutrient intake measures under restricted sampling. *Biometrics* **55**(3), 711–717.

Wood, A.M., White, I.R., Hillsdon, M., Carpenter, J. (2005). Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes. *International Journal of Epidemiology* **34**(1), 89–99.

Wu, H., Wu, L. (2001). A multiple imputation method for missing covariates in non-linear mixed-effects models with application to HIV dynamics. *Statistics in Medicine* **20**(12), 1755–1769.

Wu, H., Wu, L. (2002). Identification of significant host factors for HIV dynamics modeled by non-linear mixed-effects models. *Statistics in Medicine* **21**(5), 753–771.

Xie, F., Paik, M.C. (1997). Multiple imputation methods for the missing covariates in generalized estimating equation. *Biometrics* **53**(4), 1538–1546.

Xue, X., Shore, R.E., Ye, X., Kim, M.Y. (2004). Estimating the dose response relationship for occupational radiation exposure measured with minimum detection level. *Health Physics* **87**(4), 397–404.

Zhou, X.H., Eckert, G.J., Tierney, W.M. (2001). Multiple imputation in public health research. *Statistics in Medicine* **20**(9–10), 1541–1549.

References

Baker, S.G., Ko, C.-W., Graubard, B.I. (2003). A sensitivity analysis for nonrandomly missing categorical data arising from a national health disability survey. *Biostatistics* **4**, 41–56.

Barnard, J., Rubin, D.B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika* **86**, 948–955.

Bethlehem, J.G. (2002). Weighting nonresponse adjustments based on auxiliary information. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, pp. 275–287.

Botman, S.L., Moore, T.F., Moriarity, C.L., Parsons, V.L. (2000). Design and estimation for the National Health Interview Survey, 1995–2004. *Vital and Health Statistics*, Series 2, No. 130. National Center for Health Statistics, Hyattsville, MD.

Cook, S.R., Rubin, D.B. (2005). Multiple imputation in designing medical device trials. In: Becker, K.M., Whyte, J.J. (Eds.), *Clinical Evaluation of Medical Devices*. Humana Press, Washington, DC.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39**, 1–38, (with discussion).

Dorey, F.J., Little, R.J.A., Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in Medicine* **12**, 1589–1603.

Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association* **89**, 463–479.

Ezzati-Rice, T.M., Fahimi, M., Judkins, D., Khare, M. (1993). Serial imputation of NHANES III with mixed regression and hot-deck techniques. *American Statistical Association Proceedings of the Section on Survey Research Methods* **1**, 292–296.

Gartner, H., Rässler, S. (2005). Analyzing the changing gender wage gap based on multiply imputed right censored wages, *IAB Discussion Paper* 5/2005.

Gelfand, A.E., Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A., Carlin, J.B. (2002). Poststratification and weighting adjustment. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, pp. 289–302.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall, London.

Gelman, A., Meng, X.L. (2004). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. Wiley, New York.

Geman, S., Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6**, 721–741.

Glynn, R., Laird, N.M., Rubin, D.B. (1986). Selection modeling versus mixture modeling with non-ignorable nonresponse. In: Wainer, H. (Ed.), *Drawing Inferences from Self-Selected Samples*. Springer, New York, pp. 119–146.

Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (2002). *Survey Nonresponse*. Wiley, New York.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* **57**, 97–109.

Heckman, J.J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**, 475–492.

Heitjan, D.F., Little, R.J.A. (1991). Multiple imputation for the fatal accident reporting system. *Journal of the Royal Statistical Society C* **40**, 13–29.

Horton, N.J., Lipsitz, S.R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician* **55**, 244–254.

Jensen, U., Gartner, H., Rässler, S. (2006). Measuring overeducation with earnings frontiers and multiply imputed censored income data, *IAB Discussion Paper* 11/2006.

Kennickell, A.B. (1991). Imputation of the 1989 survey of consumer finances: Stochastic relaxation and multiple imputation. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 1–10, (discussion 21–23).

Kim, J.K., Brick, J.M., Fuller, W.A., Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **68**, 509–521.

Lee, H., Rancourt, E., Särndal, C.E. (2002). Variance estimation for survey data under single imputation. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, pp. 315–328.

Li, K.H., Meng, X.L., Raghunathan, T.E., Rubin, D.B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* **1**, 65–92.

Little, R.J.A. (1988). Adjustments in large surveys. *Journal of Business & Economic Statistics* **6**(3), 287–296.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

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

Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley, Hoboken.

Liu, C., Rubin, D.B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–648.

Liu, C., Rubin, D.B., Wu, Y.N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85**, 755–770.

Madow, W.G., Nisselson, H., Olkin, I. (1983a). *Incomplete Data in Sample Surveys, Volume 1: Report and Case Studies*. Academic Press, New York.

Madow, W.G., Olkin, I. (1983). *Incomplete Data in Sample Surveys, Volume 3: Proceedings of the Symposium*. Academic Press, New York.

Madow, W.G., Olkin, I., Rubin, D.B. (1983b). *Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliographies*. Academic Press, New York.

McLachlan, G.J., Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9**, 538–573, (with discussion).

Meng, X.L., Rubin, D.B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika* **79**, 103–111.

Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278.

Meng, X.-L., van Dyk, D. (1997). The EM algorithm – An old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society, Series B (Methodological)* **59**, 511–567.

Metropolis, N., Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association* **49**, 335–341.

Münnich, R., Rässler, S. (2005). PRIMA: A new multiple imputation procedure for binary variables. *Journal of Official Statistics* **21**, 325–341.

Raghunathan, T.E., Grizzle, J.E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association* **90**, 54–63.

Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* **27**, 85–95.

Raghunathan, T.E., Solenberger, P., Van Hoewyk, J. (2002). *IVEware: Imputation and Variance Estimation Software – User Guide. Survey Methodology Program*. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Rässler, S. (2002). Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches. In: *Lecture Notes in Statistics*, 168. Springer, New York.

Rässler, S., Rubin, D.B., Schenker, N. (2007). Incomplete data: Diagnosis, imputation, and estimation. In: de Leeuw, E., Hox, J., Dillman, D. (Eds.), *The International Handbook of Survey Research Methodology*. Sage, Thousand Oaks, in press.

Robert, C.P., Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–590.

Rubin, D.B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72**, 538–543.

Rubin, D.B. (1978a). Multiple imputation in sample surveys – A phenomenological Bayesian approach to nonresponse. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 20–40.

Rubin, D.B. (1978b). A note on Bayesian, likelihood, and sampling distribution inferences. *The Journal of Educational Statistics* **3**, 189–201.

Rubin, D.B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics* **4**, 87–95.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

Rubin, D.B. (1994). Comments on "Missing data, imputation, and the bootstrap" by Bradley Efron. *Journal of the American Statistical Association* **89**, 475–478.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473–489.

Rubin, D.B. (2003). Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica* **57**, 3–18.

Rubin, D.B. (2004a). *Multiple Imputation for Nonresponse in Surveys*, 2nd ed. Wiley, New York, 1987.

Rubin, D.B. (2004b). The design of a general and flexible system for handling nonresponse in sample surveys. *The American Statistician* **58**, 298–302, (first appeared in unpublished form in 1977, prepared under contract for the U.S. Social Security Administration).

Rubin, D.B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Statistical Science* **21**, 299–309.

Rubin, D.B., Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* **81**, 366–374.

Rubin, D.B., Schenker, N. (1987). Interval estimation from multiply imputed data: A case study using census agriculture industry codes. *Journal of Official Statistics* **3**, 375–387.

Rubin, D.B., Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* **10**, 585–598.

Rubin, D.B., Stern, H., Vehovar, V. (1995). Handling "don't know" survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association* **90**, 822–828.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, New York.

Schafer, J.L., Schenker, N. (2000). Inference with imputed conditional means. *Journal of the American Statistical Association* **95**, 144–154.

Schenker, N., Taylor, J.M.G. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis* **22**, 425–446.

Schuchat, A., Hilger, T., Zell, E., Farley, M.M., Reingold, A., Harrison, L., Lefkowitz, L., Daniła, R., Stefonek, K., Barrett, N., Morse, D., Pinner, R. (2001). Active bacterial core surveillance of the emerging infections program network. *Emerging Infectious Diseases* **7**(1), 92–99.

Shao, J. (2002). Replication methods for variance estimation in complex sample surveys with imputed data. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, pp. 303–314.

Shih, W.J. (1992). On informative and random dropouts in longitudinal studies. *Biometrics* **48**, 970–972.

Tanner, M.A., Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Van Buuren, S., Brand, J.P.L., Oudshoorn, C.G.M., Rubin, D.B. (2006). Fully conditional specification in multivariate imputation. *Statistics in Medicine* **76**, 1049–1064.

Van Buuren, S., Oudshoorn C.G.M. (2000). Multivariate imputation by chained equations: MICE V1.0 user's manual. Report PG/VGZ/00.038. Leiden: TNO Preventie en Gezondheid.

Zhang, J.L., Rubin, D.B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics* **28**, 353–368.