HHS Public Access

Author manuscript

Biom J. Author manuscript; available in PMC 2017 May 01.

Doubly robust multiple imputation using kernel-based techniques

Chiu-Hsieh Hsu[1,2], **Yulei He**[3], **Yisheng Li**[4], **Qi Long**[5], and **Randall Friese**[6]

[1]Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health and Arizona Cancer Center, University of Arizona, 1295 N. Martin A232 Campus, PO Box 245211, Tucson, AZ 85724, USA

[2]Arizona Cancer Center, College of Medicine, University of Arizona, Tucson, AZ 85724, USA

[3]National Center for Health Statistics, CDC, Hyattsville, MD 20782, USA

[4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[5]Department of Biostatistics and Bioinformatics, School of Public Health, Emory University, Atlanta, GA 48109, USA

[6]Department of Surgery, College of Medicine, University of Arizona, Tucson, AZ 85724, USA

Summary

We consider the problem of estimating the marginal mean of an incompletely observed variable and develop a multiple imputation approach. Using fully observed predictors, we first establish two working models: one predicts the missing outcome variable, and the other predicts the probability of missingness. The predictive scores from the two models are used to measure the similarity between the incomplete and observed cases. Based on the predictive scores, we construct a set of kernel weights for the observed cases, with higher weights indicating more similarity. Missing data are imputed by sampling from the observed cases with probability proportional to their kernel weights. The proposed approach can produce reasonable estimates for the marginal mean and has a double robustness property, provided that one of the two working models is correctly specified. It also shows some robustness against misspecification of both models. We demonstrate these patterns in a simulation study. In a real data example, we analyze the total helicopter response time from injury in the Arizona emergency medical service data.

Keywords

Bandwidth; Bootstrap; Local imputation; Model Misspecification; Nonparametric; Product kernel

Correspondence: pchhsu@email.arizona.edu; Phone: +1-520-6265054; Fax: +1-520-6262467.

Conflict of Interest Statement: The authors have declared no conflict of interest.

Additional supporting information including source code to reproduce the results may be found in the online version of this article at the publisher's web-site.

1. Introduction

Data collected for scientific research often contain missing values. For example, the Arizona emergency medical service (AZEMS) data contain valuable information for studying relationships between total EMS response time from injury and transportation types of traumatically injured patients. However, the variable recording the total response time from injury has around 66% of data missing. These missing data, if inadequately accounted for, might lead to invalid inferences in relevant analyses (e.g., comparing the mean total response time between different types of transportation) and misleading policy implications (e.g., misjudging which type of transportation is more efficient for sending traumatically injured patients to hospitals).

An increasing number of multiple imputation (MI) (Rubin, 1987) approaches have been proposed for dealing with missing data problems. In general, MI involves replacing each missing datum with several (M) sets of plausible values drawn from a specified imputation model, resulting in several completed datasets (i.e., data with missing values filled in by imputations). Each completed dataset is analyzed separately by a standard complete-data method. The resulting inferences, including point estimates, covariance matrices, and p-values, can then be combined to formally incorporate imputation uncertainty using the formula given in Chap. 3 of Rubin (1987) and refinements in Chap. 10 of Little and Rubin (2002). The implementation of MI in several major statistical packages including SAS (www.sas.com), R (www.r-project.org), and STATA (www.stata.com) has made this missing data strategy increasingly popular among practitioners (Harel and Zhou, 2007).

The challenges of MI often lie in constructing good imputation models. For ease of illustration, suppose Y is the incomplete variable and X contains the fully observed predictors. The development of imputation models has been largely focused on the outcome model f(Y|X). These methods usually pose parametric assumptions regarding the distribution of the missing data and the underlying regression relationships. See van Buuren (2007) for a recent review. In addition, there exist several less-parametric alternatives. One strategy is to use smoothing techniques for modeling f(Y|X). For example, Titterington and Sedransk (1989), Cheng (1994), Chu and Cheng (1995) and Aerts et al. (2002) used kernel estimators, and developed kernel-based sampling weights to create imputations. In theory, the kernel-based MI may be more robust than parametric alternatives against the misspecification of the outcome model. However in practice, these kernel-based MI might become more cumbersome to implement as the number of predictors in X increases, due to the curse of dimensionality. In addition, model misspecification can still occur, for example, if important predictors are omitted from X (Fan and Gijbels, 1996). Then the resultant kernel-based MI inference can still be invalid.

Other methods have been developed to handle the misspecification of the outcome model. The popular doubly robust methods (van Buuren, 2012; Cassel et al., 1976; Horvitz and Thompson, 1952; Robins et al., 1994; Robins et al., 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999; Wang et al., 2004; Qin et al., 2008) construct two working models, the outcome model and the model predicting the probability of missingness (propensity model). Missing-data inferences based on the information from two models can still be valid as long as one of

them is correctly specified. To weaken the parametric assumption of the outcome model for the doubly robust methods, a semi-parametric doubly robust method (Hu et al., 2010) has been developed in which nonparametric regression is proposed to estimate the outcome model. See Section 2.3 for more details.

In this paper, we combine the strength of the above two strategies and develop a kernel-based doubly robust MI approach. The remainder of the paper is organized as follows. Section 2 introduces notation and briefly reviews the ideas behind the kernel-based MI and doubly robust approaches. Section 3 describes the proposed method. Section 4 assesses its performance through a simulation study. Section 5 presents a real data example. Finally, Section 6 concludes with a discussion and points out directions for future studies.

2. Background

2.1 Setup

For simplicity, we consider one incomplete continuous outcome Y, although the proposed method can be extended to categorical variables and multivariate settings (see Section 6 for a discussion). Denote the response indicator as δ, where $\delta = 0$ (1) if Y is missing (observed). We also assume that there exist K fully observed variables $X=(X_1, X_2,\ldots,X_K)$ that are predictive of Y and/or δ, and Y is independent of δ given X, that is, missing at random (Rubin, 1987). Finally, we assume that there are n independent selections of Y, δ, and X, of those n_{obs} have observed outcome data Y. We focus on the estimation of $\mu = E(Y)$, the marginal mean of the outcome.

2.2 Kernel-based MI

The essential ingredient of this approach is the local generation/imputation of missing observations, where "local" refers to the region of the observed covariates X that is close to the cases with a missing outcome. Without loss of generality, we assume that Y_i (i=1,..,n_{obs}) is observed and Y_j (j=n_{obs}+1,…,n) is missing. In its general form, kernel-based MI imputes a missing Y_j value by drawing from the cumulative distribution function

$\sum_{i=1}^{n_{obs}} w_{ij}(X_j, X_i)I(Y_i \leq Y)$, where {$w_{ji}$} denote positive weights with $\sum_{i=1}^{n_{obs}} w_{ij}=1$ and I(.) is the indicator function satisfying $I(Y_i \quad Y) = 1$ if $Y_i \quad Y$, and 0 otherwise. That is, we replace Y_j by a random draw from observed cases {Y_i} with probabilities {w_{ji}} for j=n_{obs}+1,…,n.

More specifically, if X_j is the covariate value associated with the missing Y_j, the weights {w_{ji}} are constructed using the kernel estimator:

$$w_{ji}(X_i, X_j)=\frac{K_h(X_j-X_i)}{\sum_{k=1}^{n_{obs}} K_h(X_j-X_k)},$$

where the kernel K(.) is a symmetric unimodal probability density, and $K_h(u)=K(u/h)/h$. The use of kernel weights implies that the observed Y_i value, whose X_i value is closer to X_j, gets

a larger weight and hence has a larger probability being chosen to replace the missing Y value.

The kernel-based MI aims to put less reliance on parametric assumptions of the outcome model. It has been shown to produce consistent estimates for the marginal mean in rather general scenarios (yet with only one covariate in the outcome model) [8] due to the nice property of kernel estimators. However, its practical implementation has been largely hindered by the curse of dimensionality when X has multiple variables, which is often encountered in practice. On the other hand, if X is limited to a single variable by removing other potentially important predictors, the outcome model is generally misspecified and would lead to invalid inferences (van Buuren, 2012).

2.3 Doubly robust methods

The doubly robust methods have been developed as a general technique to handle a potential misspecification of the outcome model. The earliest version can be traced back to the calibration estimator (Cassel et al., 1976), which extended the classic inverse probability weighting method to survey estimation (Horvitz and Thompson, 1952). The idea behind the calibration estimator is to combine information from two working regression models: one predicts the missing values in Y (the outcome model), and the other predicts the response probabilities $\Pr(\delta=1)$. As a result, the estimator of μ is a sum of prediction and inverse probability-weighted prediction errors, i.e.

$$\hat{u} = n^{-1} \sum_{i=1}^{n} \hat{Y}_i(X_i) + n^{-1} \sum_{i=1}^{n} \delta_i \frac{\hat{\varepsilon}_i}{\hat{\pi}_i(X_i)},$$

where $\hat{Y}(X)$ is the prediction of Y based on a regression model for $E(Y|X)$ fitted using only the observed cases, $\hat{e} = Y - \hat{Y}$, and $\hat{\pi}$ is the estimated probability of being observed ($\Pr(\delta=1|X)$). The first term of the estimator is equivalent to imputing all values using the outcome model, and the second term is a sum of inverse probability-weighted prediction errors from the outcome model. If the outcome model is correctly specified, then the first term is consistent for μ, and the second term converges to 0, guaranteeing that the estimator converges to μ. If the propensity model is correctly specified, then it can be shown that the second term consistently removes any bias that is associated with the first term, and hence the estimator still converges to μ. As a result, the estimator is consistent if either of the two models is correctly specified.

Most of the current doubly robust methods (Robins et al., 1994; Robins et al., 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999; Wang et al., 2004; Qin et al., 2008; Hu et al., 2010) originate from the calibration estimator. In particular, Hu et al. (2010) proposed to first reduce the dimension of X by a parametric working index S(X) which summarizes $E(Y|X)$ and then estimate $E(Y)$ using S(X) via kernel regression with a bandwidth h, in which the inverse probability of the selection probability is incorporated into the estimation as a weight to account for missingness. This inverse probability weighted kernel regression approach can easily handle a situation with high dimensional X and weakens the parametric assumption of

the outcome model for the calibration estimator. It can be considered as a semi-parametric calibration estimator since the working outcome model is not fully specified.

Following the same idea, doubly-robust MI procedures have been developed (example Zhang and Little, 2009; Long et al., 2012; Hsu et al., 2014). These methods consider both the outcome and propensity models yet avoid directly performing the inverse weighting step, which is sometimes problematic in finite samples (Kang and Schafer, 2007). For example, Long et al. (2012) proposed a nearest-neighborhood MI. It develops a distance measure between the incomplete and observed cases, using a weighted sum of predictive scores from the two models. An imputing set/nearest neighborhood (i.e., observed cases with the closest distance) is selected for each missing observation, and all of the observations in the imputing set have equal chances to be drawn to replace the missing observation. This strategy can be viewed as an extension of the predictive mean-matching MI (Schenker and Taylor, 1996; Siddique and Belin, 2008), where the mean metric is derived only from the outcome model.

3. Proposed method

3.1 Motivation

Although the kernel-based MI methods in theory can be extended to handle multiple covariates, the curse of dimensionality often prohibits the practical development of effective kernel weights. We propose to reduce the covariate space into two dimensions using two predictive scores: one from the outcome model and the other from the propensity model. For simplicity, we assume that the outcome model takes a linear regression form $E(Y)=X\beta$ and the propensity model takes a logistic regression form $\text{Logit}[\Pr(\delta=1)]=X\alpha$. We choose the linear and logistic regressions in fitting the working models because they are widely used in practice. Other types of regression, however, can also be considered (Section 6). The corresponding predictive scores are $X\beta$ and $X\alpha$, respectively. A two-dimensional kernel weight function based on the two scores is derived to determine the re-sampling weights used in imputation (Section 3.2).

Motivated by the idea of doubly robust methods, here we use two models instead of only the outcome model to guard against the potential model misspecification from applying the kernel to the predictive scores rather than the original covariates. In addition, the use of unequal kernel weights might be advantageous compared to nearest neighbor-based MI, in which each donor (i.e., the observed cases in the imputing set) has an equal probability of being selected. Therefore, the use of kernel-based techniques might be more amenable to the underlying distribution of the data and link function of the missingness probabilities compared to existing doubly robust approaches.

3.2 Algorithm

We describe the imputation algorithm in the following steps; the detailed computing program is available upon request.

Step 1: Estimate the predictive scores on a bootstrap sample—A nonparametric bootstrap sample (Efron, 1979) of the original dataset, which includes both the observed and missing values, is obtained to incorporate the uncertainty of parameter estimates from the

working models. This step results in an approximately proper MI (Section 10.2 of Little and Rubin (2002)). This strategy is justified by the fact that the bootstrap distribution represents an (approximate) nonparametric, noninformative prior-based posterior distribution of model parameters (Section 8.4 of Hastie et al. (2008)). More specifically, let $(Y^{(B)}, X^{(B)}, \delta^{(B)})$ denote the bootstrap sample. The observed cases of the bootstrap sample are used to fit the outcome model to estimate the regression coefficients β, and the full data of the bootstrap sample are used to fit the propensity model to estimate the regression coefficients α. Two predictive scores, $S_Y^{(B)} = X^{(B)}\hat{\beta}$ and $S_\delta^{(B)} = X^{(B)}\hat{\alpha}$, are then calculated for each case, missing or observed, in the bootstrap sample. We further standardize these scores by subtracting their corresponding sample mean (i.e. $\overline{S}_Y^{(B)}$ or $\overline{S}_\delta^{(B)}$) and dividing by their corresponding sample standard deviation (i.e. $\sigma_{S_Y^{(B)}}$ or $\sigma_{S_\delta^{(B)}}$), respectively (denoted as $S_Y^{c(B)}$ and $S_\delta^{c(B)}$). Such a standardization step aims to weigh the contribution from each predictor relatively equally.

Step 2: Calculate the kernel weights—For a missing Y_j with fully observed X_j in the original dataset ($j=n_{obs}+1,\ldots,n$) and two standardized predictive scores $S_Y^c(j)$ and $S_\delta^c(j)$, where $S_Y^c(j) = [X_j\hat{\beta} - \overline{S}_Y^{(B)}]/\sigma_{S_Y^{(B)}}$ and $S_\delta^c(j) = [X_j\hat{\alpha} - \overline{S}_\delta^{(B)}]/\sigma_{S_\delta^{(B)}}$, the kernel weights for the $n_{obs}^{(B)}$ observed cases in the bootstrap sample are defined as

$$w_{h_1,h_2,j,i}[S_Y^c(j) - S_Y^{c(B)}(i), S_\delta^c(j) - S_\delta^{c(B)}(i)] = \frac{K_{h_1,h_2}[S_Y^c(j) - S_Y^{c(B)}(i), S_\delta^c(j) - S_\delta^{c(B)}(i)]}{\displaystyle\sum_{k=1}^{n_{obs}^{(B)}} K_{h_1,h_2}[S_Y^c(j) - S_Y^{c(B)}(k), S_\delta^c(j) - S_\delta^{c(B)}(k)]},$$

for $i=1,..,n_{obs}^{(B)}$ where h_1 (h_2) is the bandwidth parameter for S_Y^c (S_δ^c). Although K_{h_1,h_2} is in general defined as a two-dimensional kernel, for the simplicity of implementation we consider a product of two independent, one-dimensional kernels, i.e.

$K_{h_1,h_2}[S_Y^c(j) - S_Y^{c(B)}(i), S_\delta^c(j) - S_\delta^{c(B)}(i)] = K_{h_1}[S_Y^c(j) - S_Y^{c(B)}(i)] * K_{h_2}[S_\delta^c(j) - S_\delta^{c(B)}(i)]$. This strategy ignores the potential correlation of predictive scores from the two working models. The chosen independent kernel performs adequately based on the results from our simulation study (Section 4). We use the probability density function of the standard normal for each K, i.e., $K_{h_1}[S_Y^c(j) - S_Y^{c(B)}(i)] = \exp\{-1/2 * ([S_Y^c(j) - S_Y^{c(B)}(i)]/h_1)^2\}/(\sqrt{2\pi}h_1$

Step 3: Imputation step—A random draw from the set of observed cases in the bootstrap sample with the probability $w_{h_1,h_2,j,i}$ is used to replace the missing Y_j.

Step 4: Repeat Steps 1 to 3 independently M times to create M completed datasets—Once the M multiply imputed datasets are obtained, we carry out the MI analysis procedure established in (Rubin, 1987). Specifically, we calculate the complete-data estimate $\hat{u}_m$ and its sample variance W_m ($m=1,\ldots,M$) for each of the M datasets. Then the

MI point estimator for μ is $\hat{u}_{MI} = \sum_{m=1}^{M} \hat{u}_m / M$. The (average) within-imputation variance is

$W = \sum_{m=1}^{M} W_m / M$. We also calculate the between-imputation variance as

$B = \sum_{m=1}^{M} (\hat{u}_m - \hat{u}_{MI})^2 / (M-1)$. The MI variance estimator for $\hat{u}_{MI}$ is $U = W + (1 + \frac{1}{M})B$. For hypothesis testing, the MI estimator approximately follows a *t* distribution with degrees of freedom $v = (M-1)*[1 + \{U*M/(M+1)\}/B]^2$.

3.3 Additional remarks

In the scenario with only one predictor X for the missing outcome Y, Aerts et al. (2002) shows that the mean estimator derived from the kernel-based multiple imputation approach, in which the kernel weights (i.e. resampling weights) are derived using X, is consistent under missing at random (MAR) mechanism (i.e. missingness only depends on X). Here we consider a more general scenario with multiple predictors in X. We have proposed to first summarize the multiple predictors into two predictive scores and then use the two predictive scores to derive the kernel weights. In a sense, we generalize the one-dimensional kernel-based multiple imputation in Aerts et al. (2002) to a two-dimensional kernel-based multiple imputation, which uses information from both the outcome and propensity models. We have previously showed that under the MAR mechanism based on the full set of covariates, if one of the two working models is correctly specified, the MAR mechanism can be induced based on the two predictive scores (Long et al., 2012). Using this result, Long et al. (2012) showed that under certain regularity conditions and MAR, if one of the two working models is correctly specified, the final mean estimator derived from the two-dimensional nearest neighbor-based multiple imputation approach is consistent. It has been shown that under certain relationship between the bandwidth and size of the nearest neighborhood, the bias and variance of the mean estimator derived from kernel smoothers are equivalent to those for nearest neighborhood smoothers (Mack, 1981). This suggests that building on the consistency properties of the one-dimensional kernel-based multiple imputation in Aerts et al. (2002) and the two-dimensional nearest neighbor-based multiple imputation in Long et al. (2012), under the regularity conditions similar to those in Long et al. (2012) and MAR, the proposed kernel-based multiple imputation has is expected to have a double robustness property. That is, if one of the two working models is correctly specified, the final mean estimator derived from the proposed two-dimensional kernel-based multiple imputation approach is expected to be consistent.

Selecting the bandwidth parameter in the kernel-based methods has always been a complicated issue. For the kernel-based MI, Aerts et al. (2002) showed that the estimation bias depends on the selected bandwidth and derived the optimal bandwidth by minimizing the mean-squared error of the estimator. Our models are arguably more complicated, and the purpose of our research is to develop a practically appealing imputation procedure. Therefore, research on choosing the optimal bandwidth parameter in our context is beyond the scope of the current work and requires further research. However, from our experience,

we suggest that one use a small h_1 for S_Y^c compared to h_2 for S_δ^c when we have more confidence on the outcome model than the propensity model, and vice versa.

The nearest neighbor-based MI (Long et al., 2012) can be considered as a special case of the proposed approach by fixing the bandwidth parameter and using a uniform kernel, i.e.

$$K_{h_1,h_2}[S_Y^c(j)-S_Y^{c(B)}(i), S_\delta^c(j)-S_\delta^{c(B)}(i)]=I(d_{j,i} \le h)/h,$$

where $d_{j,i}=w_Y[S_Y^c(j)-S_Y^{c(B)}]^2+w_\delta[S_\delta^c(j)-S_\delta^{c(B)}]^2$ measures the distance (on the scale of predictive scores) between the incomplete and observed cases, and w_Y and w_δ are positive weights attributed to the outcome and propensity models, respectively, satisfying $w_Y+w_\delta=1$ (Long et al., 2012). Practically, w_Y and w_δ can be chosen to reflect the imputer's relative confidence in the correctness of outcome and propensity models specified, using evidence from both the empirical data and substance knowledge. On the other hand, note that using the uniform kernel assumes equal chances of selecting donors. Therefore, using unequal selection weights in the method proposed in the current paper might yield a smaller bias than the nearest neighbor-based MI since in the former approach, the observations with more "similar" predictive scores to those from the missing Y have a greater chance to be selected for the imputations.

4. Simulation

We perform several simulation assessments to investigate the finite-sample properties of the proposed method, denoted as LMI(h_1,h_2) ("L" stands for "Local"), where h_1 (h_2) is the bandwidth for S_Y^c (S_δ^c). We assess how its performance is affected by the total sample size n, misspecification of the working models, and the bandwidths chosen for deriving the kernel weights. In addition, it is compared with several alternatives. The non-imputation methods include the complete-case analysis (CC), the calibration estimator (CE) and the semi-parametric calibration estimator (SCE(h), where h is the bandwidth) introduced in Section 2.2, and the fully-observed (FO) analysis, which analyzes data before deletion and is treated as the gold standard. We also consider the nearest neighborhood MI, denoted as NNMI(NN, w_Y, w_δ), where NN is the size of the nearest neighborhood (i.e., the number of donors), and w_Y (w_δ) is the weight on S_Y^c (S_δ^c) (Section 3.3). Based on our previous studies (Long et al., 2012; Hsu et al., 2014), for NNMI we select NN=5, w_Y=0.8 and w_δ=0.2, which appears to be a generally preferable choice across tested scenarios. For all MI methods, we conduct 5 independent imputations.

4.1 Data Generation

The design is similar to the simulation study conducted in Long et al. (2012). For each of the 500 independent simulated datasets, there are 5 hypothetical predictive covariates $X_1,\ldots,X_5$ independently generated from a Uniform $(-1, 1)$ distribution. The outcome Y is generated from either N($10 + 2X_1 - 2X_2 + 3X_3 - 3X_4 + 1.5X_5$, 3) or lognormal($0.5+0.5X_1 - X_2 + 1.5X_3 - 2X_4 + 0.5X_5$,1). The response indicator δ is generated from logit ($Pr(\delta=1)$) =

$\alpha_0+0.5X_1 - X_2 + X_3 - X_4 + X_5$ or $\log[-\log(1-Pr(\delta=1))]= \alpha_0+0.5X_1 - X_2 + X_3 - X_4 + X_5$. α_0 is selected to allow the missingness rate to vary between 20% and 50%. We consider total sample sizes 200 and 800.

4.2 Specifications of the working models

We consider two types of model misspecification: inclusion of a wrong set of the covariates and misspecification of the link function for the outcome or propensity model. The working models are correctly specified if all five covariates are included and the link (distribution) functions are correctly specified; otherwise they are misspecified. For example, when Y is generated from the log-normal distribution, the working linear regression model for Y is misspecified even if it includes all five covariates. When δ is generated from a complementary log-log (cloglog) link function, the working logistic regression model for δ is misspecified even if it includes all five covariates.

We consider three scenarios for the two working regression models. In all cases, the working model for Y is a linear regression, and the working model for δ is a logistic regression. We use subscripts to indicate the number of predictors included in the working models.

> *Scenario 1:* Both models include all five predictors (CE$_{55}$, SCE$_{55}$, NNMI$_{55}$ and LMI$_{55}$). The subscript "55" indicates that all 5 predictors are used in both models.

> *Scenario 2:* The working model for Y includes all five covariates and the working model for δ only includes the first three covariates (CE$_{53}$, SCE$_{53}$, NNMI$_{53}$ and LMI$_{53}$). The subscript "53" indicates that all 5 predictors are used in the outcome model, yet only the first 3 covariates are included in the propensity model.

> *Scenario 3:* the working model for Y includes the first three covariates and the working model for δ includes all five covariates: (CE$_{35}$, SCE$_{35}$, NNMI$_{35}$ and LMI$_{35}$).

We summarize various model misspecifications in the simulation using LMI as an example in Table 1. Misspecifications of the models can happen on the distributional forms (normal vs. lognormal), link functions (logit vs. cloglog), and inclusion of predictors (3 vs. 5). Similar patterns can be found for CE, SCE and NNMI. Note that doubly-robust methods are expected to perform well if at least one of the two working models is correctly specified, yet here we consider additional situations in which both working models are misspecified and assess the performance of the proposed method in those situations.

4.3 Results

Tables 2–7 include the simulation results. We highlight the three doubly robust methods with the smallest (including ties) root mean squared error (RMSE) measures. In all scenarios, CC produces biased estimates (bias increases with missing rate) and results in a low coverage rate as expected. In the situation where Y is generated from a normal distribution and δ is generated based on a logit link function (Tables 2 and 3), CE produces estimates almost identical to FO as long as one of the working models is correctly specified (i.e. including all

the five predictive covariates in the model). SCE produces estimates with smaller biases than LMI and NNMI when the working model for Y is correctly specified. When the working model for Y is misspecified, in some situations SCE produces estimates with larger biases than LMI. Both LMI and NNMI produce estimates with larger biases than CE produces. The bias is much smaller when both working models are correctly specified, i.e. LMI_{55} and $NNMI_{55}$. For NNMI, when the working model for Y is misspecified, i.e. $NNMI_{35}$, the bias is larger than when the working model for Y is correctly specified, i.e. $NNMI_{53}$. In contrast, LMI_{53} and LMI_{35} have similar biases. This indicates that LMI is less affected by the misspecification of the working model for Y than NNMI. For LMI, when the bandwidths are small, especially for the bandwidth of the predictive score derived from the working model for Y, i.e. h_1, LMI produces estimates with much smaller biases than NNMI even if the working model for Y is misspecified. For example, in Table 2, $LMI_{35}(0.10,0.10)$ has a percent bias of 0.51% compared to a percentage bias of 1.19% for $NNMI_{35}(5,0.8,0.2)$. The bias for LMI increases with h_1 and h_2, especially with h_1. For example, in Table 2, $LMI_{55}(0.10,0.10)$ has a percentage bias of 0.15%, $LMI_{55}(0.10,0.25)$ has a percentage bias of 0.24%, and $LMI_{55}(0.25,0.10)$ has a percentage bias of 0.24%. As the total sample size increases to 800 (Table 3), the bias for both LMI and NNMI decreases. For example, the percentage bias for $NNMI_{35}(5, 0.8, 0.2)$ decreases from 1.19% to 0.41%, and the percentage bias for $LMI_{35}(0.10,0.10)$ decreases from 0.51% to 0.18%. This indicates that if one of the working models is correctly specified, both LMI and NNMI can produce reasonable estimates. Based on the simulation results not provided here, we also observe the bias-variance trade-off when the bandwidth parameters in LMI increase. That is, with smaller h_1 or h_2, LMI tends to have smaller bias yet larger variance (or SD) than those with larger h_1 or h_2. Such pattern is consistent with the property of kernel regression-based estimates. We also note that the coverage rates for LMI can be somewhat off from the nominal level, i.e. 95%. With large bandwidth parameters, the under-coverage can be largely caused by the bias. With small bandwidth parameters, in which the bias effect is small, this might be caused by the underestimation of standard error (SE) compared to the empirical standard error, i.e. SD. The under-coverage seems to be less severe with increased sample size.

In the situation where Y is generated from a normal distribution and δ is generated from a cloglog link function (Tables 4 and 5), the working model for δ (a logit link) is always misspecified. When the working model for Y is correctly specified (e.g., LMI_{55} or LMI_{53}), the proposed methods yield in general satisfactory results (e.g., reasonable bias and coverage rates), with small h_1 and h_2. If both working models are misspecified, the double robustness property is not expected to hold generally for all the considered methods. Note that CE_{35} produces an estimate with large bias (larger than LMI especially when both bandwidths are small), and the bias increases with sample size. SCE_{35}, LMI_{35} and $NNMI_{35}$ also produce estimates with large bias, but the bias decreases with sample size. For example, when sample size is 200, CE_{35} has a percentage bias of -1.85%, $SCE_{35}(0.10)$ has a percentage bias of 2.39%, $LMI_{35}(0.10,0.10)$ has a percentage bias of 1.36%, and $NNMI_{35}(5,0.8,0.2)$ has a percentage bias of 2.99%. When sample size increases to 800, CE_{35} has a percentage bias of -2.24%, $SCE_{35}(0.10)$ has a percentage bias of 0.17%, $LMI_{35}(0.10,0.10)$ has a percentage bias of 0.53%, and $NNMI_{35}(5,0.8,0.2)$ has a percentage bias of 1.31%. The overall performance of LMI with small h_1 and h_2 are the most satisfactory when both working

models are mis-specified. These results indicate that LMI might be more robust to the misspecification of both working models than CE or NNMI.

In the situation where Y is generated from a log-normal distribution and δ is generated from a logit link function (Tables 6 and 7), the working model for Y (a normal model) is always mis-specified. When the working model for δ is correctly specified (e.g., LMI_{55} or LMI_{35}), LMI methods yield in general satisfactory results (e.g., reasonable bias and coverage rates), with small h_1 and h_2. When both working models are mis-specified, the bias for CE_{53} is again much larger than that for SCE_{53}, LMI_{53} and $NNMI_{53}$. Specifically, in Table 6, CE_{53} has a percentage bias of -29.82%, $SCE_{53}(0.10)$ has a percentage bias of 3.53%, $LMI_{53}(0.10,0.10)$ has a percentage bias of 0.26%, and $NNMI_{53}(5,0.8,0.2)$ has a percentage of 1.05%. As the sample size increases to 800 (Table 7), the bias for CE_{53} does not decrease. The relative performance of LMI vs. CE or NNMI is similar to that shown in Tables 4 and 5, again indicating that LMI might be more robust against the misspecification of both working models than CE or NNMI.

In summary, CC tends to produce biased estimates as expected. The doubly robust methods including CE, SCE, NNMI, and LMI can produce reasonable estimates if one of the two working regression models is correctly specified. Among the doubly robust methods, LMI with small bandwidth parameters in general performs either the same or better than CE, SCE or NNMI, with respect to RMSE. We also find that LMI with small bandwidth parameters seems to be more robust when both working models are mis-specified. Consistent with the existing literature, choosing small bandwidth parameters tends to yield smaller bias for LMI. Therefore, through the tuning of bandwidths, LMI can produce estimates with smaller biases than NNMI since LMI assigns larger re-sampling weights to the observations more "similar" to the missing observations.

On the other hand, we note that in certain cases, the SE of LMI can be appreciably lower than SD, which might largely cause the under-coverage. We do not have a good explanation for this. Since the variance formulae for MI (Rubin, 1987) assumes that the imputation model is correctly specified, yet here the proposed methods are based on predictive scores derived from working models which can be mis-specified, we surmise that the direct use of MI variance formulae might lead to sizable bias in certain scenarios. Such bias (for the variance estimation) is hard to quantify but it might be related to the extent to which the working model is mis-specified and to the sample size. For the latter, if we compare the results from N=200 with those from N=800 (e.g. Tables 2 vs. 3), we can see the variance underestimation is less severe with an increased sample size. Further investigation of this issue is beyond the scope of this paper, but some of the related references on the bias of MI variance formulae can be found in Kim and Shao (2013). A practical solution to this potential problem is to conduct a sensitivity analysis including alternative methods, paying specific attention to confidence intervals that appear to be too narrow or too wide.

5. Application

We apply various missing data methods to the 2012 Arizona emergency medical service data. The example focuses on estimating the mean total response time from injury

(transported by helicopters) by radius distance from the regional level I hospitals. After excluding a few cases with an extremely long response time, say >1000 minutes (those observations are more likely due to some technical recording errors), there are 1160 patients left, among whom around 66% of the cases lack the total response time from injury.

We run some exploratory analyses to determine the set of predictive covariates for both the outcome and propensity models. Based on a logistic regression analysis, younger, white patients, and patients with lower ICD9 severity score are more likely to have a missing response time, and patients further away from the regional hospitals are less likely to have a missing response time. Based on a linear regression using only the observed cases, the distance is positively correlated with the total response time. Assuming the missingness is at random, the five variables (i.e., age, gender, distance, ICD9 severity score and race [white vs non-white]) are included as covariates in the two working regression models.

Similar to our simulation study, we consider three scenarios for the two working regression models:

> *Scenario 1:* both models include all five predictive covariates (CE_{55}, SCE_{55}, $NNMI_{55}$ and LMI_{55}).

> *Scenario 2:* the working model for total response time includes all five covariates, and the working model for missingness probabilities only includes the first three covariates, i.e. age, male and distance (CE_{53}, SCE_{53}, $NNMI_{53}$ and LMI_{53}).

> *Scenario 3:* the working model for total response time includes the first three covariates and the working model for δ includes all five covariates (CE_{35}, SCE_{35}, $NNMI_{35}$ and LMI_{35}).

The results are displayed in Table 8. For simplicity, we only show the LMI results when both h_1 and h_2 are 0.1, as those with other choices of bandwidth parameters remain largely similar. The results from CC, CE and $NNMI_{35}$ show that the total response time from injury does not increase with radius distance until the distance is greater than 50 miles. The results from SCE, $NNMI_{55}$ and $NNMI_{53}$ indicate that the total response time does not increase with radius distance until the distance is greater than 40 miles. The results from all three LMI indicate that the total response time from injury consistently increases with a longer distance. However, the rate of increase is slower when the distance is shorter than 40 miles. The results from LMI are consistent with our intuition. In addition, the observed total response time from injury is highly right skewed with a skewness of 3.10. The Shapiro-Wilk test of normality for the observed total response time data produces a p-value <0.0001 (results not shown). Based on this fact and the simulation results when Y is generated from a log-normal (right skewed) distribution (Section 4), the application results obtained from LMI might be more reliable.

6. Discussion

This paper describes a MI method that combines the strength from the doubly robust methods and kernel-smoothing methods. It first uses fully observed predictors to build the

outcome and propensity model to obtain two summarizing predictive scores, and then constructs kernel weights on the difference of the scores from observed and incomplete cases, and finally resamples the observed cases using these kernel weights for imputation. Its reliance on the correct specification of the working models is weak, because only the predictive scores, which effectively reduce the dimension of the covariates, are used in the imputation. As demonstrated by our simulation results, the problem of the misspecification of the working models can be effectively alleviated by this double robustness setup. In addition, the role of predictive scores is to construct a distance measure between the incomplete and observed cases. This is in contrast with most of the methods in the literature that directly use the information from the predictive covariates for estimation and that, therefore, result in performance highly dependent on the correctness of the model specification. In addition, the proposed method appears to be more robust to the violations of distributional assumptions than the existing doubly robust methods such as the calibration estimator and nearest-neighborhood MI method.

Our research can be extended in several directions. First, since kernel smoothing is used in the imputation, the performance of the imputation procedure is affected by the selection of bandwidths. When the bandwidths are too wide, more observations will have non-zero sampling weights. Therefore, some observations that are not very "similar" to the observation with missing outcome might be drawn to replace the latter. This could result in an increased bias in estimation. In this paper, we select a range of bandwidths and compare their performances. This can be easily implemented by practitioners as a strategy for data analysis. Specifically, one needs to first select a range of bandwidths to select from. The two predictive scores are standardized and expected to range from −3 to +3, respectively. Therefore, a small range (e.g. 0.10~0.30) should be sufficient. For given bandwidths one can first derive the mean estimate and the associated standard error estimate. One can then specify a range of potential true mean values based on knowledge about the background of the data. A squared error can then be calculated for the mean estimate corresponding to each true mean, from which one can calculate an average squared error value across the assumed different true mean values. The bandwidth with the smallest averaged squared error can then be chosen as the optimal bandwidth. More rigorously, simulations can be performed to calculate the MSE (rather than the squared error) under each scenario and use the average MSE across scenarios to evaluate this strategy for selecting an optimal bandwidth. More sophisticated methods to effectively construct correlated bivariate kernels in the same spirit of that of Aerts et al. (2002) based on Jackknife or related methods are another future research area in terms of bandwidth selection.

Second, more options for the working models might be considered. In this paper, we simply use linear regression to predict the outcome with missing observations and logistic regression to predict the missingness probability. Potentially, when the outcome variable is not normally distributed, a transformation may be performed to better approximate a normal distribution for a continuous outcome, or a generalized linear model may be fitted for categorical outcomes. Similarly, binary regression models with link functions other than logit can also be explored for the propensity models.

Third, the proposed method might be generalized to impute missing data from multiple variables, which are often encountered in practice. For these problems, this strategy may be implemented for one missing outcome at a time and then sequentially over all of the incomplete variables, fitting in the framework of the sequential regression multiple imputation or multiple imputation by chained equations (Raghunathan et al., 2001; van Buuren et al., 1999).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aerts M, Claeskens G, Hens N, Molenberghs G. Local multiple imputation. Biometrika. 2002; 89:375–388.

Cassel CM, Sarndal CE, Wretman JH. Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika. 1976; 63:615–620.

Cheng PE. Nonparametric estimation of mean functionals with data missing at random. Journal of American Statistical Association. 1994; 89:81–87.

Chu CK, Cheng PE. Nonparametric regression estimation with missing data. Journal of Statistical planning and Inference. 1995; 48:85–99.

Efron B. Bootstrap methods: another look at the jackknife. Annals of Statistics. 1979; 7:1–26.

Fan, J.; Gijbels, I. Local Polynomial Modelling and Its Applications. Chapman and Hall; London: 1996.

Harel O, Zhou XH. Multiple imputation: review of theory, implementation, and Software. Statistics in Medicine. 2007; 26:3057–3077. [PubMed: 17256804]

Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. Springer; New York: 2008.

Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. 1952; 47:663–685.

Hsu CH, Long Q, Li Y, Jacobs E. A nonparametric multiple imputation approach for data with missing covariate values with application to colorectal adenoma data. Journal of Biopharmaceutical Statistics. 2014; 24:634–648. [PubMed: 24697618]

Hu Z, Follmann D, Qin J. Semiparametric dimension reduction estimation for mean response with missing data. Biometrika. 2010; 97:305–319. [PubMed: 23049121]

Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science. 2007; 22:523–539.

Kim, JK.; Shao, J. Statistical Methods for Handling Incomplete Data. CRC Press; Boca Faton, FL: 2013.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. Wiley; New York: 2002.

Long Q, Hsu CH, Li Y. Doubly robust nonparametric multiple imputation for ignorable missing data. Statistica Sinica. 2012; 22:149–172. [PubMed: 22347786]

Mack YP. Local properties of k-nn regression estimates. SIAM, J Alg Disc Meth. 1981; 2:311–23.

Qin J, Shao J, Zhang B. Effcient and doubly robust imputation for covariate-dependent missing responses. Journal of the American Statistical Association. 2008; 103:797–809.

Raghunathan TE, Lepkowski JM, van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology. 2001; 27:85–95.

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association. 1994; 89:846–86.

Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995; 90:106–121.

Rotnitzky A, Robins J, Scharfstein D. Semiparametric regression for repeated outcomes with nonignorable nonresponse. Journal of the American Statistical Association. 1998; 93:1321–1339.

Rubin, DB. Multiple Imputation for Nonresponse in Surveys. Wiley; New York: 1987.

Scharfstein D, Rotnitzky A, Robins J. Adjusting for nonignorable dropout using semiparametric nonresponse models. Journal of the American Statistical Association. 1999; 94:1096–1120.

Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. Computational Statistics and Data Analysis. 1996; 22:425–446.

Siddique J, Belin TR. Multiple imputation using an iterative hot-deck with distance-based donor selection. Statistics in Medicine. 2008; 27:83–102. [PubMed: 17634973]

Titterington DM, Sedransk J. Imputation of missing values using density estimation. Statistics & Probability Letter. 1989; 8:411–418.

van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine. 1999; 18:681–694. [PubMed: 10204197]

van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research. 2007; 16:219–42. [PubMed: 17621469]

van Buuren, S. Flexible Imputation of Missing Data. CRC Press; Boca Raton: Florida: 2012.

Wang Q, Linton O, Härdle W. Semiparametric regression analysis with missing response at random. Journal of the American Statistical Association. 2004; 99:334–345.

Zhang G, Little RJA. Extensions of the Penalized Spline of Propensity Prediction Methodof Imputation. Biometrics. 2009; 65:911–918. [PubMed: 19053998]

Table 1

Model Misspecfications for LMI

Data Generation		LMI_{55}		LMI_{53}		LMI_{35}	
Distribution for Y	Link function for $Pr(\delta=1)$	Outcome model	Propensity model	Outcome model	Propensity model	Outcome model	Propensity model
Normal	Logit	Yes[a]	Yes	Yes	No	No	Yes
Normal	Cloglog	Yes	No	Yes	No	No	No
Lognormal	Logit	No	Yes	No	No	No	Yes

[a]Yes/No indicates correctly specified/misspecified.

Table 2

Monte Carlo Simulation Results for Data Missing at Random: Y~Normal Distribution and logit link for $Pr(\delta=1)$ with $\alpha_0=1.5$; N=200; Replicates=500; M=5; Missing Rate: 23%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	9.992	−0.08	0.304	0.304	0.303	95.2
CC	10.66	6.60	0.330	0.738	0.334	48.6
Correct working models for both Y and δ						
CE$_{55}$	9.991	−0.09	0.328	0.328	0.336	94.6
SCE$_{55}$(0.10)	10.009	0.09	0.333	0.333	0.351	95.2
SCE$_{55}$(0.25)	10.006	0.06	0.330	0.330	0.342	95.0
NNMI$_{55}$(5,0.8,0.2)	10.04	0.40	0.326	0.328	0.331	94.4
LMI$_{55}$(0.10,0.10)	10.015	0.15	0.337	0.337	0.339	93.4
LMI$_{55}$(0.10,0.25)	10.013	0.13	0.332	0.332	0.339	94.4
LMI$_{55}$(0.25,0.10)	10.024	0.24	0.333	0.334	0.339	94.0
LMI$_{55}$(0.25,0.25)	10.027	0.27	0.331	0.332	0.340	93.2
Wrong working model for δ (including a reduced set of predictors) only						
CE$_{53}$	9.993	−0.07	0.321	0.321	0.327	94.2
SCE$_{53}$(0.10)	10.009	0.09	0.335	0.335	0.343	95.4
SCE$_{53}$(0.25)	10.009	0.09	0.330	0.330	0.334	94.2
NNMI$_{53}$(5,0.8,0.2)	10.045	0.45	0.324	0.327	0.333	93.8
LMI$_{53}$(0.10,0.10)	10.03	0.30	0.322	0.323	0.336	94.8
LMI$_{53}$(0.10,0.25)	10.02	0.20	0.327	0.328	0.336	94.6
LMI$_{53}$(0.25,0.10)	10.063	0.63	0.326	0.332	0.334	94.0
LMI$_{53}$(0.25,0.25)	10.049	0.49	0.325	0.329	0.334	94.8
Wrong working model for Y (including a reduced set of predictors) only						
CE$_{35}$	9.994	−0.06	0.335	0.335	0.359	95.6
SCE$_{35}$(0.10)	10.084	0.84	0.326	0.337	0.362	95.0
SCE$_{35}$(0.25)	10.038	0.38	0.326	0.328	0.353	95.0

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
$NNMI_{35}(5,0.8,0.2)$	10.119	1.19	0.318	0.340	0.332	93.4
$LMI_{35}(0.10,0.10)$	10.051	0.51	0.333	0.337	0.341	94.0
$LMI_{35}(0.10,0.25)$	10.078	0.78	0.320	0.329	0.338	93.8
$LMI_{35}(0.25,0.10)$	10.04	0.40	0.338	0.340	0.344	93.8
$LMI_{35}(0.25,0.25)$	10.065	0.65	0.328	0.334	0.341	94.4

[a] average of the 500 point estimates

[b] 100*(estimate- true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 3

Monte Carlo Simulation Results for Data Missing at Random: Y~Normal Distribution and logit link for Pr(δ=1) with $α_0$=1.5; N=800; Replicates=500; M=5; Missing Rate: 23%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	9.992	−0.08	0.148	0.148	0.152	95.6
CC	10.661	6.61	0.165	0.681	0.167	1.4
Correct working models for both Y and δ						
CE_{55}	9.992	−0.08	0.162	0.162	0.167	95.0
SCE_{55}(0.10)	9.994	−0.06	0.168	0.168	0.170	94.8
SCE_{55}(0.25)	9.994	−0.06	0.165	0.165	0.168	94.8
$NNMI_{55}$(5,0.8,0.2)	10.004	0.04	0.168	0.168	0.169	94.4
LMI_{55}(0.10,0.10)	9.997	−0.03	0.167	0.167	0.172	95.4
LMI_{55}(0.10,0.25)	9.996	−0.04	0.166	0.166	0.172	94.6
LMI_{55}(0.25,0.10)	10.001	0.01	0.168	0.168	0.174	94.6
LMI_{55}(0.25,0.25)	10.01	0.10	0.165	0.165	0.170	95.8
Wrong working model for δ (including a reduced set of predictors) only						
CE_{53}	9.99	−0.10	0.161	0.161	0.164	95.0
SCE_{53}(0.10)	9.993	−0.07	0.167	0.167	0.167	94.0
SCE_{53}(0.25)	9.994	−0.06	0.165	0.165	0.165	94.8
$NNMI_{53}$(5,0.8,0.2)	10.012	0.12	0.168	0.168	0.169	93.8
LMI_{53}(0.10,0.10)	10.01	0.10	0.166	0.166	0.170	95.6
LMI_{53}(0.10,0.25)	10.004	0.04	0.167	0.167	0.171	95.4
LMI_{53}(0.25,0.10)	10.045	0.45	0.169	0.175	0.169	94.2
LMI_{53}(0.25,0.25)	10.036	0.36	0.166	0.170	0.168	94.6
Wrong working model for Y (including a reduced set of predictors) only						
CE_{35}	9.993	−0.07	0.163	0.163	0.178	96.2
SCE_{35}(0.10)	10.021	0.21	0.164	0.165	0.175	95.6
SCE_{35}(0.25)	10.005	0.05	0.163	0.163	0.173	95.6

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
NNMI$_{35}$(5,0.8,0.2)	10.041	0.41	0.165	0.170	0.169	94.8
LMI$_{35}$(0.10,0.10)	10.018	0.18	0.167	0.168	0.171	95.4
LMI$_{35}$(0.10,0.25)	10.043	0.43	0.163	0.169	0.170	94.6
LMI$_{35}$(0.25,0.10)	10.008	0.08	0.170	0.170	0.174	95.4
LMI$_{35}$(0.25,0.25)	10.037	0.37	0.165	0.169	0.171	95.6

[a] average of the 500 point estimates

[b] 100*(estimate- true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 4

Monte Carlo Simulation Results for Data Missing at Random: $Y \sim$ Normal Distribution and complementary log-log link for $\Pr(\delta=1)$ with $\alpha_0=0$; N=200; Replicates=500; M=5; Missing Rate: 39%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	9.99	−0.1	0.304	0.304	0.303	95.4
CC	11.354	13.54	0.360	1.401	0.362	4.0
Wrong working models for δ (a misspecified link function) only						
CE_{55}	9.995	−0.05	0.506	0.506	0.481	96.4
$SCE_{55}(0.10)$	10.064	0.64	0.424	0.429	0.490	96.0
$SCE_{55}(0.25)$	10.056	0.56	0.416	0.420	0.483	97.0
$NNMI_{55}(5,0.8,0.2)$	10.12	1.2	0.395	0.413	0.394	94.0
$LMI_{55}(0.10,0.10)$	10.067	0.67	0.404	0.410	0.421	94.2
$LMI_{55}(0.10,0.25)$	10.071	0.71	0.409	0.415	0.409	93.8
$LMI_{55}(0.25,0.10)$	10.081	0.81	0.420	0.428	0.422	92.6
$LMI_{55}(0.25,0.25)$	10.086	0.86	0.405	0.414	0.403	92.2
Wrong working model for δ (including a reduced set of predictors and a misspecified link function) only						
CE_{53}	9.986	−0.14	0.367	0.367	0.370	94.6
$SCE_{53}(0.10)$	10.056	0.56	0.428	0.432	0.385	91.8
$SCE_{53}(0.25)$	10.054	0.54	0.417	0.420	0.375	93.0
$NNMI_{53}(5,0.8,0.2)$	10.161	1.61	0.393	0.425	0.389	92.6
$LMI_{53}(0.10,0.10)$	10.121	1.21	0.393	0.411	0.394	93.6
$LMI_{53}(0.10,0.25)$	10.094	0.94	0.400	0.411	0.398	93.4
$LMI_{53}(0.25,0.10)$	10.21	2.1	0.389	0.442	0.388	92.8
$LMI_{53}(0.25,0.25)$	10.155	1.55	0.389	0.419	0.391	91.4
Wrong working model for both Y (including a reduced set of predictors) and δ (a misspecified link function)						
CE_{35}	9.815	−1.85	0.701	0.725	0.634	95.4
$SCE_{35}(0.10)$	10.239	2.39	0.386	0.454	0.537	97.4
$SCE_{35}(0.25)$	10.099	0.99	0.410	0.422	0.531	97.8

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
$NNMI_{35}(5,0.8,0.2)$	10.299	2.99	0.385	0.487	0.380	86.6
$LMI_{35}(0.10,0.10)$	10.136	1.36	0.389	0.412	0.411	93.8
$LMI_{35}(0.10,0.25)$	10.221	2.21	0.386	0.445	0.390	90.2
$LMI_{35}(0.25,0.10)$	10.112	1.12	0.411	0.426	0.418	93.4
$LMI_{35}(0.25,0.25)$	10.171	1.71	0.394	0.430	0.399	92.4

[a] average of the 500 point estimates

[b] 100*(estimate- true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 5

Monte Carlo Simulation Results for Data Missing at Random: Y~Normal Distribution and complementary log-log link for $Pr(\delta=1)$ with $\alpha_0=0$; N=800; Replicates=500; M=5; Missing Rate: 39%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	9.993	−0.07	0.149	0.149	0.152	95.4
CC	11.357	13.57	0.177	1.368	0.181	0.0
Wrong working models for δ (a misspecified link function) only						
CE_{55}	9.994	−0.06	0.251	0.251	0.251	96.0
$SCE_{55}(0.10)$	10.012	0.12	0.211	0.211	0.228	97.2
$SCE_{55}(0.25)$	10.01	0.10	0.211	0.211	0.227	96.8
$NNMI_{55}(5,0.8,0.2)$	10.041	0.41	0.214	0.218	0.203	93.6
$LMI_{55}(0.10,0.10)$	10.018	0.18	0.219	0.220	0.209	93.2
$LMI_{55}(0.10,0.25)$	10.029	0.29	0.220	0.222	0.208	92.6
$LMI_{55}(0.25,0.10)$	10.029	0.29	0.226	0.228	0.212	93.0
$LMI_{55}(0.25,0.25)$	10.048	0.48	0.208	0.213	0.207	94.4
Wrong working model for δ (including a reduced set of predictors and a misspecified link function) only						
CE_{53}	9.991	−0.09	0.194	0.194	0.186	94.0
$SCE_{53}(0.10)$	10.008	0.08	0.208	0.208	0.187	93.6
$SCE_{53}(0.25)$	10.011	0.11	0.208	0.208	0.185	92.8
$NNMI_{53}(5,0.8,0.2)$	10.06	0.60	0.204	0.213	0.199	93.0
$LMI_{53}(0.10,0.10)$	10.048	0.48	0.209	0.214	0.201	94.0
$LMI_{53}(0.10,0.25)$	10.038	0.38	0.208	0.211	0.203	92.6
$LMI_{53}(0.25,0.10)$	10.126	1.26	0.209	0.244	0.200	89.6
$LMI_{53}(0.25,0.25)$	10.111	1.11	0.201	0.230	0.200	91.6
Wrong working model for both Y (including a reduced set of predictors) and δ (a misspecified link function)						
CE_{35}	9.776	−2.24	0.337	0.405	0.348	97.2
$SCE_{35}(0.10)$	10.017	0.17	0.228	0.229	0.249	96.4
$SCE_{35}(0.25)$	9.944	−0.56	0.239	0.245	0.248	95.2

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
$NNMI_{35}(5,0.8,0.2)$	10.131	1.31	0.208	0.246	0.199	90.2
$LMI_{35}(0.10,0.10)$	10.053	0.53	0.218	0.224	0.206	93.0
$LMI_{35}(0.10,0.25)$	10.129	1.29	0.212	0.248	0.201	88.0
$LMI_{35}(0.25,0.10)$	10.038	0.38	0.222	0.225	0.211	92.2
$LMI_{35}(0.25,0.25)$	10.111	1.11	0.208	0.236	0.203	90.6

[a] average of the 500 point estimates

[b] 100*(estimate- true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 6

Monte Carlo Simulation Results for Data Missing at Random: $Y \sim$ Log-Normal Distribution and logit link for $Pr(\delta=1)$ with $\alpha_0=0$; N=200; Replicates=500; M=5; Missing Rate: 50%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	8.858	−0.83	2.252	2.253	1.959	87.2
CC	13.832	54.86	4.115	6.399	3.578	84.6
Wrong working models for Y (a misspecified distribution) only						
CE_{55}	8.609	−3.62	3.005	3.022	2.936	87.8
$SCE_{55}(0.10)$	9.231	3.35	2.683	2.700	4.414	96.4
$SCE_{55}(0.25)$	9.028	1.08	2.552	2.554	4.380	96.4
$NNMI_{55}(5,0.8,0.2)$	8.873	−0.66	2.307	2.308	2.025	87.2
$LMI_{55}(0.10,0.10)$	8.819	−1.26	2.382	2.385	2.037	87.0
$LMI_{55}(0.10,0.25)$	8.861	−0.79	2.335	2.336	2.030	86.6
$LMI_{55}(0.25,0.10)$	8.774	−1.77	2.343	2.348	2.026	87.4
$LMI_{55}(0.25,0.25)$	8.834	−1.10	2.350	2.352	2.046	86.6
Wrong working model for Y (including a reduced set of predictors and a misspecified distribution) only						
CE_{53}	6.268	−29.82	2.173	3.438	2.689	74.2
$SCE_{53}(0.10)$	9.247	3.53	2.680	2.698	3.886	95.8
$SCE_{53}(0.25)$	9.065	1.49	2.556	2.559	3.848	95.6
$NNMI_{53}(5,0.8,0.2)$	9.026	1.05	2.378	2.380	2.044	88.2
$LMI_{53}(0.10,0.10)$	8.955	0.26	2.380	2.380	2.034	88.2
$LMI_{53}(0.10,0.25)$	8.958	0.29	2.381	2.381	2.033	88.0
$LMI_{53}(0.25,0.10)$	9.044	1.26	2.431	2.434	2.063	89.0
$LMI_{53}(0.25,0.25)$	9.007	0.84	2.406	2.407	2.050	88.6
Wrong working model for both Y (a misspecified distribution) and δ (including a reduced set of predictors)						
CE_{35}	8.825	−1.20	2.519	2.521	2.352	89.0
$SCE_{35}(0.10)$	10.252	14.78	3.049	3.323	4.707	97.4
$SCE_{35}(0.25)$	9.792	9.63	2.950	3.073	4.623	97.0

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
NNMI$_{35}$(5,0.8,0.2)	9.28	3.90	2.475	2.499	2.158	91.0
LMI$_{35}$(0.10,0.10)	9.012	0.90	2.410	2.411	2.129	88.6
LMI$_{35}$(0.10,0.25)	9.15	2.44	2.448	2.458	2.118	89.8
LMI$_{35}$(0.25,0.10)	8.956	0.27	2.353	2.353	2.098	88.4
LMI$_{35}$(0.25,0.25)	9.064	1.48	2.368	2.372	2.099	88.4

[a] average of the 500 point estimates

[b] 100*(estimate- true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 7

Monte Carlo Simulation Results for Data Missing at Random: Y~Log-Normal Distribution and logit link for $\Pr(\delta=1)$ with $\alpha_0=0$; N=800; Replicates=500; M=5; Missing Rate: 50%

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
FO	8.859	−0.82	1.098	1.100	1.037	90.0
CC	13.764	54.10	2.021	5.238	1.891	16.6
Wrong working models for Y (a misspecified distribution) only						
CE_{55}	8.801	−1.47	1.411	1.417	1.511	93.0
$SCE_{55}(0.10)$	8.895	−0.41	1.206	1.207	2.272	98.0
$SCE_{55}(0.25)$	8.839	−1.04	1.18	1.184	2.244	98.0
$NNMI_{55}(5,0.8,0.2)$	8.812	−1.34	1.147	1.153	1.076	89.8
$LMI_{55}(0.10,0.10)$	8.792	−1.57	1.15	1.158	1.081	90.0
$LMI_{55}(0.10,0.25)$	8.801	−1.47	1.156	1.163	1.075	88.6
$LMI_{55}(0.25,0.10)$	8.762	−1.90	1.136	1.149	1.078	90.8
$LMI_{55}(0.25,0.25)$	8.794	−1.54	1.147	1.155	1.079	89.0
Wrong working model for Y (including a reduced set of predictors and a mis-specified distribution) only						
CE_{53}	6.401	−28.34	0.951	2.704	1.388	50.4
$SCE_{53}(0.10)$	8.92	−0.13	1.211	1.211	2.044	97.6
$SCE_{53}(0.25)$	8.875	−0.64	1.190	1.191	2.013	97.0
$NNMI_{53}(5,0.8,0.2)$	8.846	−0.96	1.147	1.150	1.078	90.4
$LMI_{53}(0.10,0.10)$	8.828	−1.16	1.145	1.150	1.076	90.4
$LMI_{53}(0.10,0.25)$	8.833	−1.11	1.155	1.159	1.079	89.8
$LMI_{53}(0.25,0.10)$	8.892	−0.45	1.155	1.156	1.082	90.4
$LMI_{53}(0.25,0.25)$	8.891	−0.46	1.159	1.160	1.083	90.4
Wrong working model for both Y (a misspecified distribution) and δ (including a reduced set of predictors)						
CE_{35}	8.829	−1.15	1.195	1.199	1.195	91.0
$SCE_{35}(0.10)$	9.456	5.87	1.381	1.477	2.345	99.2
$SCE_{35}(0.25)$	9.124	2.15	1.270	1.284	2.301	98.6

Method	Estimate[a]	% Bias[b]	SD[c]	RMSE[d]	SE[e]	CR(%)[f]
NNMI$_{35}$(5,0.8,0.2)	8.941	0.10	1.168	1.168	1.104	91.2
LMI$_{35}$(0.10,0.10)	8.897	−0.39	1.184	1.185	1.106	90.2
LMI$_{35}$(0.10,0.25)	8.986	0.61	1.175	1.176	1.115	91.4
LMI$_{35}$(0.25,0.10)	8.888	−0.49	1.155	1.156	1.103	91.2
LMI$_{35}$(0.25,0.25)	8.965	0.37	1.175	1.175	1.107	91.6

[a] average of the 500 point estimates

[b] 100*(estimate− true value)/true value

[c] sample standard deviation of the 500 point estimates

[d] square root of mean-squared-errors

[e] average of the 500 standard error estimates

[f] coverage rate of the 500 95% confidence intervals

Table 8

Total response time from injury (min) by distance (mile) for air transportation

Method	Distance			
	5–10 (37[a]; 75.7%[b])	11–20 (187; 74.9%)	21–30 (213; 73.2%)	31–40 (130; 67.7%)
CC	82.78±20.98[c]	82.02±9.72	103.86±11.60	76.71±4.59
Include all 5 predictors in both working models				
CE_{55}	86.97±20.43	82.92±8.90	101.10±11.23	75.32±5.00
$SCE_{55}(0.10)$	84.22±26.39	74.93±11.09	85.01±10.10	91.38±12.36
$NNMI_{55}(5,0.8,0.2)$	76.22±15.33	80.03±7.25	91.08±8.00	80.43±9.46
$LMI_{55}(0.10,0.10)$	92.03±18.04	88.21±9.25	98.45±8.62	88.12±6.41
The working model for Y (δ) contains 5(3) predictors				
CE_{53}	86.61±20.61	80.57±8.61	103.09±10.79	77.70±5.42
$SCE_{53}(0.10)$	84.66±26.57	75.14±10.98	85.27±9.89	91.94±12.05
$NNMI_{53}(5,0.8,0.2)$	71.04±9.72	81.49±7.50	88.80±8.36	82.50±10.11
$LMI_{53}(0.10,0.10)$	83.44±10.26	87.74±17.66	101.61±16.09	91.16±8.49
The working model for Y (δ) contains 3(5) predictors				
CE_{35}	87.25±20.54	82.94±8.91	101.16±11.22	75.46±4.91
$SCE_{35}(0.10)$	87.74±26.32	73.41±10.99	89.80±10.08	90.79±12.22
$NNMI_{35}(5,0.8,0.2)$	79.46±15.33	81.83±9.39	92.36±8.91	77.87±4.46
$LMI_{35}(0.10,0.10)$	85.80±14.11	85.78±6.51	103.01±8.20	92.98±8.44

Method	Distance		
	41–50 (156; 61.5%)	51–60 (132; 58.3%)	61–70 (212; 59.4%)
CC	91.50±6.01	108.40±12.63	121.79±8.83
Include all 5 predictors in both working models			
CE_{55}	89.96±6.59	110.19±14.54	121.24±8.52
$SCE_{55}(0.10)$	101.70±10.81	116.24±12.56	119.49±9.14
$NNMI_{55}(5,0.8,0.2)$	95.14±6.31	105.59±10.58	122.48±11.99
$LMI_{55}(0.10,0.10)$	95.38±7.70	104.10±12.20	123.11±14.81
The working model for Y (δ) contains 5(3) predictors			
CE_{53}	89.72±6.56	109.84±13.97	121.45±8.24
$SCE_{53}(0.10)$	101.79±10.73	115.92±12.40	119.80±9.03
$NNMI_{53}(5,0.8,0.2)$	94.36±5.12	110.10±11.96	121.23±7.56
$LMI_{53}(0.10,0.10)$	95.79±5.59	104.69±8.66	116.29±9.25
The working model for Y (δ) contains 3(5) predictors			
CE_{35}	90.04±6.58	110.19±14.56	121.05±8.52
$SCE_{35}(0.10)$	97.79±10.47	102.28±11.09	116.34±8.65
$NNMI_{35}(5,0.8,0.2)$	93.24±5.60	109.25±13.87	117.61±13.20

Method	Distance			
	5–10 (37[a]; 75.7%[b])	**11–20 (187; 74.9%)**	**21–30 (213; 73.2%)**	**31–40 (130; 67.7%)**
LMI$_{35}$(0.10,0.10)	97.22±7.06	113.23±11.50	119.38±8.12	

[a] sample size

[b] missingness rate

[c] mean±standard error