Supplementary Materials for "Accounting for not-at-random missingness through imputation stacking" by

Lauren J. Beesley[*1] **and Jeremy Taylor**[1]
[1]University of Michigan, Department of Biostatistics
*Corresponding Author: lbeesley@umich.edu

A An extension to multiple variable MNAR missingness

In this section, we provide a generalization of the proposed method that allows for MNAR missingness in *multiple* variables, with some restrictions. Suppose the columns of Z are ordered such that the first d variables are assumed to be MNAR. We make the following assumptions.

Suppose we partition the joint model for missingness as

$$\begin{aligned} f(\mathcal{R}_{i1},\ldots,\mathcal{R}_{ik}|Z_{i.}) &= f(\mathcal{R}_{i,d+1},\ldots,\mathcal{R}_{ik}|Z_{i.},\mathcal{R}_{i1},\ldots\mathcal{R}_{id}) \\ &\times f(\mathcal{R}_{id}|Z_{i.},\mathcal{R}_{i1},\ldots,\mathcal{R}_{i,d-1}) \times \ldots \times f(\mathcal{R}_{i2}|Z_{i.},\mathcal{R}_{i1})f(\mathcal{R}_{i1}|Z_{i.}). \end{aligned}$$

where f denotes the distribution function for the corresponding variables. We will assume the following:

1. $Z_{i,d+1},\ldots,Z_{ik}$ are MAR, with
   $f(\mathcal{R}_{i,d+1},\ldots,\mathcal{R}_{ik}|Z_{i.},\mathcal{R}_{i1},\ldots,\mathcal{R}_{id}) = f(\mathcal{R}_{i2},\ldots,\mathcal{R}_{ik}|W_{i.})$.
2. For $j=1,\ldots,d$, Z_{ij} may be MNAR, with $f(\mathcal{R}_{ij}|Z_{i.},\mathcal{R}_{i1},\ldots,\mathcal{R}_{i,j-1}) = f(\mathcal{R}_{ij}|Z_{ij},W_{i.})$.

In *Assumption 2*, we allow Z_{ij} to be MNAR such that its missingness depends on the true value of Z_{ij} but does not depend on the other variables with missingness or their missingness indicators given Z_{ij} and $W_{i.}$.

A.1 Imputation and weights under *Assumptions 1-2*

Imputation of $Z_{i,d+1},\ldots,Z_{ik}$ will be the same as before, since these variables are assumed to be MAR and independent of missingness in the MNAR variables. Since we assume that missingness is independent between the MNAR variables and that missingness in each variable is independent of *other* variables with missingness, we can again re-write the imputation distribution for each MNAR variable $Z_{ij}, j=1,\ldots,d$ as follows:

$$f(Z_{ij}|Z_{i,-j},\mathcal{R}_{ij}=0) \propto \frac{P(\mathcal{R}_{ij}=0|Z_{ij},W_{i.})}{1-P(\mathcal{R}_{ij}=0|Z_{ij},W_{i.})} f(Z_{ij}|Z_{i,-j},\mathcal{R}_{ij}=1) \qquad (Eq.\ S1)$$

Suppose that instead of imputing from *Eq. S1* directly, we impute each Z_{ij} from $f(Z_{ij}|Z_{i,-j},\mathcal{R}_{ij}=1)$. We can collect the "weight" terms for each one of the imputed variables to obtain an aggregate weight to use for the final analysis as follows:

$$\omega_{im} \propto \prod_{j=1}^{d} \frac{P(\mathcal{R}_{ij}=0|Z_{ijm},W_{i.})}{1-P(\mathcal{R}_{ij}=0|Z_{ijm},W_{i.})} \qquad (Eq.\ S2)$$

where Z_{ijm} is the m^{th} imputation of Z_{ij}. Suppose, now, that we can reasonably approximate each missingness model with a logistic regression model as follows:

$$\text{logit}\left(P(\mathcal{R}_{ij}=1|Z_{ij},W_{i.})\right) = \phi_0 + \phi_j Z_{ij} + \phi_{Wj}^T W_{i.} \qquad (Eq.\ S3)$$

In this case, we can re-write the weight as

$$\omega_{im} \propto \prod_{j=1}^{d} \exp(-\phi_j Z_{ijm}) = \exp(-\sum_{j=1}^{d} \phi_j Z_{ijm}) \qquad (Eq.\ S4)$$

This weight is a very simple function of the imputed data and d sensitivity parameters, $\phi_1, \ldots, \phi_d$. Analysis can proceed as described in the main paper, where the form of the weights now depends on *multiple* sensitivity parameters, each describing the strength of the MNAR dependence of each variable and its own missingness, adjusting for W.

A.2 Approximated method for more complicated MNAR mechanisms

Suppose we want to consider more general MNAR missingness, where we no longer make *Assumptions 1-2.* We examine the conditional distribution we want to impute from under a chained equations imputation philosophy: $f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{i.} = R_{i.})$. Recall, $R_{i.}$ is the data realization of random variable $\mathcal{R}_{i.}$ related to whether each of the variables is observed for subject i. We have that

$$\begin{aligned} f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{i.} = R_{i.}) &= \frac{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.})}{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.,-j})} f(Z_{ij}|Z_{i,-j}) \\ &= \frac{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.})}{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.,-j})} \frac{P(\mathcal{R}_{ij} = 1|Z_{i,-j})}{P(\mathcal{R}_{ij} = 1|Z_{i.})} f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{ij} = 1) \\ &\propto \frac{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.})}{P(\mathcal{R}_{ij} = 1|Z_{i.})} f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{ij} = 1) \qquad (Eq.\ S5) \end{aligned}$$

Without *Assumptions 1-2*, we note that the proportionality term in *Eq. S5* may depend on *multiple* variables with missingness in general, not just the value of Z_{ij}. If we construct global weights by multiplying together all of the proportionality terms for all the imputed variables, a given imputed covariate may appear in multiple proportionality terms. The logic motivating the importance sampling and stacking approach starts to break down in this case, and the resulting weighted imputations may no longer correspond to a draw from the correct *joint* posterior predictive distribution, even with correctly specified imputation and missingness models. In spite of this technical limitation, we propose the following weighting strategy as an approximate approach to handle the MNAR missingness.

The expression in *Eq. S5* suggests that we might impute each Z_{ij} from $f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{ij} = 1)$ as discussed in the main paper and weight the multiple imputations proportional to the product of the remaining components of *Eq. S5*. After pulling the term constant in j out of the product, we define weights as follows

$$\omega_{im} \propto \frac{P(\mathcal{R}_{i.} = R_{i.}|Z_{i.m})}{\prod_{j:R_{ij}=0} P(\mathcal{R}_{ij} = 1|Z_{i.m})} \qquad (Eq.\ S6)$$

where $Z_{i.m}$ is the m^{th} imputation of $Z_{i.}$. In general, this form of the weight could be complicated and may depend on the association between missingness and fully-observed variables.

We consider the special case where $\mathcal{R}_{ij} \perp \mathcal{R}_{ik}|Z_{i.}$ for all $k \neq j$. Restated, assume that missingness for each variable is conditionally independent of the missingness of other variables, given the true values for $Z_{i.}$. In this case, we can simplify

$$\omega_{im} \propto \left[\prod_{j:R_{ij}=1} P(\mathcal{R}_{ij} = 1|Z_{i.m})\right] \times \left[\prod_{j:R_{ij}=0} \frac{P(\mathcal{R}_{ij} = 0|Z_{i.m})}{P(\mathcal{R}_{ij} = 1|Z_{i.m})}\right] \qquad (Eq.\ S7)$$

Under logistic regression models for the missingness in each variable as in *Eq. S4*, the second product in *Eq. S7* will have a simple, convenient form similar to *Eq. S4*, but the first product will not. In order to use the weight in *Eq. S7*, we will need to posit a model for each MNAR missingness mechanism. *Eq. S7* further reduces to *Eq. S2* when *Assumptions 1-2* are satisfied.

A.3 Exact method for more complicated MNAR mechanisms

The imputation and weighting approach proposed in A.2 is obtained by defining importance sampling weights for each imputed variable separately and multiplying them together. This strategy ignores that several of the imputed variables may contribute to each component of product. Through simulation, we demonstrate later on that this merged weight may result in some residual bias in downstream analysis for certain missingness scenarios. In this section, we propose an alternative method motivated by the target joint posterior predictive distribution directly.

Let $Z_{i,mis}$ denote the missing elements of $Z_{i.}$ and let $Z_{i,obs}$ denote the observed values of $Z_{i.}$ for a given subject i. Similarly, define $\mathcal{R}_{i,mis}$ to be the collection of random missingness indicators such that $R_{ij} = 0$ and define $\mathcal{R}_{i,obs}$ to be the collection of random missingness indicators such that $R_{ij} = 1$ for subject i. We want to impute values for $Z_{i,mis}$ from the following distribution:

$$f(Z_{i,mis}|Z_{i,obs}, \mathcal{R}_{i,obs} = \vec{1}, \mathcal{R}_{i,mis} = \vec{0}) = \frac{P(\mathcal{R}_{i,obs} = \vec{1}|Z_{i.}, \mathcal{R}_{i,mis} = \vec{0})f(Z_{i,mis}|Z_{i,obs}, \mathcal{R}_{i,mis} = \vec{0})}{P(\mathcal{R}_{i,obs} = \vec{1}|Z_{i,obs}, \mathcal{R}_{i,mis} = \vec{0})}$$

$$\propto P(\mathcal{R}_{i,obs} = \vec{1}|Z_{i.}, \mathcal{R}_{i,mis} = \vec{0})\frac{P(\mathcal{R}_{i,mis} = \vec{0}|Z_{i.})}{P(\mathcal{R}_{i,mis} = \vec{1}|Z_{i.})}f(Z_{i,mis}|Z_{i,obs}, \mathcal{R}_{i,mis} = \vec{1}) \qquad (Eq.\ S8)$$

where the notation $\mathcal{R}_{i,mis} = \vec{c}$ means that each random indicator in $\mathcal{R}_{i,mis}$ takes the value c.

Using the importance sampling logic in the main paper, we propose first obtaining multiple draws of $Z_{i,mis}$ from $f(Z_{i,mis}|Z_{i,obs}, \mathcal{R}_{i,mis} = \vec{1})$. We can implement this step using a chained equations imputation philosophy and iteratively draw each Z_{ij} from $f(Z_{ij}|Z_{i,-j}, \mathcal{R}_{i,mis} = \vec{1})$. Consider a more concrete example where we want to impute Z_{i1} for a subject with Z_{i2} observed and Z_{i3} missing. We would impute Z_{i1} from $f(Z_{i1}|Z_{i2}, Z_{i3}, \mathcal{R}_{i1} = 1$ and $\mathcal{R}_{i3} = 1)$. For a different subject with only Z_{i1} missing, we would impute Z_{i1} from $f(Z_{i1}|Z_{i2}, Z_{i3}, \mathcal{R}_{i1} = 1)$. This approach translates into different imputation distributions (or corresponding parameter draws) for each *pattern* of missing values in the data.

After obtaining the multiple imputations, we then weight the resulting draws using

$$\omega_{im} \propto P(\mathcal{R}_{i,obs} = \vec{1}|Z_{i.m}, \mathcal{R}_{i,mis} = \vec{0})\frac{P(\mathcal{R}_{i,mis} = \vec{0}|Z_{i.m})}{P(\mathcal{R}_{i,mis} = \vec{1}|Z_{i.m})} \qquad (Eq.\ S9)$$

In the special case where $\mathcal{R}_{ij} \perp \mathcal{R}_{ik}|Z_{i.}$ for all $k \neq j$, the form of the weight simplifies to match *Eq. S7* exactly. In this case, the only difference between this approach and the approach in **Section A.2** is how the multiple imputations are obtained.

B Eliciting Sensitivity Parameters

In the main paper and **Section A**, we describe how we can account for MNAR missingness through analysis of stacked and weighted multiple imputations, where the form of the weights depends on the missingness mechanism. In practice, however, this mechanism will not be known and the goal of "correcting for" MNAR missingness is not reasonable. Instead, the goal of analysis may be to characterize the impact of various levels of deviation from MAR assumptions, where repeated analyses are performed and interpreted as a sensitivity analysis. The challenge, then, is in defining the space of plausible MNAR mechanisms over which to perform the sensitivity analysis. In this section, we highlight several existing strategies for performing this sensitivity analysis in the literature and describe how they could be modified and implemented in the proposed modeling framework.

Existing methods for this type of sensitivity analysis (e.g. Tompsett et al. (2018)) usually conceptualize the problem in a pattern mixture model framework, where deviations from MNAR are defined in terms of the differences between the distribution of the missing data for subjects with observed and missing data. A common implementation is to posit a generalized linear imputation model for $Z_{.j}$ including $Z_{.,-j}$ and $R_{.,-j}$ as predictors and including the offset $\delta_j(1 - R_{.j})$ in the linear predictor. The parameter δ_j is not identified, and its fixed value can be interpreted in terms of an assumed association between missingness in $Z_{.j}$ and the value of $Z_{.j}$, adjusting for $Z_{.,-j}$ and $R_{.,-j}$. The impact of deviations from MAR assumptions are then evaluated by repeating analysis across plausible values for δ_j.

Our proposed method instead casts the problem in terms of the selection modeling framework, where the imputation model for $Z_{.j}$ is the fully identified distribution $f(Z_{.j}|Z_{.,-j}, \mathcal{R}_{.j} = 1)$ and the selection/missingness mechanism, $P(R_{.j} = 1|Z)$, is modeled directly. Parameters in this selection model are not fully identified, and we instead posit a generalized linear model for $R_{.j}$ adjusting for $Z_{.,-j}$ and including the offset $\phi_j Z_{.j}$ in the linear predictor. Parameter ϕ_j is not identified and its fixed value can be interpreted in terms of an assumed association between covariate $Z_{.j}$ and its own missingness, adjusting for $Z_{.,-j}$.

In both pattern mixture and selection modeling frameworks, reasonable values of sensitivity parameters can be difficult to determine, and several strategies for defining a plausible set of sensitivity parameters have emerged in the literature. One commonly-used strategy that can be easily implemented under the proposed modeling framework is called "tipping point" analysis (e.g. Tompsett et al. (2018); Ratitch et al. (2013)), where analysis is repeated for a wide interval of sensitivity parameters and we bounds of the sensitivity parameter for which our study conclusions are changed in a meaningful way. The extent of concern about deviations from MAR is then converted into a question of the plausibility of these bounds.

Given that these sensitivity parameters are defined in terms of adjusted associations, it can still be difficult to determine whether a single fixed value of the sensitivity parameter/s is reasonable. One solution is to reformulate the sensitivity problem in terms of more easily interpretable parameters. In Tompsett et al. (2020), sensitivity parameters δ_j are converted into more easily interpretable parameters π_j (e.g. the unadjusted association between $Z_{.j}$ and its own missingness) by calculating π_j using imputed datasets obtained for a fixed δ_j. This defines a mapping between δ_j and more easily interpreted π_j, and the association between π_j and the downstream inference for outcome model parameters can be directly assessed. This same approach can be applied in the proposed modeling framework, where instead of estimating π_j using Rubin's rules applied to multiply imputed datasets obtained using the pattern mixture model-type offset method, we can instead estimate π_j using a stacked and weighted analysis of the MAR-based multiple imputations where weights are defined using sensitivity parameter ϕ_j. In this way, we can convert the problem of specifying plausible values of ϕ_j to the assessment of plausible ranges of some more interpretable parameter π_j.

When determining plausible values of the sensitivity parameter, it is often desirable to consult subject matter experts. As discussed in Rezvan et al. (2018) and Tompsett et al. (2020) among others, expectations for marginal summary statistics obtained from subject matter experts can help inform values for more complicated sensitivity parameters. While it may be difficult to conceptualize values for sensitivity parameters δ_j or ϕ_j, subject matter experts may be able to more easily provide expected outcomes or differences between people with missing and observed data on average. Expectations obtained from several different subject matter experts can be combined to formulate a distribution for the unknown sensitivity parameter, and this can be used to guide sensitivity analyses. For a concrete example, suppose we want to elicit parameter δ_j associated with the conditional distribution of binary covariate Z_j. We could perform our analysis across several values of δ_j and for each estimate a corresponding value for the mean of Z_j given $\mathcal{R}_j = 0$. We can then use the expected distribution $Z_j|\mathcal{R}_j = 0$ obtained from subject matter experts to define plausible values for δ_j, using the mapping estimated from the imputed data. This same approach can just as easily be applied to our selection modeling setting, where Z_j given $\mathcal{R}_j = 0$ can be estimated using the stacked imputed data analyzed using weights parameterized by ϕ_j.

In this way, existing literature guiding sensitivity parameter elicitation for the pattern mixture modeling framework can often be applied in the proposed selection modeling framework as well and can be leveraged to guide choices for sensitivity parameters ϕ.

C Simulation 1: additional results for single variable missingness

Figure C.1: Distribution of linear regression Z_2 coefficient for 100 imputed datasets under MAR assumptions. Points correspond to corresponding re-weighting of these estimates using the method in Carpenter et al. (2007) with *assumed* values of ϕ_1 in (-1.5, 1.5). The true coefficient value for the regression for $Z_2|Z_1$ is 0.50, denoted by the vertical line. The true missingness model log-odds ratio, ϕ_1, is 1.

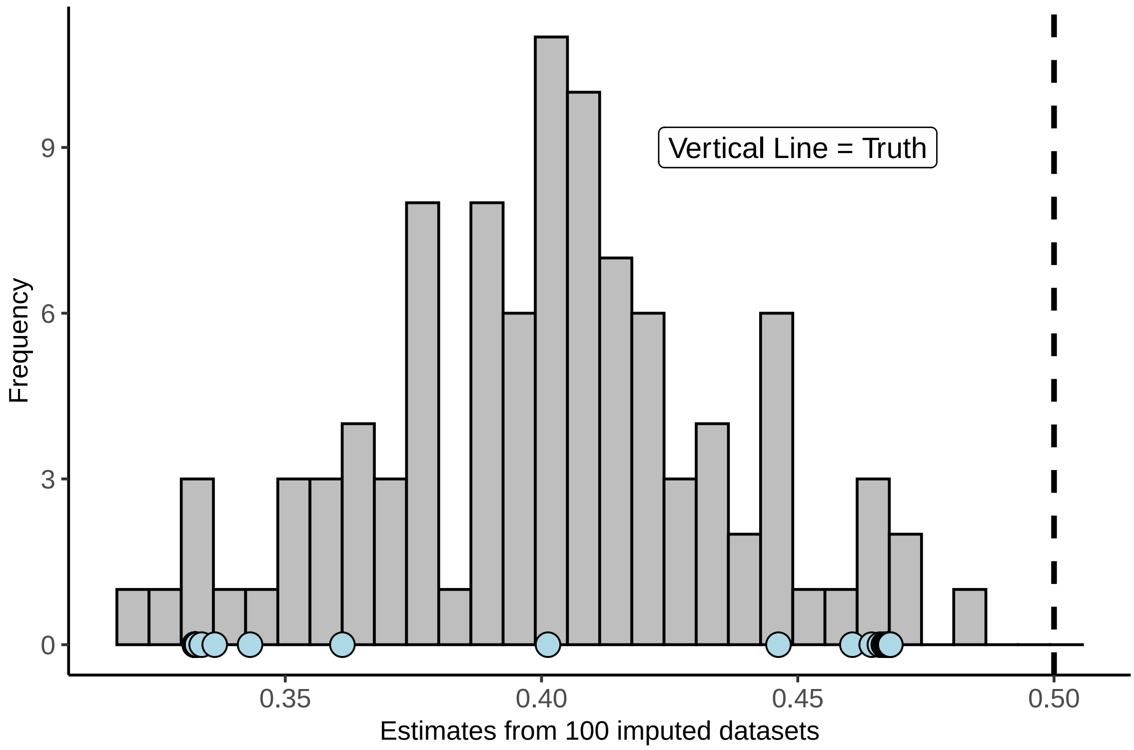


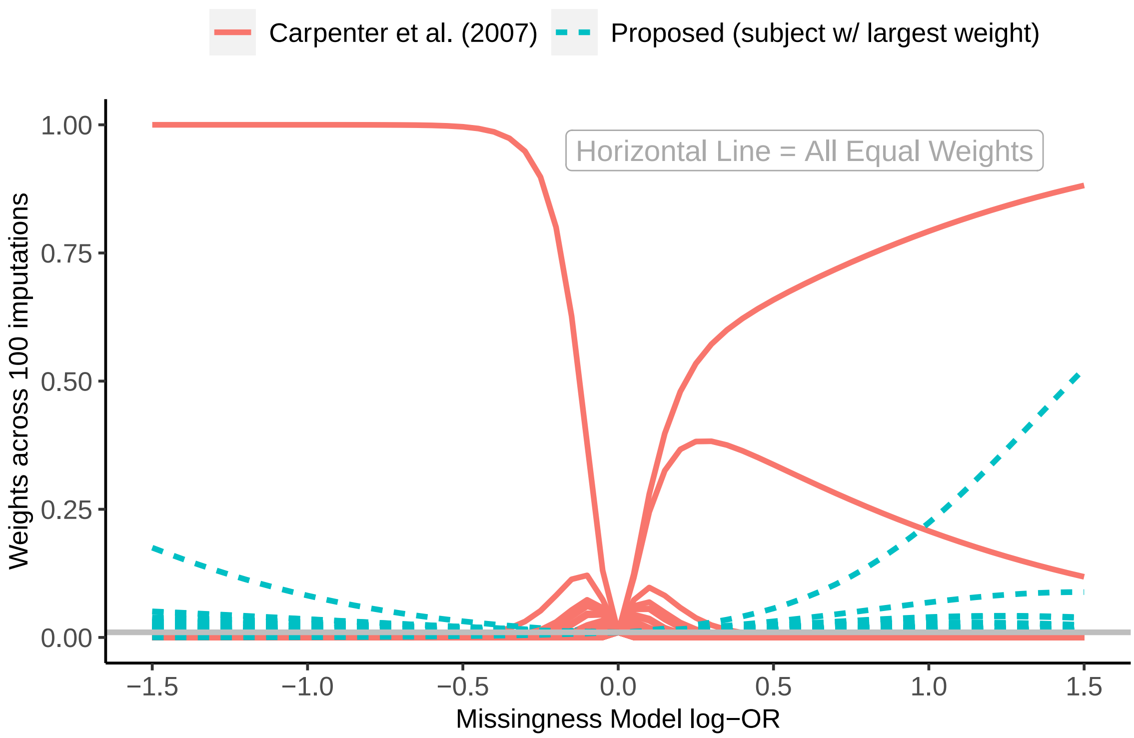
Table C.1: Bias in estimated $Z_1|Z_2$ model parameters across 1000 simulated datasets for alternative MNAR-based chained equations methods and the proposed method (n=1000, M=100). [1]

Method	ϕ_1	Intercept	Coefficient of Z_2
Truth	1	0	0.5
Complete Case Analysis	1	0.323	0.342
Proposed Method	1	-0.009	0.507
Tompsett et al. (2018)	1	0.003	0.500
Jolani (2012)	1	0.411	0.297
Truth	0.5	0	0.5
Complete Case Analysis	0.5	0.180	0.416
Proposed Method	0.5	-0.001	0.501
Tompsett et al. (2018)	0.5	0.005	0.498
Jolani (2012)	0.5	0.266	0.376

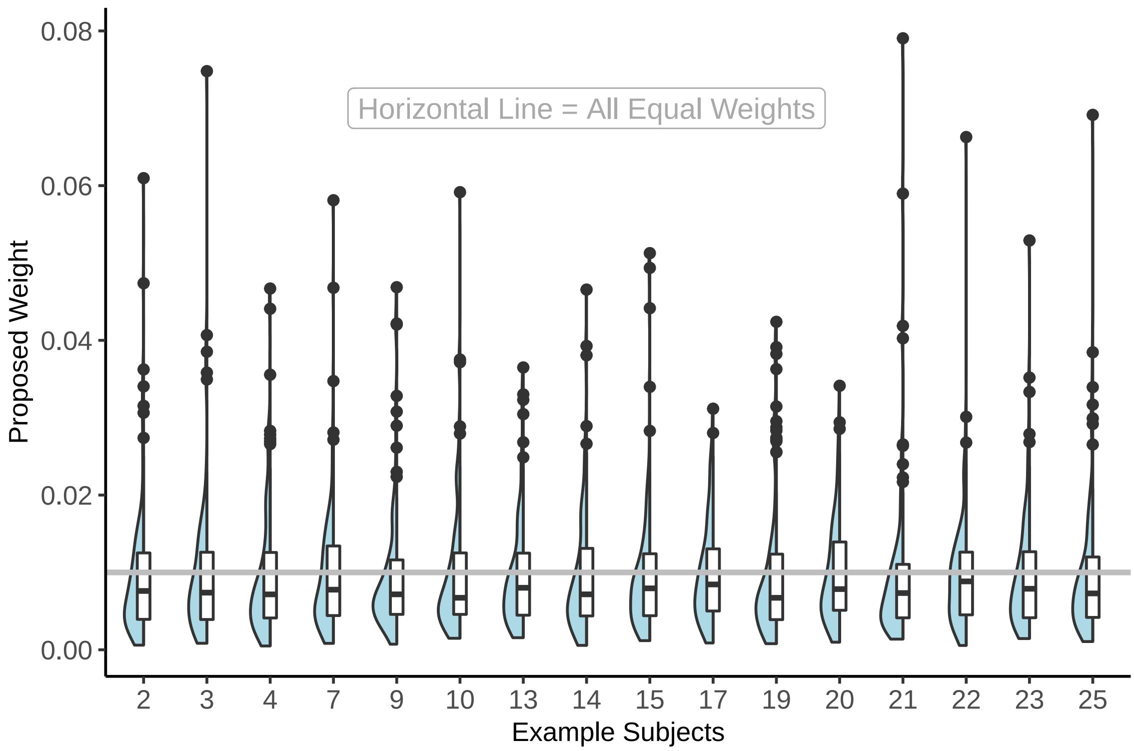
[1] For the method in Tompsett et al. (2018), we did not adjust for missingness indicators for *other* variables when imputing $Z_{\cdot 1}$ and instead only incorporated the fixed offset as a function of $R_{\cdot 1}$. This method was implemented using the *mice.impute.mnar.norm* imputation method in R. The method in Jolani (2012) was implemented using *mice.impute.ri* imputation method in R. For the proposed method and the method in Tompsett et al. (2018), sensitivity parameters were fixed to the best possible value (which differs between the two methods).

Figure C.2: Weights assigned to each of 100 multiple imputations across different *assumed* values for ϕ_1 in (-1.5, 1.5) for normally-distributed Z_1. The true missingness model log-odds ratio, ϕ_1, is 1.[1].

(a) Weights assigned to each of 100 imputed datasets (lines) for proposed method and Carpenter et al. (2007) method

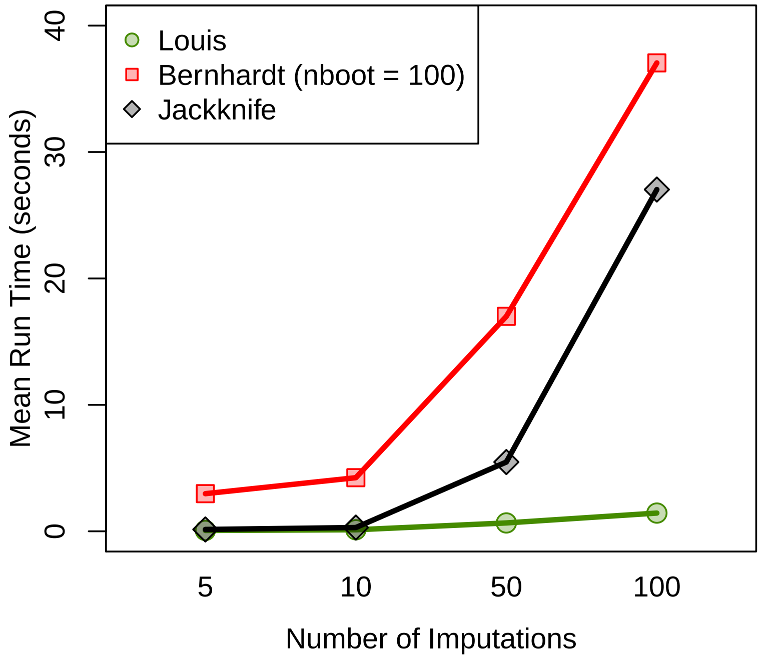


(b) Boxplots and densities of *proposed* weights across 100 multiple imputations for some example subjects



[1] For the method in Carpenter et al. (2007), weights are scaled to sum to 1 across imputed datasets. For the proposed methods, weights are scaled to sum to 1 across imputed datasets *within subjects*. The horizontal gray line in both figures corresponds to the setting with equal weights assigned to all imputations.

Figure C.3: Time to compute standard errors for linear regression based on stacked multiple imputations as a function of M (n = 1000)



D Simulation 2: missingness in multiple covariates

We now consider the case where we have missingness in multiple covariates. In each simulation setting, we generate 1000 simulated datasets of N = 2000 subjects. In all settings, we generate covariates (Z_1, Z_2, Z_3) following a multivariate normal distribution with mean zero, standard deviation 1, and covariances of 0.3. We then generate Z_4 under linear regression $Z_4 \sim N(0 + 0.5Z_1 + 0.5Z_2 + 0.5Z_3, 1)$. We impose MNAR missingness in Z_1 under the following model: $\text{logit}(P(R_1 = 1|Z)) = \phi_0 + \phi_1 Z_1 + 0.5Z_4$. Missingness model parameters were specified to generate different degrees of deviation from MAR, with true missingness model log-odds ratio ϕ_1 taking values in -0.5, 0, 0.5, and 1. ϕ_0 was chosen to give roughly a 50% missingness rate for Z_1. We also generate 50% MCAR missingness in Z_2.

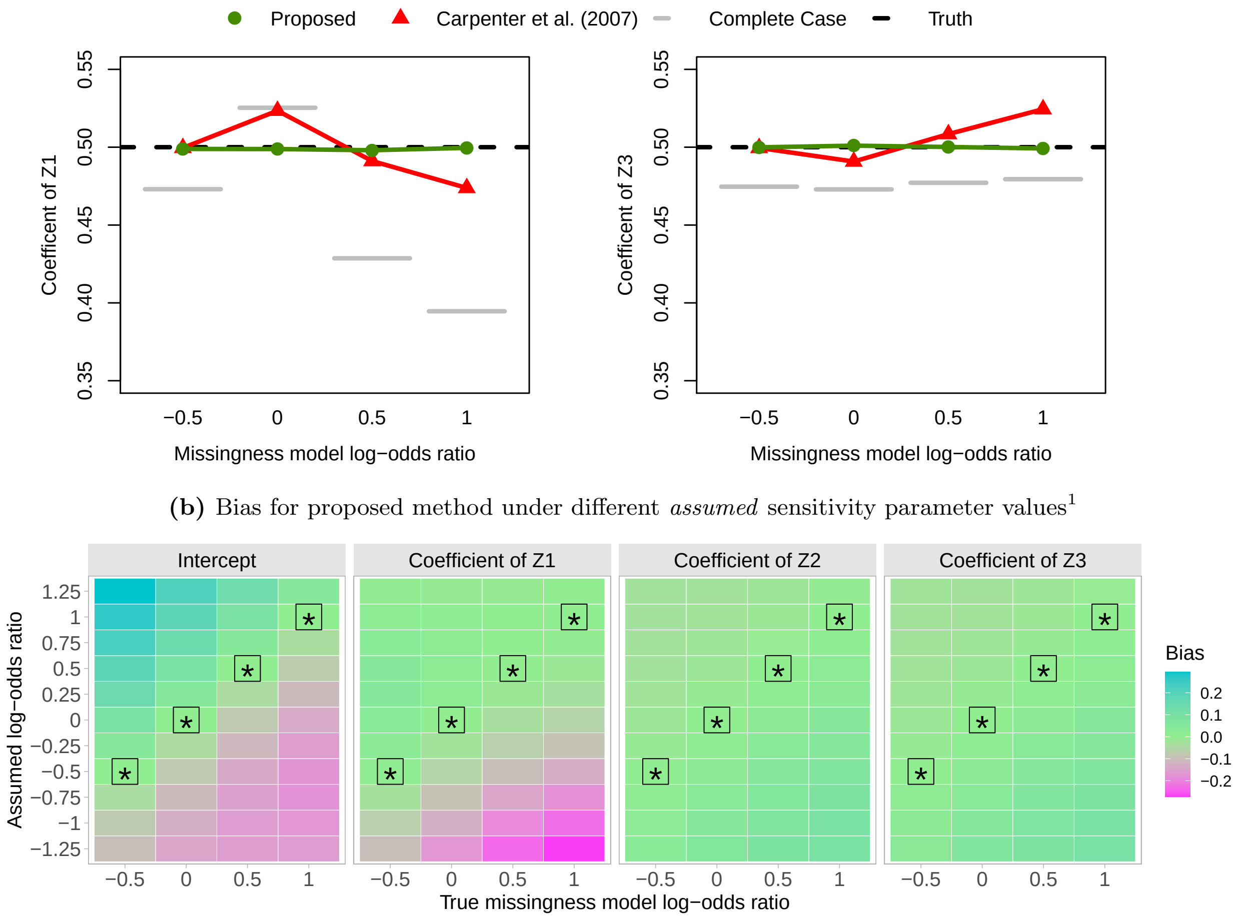
We then obtain 50 multiple imputations for missing values of Z_1 and Z_2 using the package *mice* in R assuming linear regression imputation models for each variable. Using these multiple imputations, we then estimate parameters in the model for $Z_4|Z_1, Z_2, Z_3$ either using the method in Carpenter et al. (2007) or the proposed stacking and weighting method. Since ϕ_1 would be usually unknown in practice, we perform this estimation for different *assumed* values of log-odds ratio ϕ_1.

Figure D.1a shows the bias in linear regression parameter estimates (model for $Z_4|Z_1, Z_2, Z_3$) when ϕ_1 is *correctly specified*. As before, we find that our proposed approach can do a good job at estimating model parameters and that the method in Carpenter et al. (2007) can result in substantial residual bias. **Figure D.1b** provides the estimates across different *assumed* values for ϕ_1, with the true value of ϕ_1 highlighted with a star.

Figure D.1: Average bias in estimated parameters from linear regression of $Z_4|Z_1, Z_2, Z_3$ under MNAR missingness in covariate Z_1 and MCAR missingness in covariate Z_2.

(a) Bias with correctly-specified sensitivity parameter, ϕ_1

(b) Bias for proposed method under different *assumed* sensitivity parameter values[1]



[1] The true value of ϕ_1 is highlighted for each simulation setting by '*'.

An alternative imputation strategy is to apply the Tompsett et al. (2018) pattern mixture model approach, where imputation of Z_{i1} uses $Z_{i,-1}$ and $1 - R_{.,-1}$ as covariates and the coefficient for $1 - R_{.1}$ is a fixed sensitivity parameter. Since we assume that Z_{i1} is independent of $1 - R_{.,-1}$, this approach reduces to performing imputation with a fixed offset proportional to $(1 - R_{.1})$ with corresponding sensitivity parameter controlling the degree of deviation from MAR. This coefficient does not directly correspond to the coefficient in the missingness model. However, we can determine the best possible value of this offset parameter for our simulated data by fitting a regression model for true Z_{i1} given true $Z_{i,-1}$ and R_{i1}. We use the resulting ideal parameter, which would not be known in real data analyses, to benchmark the performance of our proposed method relative to the strategy of including an offset in the imputation model. For selection model sensitivity parameter ϕ in -0.5, 0, 0.5, and 1, the ideal Z_{i1} pattern mixture model model parameter should be roughly 0.34, 0, -0.34, and -0.62, respectively.

Figure D.2 provides bias in outcome model parameter estimates across different values of these alternative pattern mixture model sensitivity parameters, where the ideal offset sensitivity parameter is shown for each of the 4 simulation settings along the x-axis. Generally, if we posit values of the sensitivity parameter near the best possible value, this method can produce low bias. This simulation demonstrates that the Tompsett et al. (2018) approach can work in ideal settings, as can our method. The implementation is the primary difference here, where our sensitivity parameters correspond to the missingness model directly, and we only have to impute once rather than separately for each value of the sensitivity parameters.

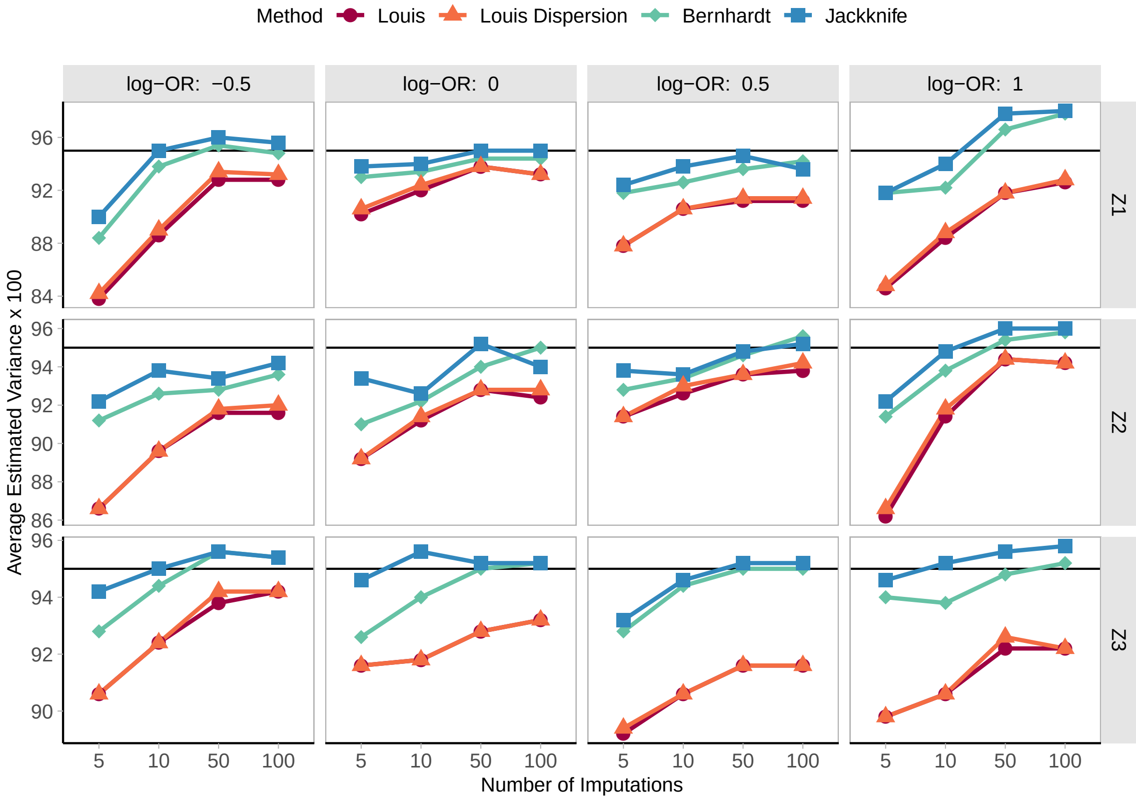
We also compare the proposed method and the method in Tompsett et al. (2018) to an alternative method proposed in Jolani (2012), which aims to avoid the need to specify the sensitivity parameter entirely. In **Table D.1**, we demonstrate that the current *mice* implementation of this method performs poorly relative to the proposed method and the method in Tompsett et al. (2018) under ideal settings.

Figure D.2: Average bias from fixed offset imputation method

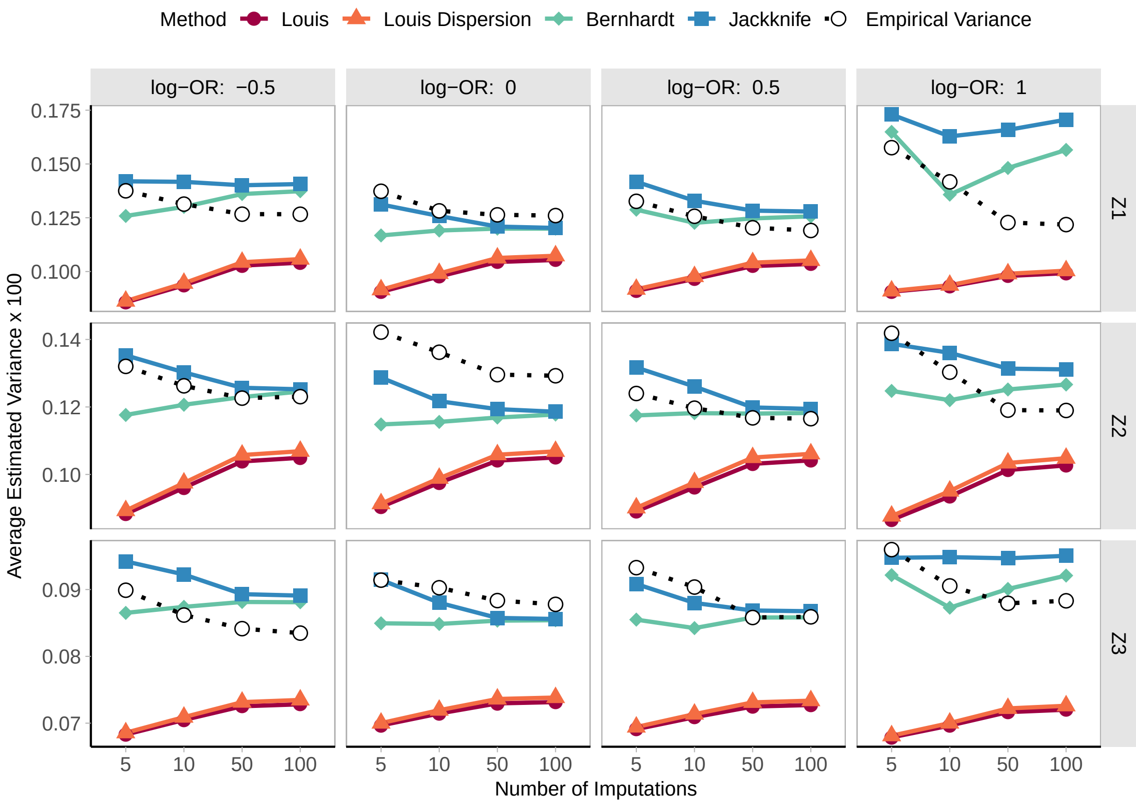
Figure D.3 provides the average estimated variances and coverage of 95% confidence intervals for the proposed method using each of the three variance estimation strategies. We again find that the method in *Eq. 9* tends to produce slight under-coverage. Additionally, for large values of true ϕ_1, the method from Bernhardt (2019) and our jackknife modification both produce slight over-coverage. Coverage rates for logistic regression model parameters are near-nominal for all methods (not shown). Under-coverage of the method in *Eq. 9* is not a result of estimating the linear regression dispersion parameter, since the under-coverage persists even when we fix the dispersion parameter at the simulation truth rather than estimating it (**Figure D.3**, "Louis Dispersion" method).

Figure D.3: Coverage of 95% confidence intervals and average estimated variances for linear regression coefficients from stacked and weighted analysis across 1000 simulated datasets, assuming true ϕ_1 is known

(a) Coverage of 95% confidence intervals



(b) Average estimated variances



The "Louis Dispersion" method uses the estimator in *Eq. 9* applied to imputed data obtained after fixing the imputation model dispersion parameters to the simulation truths.

Table D.1: Bias in estimated $Z_4|Z_1, Z_2, Z_3$ model parameters across 1000 simulated datasets for alternative MNAR-based chained equations method and the proposed method (M=100). [1]

Method	ϕ_1	Intercept	Coefficient of Z_1	Coefficient of Z_2	Coefficient of Z_3
Truth	0	0	0.5	0.5	0.5
Complete Case Analysis	0	0.236	0.473	0.375	0.475
Proposed Method	0	0.000	0.499	0.499	0.500
Tompsett et al. (2018)	0	0.000	0.499	0.500	0.500
Jolani (2012)	0	-0.028	0.473	0.510	0.510
Truth	-0.5	0	0.5	0.5	0.5
Complete Case Analysis	-0.5	0.234	0.525	0.472	0.473
Proposed Method	-0.5	0.000	0.499	0.498	0.501
Tompsett et al. (2018)	-0.5	0.000	0.499	0.499	0.501
Jolani (2012)	-0.5	0.067	0.521	0.488	0.490
Truth	0.5	0	0.5	0.5	0.5
Complete Case Analysis	0.5	0.236	0.429	0.477	0.477
Proposed Method	0.5	0.001	0.498	0.500	0.500
Tompsett et al. (2018)	0.5	0.000	0.498	0.501	0.500
Jolani (2012)	0.5	-0.101	0.429	0.535	0.533
Truth	1	0	0.5	0.5	0.5
Complete Case Analysis	1	0.235	0.395	0.482	0.479
Proposed Method	1	0.002	0.500	0.500	0.499
Tompsett et al. (2018)	1	0.000	0.498	0.501	0.500
Jolani (2012)	1	-0.148	0.397	0.553	0.551

[1]For the method in Tompsett et al. (2018), we did not adjust for missingness indicators for *other* variables when imputing $Z_{\cdot 1}$ and instead only incorporated the fixed offset as a function of $R_{\cdot 1}$. This method was implemented using the *mice.impute.mnar.norm* imputation method in R. The method in Jolani (2012) was implemented using *mice.impute.ri* imputation method in R. For the proposed method and the method in Tompsett et al. (2018), sensitivity parameters were fixed to the best possible value (which differs between the two methods).

E Simulation 3: missingness with multiple MNAR variables

We now consider the case where we have MNAR missingness in multiple multivariate normal covariates with (Z_1, Z_2, Z_3, Z_4) generated as in **Section D**. We impose MNAR missingness in Z_1 under the following model: $\text{logit}(P(R_1 = 1|Z)) = \phi_1 Z_1 + \phi_2 Z_2 + 0.5Z_4$. We impose MNAR missingness in Z_2 under $\text{logit}(P(R_2 = 1|Z)) = \beta_1 Z_1 + \beta_2 Z_2 + 0.5Z_4$. We consider 5 different missingness mechanism scenarios as detailed in **Table E.1**. Scenario 1 corresponds to a setting where *Assumptions 1-2* are satisfied and the method in **Supplementary Section A.1** can be applied. For Scenarios 2-5, the critical assumption that $Z_{ij} \perp \mathcal{R}_{i,-j}|Z_{i,-j}$ is violated. For each simulated dataset, we apply the imputation and weighting strategies proposed in **Supplementary Section A.2** (denoted "Proposed, Approx.") or **Supplementary Section A.3** (denoted "Proposed, Exact") using 50 imputed datasets per method, assuming linear regression imputation models for each variable.

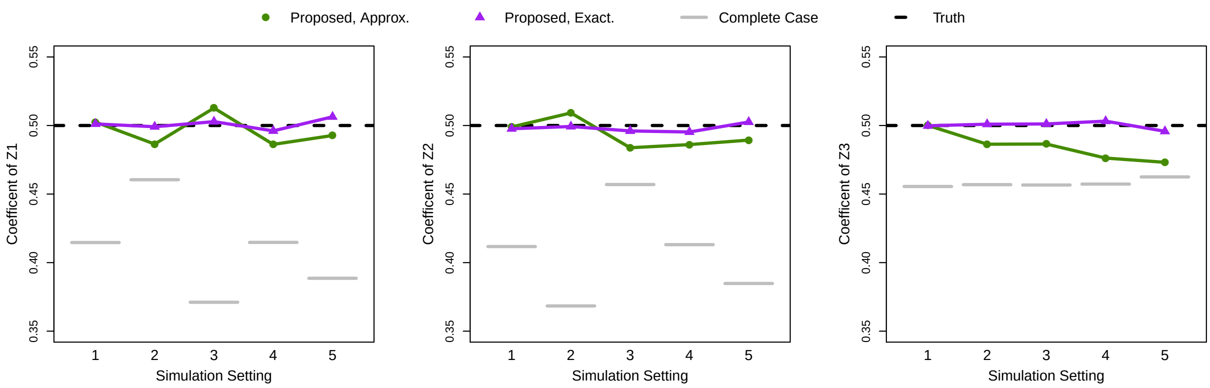
Table E.1: Scenarios considered in Simulation 3 with MNAR missingness in Z_1 and Z_2[1]

	Missingness in Z_1		Missingness in Z_2			
Setting	ϕ_1	ϕ_2	β_1	β_2	**Description of Missingness Mechanisms**	***Assumption 2***
1	0.5	0	0	0.5	Each mechanism depends on its *own* value	Satisfied
2	0	0.5	0	0.5	Each mechanism depends on *other* variable.	Violated
3	0.5	0	0.5	0	Both mechanisms depend only on Z_1.	Violated
4	0	0.5	0.5	0	Both mechanisms depend only on Z_2.	Violated
5	0.5	0.5	0.5	0.5	Both mechanisms depend on Z_1 and Z_2.	Violated

[1] *Assumption 1* is trivially satisfied for all simulation settings considered.

Figure E.1 shows the resulting bias in linear regression parameter estimates in a model for $Z_4|Z_1, Z_2, Z_3$. We also show the bias of complete case parameter estimates. We find that both varieties of the proposed method perform well when *Assumptions 1-2* are satisfied (Setting 1). Although not shown, this good performance was seen across a variety of simulation settings. When *Assumptions 1-2* are violated (Settings 2-5), both methods result in reduced bias compared to complete case analysis. However, we find that the "exact" method in **Supplementary Section A.3** tends to produce negligible bias, while the "approximate" method in **Supplementary Section A.2** results in some residual bias even under ideal settings.

Figure E.1: Average bias in outcome model parameters across 1000 simulated datasets in settings with MNAR missingness in multiple variables (M=50)[1]



[1] Assuming true missingness mechanisms are known.

F Additional information for oropharynx cancer example

In this section, we provide some additional information about HPV missingness and the implementation of imputation assuming MAR. Missing values in HPV status (positive/negative), smoking status (current/former/never), T-stage (T1/T2/T3/T4), overall cancer stage (I/II/III/IV), and ACE7 comorbidities (none/mild/moderate/severe) were imputed 50 times using the method in White and Royston (2009) to generate the multiple imputations. These imputations are then stacked, and analysis proceeds using the proposed method.

For generating multiple imputations, we first obtained the Nelson-Aalen estimate for the cumulative hazard of all-cause survival, denoted $\Lambda(T_i)$, using the censored overall survival outcome and event indicator, δ, in the data. Each covariate was then imputed using a regression model adjusting for other covariates along the $\Lambda(T_i)$ and δ. HPV status was imputed using a logistic regression model adjusting for gender, smoking status, age at diagnosis, overall cancer stage, ACE27 comorbidities, year of study enrollment, $\Lambda(T_i)$ and δ. Smoking status and ACE27 score were imputed using multinomial logistic regression adjusting for gender, HPV status, overall cancer stage, age at diagnosis, $\Lambda(T_i)$, δ, and either ACE27 score or smoking status, respectively. T-stage and overall cancer stage were both imputed using proportional odds regression adjusting for gender, smoking status, HPV status, comorbidities, age at diagnosis, $\Lambda(T_i)$, δ, and either overall cancer stage or T-stage, respectively.

Figure F.1: Observed and missing values of baseline HPV status by year of enrollment

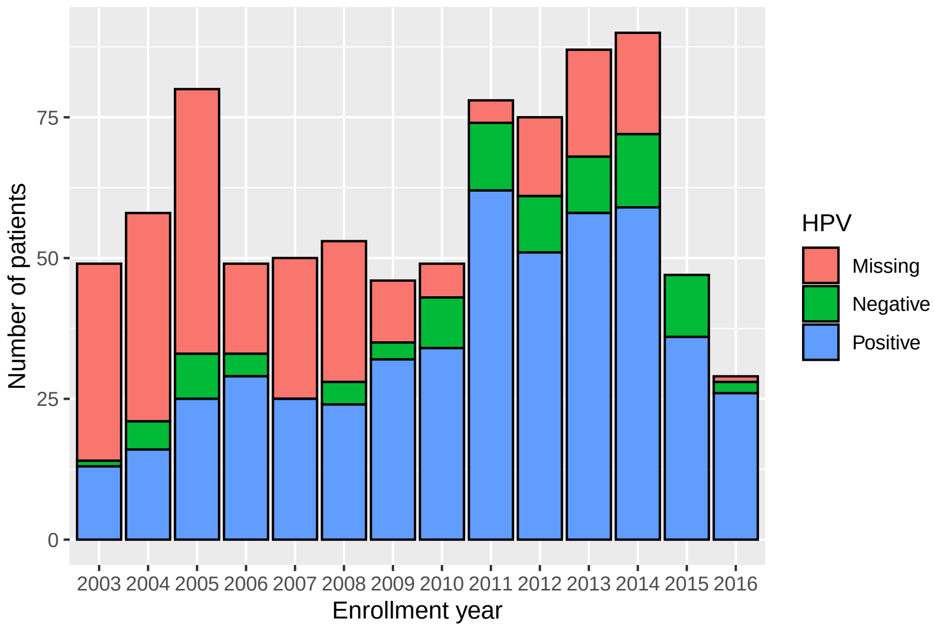


Table F.1: Logistic regression model estimates for the probability of observing HPV status [1]

Characteristic	log-odds ratio (95% CI)
Smoking	
Never	reference
Former (> 12 months)	-0.05 (-0.14, 0.03)
Current (< 12 months)	-0.13 (-0.21, -0.04)
ACE27 comorbidities	
None	reference
Mild	0.00 (-0.07, 0.09)
Moderate	0.04 (-0.07, 0.14)
Severe	-0.17 (-0.32, -0.03)
Age at diagnosis (decades)	-0.04 (-0.08, 0.00)
AJCC Cancer Stage	
I	reference
II	0.04 (-0.30, 0.37)
III	0.03 (-0.27, 0.34)
IV	0.08 -0.20, 0.37)
Year of enrollment (2012-2016)	0.06 (0.05, 0.07)

[1] among the $N = 612$ subjects with fully-observed data for comorbidities, smoking status, and cancer stage.

References

Paul Bernhardt. A comparison of stacked and pooled multiple imputation. *Joint Statistical Meetings Poster Presentation*, 2019.

James R Carpenter, Michael G Kenward, and Ian R White. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research*, 16(3):259–275, 2007.

Shahab Jolani. *Dual Imputation Strategies for Analyzing Incomplete Data*. PhD thesis, Utrecht University, 2012.

Bohdana Ratitch, Michael O Kelly, and Robert Tosiello. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, 12(6):337–347, 2013. doi: 10.1002/pst.1549.

Panteha Hayati Rezvan, Katherine J Lee, and Julie A Simpson. Sensitivity analysis within multiple imputation framework using delta-adjustment: Application to Longitudinal Study of Australian Children. *Longitudinal and Life Course Studies*, 9(3):259–278, 2018.

Daniel Tompsett, Stephen Sutton, Shaun R Seaman, and Ian R White. A general method for elicitation, imputation, and sensitivity analysis for incomplete repeated binary data. *Statistics in Medicine*, 39(22):2921–2935, 2020. doi: 10.1002/sim.8584.

Daniel Mark Tompsett, Finbarr Leacy, Margarita Moreno-Betancur, Jon Heron, and Ian R White. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. *Statistics in Medicine*, 37(15):2338–2353, 2018. doi: 10.1002/sim.7643.

Ian R White and Patrick Royston. Imputing missing covariate values for the Cox model. *Statistics in Medicine*, 28(15):1982–1998, 2009.