The International Journal of Biostatistics

Volume 4, *Issue* 1 2008 *Article* 13

Missing Confounding Data in Marginal Structural Models: A Comparison of Inverse Probability Weighting and Multiple Imputation

Erica E. M. Moodie* Joseph A.C. Delaney†

Geneviève Lefebvre‡ Robert W. Platt**

*McGill University, erica.moodie@mcgill.ca

†University of Washington, jacd@u.washington.edu

‡University of British Columbia, lefebvge@hotmail.com

**McGill University, robert.platt@mcgill.ca

Missing Confounding Data in Marginal Structural Models: A Comparison of Inverse Probability Weighting and Multiple Imputation*

Erica E. M. Moodie, Joseph A.C. Delaney, Geneviève Lefebvre, and Robert W. Platt

Abstract

Standard statistical analyses of observational data often exclude valuable information from individuals with incomplete measurements. This may lead to biased estimates of the treatment effect and loss of precision. The issue of missing data for inverse probability of treatment weighted estimation of marginal structural models (MSMs) has often been addressed, though little has been done to compare different missing data techniques in this relatively new method of analysis. We propose a method for systematically dealing with missingness in MSMs by treating missingness as a cause for censoring and weighting subjects by the inverse probability of missingness. We developed a series of simulations to systematically compare the effect of using case deletion, our inverse weighting approach, and multiple imputation in a MSM when there is missing information on an important confounder. We found that multiple imputation was slightly less biased and considerably less variable than the inverse probability approach. Thus, the lower variability achieved through multiple imputation makes it desirable in most practical cases where the missing data are strongly predicted by the available data. Inverse probability weighting is, however, a superior alternative to naive approaches such as complete-case analysis.

KEYWORDS: causal inference, marginal structural models, confounding, missing data, inverse probability weighting, double robustness, multiple imputation, simulations

---

*This research was supported by Discovery Grants to Robert Platt and Erica Moodie from the Natural Sciences and Engineering Research Council of Canada (NSERC) and core funding to the Montreal Children's Hospital from the Fonds de recherche en santé du Quebec. Robert Platt is a Chercheur-boursier of the Fonds de recherche en santé du Quebec. Erica Moodie is supported by a University Faculty Award from NSERC. Genevieve Lefebvre's research was supported by a Canada Graduate Scholarship from NSERC. The authors would like to thank the anonymous referees for their helpful comments and suggestions.

1 INTRODUCTION

In observational studies, estimation of the causal effect of a treatment may be biased either because of confounding, selection bias, or because of missing information. In longitudinal studies with a time-varying exposure, this is further complicated as time-varying covariates frequently act as both confounders and intermediate variables, and missing information is likely.

Marginal structural models (MSMs) have been proposed as an unbiased alternative to traditional regression models for estimating the causal effect of a time-varying exposure in the presence of time-varying confounding (Robins et al., 2000). MSMs are particularly useful for the analysis of observational drug data (Robins et al., 2000; Hernán et al., 2000). However, more commonly than not, some individuals in longitudinal studies have missing covariate information, and omission of an important variable from the probability of treatment model can lead to biased inference (Brumback et al., 2004). In general, inappropriate handling of the missing data in the analysis can lead to incorrect conclusions, either because of biased treatment effect estimates or reduced power (or both). Recently, a survey of the handling of missing data in the analyses of 63 randomized trials published in general medical journals found that 65% used complete-case analysis, while fewer than 4% used a more sophisticated approach such as multiple imputation (Wood et al., 2004).

Multiple imputation and inverse probability weighting are two approaches to handling missing data that provide unbiased estimates under relatively weak assumptions. A recent comparison of the methods in a cross-sectional setting found the performance of these methods to be similar, with multiple imputation only slightly more efficient than the inverse weighting (Carpenter et al., 2006).

We propose a method of accounting for missing confounding data that respects the assumed causal structure of the problem using inverse weighting. We use a series of Monte Carlo simulations on a problem of moderate complexity to compare different missing data techniques in MSMs. More specifically, we investigate the impact of varying levels of missingness of a confounding variable, and of different assumptions regarding the nature of missingness under three missing data strategies: complete-case analysis, our inverse probability weighting method, and multiple imputation. We allow missing data only in the confounding variable, as previous research has shown that this is the only variable without which the treatment model in the MSM

cannot be estimated without introducing bias in the estimate of treatment effect (Lefebvre et al., 2008). We consider only scenarios where time-dependent confounding is observed, as this is the context in which MSMs or other related approaches are necessary (Robins et al., 2000; Petersen et al., 2006).

This paper is organized as follows: In Section 2 we develop the problem and proposed approaches for missing data. MSMs are introduced in Section 2.1 and missing data mechanisms and strategies for dealing with missing information in analyses are discussed in Sections 2.2-2.4. Details of the simulation study can be found in Section 3, with results presented in Section 4. We applied the three methods considered in Section 3, namely complete-case, inverse probability weighting, and multiple imputation, to examine the effects of each method in a practical setting in Section 5. In the example, the effect of beta blocker use following an acute myocardial infarction on mortality is investigated using the General Practice Research Database. Section 6 discusses and concludes.

2 MARGINAL STRUCTURAL MODELS AND MISSING DATA

2.1 Marginal structural models

A marginal structural model (Robins et al., 2000) is a model for the marginal expectation of a counterfactual outcome under a specified static treatment regime. For example, if Y is a continuous outcome and A is a time-varying treatment, then an MSM is specified as

$$E[Y_{\overline{a}}] = f(\overline{a}) \tag{1}$$

where $\overline{a}$ refers to the history of treatment A and f is a defined function, typically a linear combination of components of $\overline{a}$. To estimate the parameters of a MSM, we first fit a model for the probability of receiving treatment. The treatment model is then used to weight individuals by the inverse probability of receiving the observed treatment (given history) in an unadjusted model for the outcome as a function of treatment. Stabilization of the inverse probability of treatment weights is commonly used to reduce the variability of MSM estimates that can arise when some combinations of covariates are rare (Sturmer et al., 2005). If the treatment model is correctly specified, the estimate of the marginal effect of treatment has a causal interpretation.

As with all models for observational data, MSMs require strong assumptions to be appropriately specified (Robins et al., 2000). Specifically, we require *no unmeasured confounding* (sequential randomization), *time ordering* (exposure precedes outcome) and *consistency* (Mortimer et al., 2005; Brumback et al., 2004). Also, we require that all treatments are possible (i.e., the probability of receiving treatment is neither zero nor one) given the covariates (the *experimental treatment assumption*) (Robins et al., 2000; Mortimer et al., 2005).

2.2 Missing data mechanisms

To draw meaningful inference from a study with missing data requires assumptions regarding the nature of the missingness. The strongest assumption that could be made is that the data are *missing completely at random* (MCAR), meaning that the probability of an observation being missing does not depend on any variables, observed or otherwise (Little and Rubin, 2002). A weaker assumption is that the data are *missing at random* (MAR); this requires that the probability of an observation being missing depends on observed covariates only, but that – conditional on the observed covariates – the probability does not depend on the true value of the unobserved variable (Little and Rubin, 2002). When the probability of an observation being missing does depend on its true, unobserved value or on the values of some other unmeasured variables, the data are said to be *not missing at random* (NMAR) or informatively missing.

As noted above, a common approach to missing information is to use only individuals with complete data ("complete-case" analysis). Complete-case analysis and mean imputation only produce unbiased parameter estimates if data are MCAR (Little and Rubin, 2002). Even when data are MCAR, complete-case analysis uses the available information inefficiently, which reduces power. Single imputation by the mean or the last observed measurement, on the other hand, produce standard errors that underestimate the variability of the estimates as they do not take into account the missing information.

To clarify the issues under study, assume that the data were generated from a structure as described by the directed acyclic graphs (DAGs) in Figures 1 and 2. A DAG provides a visual representation of the causal structure of a dataset, where an arrow (directed edge) from one variable (node) into another indicates that the first variable causes changes in the second (Pearl,

1995). DAGs are acyclic, that is, one cannot begin at a variable and follow the arrows through the graph to return to that variable; that is, information flows in one direction rather than in cycles. DAGs encode the conditional dependencies between variables and may be used to determine the variables on which to condition in order to achieve unbiased effect estimation (see, for example Greenland et al. (1999); VanderWeele and Robins (2007)).

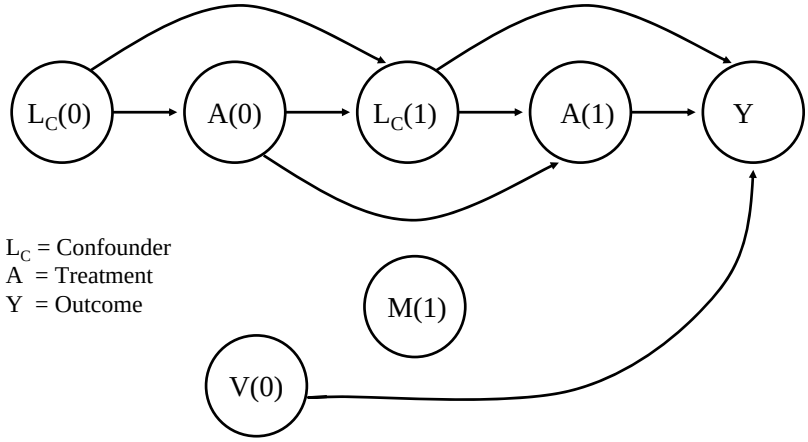

Figure 1: Directed acyclic graph (DAG) of the complete data under the missing completely at random assumption.

Specifically, we assume a simple longitudinal setting in which there are two discrete time points, $T = 0, 1$, and we consider a single confounding variable $L_C(T)$, treatment $A(T)$, and response Y. We use $V(0)$ to denote the predictor of the outcome, and $M(1)$ to denote an indicator of missingness for the confounder $L_C(1)$. Missing data are considered only in the confounding variable $L_C(1)$. We therefore consider the sequence $(L_C(0), A(0), V(0), M(1), L_C(1), A(1), Y)$. Figure 1 describes the situation when $L_C(1)$ is MCAR; that is, $M(1)$ is external to the causal system, and subjects with missing data can be seen as a random sub-sample of the data. Figure 2 describes the MAR case; here, missingness at time 1 is caused by $V(0)$, which is also a cause of outcome, and by $A(0)$. The data generation for the NMAR situation may be conceived as identical to the MAR case (Figure 2), but $V(0)$ is now unmeasured. This is a simplification of the more general missing data problem; however, in the context of causal inference it is helpful

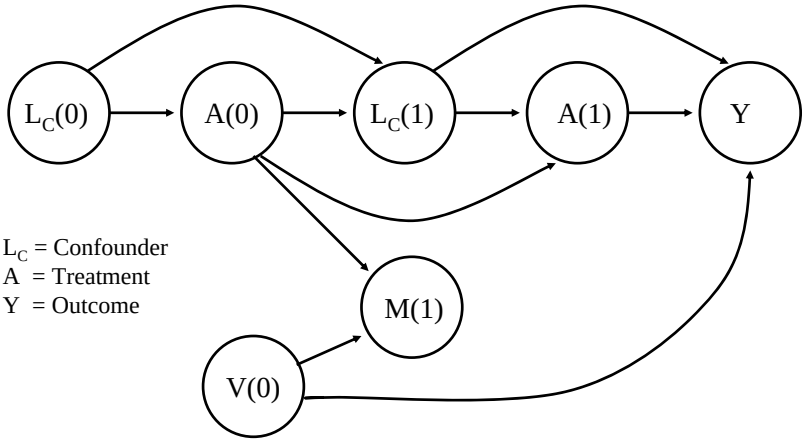

Figure 2: Directed acyclic graph (DAG) of the complete data under the missing at random ($V(0)$ observed) and the not missing at random assumptions ($V(0)$ unobserved).

to assume a causal structure for the missingness. This causal structure can be thought of as a form of selection bias (Hernán et al., 2004).

2.3 Missing data techniques: inverse probability weighting

Inverse probability of missingness weighting is not a new approach to missing data. However, it is only relatively recently that improvements to its efficiency (Robins and Rotnitzky, 1992; Robins et al., 1994) have brought it greater attention and utility. This technique is particularly natural to consider in a MSM setting, as it is similar to the weighting by the inverse probability of observed treatment that is performed when estimating parameters of MSMs.

In a simple regression setting, inverse probability weighting proceeds by calculating the probability of having complete data for each individual in the study and then performing a regression where the individual contributions are weighted by the inverse probability of having complete data (conditional on covariates). The probability that an individual observation is complete is typically estimated via a logistic regression model, which requires the data

to be MCAR or MAR.

Extending this idea to the longitudinal setting of MSMs, the probability of having complete data up to a given interval is calculated for each interval for each person, and a *full weight* is calculated, where the full weight is composed of the product of two weights, one which takes into account the probability of receiving the observed treatment and the other which accounts for the probability of having missing data. This approach is essentially the same as the use of inverse probability of censoring weighting considered, for example, by Hernán et al. (2000) and others. (See Bodnar et al. (2004) for a clear illustration of the technique in a two-interval example where information is lost due to drop-out or censoring.) These weights are then used in the MSM to, at least heuristically, attempt to approximate the results of an unblinded randomized controlled trial with no missing information (Cole et al., 2003; Hernán et al., 2001). This method also bears close analogy to an approach to dynamic treatment regimes using MSMs that was recently proposed by Hernán et al. (2006). To compare two dynamic regimes using MSMs, Hernán et al. proposed censoring subjects when they deviate from one of the two regimes. Our approach treats missing observations as deviating from one of the regimes under study, and censors at that time point.

We considered a second approach for the NMAR situation. When it was assumed that $V(0)$ was unmeasured but that $V(0)$ caused changes in Y, we used Y in the inverse probability weighting models as a proxy for the unmeasured $V(0)$. This approach may seem counter-intuitive, in that Y is being used to predict A although Y follows A temporally and causally; however, the outcome is frequently used to predict other variables in multiple imputation (Moons et al., 2006). Further, the approach seems less unusual when Y is thought of as a surrogate for (or strong correlate of) $V(0)$.

2.4 Missing data techniques: multiple imputation

Multiple imputation has been recognized as an attractive method for handling missing data, and has become more practical in the last several years as functions to perform the imputation in both cross-sectional and longitudinal settings have become more widely available in statistical packages (Schafer, 1999). Multiple imputation proceeds by generating m complete data-sets where missing values in the incomplete, observed data-set are filled, typically via a regression method. Each of the m data-sets is then analyzed using the same model and estimation method. The estimates from the m

analyses are then combined to produce a single estimate that incorporates the usual sampling variability as well as the variability due to the missing data (Rubin, 2004). Conventional wisdom suggests taking $m = 5$ (see, for example, Schafer (1999)), however with the increased power and speed of computers, there is often little to be lost by considering a larger number of completed data-sets.

Multiple imputation does not require data to be MAR (Schafer, 1999), as the imputation phase of the analysis is distinct from the analysis phase. However if data are NMAR, considerable subject-area knowledge is required to create a reasonable model for the distribution of the missing variables. Provided the model used to impute the missing values is correctly specified, estimates from an analysis using multiple imputation are consistent. Herein lies the greatest challenge of multiple imputation: the specification of the model to use for the data augmentation may be difficult, particularly if the data contain missing covariates of different types (e.g., Normal, skewed, discrete and so on). Much of the standard software for imputation is not well-suited to handle discrete data.

Other more sophisticated approaches to missing data such as the EM algorithm have many of the desirable properties of multiple imputation and inverse probability weighting (Dempster et al., 1977). However, these may be difficult to implement under general distributional assumptions for the observed data. Multiple imputation and inverse probability of missingness weighting are both very flexible can be used for virtually any statistical problem where MAR may be assumed.

3 SIMULATION STUDY

We developed a simulation study to investigate complete-case analysis, inverse probability weighting, and multiple imputation in MSMs over increasingly problematic scenarios of missingness, examining both the degree of missingness and each of the three missingness mechanisms.

Marginal structural models find their strengths in the analysis of repeated measures data, where variables may introduce confounding and be intermediate variables. When variables act in this dual fashion, the use of standard regression models may cause considerable bias in the estimated treatment effect (Blais et al., 1996).

Sample sizes of 250, 500, and 1000 were used. We consider analyses

where 10, 20, and 50% of the data were removed assuming MCAR, MAR, and NMAR. Thus, the three methods of analysis were employed for the three different sample sizes, in each of the nine possible degree and type of missingness pairs.

All simulations were performed in `R` version 2.31. Multiple imputation was carried out using the `mice` package (Van Buuren and Oudshoorn, 2000).

Simulation: data generating models

For the MCAR simulations, data were generated according to the causal structure in Figure 1 using the following conditional models:

$$\begin{aligned} L_C(0) &\sim \mathcal{N}(10,1) && (2)\\ A(0) &\sim \textit{Bernoulli}(p_{A(0)}) && (3)\\ V(0) &\sim \mathcal{N}(3,1) && (4)\\ L_C(1) &\sim \mathcal{N}(L_C(0)+\beta_0 A(0),1) && (5)\\ A(1) &\sim \textit{Bernoulli}(p_{A(1)}) && (6)\\ Y &\sim \mathcal{N}(L_C(1)+\beta_1 A(1)+6V(0),1) && (7) \end{aligned}$$

where the treatment effect parameters are $\beta_0 = -7.5$, $\beta_1 = -5.1$; the treatment probabilities are $p_{A(0)} = \texttt{expit}(-2.7{+}0.25L_C(0))$, $p_{A(1)} = \texttt{expit}(-2.7{+}.25L_C(1) + 0.1A(0))$. The missingness mechanism is completely at random, so that

$$M(1) \sim \textit{Bernoulli}(p_{NA})$$

with the fraction of missing information given by $p_{NA} \in \{0.1, 0.2, 0.5\}$.

For the MAR and NMAR simulations, data were generated according to the causal structure in Figure 2 using the conditional models (2)-(7), with the missingness mechanism now given by

$$p_{NA} = 1 - \texttt{expit}(\alpha - V(0) + A(0))$$

for $\alpha \in \{5.10, 4.15, 2.45\}$; the values of α were selected to give approximately 10, 20, and 50% missing data, respectively. In the NMAR setting, we assume $V(0)$ is unavailable to the analyst.

Simulation: data analysis models

The MSM requires models for the treatment mechanism at each time interval. For both MCAR and MAR simulations, the following mean models were

assumed in a pair of logistic regressions: (i) $A(0)$ depends on $L_C(0)$ and (ii) $A(1)$ depends on $L_C(0)$, $A(0)$, $V(0)$, and $L_C(1)$. Note that the model for $A(1)$ includes an additional variable, $V(0)$, that is a risk factor for the outcome but is not a cause of the treatment being modeled. This model conditions on the available past information and is therefore both a natural and a useful model to consider, as the inclusion of predictors of the outcome in the treatment model improves the accuracy of the treatment effect estimates (Lefebvre et al., 2008).

The inverse probability weighting approach to missing data for the MSM further requires a model for the missingness mechanism. In the MCAR and MAR settings, inverse probability weighting models the missingness mechanism via a logistic regression of $M(1)$ on $L_C(0)$, $A(0)$, and $V(0)$.

In the NMAR setting, where $V(0)$ is not available to the analyst, the most natural model to consider is that which conditions only on the past, that is, a logistic model regressing $M(1)$ on $L_C(0)$ and $A(0)$. However, if knowledge of the causal structure (the DAG) was available, the analyst might attempt to reduce the bias induced by data being missing not at random by including variables that are correlated with the missing covariate. For example, in examining Figure 2, we observe that $V(0)$ predicts Y. The response Y could, therefore, be considered a *surrogate* for $V(0)$. Simulations in which the probability of having complete data is fit as a logistic regression of $M(1)$ on $L_C(0)$, $A(0)$, and Y were also therefore considered in an additional set of NMAR simulations.

In contrast to the inverse probability weighting approach, multiple imputation requires the analyst to specify which variables are to be used as regressors in the imputation model. In cross-sectional settings, current research (Moons et al., 2006) for multiple imputation suggests using all available data (including the response) to predict the missing values. We take this approach, using $L_C(0)$, $A(0)$, $V(0)$, $A(1)$, and Y to impute $L_C(1)$. In the NMAR setting, $L_C(1)$ is imputed using a linear model which depends on $L_C(0)$, $A(0)$, $A(1)$, and Y.

4 SIMULATION RESULTS

In all tables, the bias (average deviation from the known, data-generating parameter value over the 1000 simulations) and the root mean squared error (rMSE, or the square root of the sum of the variance of the estimates over

the 1000 simulations and the squared bias) of the treatment effect estimates at both time points are used to summarize the performance of the three approaches to missing data.

Table 1 presents the results from the MCAR case. In this and subsequent tables, IPW-Y refers to the inverse probability weighting approach using Y in the weighting model. We note that all methods give essentially unbiased results; the rMSE is lowest for multiple imputation, with the other methods providing similar, slightly larger rMSEs.

Table 2 presents the results from the MAR case. Not surprisingly, the complete-case results show significant bias and larger rMSE than the other approaches. Inverse probability weighting gives slightly larger bias and rMSE than does multiple imputation, which is essentially unbiased.

Table 3 presents the results from the NMAR case. The complete-case results show significant bias in the estimate of the treatment effect of $A(0)$, but relatively little bias in the estimate of $A(1)$. Multiple imputation is essentially unbiased, and estimates are less variable than complete-case estimates. Not surprisingly, inverse probability weighting – whose missingness model is misspecified in that it does not include the (unobserved) predictor of missingness $V(0)$ – performs poorly. In fact, inverse probability weighting estimates are comparable to complete-case estimates, demonstrating considerable bias in the estimate of the treatment effect of $A(0)$, which predicts missingness, but little bias in the estimate of the effect of $A(1)$. The bias of treatment effect estimates that arises under NMAR settings using complete-case analyses and inverse probability weighting is exaggerated in instances where treatment predicts missingness. This observation is backed up by further simulations in which missingness occurred in the second interval rather than first and was predicted by $A(1)$ in place of $A(0)$ (results not shown).

We wished to further explore the possibility of achieving comparable results using inverse probability weighting and multiple imputation under NMAR. To that end, we performed another simulation using the same distributional assumptions as above, and included an additional variable, V^*, in the model which was simulated as a function of $V(0)$:

$$V^* \sim \mathcal{N}(2 + .85V(0), 1).$$

See Figure 3. Table 4 presents the results from the NMAR case. Including a surrogate for the unmeasured predictor of missingness that is external to

Table 1: Bias and root mean squared error (rMSE) of treatment effects estimated with full data (no missing values), a complete-case analysis (CC), multiple imputation (MI), and inverse probability weighting (IPW) under a missingness mechanism of data missing completely at random (MCAR). Summaries based on 1000 simulated data-sets, for sample sizes $n = 250, 500, 1000$ and the fraction of missing data equal to 10, 20, or 50%.

n	% NA	Full data		CC		MI		IPW	
		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
	$A(0)$								
250	10	0.008	0.852	-0.005	0.893	0.008	0.850	-0.002	0.894
	20	0.005	0.835	0.010	0.961	0.005	0.836	0.018	0.968
	50	0.038	0.824	0.069	1.202	0.038	0.819	0.079	1.217
500	10	0.027	0.577	0.025	0.611	0.028	0.577	0.027	0.611
	20	-0.020	0.568	-0.023	0.643	-0.019	0.566	-0.023	0.647
	50	0.017	0.586	0.011	0.856	0.018	0.586	0.016	0.866
1000	10	0.011	0.412	0.007	0.438	0.012	0.411	0.008	0.438
	20	0.005	0.407	0.008	0.454	0.005	0.408	0.011	0.457
	50	-0.020	0.401	-0.025	0.564	-0.020	0.401	-0.027	0.571
	$A(1)$								
250	10	-0.032	0.556	-0.028	0.588	-0.035	0.556	-0.032	0.586
	20	0.004	0.573	0.010	0.659	0.007	0.571	0.013	0.662
	50	-0.025	0.568	-0.013	0.906	-0.022	0.565	-0.025	0.926
500	10	-0.001	0.363	0.003	0.376	-0.002	0.360	0.003	0.377
	20	-0.007	0.373	-0.008	0.424	-0.007	0.370	-0.008	0.430
	50	-0.007	0.372	-0.028	0.595	-0.002	0.379	-0.027	0.596
1000	10	0.000	0.253	-0.003	0.268	-0.002	0.253	-0.002	0.269
	20	-0.012	0.243	-0.015	0.279	-0.014	0.243	-0.016	0.279
	50	0.005	0.253	0.013	0.367	0.003	0.257	0.015	0.368

Table 2: Bias and root mean squared error (rMSE) of treatment effects estimated with a complete-case analysis (CC), multiple imputation (MI), and inverse probability weighting (IPW) under a missingness mechanism of data missing at random (MAR). Summaries based on 1000 simulated datasets, for sample sizes $n = 250, 500, 1000$ and the fraction of missing data approximately equal to 10, 20, or 50%.

n	% NA	CC		MI		IPW	
	(mean)	Bias	rMSE	Bias	rMSE	Bias	rMSE
	$A(0)$						
250	9.9	0.413	0.949	0.003	0.834	0.011	0.963
	20.1	0.666	1.094	0.007	0.826	0.022	1.033
	50.0	1.035	1.545	0.015	0.834	0.203	1.746
500	10.0	0.422	0.749	0.004	0.606	0.001	0.679
	20.2	0.674	0.898	0.012	0.557	0.033	0.701
	50.0	1.048	1.328	0.000	0.569	0.061	1.314
1000	10.0	0.442	0.604	0.021	0.406	0.028	0.448
	20.1	0.663	0.794	0.002	0.409	-0.001	0.526
	50.0	1.003	1.140	-0.009	0.399	-0.014	0.913
	$A(1)$						
250	9.9	-0.003	0.626	0.008	0.592	0.020	0.670
	20.1	-0.005	0.651	-0.001	0.583	0.021	0.763
	50.0	-0.008	1.000	-0.019	0.635	0.081	1.480
500	10.0	0.013	0.393	0.009	0.375	0.023	0.418
	20.2	-0.004	0.415	0.002	0.382	-0.011	0.501
	50.0	-0.026	0.543	-0.011	0.370	0.005	0.983
1000	10.0	-0.001	0.254	-0.008	0.249	0.002	0.266
	20.1	-0.016	0.266	-0.008	0.248	-0.005	0.319
	50.0	-0.011	0.362	0.007	0.259	-0.010	0.742

Table 3: Bias and root mean squared error (rMSE) of treatment effects estimated with a complete-case analysis (CC), multiple imputation (MI), inverse probability weighting (IPW), and inverse probability weighting using response Y as a surrogate for the unobserved variable $V(0)$ in the model for missingness (IPW-Y) under a missingness mechanism of data missing not at random (NMAR). Summaries based on 1000 simulated data-sets, for sample sizes $n = 250, 500, 1000$ and the fraction of missing data approximately equal to 10, 20, or 50%.

n	% NA	CC		MI		IPW		IPW-Y	
		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
	$A(0)$								
250	10.0	0.418	0.969	0.004	0.856	0.425	0.972	0.186	0.943
	20.0	0.635	1.145	-0.027	0.871	0.635	1.144	0.247	1.057
	49.9	1.006	1.558	-0.007	0.890	1.001	1.557	0.387	1.704
500	10.0	0.462	0.784	0.032	0.611	0.464	0.786	0.228	0.709
	20.2	0.649	0.920	-0.017	0.608	0.649	0.921	0.258	0.765
	49.9	1.036	1.324	-0.006	0.608	1.041	1.325	0.450	1.242
1000	10.0	0.387	0.586	-0.019	0.429	0.386	0.586	0.141	0.478
	20.2	0.672	0.823	0.000	0.446	0.674	0.826	0.291	0.598
	50.0	1.034	1.197	-0.011	0.432	1.036	1.202	0.422	0.944
	$A(1)$								
250	10.0	0.011	1.097	0.036	1.059	0.010	1.079	-0.288	1.192
	20.0	0.067	1.205	0.090	1.088	0.071	1.163	-0.410	1.367
	49.9	0.056	1.502	0.067	1.134	0.065	1.418	-0.519	1.926
500	10.0	-0.015	0.759	-0.004	0.742	-0.016	0.747	-0.325	0.861
	20.2	0.007	0.844	0.032	0.767	0.005	0.817	-0.485	1.037
	49.9	-0.018	1.004	0.013	0.755	-0.024	0.937	-0.736	1.529
1000	10.0	0.035	0.548	0.047	0.531	0.036	0.538	-0.257	0.623
	20.2	0.020	0.556	0.039	0.536	0.021	0.537	-0.477	0.774
	50.0	-0.022	0.737	0.032	0.543	-0.017	0.695	-0.691	1.195

the causal pathway of treatment, covariates, and response in the imputation model makes very little difference to the estimates obtained after correcting for missing data by multiple imputation. In inverse probability weighting, on the other hand, including a surrogate of a predictor of missingness reduces bias of treatment effect estimates with either effectively no change or a slight increase in efficiency of the estimates, as evidenced by the comparable or smaller rMSEs.

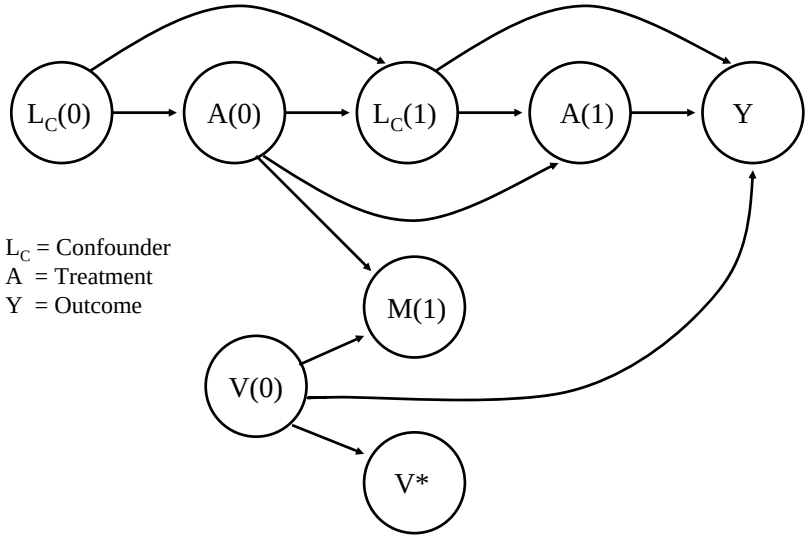

Figure 3: Directed acyclic graph (DAG) of the complete data under the not missing at random assumption ($V(0)$ unobserved, but a correlate of $V(0)$, V^*, is observed).

Table 4: Bias and root mean squared error (rMSE) of treatment effects under multiple imputation (MI), multiple imputation (MI-V^*) using V^* in the imputation model, inverse probability weighting (IPW), and inverse probability weighting using V^* as a surrogate for the unobserved variable $V(0)$ in the model for missingness (IPW-V^*) under a missingness mechanism of data missing not at random (NMAR). Summaries based on 1000 simulated data-sets, for sample sizes $n = 250, 500, 1000$ and the fraction of missing data approximately equal to 10, 20, or 50%.

n	% NA	MI		MI-V^*		IPW		IPW-V^*	
		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
	$A(0)$								
250	10.0	-0.027	0.887	-0.026	0.885	0.395	0.992	0.229	0.945
	20.0	0.007	0.860	0.009	0.860	0.683	1.167	0.427	1.056
	50.1	-0.044	0.894	-0.041	0.897	0.967	1.536	0.633	1.459
500	10.0	-0.027	0.640	-0.027	0.640	0.373	0.761	0.209	0.704
	20.1	-0.025	0.608	-0.023	0.608	0.629	0.907	0.379	0.770
	49.9	-0.006	0.589	-0.004	0.588	1.000	1.293	0.628	1.109
1000	10.0	0.024	0.523	0.023	0.524	0.004	0.529	0.003	0.536
	20.1	0.030	0.514	0.026	0.513	-0.004	0.546	-0.001	0.557
	50.1	0.077	0.534	0.073	0.535	0.018	0.656	0.015	0.711
	$A(1)$								
250	10.0	0.037	1.077	0.035	1.074	0.007	1.096	0.010	1.110
	20.0	0.050	1.033	0.045	1.031	0.021	1.099	0.021	1.122
	50.1	0.091	1.146	0.087	1.143	0.049	1.418	0.065	1.519
500	10.0	0.037	0.771	0.037	0.770	0.024	0.790	0.027	0.797
	20.1	0.051	0.754	0.050	0.753	0.029	0.782	0.033	0.810
	49.9	0.043	0.752	0.042	0.749	-0.024	0.954	-0.011	1.046
1000	10.0	-0.019	0.433	-0.019	0.433	0.407	0.596	0.244	0.504
	20.1	-0.005	0.431	-0.004	0.432	0.669	0.819	0.410	0.635
	50.1	-0.037	0.450	-0.036	0.449	1.000	1.171	0.637	0.916

5 EXAMPLE: MORTALITY AFTER ACUTE MYOCARDIAL INFARCTION

We employed the three methods of handling missing data considered in the simulations of the previous section, namely complete-case, inverse probability

weighting, and multiple imputation, to investigate the effect of beta blocker use after an acute myocardial infarction on mortality using the General Practice Research Database (GPRD).

5.1 Methods

The GPRD is a large clinical database based on information generated from general practices in the United Kingdom. The GPRD has been previously validated for studies of pharmacological effects on blood pressure (Delaney et al., 2008). Using this database, a cohort was constructed which included all members of the database who survived a first acute myocardial infarction between 1 January, 2002, and 31 December, 2004. Cohort inclusions criteria were 90-day survival post myocardial infarction (see Zhou et al. (2005) for a discussion of this approach to cohort selection criteria), age of 20 years or older, and participation in the GPRD for at least three years to ensure adequate time for data collection.

The treatment variables were defined as: $A(0)$ was exposure to beta blockers in the 45 days after an acute myocardial infarction and $A(1)$ was exposure to beta blockers in following 45 days, i.e. 46 to 90 days after myocardial infarction. The outcome of interest, Y, was all-cause mortality between 90 days and one year after the acute myocardial infarction. Subjects were followed for one year after myocardial infarction or until the occurrence of the outcome (death).

Blood pressure is a time-varying confounder, L_C, that may mediate the relationship between treatment after myocardial infarction and the outcome, death. Both systolic and diastolic blood pressure were considered. Due to the irregular recording of blood pressure in the GPRD, it was not possible to obtain blood pressure measurements immediately after acute myocardial infarction or at exactly 45 days after recovery from the myocardial infarction. Instead, we approximate these values with $L_C(0)$, the mean of all blood pressure recordings in the 90 days prior to myocardial infarction and $L_C(1)$, the mean of all blood pressure recordings in the 45 days following the myocardial infarction. Missing data are common, and many participants did not have any blood pressure readings in either interval (Delaney et al., 2008). The postulated DAG for this GPRD example is given in Figure 4.

The ratios of the odds of death associated with beta blocker use were estimated via a marginal structural model, using three methods of accounting for missing data: complete-case, multiple imputation, and inverse probability

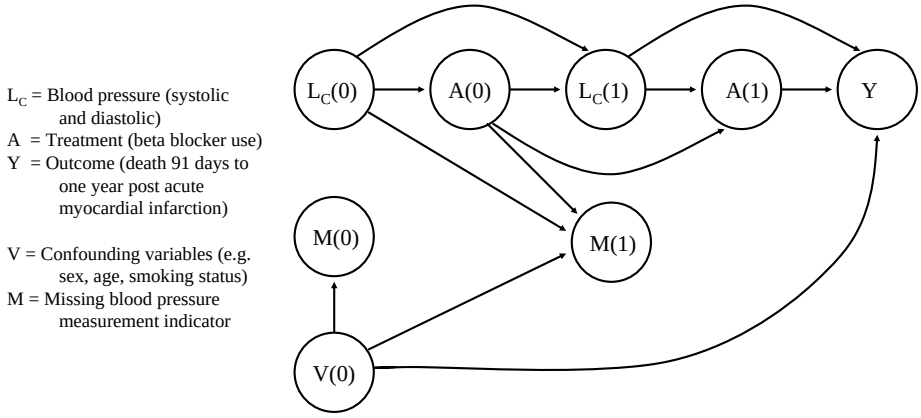

Figure 4: Postulated Directed acyclic graph (DAG) of the GPRD data under.

weighting. Baseline risk factors for all-cause mortality, such as age, sex, smoking, alcohol use, obesity, number of hospitalizations in the past year and serious medical conditions (e.g. cancer and respiratory disease), were included in the probability of treatment models. Similarly, models that were rich in baseline covariates were used to estimate the probability of having missing blood pressure data for the inverse probability weighting approach, and to model blood pressure itself in multiple imputation. Confidence intervals were calculated via bootstrap using 1000 resamples of the data.

5.2 Results

There were 7749 individuals who met the cohort inclusion criteria and, among these, 469 deaths. After myocardial infarction, 600 individuals used beta blockers within 45 days of the acute myocardial infarction but not after that ($A(0) = 1$, $A(1) = 0$), 605 individuals used beta blockers only in the second interval ($A(0) = 0$, $A(1) = 1$), and 3582 individuals used beta blockers in both intervals ($A(0) = A(1) = 1$).

There was considerable missing information on blood pressure: 2489 (32%) had no blood pressure measurements recorded, while a further 2497 (32%) and 1270 (16%) had missing values exclusively at the first or second interval, respectively. The resulting estimates of the effect of beta blocker treatment on all-cause mortality are presented in Table 5.

Table 5: Estimates of the effect of treatment with beta blockers on all cause mortality (Y) as found by marginal structural modelling (estimating via inverse probability of treatment weighting). Missing blood pressure data are accounted for by three methods: complete-case, inverse probability weighting, and multiple imputation.

Missing data approach	Rate ratio	95% CI
Effect estimates of $A(0)$		
Complete-case	1.02	0.66 to 2.48
Inverse probability weighting	0.88	0.54 to 2.12
Multiple imputation	0.68	0.59 to 1.15
Effect estimates of $A(1)$		
Complete-case	0.92	0.39 to 1.67
Inverse probability weighting	0.79	0.34 to 1.41
Multiple imputation	0.73	0.49 to 1.02
Post-MI treatment with beta blockers		
RCT meta-analysis(Freemantle et al., 1999)	0.77	0.69 to 0.85

The average causal effect of treatment with beta blockers after acute myocardial infarction is well-studied; an estimate derived from a meta-analysis of randomized controlled trials conducted over a (single) similar time period (Freemantle et al., 1999) is used as a point of comparison for the average causal effect of treatment post-myocardial infarction. The trials included in the meta-analysis were conducted in very broad post-myocardial infarction populations and, therefore, the meta-analysis average effect was estimated in a population that was likely relevant and comparable to the individuals captured by the GPRD.

Inverse probability weighting and multiple imputation both provide estimates that are closer to the randomized trial meta-analysis estimates than complete-case approach, which yields estimates of the effect of the drug on all-cause mortality that are closer to the null value of 1.0. There appears to be little difference in this example between accounting for missing data with inverse probability weighting versus multiple imputation in terms of the unbiasedness of the estimates, however the multiple imputation approach gave narrower confidence intervals.

6 DISCUSSION

Accounting for missing information on key variables is an important component of all observational analyses. Both inverse probability weighting and multiple imputation are readily implemented in standard statistical software packages (such as `R` and SAS) and yield improved estimates over naive approaches (such as complete case analyses) in the context of marginal structural model analyses.

The need to carefully account for missing data in order to prevent biased estimates in epidemiologic studies is well known (Wood et al., 2004; Schafer, 1999). However, previous investigations into the consequences of improper handling of missing data have not considered the case of marginal structural models. Marginal structural models rely on different assumptions than standard regression models (Robins et al., 2000) and it is not clear how analytic approaches developed for the latter will perform in the context of the former.

We have demonstrated the use of multiple imputation and inverse weighting to address missing data in the context of marginal structural models. When the missingness is completely at random, all methods give similar results. Unfortunately, it is seldom the case that data are MCAR. However, when the mechanism for data missingness is not completely at random and yet is well understood – i.e., when data are MAR – multiple imputation was slightly less biased and considerably less variable than the inverse probability approach. Thus, the lower variability achieved through multiple imputation makes it desirable in most practical cases.

In the simulations considered in this study, baseline measures of the confounding variables were available to the analyst. This may have provided multiple imputation with an advantage over inverse probability weighting as the baseline confounder was predictive of the missing variable. The goal of multiple imputation is to model the missing value while inverse probability weighting focuses on predicting the missing data mechanism (Schafer and Graham, 2002). In other situations, such as those where strong predictors of the missing values are not present or the missingness mechanism is well-understood, inverse probability weighting may perform better relative to multiple imputation. In cases where the missingness mechanism is not well-understood or is unlikely to be strongly predicted by the available covariates, a doubly-robust procedure (Bang and Robins, 2005) that does not require correct modelling of the missing data mechanism (and requires no modelling of the missing values themselves) may be preferable.

It is important to note that our simulations are based on a small number of time points and a single, non-time-varying outcome. This is not an unrealistic setting, as many applications of marginal structural models occur in a two-interval context. However, when the number of observation periods is substantial, the potential for feedback between the intermediate/confounding variables and the exposure may be more important and our results may not generalize to these cases. It may, for example, be difficult to correctly model a covariate that is missing for a number of intervals, as is required by multiple imputation. On the other hand, a small number of data points that are missing at early intervals may lead to considerable loss of information since the inverse probability weighting approach requires censoring of individuals at the first instance of a missing value.

Inverse probability weighting has a natural and intuitive appeal in the context of marginal structural models, where one attempts to make valid inference from complex data without requiring full specification of the likelihood of the data. However our results suggest that in cases where the missing data are strongly predicted by the available data, inverse weighting cannot be recommended over multiple imputation. Nevertheless, inverse probability weighting remains a superior alternative to more naive approaches to accounting for the presence of missing data.

References

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics 61*, 926–972.

Blais, L., P. Ernst, and S. Suissa (1996). Confounding by indication and channeling over time: The risks of beta 2-agonists. *American Journal of Epidemiology 144*, 1161–1169.

Bodnar, L. M., M. Davidian, A. M. Siega-Riz, and A. Tsiatis (2004). Marginal structural models for analyzing causal effects of time-dependent treatments: An application in perinatal epidemiology. *American Journal of Epidemiology 159*, 926–934.

Brumback, B., M. A. Hernán, J. M. Robins, K. Anastos, J. Chmiel, R. Detels, and *et al.* (2004). Sensitivity analyses for unmeasured confounding

assuming a marginal structural model for repeated measures. *Statistics in Medicine 23*, 749–767.

Carpenter, J., M. Kenwood, and S. Versteelandt (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A 169*, 571–584.

Cole, S. R., M. A. Hernán, J. M. Robins, K. Anastos, J. Chmiel, R. Detels, and *et al.* (2003). Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *American Journal of Epidemiology 158*, 687–694.

Delaney, J. A. C., E. E. M. Moodie, and S. Suissa (2008). Validating the effects of drug treatment on blood pressure in the General Practice Research Database. *Pharmacoepidemiology and Drug Safety 17*, 535–545.

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

Freemantle, N., J. Cleland, P. Young, J. Mason, and J. Harrison (1999). β-blockade after myocardial infarction: Systematic review and meta regression analysis. *British Medical Journal 318*, 1730–1737.

Greenland, S., J. Pearl, and J. M. Robins (1999). Causal diagrams for epidemiologic research. *Epidemiology 10*, 37–48.

Hernán, M. A., B. Brumback, and J. M. Robins (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology 11*, 561–70.

Hernán, M. A., B. Brumback, and J. M. Robins (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association 96*, 440–408.

Hernán, M. A., S. Hernández-Díaz, and J. M. Robins (2004). A structural approach to selection bias. *Epidemiology 15*, 615–625.

Hernán, M. A., E. Lanoy, D. Costagliola, and J. M. Robins (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology 98*, 237–242.

Lefebvre, G., J. A. C. Delaney, and R. W. Platt (2008). Impact of misspecification of the treatment model on estimates from a marginal structural model. *Statistics in Medicine*, DOI 10.1002/sim.3200.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. Wiley-Interscience.

Moons, K. G. M., R. A. R. T. Dondersa, T. Stijnenc, and F. E. Harrell (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology 59*, 1092–1101.

Mortimer, K. M., R. Neugebauer, M. J. van der Laan, and I. B. Tager (2005). An application of model-fitting procedures for marginal structural models. *American Journal of Epidemiology 162*, 382–308.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*, 669–688.

Petersen, M. L., S. E. Sinisi, and M. J. van der Laan (2006). Estimation of direct causal effects. *Epidemiology 17*, 276–284.

Robins, J. M., M. A. Hernán, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*, 550–60.

Robins, J. M. and A. Rotnitzky (1992). *AIDS Epidemiology: Methodological Issues*, Chapter Recovery of information and adjustment for dependent censoring using surrogate markers, pp. 846–866. Boston, MA: Birkhäuser.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*, 846–866.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. Wiley-Interscience.

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

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

Sturmer, T., S. Schneeweiss, M. A. Brookhart, K. J. Rothman, J. Avorn, and R. J. Glynn (2005). Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: Nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *American Journal of Epidemiology 161*, 891–808.

Van Buuren, S. and C. G. M. Oudshoorn (2000). *Multivariate Imputation by Chained Equations: MICE V1.0 User's manual*. TNO Prevention and Health.

VanderWeele, T. J. and J. M. Robins (2007). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *American Journal of Epidemiology 166*, 1096–1104.

Wood, A., I. White, and S. Thompson (2004). Are missing outcome data adequately handled? A review of published randomised controlled trials in major medical journals. *Clinical Trials 1*, 368–376.

Zhou, Z., E. Rahme, M. Abrahamowicz, and L. Pilote (2005). Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: A comparison of methods. *American Journal of Epidemiology 162*, 1016–1023.