Common Methods for Missing Data in Marginal Structural Models: What Works and Why

Clémence Leyrat[1*], James R. Carpenter[1,2], Sébastien Bailly[3], and Elizabeth J. Williamson[1,4]

Correspondence to Dr Clémence Leyrat, Department in Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, UK.

Phone : +44 (0)207 927 2169 ;

Fax : +44 (0)20 7436 5389 ;

e-mail: clemence.leyrat@lshtm.ac.uk

Author affiliations:

London School of Hygiene and Tropical Medicine, Department of Medical Statistics, London, UK (Clémence Leyrat, James R. Carpenter, and Elizabeth J. Williamson); London Hub for Trials Methodology Research, MRC Clinical Trials Unit, UCL, London, UK (James R Carpenter); HP2, Inserm 1042, Université Grenoble Alpes, Grenoble, France (Sébastien Bailly); Health Data Research London site, UK (Elizabeth J. Williamson)

Funding: This work was supported by the UK Medical Research Council (Project Grant MR/M013278/1 (Dr Elizabeth J. Williamson), and MC UU 12023/21 and MC UU 12023/29 (Prof James R. Carpenter)). Dr Sébastien Bailly is supported by the "e-health and integrated care" Chair of excellence of the University Grenoble Alpes Foundation and the ANR 3IA MIAI@Grenoble Alpes project.

Conflict of interest: none declared

Running head: Missing data in Marginal Structural Models

Abstract

Marginal structural models (MSMs) are commonly used to estimate causal intervention effects in longitudinal non-randomised studies. A common challenge when using MSMs to analyse observational studies is incomplete confounder data, where a poorly informed analysis method will lead to biased intervention effect estimates. Despite a number of approaches described in the literature to handle missing data in MSMs, there is little guidance on what works in practice and why. We reviewed existing missing data methods for MSMs and discussed the plausibility of their underlying assumptions. We also performed realistic simulations to quantify the bias of five methods used in practice: complete case analysis, the last observation carried forward, the missingness pattern approach, multiple imputation and inverse-probability-of-missingness weighting. We considered three mechanisms for non-monotone missing data encountered in electronic health record data research. Further illustration of the strengths and limitations of these analysis methods are provided through an application using a cohort of individuals with sleep apnoea, the research database of the French "Observatoire Sommeil de la Fédération de Pneumologie" (OSFP). We recommend a careful consideration of (i) the reasons for missingness, (ii) whether missingness modifies the existing relationships among observed data and (iii) the scientific context and data source, to inform the choice of the appropriate method(s) to handle partially observed confounders in MSMs.

Keywords: time-varying confounding; propensity score; multiple imputation; inverse probability weighting; missingness pattern approach; last observation carried forward; complete cases.

List of abbreviations: CC: Complete Case; CPAP: Continuous Positive Airway Pressure; IPMW: Inverse Probability of Missingness Weighting; LOCF: Last Observation Carried Forward; MAR: Missing At Random; MCAR: Missing Completely At Random; MNAR: Missing Not At Random; MSM: Marginal Structural Model; MI: Multiple Imputation; MPA: Missingness Pattern Approach;

BACKGROUND

Although randomised trials are the gold standard to establish causal effects of treatments and non-pharmacological interventions on health outcomes, observational data are increasingly used for causal inference (1). The enormous potential offered by the wealth of routinely collected medical data available and the need for real-world evidence to assess the efficacy and safety of treatments have contributed to this phenomenon.

These routinely collected data typically have a longitudinal structure, following individuals over time, allowing the measurement of dynamic treatment patterns, including treatment switching or delay to treatment initiation. Patients with chronic conditions often have a non-linear treatment history: treatment prescription might be updated based on the occurrence of new health events, changes in individual factors or side-effects induced by previous treatments. The newly prescribed treatment might, in turn, affect future health events and individual factors, themselves potentially associated with the outcome of interest (2). In such settings, specific statistical methods are required to account for confounding bias induced by time-varying variables (3). Indeed, adjusting for the confounders and treatment history is not sufficient, and often leads to biased estimates of the causal treatment effects (4). This is because the effect of a treatment received at a specific time on the outcome is mediated by subsequent treatments. Propensity scores - the individual probabilities of receiving the treatment of interest conditionally on individual characteristics - have been extended to situations with time-varying treatment and confounders (5), with scores estimated at each time point (6). The cumulative product of the inverse of these scores over time can be used as a weight to account for confounding in the

estimation of the treatment effect in a marginal structural model (MSM) (7). This method of adjusting for time-varying confounders is by far the most common in practice (8).

A challenge when analysing observational data is incomplete confounder information. In routinely collected data, missingness is particularly prevalent on covariates. This can happen if some information is not recorded at a given time point or the frequency of the measurement varies from one patient to another (e.g. asynchronous medical visits (9)). This might jeopardize the validity of the results if the issue is ignored in the analysis, depending on the underlying missingness mechanisms. In practice, despite the STROBE recommendations to report the amount of missing data and the way they are handled (10) in observational studies, reporting is often suboptimal. A review of reporting of missing exposure data in a longitudinal cohort showed that 43% of identified publications adhered to these guidelines (11). Importantly, when the method for handling missing data was reported, it was often done using inadequate methods. Although several methods to handle missing data on covariates have been used in the context of time-varying exposures (11), the most common approaches – complete case (CC) analysis and last observation carried forward (LOCF) – have been criticized. Use of more complex approaches such as multiple imputation (MI) or inverse-probability of-missingness-weighting (IPMW) have been suggested, but their performance is yet to be fully explored. Another promising approach, which has not been extended to the context of MSMs, is the missingness pattern approach (MPA).

To our knowledge, there are no published guidelines for the choice of methods to handle missing confounder data in MSMs. Published studies on missing data in MSMs focused on missing data in the exposure (11,12) or compared the performances of a few methods only (13,14). Moodie et al. (14) compared the use of IPMW and MI, finding that MI outperformed IPMW, but they did not investigate the performance of MPA and LOCF. Moreover, only one covariate and two time points were considered, limiting the generalisability of the results. Vourli and Touloumi (15) investigated the performance of MI, IPMW and LOCF but found opposite conclusions in their setting, finding that IPMW usually performed better than MI. This might be explained by the omission of the outcome from the imputation model. A

recent plasmode simulation (13) suggested superiority of MI over IPMW, but surprisingly, CC analysis was the least biased. A limitation of these published studies is their focus on missingness mechanisms described under Rubin's taxonomy of missing data (16); this taxonomy may be too restrictive to describe complex missingness scenarios encountered in routinely collected data (17).

The aims of this paper are, first, to provide an overview of existing methods to handle missing data on confounders in MSMs and, second, to recommend practical guidelines. These guidelines will rely on the understanding of the assumptions and missingness mechanisms under which these methods are valid. We focus on situations where some variables are not recorded during the visit, rather than missing data introduced because of sparse follow-up. These two scenarios differ both in terms of underlying missingness mechanisms and required statistical methods. The challenges of sparse follow-up has been discussed by Mojaverian *et al.* (9) and Kreif *et al.* (18). We present a simulation study comparing the performance of CC analysis, LOCF, MI, IPMW and MPA to handle partially observed confounders under common missingness mechanisms encountered in observational studies. Finally, we illustrate the implementation of these methods by investigating the impact of observance to treatment on sleepiness in patients with sleep apnoea.

CAUSAL INFERENCE IN THE PRESENCE OF TIME-VARYING TREATMENT AND CONFOUNDERS

When time-varying confounding occurs, standard regression approaches fail because of treatment-confounder feedback (19), even when past treatment and confounders values are adjusted for (3). MSMs were developed (7) to estimate causal effects in this setting. MSMs rely on an extension of inverse-probability-of-treatment weighting, a propensity score approach, for multiple time points. Details about this framework and underlying assumptions for a single time point are in Web Appendix 1. Similar to propensity score approaches, MSMs are a two-stage process. In the first stage, (19), weights – based on the inverse of the probability of a patient receiving the treatment they actually received – are

estimated to create a pseudo-population in which treatment and confounders are independent. In the second stage, a weighted regression (using the weights derived in the first stage) including only the treatment history can be used to obtain estimate the causal effect of the treatment regimens of interest. Under the assumptions of no interference, consistency, exchangeability and positivity extended to time-varying settings, and assuming the model used to obtain the weights is correctly specified, MSMs lead to unbiased estimates of the marginal causal effect of the treatment regimen.

In practice, the weights can be estimated using pooled logistic regression (6), in which each person-time interval is considered as an observation. This pooled logistic regression model must include the confounders and their relevant interactions to ensure the distributions of confounders are balanced between treatment groups in the weighted pseudo-population at each time point. Further details about the implementation of MSMs can be found in Web Appendix 2.

MISSING DATA IN MSMs: mechanisms and methods

The choice of an appropriate missing data method relies on the characterization of the missingness patterns and the missingness mechanisms. The missingness patterns simply define which values of the covariates are observed and which are missing. For example, if the dataset contains only two time-fixed covariates L1 and L2, there are four missing data patterns: L1 and L2 can be both observed, both missing, L1 can be observed and L2 missing, or L1 can be missing and L2 observed. Similarly, if L1 and L2 are measured at two time-points, there are 16 patterns. Some methods for handling missing data, such as multiple imputation, apply to any pattern of missing data; other methods apply only to specific structures of missing data, the most common being the monotone missing data pattern. In longitudinal data, monotone missingness patterns occur when, once a patient has a missing observation at one time point, values for all subsequent time points are also missing. This is typically what happens when patients are lost to follow-up. Whereas methods based on inverse weighting have been proposed to address this type of missing data in MSMs, there is no guidance on how to handle arbitrary patterns (not monotone) in MSMs. This is however the most common pattern found in routinely collected data where

data are not collected for research purposes, and the quality of recording may vary from one visit to another.

Little and Rubin's classification is often used to classify the missing data as being (i) missing completely at random (MCAR) when the probability of data being missing does not depend on the observed or unobserved data, (ii) missing at random (MAR) if the probability of data being missing does not depend on the unobserved data, conditional on the observed data or (iii) missing not at random (MNAR) if the probability of data being missing depends on the unobserved data, even after conditioning on the observed data (20).

A variety of methods to handle missing data on covariates have been used in the context of time-varying exposures (11). The most common approach is complete case (CC) analysis, in which only patients with a complete record for all the covariates are included in the analysis. Another simple and popular approach is the last observation carried forward (LOCF): when a measurement is missing, the most recent past value observed for this patient is used to impute the missing value. Multiple imputation (MI) uses relationships existing among the observed variables to draw multiple times plausible values for the missing data; the standard error of the treatment effect estimates accounts for the uncertainty in these predictions. In MSMs, Robins and Hernán proposed to use censoring weights to account for patients lost to follow-up. Complete cases are re-weighted by the inverse of their probability of remaining in the study. Loss to follow-up can be viewed as a missing data problem, and therefore, these weights can be accommodated to account for missing data. This method is called inverse-probability of-missingness-weighting (IPMW). Another promising approach to handle partially observed confounders that should be extended to MSMs is the missingness pattern approach (MPA). The MPA has been proposed for the estimation of propensity score weights in studies with a single time point (17). The sample is split into subgroups of patients having missing information on the same set of covariates and the weights are derived in each subgroup from the covariates available in that pattern. More details of these methods are in Web Appendix 3. The approaches rely on different assumptions; their validity depends on the

missingness mechanisms in the data at hand. These assumptions along with the strengths and limitations of each method are summarised in Table 1.

METHODS

We performed a simulation study to (i) illustrate the impact on bias of violations of the assumptions required for each method to be valid, and the relative precision of these methods when assumptions hold and (ii) highlight existing challenges in their implementation in practice. Data were simulated to mimic an observational study looking at the effect of a time-varying binary treatment on a continuous outcome, in the presence of time-varying confounding. We focused on four plausible types of missingness mechanisms (Figure 1):

- MCAR mechanism: missingness is not dependent on either observed or unobserved variables
- MAR mechanism: we consider 3 situations in which missing data depends on observed past treatment and confounder values (MAR|A,L), past treatment and confounder values and outcome (MAR|A,L,Y) or considering an association between missingness and the outcome introduced through the independent risk factor (MAR|A,L,V).
- "Constant" mechanism: confounder values are missing if they have remained constant since the last visit (a mechanism under which LOCF is expected to perform well).
- "Differential" mechanism: the missingness mechanism itself is MAR, but missingness affects the subsequent association between the true value of the confounder, and the treatment (the mechanism implicitly assumed by the MPA). In other words, the past observed values of the confounders and treatment predict missingness, but among individuals with a missing covariate value at a given time point, there is no association between the true (but unmeasured) value and the subsequent treatment received.

We compared the performance of CC analysis, LOCF, MPA, MI and IPMW to estimate the causal effect of the intervention at each time point. The analysis model was the additive model proposed in (3,7):

$$Y = \beta_{int} + \beta_0 a_0 + \beta_1 a_1 + \beta_2 a_2,$$

Where *Y* is a continuous outcome, a_k are the binary treatment indicators at time *k* (*k*=0,1,2) and the β coefficients are the parameters of the marginal structural MSM. This model allows the estimation of the contrast between any treatment strategy of interest. The data-generating mechanisms, methods, estimands and performance measures for our simulations are presented in Web Appendix 4 and the R code to generate the data is available in Web Appendix 5. In the main scenario, the proportion of missing data was around 40%, and the sample size was n=10,000. We also investigated the impact of a smaller proportion of missing data (5%) and smaller sample size (n=500).

RESULTS

The results of the main simulation study (n=10,000 and 40% of missing data) are presented as boxplots (Figures 2-3), showing the distribution of the absolute bias for each method, and summarised in Table 1. Full results are presented in Web Tables 1- 5 and Web Figures 1-4.

MCAR

Whereas CC, MI and IPMW lead to unbiased estimates at the three time points, the MPA estimates are biased at each time point and LOCF estimates are biased at times 1 and 2 (Figure 2). The bias for the MPA arises from the direct associations existing between the confounders and the treatment allocation at subsequent time points even among participants with missing covariate values. For LOCF, the bias arises because the missing values were generally different from the observed previous value because confounder values were affected by prior treatment.

MAR

Except MI, which led to unbiased estimates at each time point for the three MAR scenarios, the performance of the other analysis strategies relied on the variables that were predictive of missingness (Figure 3). When missingness depended on the values of past treatment assignment and confounders,

IPMW estimates were unbiased at the three time points. A small bias was observed for complete case analysis and larger biases were obtained when using LOCF and MPA, for similar reasons as in MCAR scenarios. When the outcome was directly related to missingness, the only unbiased approached was MI. However, when an indirect association between the outcome and missingness existed, the IPMW led to unbiased estimates, but with a lower precision than MI.

Missingness on constant values

Only LOCF was unbiased (Figure 3) and the bias was worse with MI and IPMW than with CC. This is because they both use the existing relationships between the confounders, treatment and outcome in the observed data, but in this scenario, these relationships do not reflect the associations existing between the true (missing) confounder values and the other variables.

Missingness affecting the subsequent covariate-treatment associations

MPA was the only appropriate method to obtain unbiased treatment effect estimates, although the bias of the other approaches was quite small. As in the previous scenario, the associations between confounders and treatment among the complete cases cannot be used to make inferences about the relationship among participants with missing confounder values. Therefore, CC, MI and IPMW are biased. LOCF estimates are unbiased when missingness on a variable depends only on the previous measurement for that variable. However, in the current scenario, missingness depends on past values of the treatment and confounders.

When the sample size was small (n=500), the magnitude of bias was similar to that observed for n=10,000, but the standard errors of the treatment effect estimates were very large, illustrating the lack of efficiency of MSMs in small samples (Web Figures 3-4).

When only 5% of the data were missing, biases were smaller in magnitude and were, in our setting, negligible for non-MAR situations. However, we would not recommend the implementation of a method known to be biased when unbiased alternatives exist.

ILLUSTRATIVE EXAMPLE

We used the data from a prospective national cohort, using the research database of the "Observatoire Sommeil de la Fédération de Pneumologie" (OSFP). The OSFP registry is a standardized web-based report, containing anonymized longitudinal data from patients with sleep disorders (21). We aimed to estimate the causal effect of compliance in the use of continuous positive airway pressure (CPAP) device on sleepiness symptoms among patients diagnosed with obstructive sleep apnoea. In the OSFP registry, the number of recorded visits per patient varies, so for simplicity, we focused on patients who had a visit within 3 months, 6 months and 1 year following the initiation of CPAP treatment in order to focus on the problem of missing records rather than sparse follow-up. Compliance was determined as an average use of CPAP >4hours per night within each time interval. The outcome was a continuous sleepiness score measured during the last visit using the Epworth Sleepiness Scale. Age, sex, body mass index, nocturia and presence of depression were considered as potential confounders in this study. The investigators intended to record updated values of body mass index, nocturia and depression at each visit; however, measurement was not always undertaken as planned. Patients could have one (or more) missing measurement(s) on at least one of these variables. Patients' characteristics are described in appendix (Web Table 6).

Out of 1,169 included patients, only 263 (22.8%) patients had a complete record (Web Table 7). Data is not MCAR since associations were observed between all potential confounders and the probability of having a complete record. However, a MAR mechanism is plausible in this setting. The MPA is not a suitable method for this analysis because of missing data on the CPAP exposure.

Results are presented in Figure 4 and Web Table 8. Overall, we found no causal effect of CPAP compliance on sleepiness. Due to the relatively small sample size, IPMW led to very wide 95% confidence intervals. The 95% CI for LOCF is narrower, but does not account for the uncertainty around the imputed values. Furthermore, the assumption underlying the validity of LOCF is unlikely to hold here. As expected, all approaches gave similar results because confounding was not very strong (Web

Table 7). However, it illustrates the inefficiency of IPMW and the limitations of the MPA approach when exposure data are missing.

DISCUSSION

In this paper, we presented five methods to handle missing values in partially observed time-varying covariates in MSMs, identified situations in which they are appropriate to estimate unbiased causal effects, and illustrated how to implement these approaches in practice. We showed that, for the estimation of causal effects, CC analysis is often biased, unless data are MCAR. The validity of this assumption cannot be tested from the data (22), but violations of these assumptions can be detected by looking at associations between the probability of being a complete case and the variables available in the dataset. While the MCAR assumption is rarely plausible, we also showed that when missing values are MAR given treatment history and confounders, the bias of the CC estimates is usually small.

LOCF leads to biased estimations of the treatment effects, unless missing values are in truth missing because they remained constant over time or when the previous measurement is used to adapt treatment (rather than the true - but missing - measurement). This assumption may hold in routinely collected data. For instance, GPs might not record a patient's weight during a visit if it has not changed since the previous consultation. This assumption cannot be tested from the data but the plausibility of the assumption can be assessed using expert opinion, building on what has been proposed in randomised trials (23). Moreover, when using LOCF, the uncertainty around the single imputation of missing values is not accounted for (24). Although this is not an issue for categorical variables, it is problematic for continuous confounders where imputing exactly the previous measurement may lead to inappropriate certainty.

As expected, MI led to unbiased estimates of the treatment effect when data were MCAR or MAR. When implementing MI, the outcome must be included in the imputation model and the treatment effect estimated in each imputed dataset and combined using Rubin's rules, as recommended in settings

with a single time point (25). In our simulations and example, treatment and covariate values at all time points were included in the imputation model. With an increasing number of time points, issues of overfitting may arise. Two-fold multiple imputation has been proposed to circumvent this problem (26). Instead of using all the time-blocks in the imputation model, only the current and adjacent times are used, and therefore fewer parameters are to be estimated in the imputation model.

MPA and IPMW have never been investigated in the context of MSMs. The MPA is unbiased when either the association between the partially observed confounder and the outcome or between the partially observed confounder and the subsequent treatment disappears among patients with a missing value for that confounder. Hence, the validity of this approach does not depend on the missingness mechanism but instead the relationship between covariates, treatment and outcome among patients with missing data. In routinely collected primary care data, this assumption is plausible when, for instance, the results of a blood test are used by the general practitioner to adapt treatment prescription. If these results are missing (i.e. not available to the general practitioner), they will not be used in the treatment decision: among patients with missing blood test results, the true (but unmeasured) value of the biological parameter is not directly associated with the treatment, and therefore the test result is no longer a confounder. The MPA is straightforward to implement but issues may arise when there are many missingness patterns. The MPA's inability to accommodate missing data on the treatment and the outcome might limit its applicability, unlike IPMW, which includes only the complete records in the analysis, regardless of which variables have missing values. IPMW generally leads to unbiased estimates when data are MCAR and MAR. The exception is when missingness is directly affected by the outcome, because the outcome might be associated with treatment and confounder values at later time points, which are not accounted for in the missingness model to compute the weights. IPMW is also unbiased in scenarios where the MPA is unbiased because patients are censored at the first missing data, and therefore, no use is made of the information measured at later time points. However, IPMW is somewhat inefficient. This is explained by a loss in sample and by the multiplication of two weights (the

treatment weight and the missingness weight) that are both estimated with uncertainty, leading to highly variable treatment effect estimates. We recommend the use of IPMW in very large datasets, a situation in which MI would be highly computationally intensive. Furthermore, a limitation in the current implementation of IPMW is that missing data were considered as being monotone, that is, patients were excluded from the analysis even when the outcome was available at the end of follow-up. Recent developments on inverse weighting have included an extension to non-monotone missingness patterns (27), but it remains unclear how it could be transposed to MSMs.

It is clear that no single missing data method can simultaneously handle different types of missingness mechanisms. However, in practice, missing values can occur on several variables according to different mechanisms. In such situations, it is crucial to understand the reasons for missingness to identify groups of variables with similar missingness mechanisms that could be handled altogether. For instance, in routinely collected data, some variables might not have been updated because their values remained unchanged, and some variables might be missing at random. A pragmatic approach would be to first use the LOCF on the first group of variables, and then multiply impute the variables from the second group. A more principled combination of methods has been proposed in simpler settings. Qu and Lipkovich (28) combined the MPA and MI for propensity score analysis with a single time point. Seaman and White proposed to combine MI and IPMW (29) but further investigation is needed before implementing these methods in MSMs.

Although the role of the simulation was not to investigate the statistical properties of the five approaches in a broad range of settings, but to empirically illustrate the theoretical findings, the design of the simulations has several limitations. First, we focussed on a relatively simple setting with three time points and a few covariates. A plasmode simulation approach based on the sleep apnoea study would have been more realistic but would have not allowed us to investigate the “constant” and “differential” mechanisms of missing data. Second, we generated data with a continuous outcome only. While this was chosen because bias is more easily observed with continuous outcomes, our conclusions

will apply to binary and time-to-event outcomes; the validity of the methods relies on the missingness mechanism which is independent of the nature of the outcome. Similar conclusions would also hold had the outcome been measured repeatedly. With a larger number of partially observed confounders, sparse data in some missingness patterns may preclude use of the MPA approach. Moreover, in the presence of numerous confounders with interactions and non-linear effects, the functional form of the weight model might be harder to specify parametrically, and the obtained weights could be unstable. These problems may be alleviated by using more robust approaches (30) or statistical learning methods (31). Finally, the standard error of treatment effect estimates in our simulation study did not account for the uncertainty in the weight estimation, resulting in overly wide 95% confidence intervals. Non-parametric bootstrap was used in our illustrative example but was too computationally demanding for use in simulations.

In conclusion, the choice of the appropriate method(s) to handle partially observed confounders in MSMs must rely on a careful consideration of the reasons for missingness and whether missingness modifies the existing relationships among observed data. Causal diagrams may help in understanding the structure of the data and the relationships between variables when data are missing, and when data are observed. Although MI outperforms the other approaches when data are MAR, we presented two scenarios, encountered in routinely collected data, where MI leads to biased estimates of the treatment effect estimates but LOCF and the MPA might be suitable alternatives. Any analysis with missing data inevitably relies on assumptions about the missingness mechanisms or missingness patterns, which are often not made explicit. We therefore encourage researchers to clearly describe the assumptions under which their primary analysis is valid, and to perform sensitivity analyses to assess robustness of their results to departures from these postulated missingness mechanisms.

Acknowledgments: Author affiliations: London School of Hygiene and Tropical Medicine, Department of Medical Statistics, London, UK (Clémence Leyrat, James R. Carpenter, and Elizabeth J. Williamson); London Hub for Trials Methodology Research, MRC Clinical Trials Unit, UCL, London, UK (James R Carpenter); HP2, Inserm 1042, Université Grenoble Alpes, Grenoble, France (Sébastien Bailly); Health Data Research London site, UK (Elizabeth J. Williamson). This work was supported by the UK Medical Research Council (Project Grant MR/M013278/1 (Dr Williamson), and MC UU 12023/21 and MC UU 12023/29 (Prof Carpenter)). Dr Bailly is supported by the "e-health and integrated care" Chair of excellence of the University Grenoble Alpes Foundation and the ANR 3IA MIAI@Grenoble Alpes project. Conflict of interest: none declared.

REFERENCES

1. Rubin DB. The designversus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Stat Med. 2007 Jan;26(1):20–36.
2. Platt RW, Schisterman EF, Cole SR. Time-modified Confounding. Am J Epidemiol. 2009 Sep 15;170(6):687–94.
3. Daniel R m., Cousens S n., De Stavola B l., Kenward MG, Sterne J a. C. Methods for dealing with time-dependent confounding. Stat Med. 2013 Apr 30;32(9):1584–618.
4. Rosenbaum PR. The Consquences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment. J R Stat Soc Ser Gen. 1984 Jan 1;147(5):656–66.
5. Robins JM. Marginal Structural Models. In: Proceedings of the American Statistical Association. 1997. p. 1–10. (Section on Bayesian Statistical Science).
6. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008 Sep 15;168(6):656–64.
7. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology. 2000 Sep;11(5):550–60.
8. Clare PJ, Dobbins TA, Mattick RP. Causal models adjusting for time-varying confounding-a systematic review of the literature. Int J Epidemiol. 2019 Feb 1;48(1):254-265.

9. Mojaverian N, Moodie EEM, Bliu A, Klein MB. The Impact of Sparse Follow-up on Marginal Structural Models for Time-to-Event Data. Am J Epidemiol. 2015 Dec 15;182(12):1047–55.

10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344–9.

11. Karahalios A, Baglietto L, Carlin JB, English DR, Simpson JA. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Med Res Methodol. 2012;12:96.

12. Shortreed SM, Forbes AB. Missing data in the exposure of interest and marginal structural models: A simulation study based on the Framingham Heart Study. Stat Med. 2010 Feb 20;29(4):431–43.

13. Liu S-H, Chrysanthopoulou SA, Chang Q, Hunnicutt JN, Lapane KL. Missing Data in Marginal Structural Models: A Plasmode Simulation Study Comparing Multiple Imputation and Inverse Probability Weighting. Med Care. 2019 Mar;57(3):237–43.

14. Moodie EEM, Delaney JAC, Lefebvre G, Platt RW. Missing Confounding Data in Marginal Structural Models: A Comparison of Inverse Probability Weighting and Multiple Imputation. Int J Biostat [Internet]. 2008 [cited 2017 May 22];4(1). Available from: https://www.degruyter.com/view/j/ijb.2008.4.1/ijb.2008.4.1.1106/ijb.2008.4.1.1106.xml

15. Vourli G, Touloumi G. Performance of the marginal structural models under various scenarios of incomplete marker's values: A simulation study. Biom J. 2015 Mar 1;57(2):254–70.

16. Little R. Statistical analysis with missing data. 2nd ed. Hoboken N.J.: Wiley; 2002.

17. Blake HA, Leyrat C, Mansfield KE, Seaman S, Tomlinson LA, Carpenter J, et al. Propensity scores using missingness pattern information: a practical guide. Statistics in Medicine. 2020; 39: 1641– 1657.

18. Kreif N, Sofrygin O, Schmittdiel JA, Adams A, Grant R, Zhu Z, et al. Evaluation of adaptive treatment strategies in an observational study where time-varying covariates are not monitored systematically. preprint. 2018. (doi: 1806.11153). Accessed September 13, 2020.

19. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, Florida. 2020.

20. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Jun 29;338:b2393.

21. Bailly S, Destors M, Grillet Y, Richard P, Stach B, Vivodtzev I, et al. Obstructive Sleep Apnea: A Cluster Analysis at Time of Diagnosis. PLOS ONE. 2016 Jun 17;11(6):e0157318.

22. Potthoff RF, Tudor GE, Pieper KS, Hasselblad V. Can one assess whether missing data are missing at random in medical studies? Stat Methods Med Res. 2006 Jun;15(3):213–34.

23. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clin Trials. 2007 Apr 1;4(2):125–39.

24. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. Biostat Oxf Engl. 2004 Jul;5(3):445–64.

25. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used? Stat Methods Med Res. 2019 Jan;28(1):3-19.

26. Nevalainen J, Kenward MG, Virtanen SM. Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. Stat Med. 2009 Dec 20;28(29):3657–69.

27. Sun B, Tchetgen EJT. On Inverse Probability Weighting for Nonmonotone Missing at Random Data. J Am Stat Assoc. 2018 Jan 2;113(521):369–79.

28. Qu Y, Lipkovich I. Propensity score estimation with missing values using a multiple imputation missingness pattern (MIMP) approach. Stat Med. 2009 Apr 30;28(9):1402–14.

29. Seaman SR, White IR, Copas AJ, Li L. Combining Multiple Imputation and Inverse-Probability Weighting. Biometrics. 2012 Mar;68(1):129–37.

30. Imai K, Ratkovic M. Robust Estimation of Inverse Probability Weights for Marginal Structural Models. J Am Stat Assoc. 2015 Jul 3;110(511):1013–23.

31. Karim ME, Petkau J, Gustafson P, Tremlett H, Group TBS. On the application of statistical learning approaches to construct inverse probability weights in marginal structural Cox models: Hedging against weight-model misspecification. Commun Stat - Simul Comput. 2017 Nov 26;46(10):7668–97.

Table 1. Characteristics of the 5 missing data methods for partially observed time-varying confounders

Method	Missing data on …	Assumptions	Unbiased in MSMs when…	Advantages	Limitations
Complete case	Covariates Treatment Outcome	Missing data are MCAR	MCAR	Straightforward	May be inefficient because of the loss in sample size
Last Observation Carried Forward	Covariates Treatment Outcome except at baseline	The true, but missing, value is the same as the last available measurement OR The treatment decision depends on the previous available measurement rather than the true (unobserved) one	Constant	Straightforward Discards fewer patients from the analysis than CC	It can lead to too narrow confidence intervals Patients are discarded if baseline measurements are missing
Multiple imputation	Covariates Treatment Outcome	Missing data are MAR[a] The imputation model is correctly specified	MCAR MAR|A,L MAR|A,L,Y MAR|A,L,V	Maintains the original sample size	May be computationally intensive Challenging for a large number of time points
Inverse-probability-of-missingness weighting	Covariates Treatment Outcome	Missing data are MAR given the treatment and the covariates, but not the outcome The weight model is correctly specified	MCAR MAR|A,L MAR|A,L,V Constant	Faster than MI for large datasets Weights simultaneously address confounding and missing data	May be inefficient for small and moderate sample size
Missingness Pattern Approach	Covariates	The partially observed covariate is no longer a confounder once missing e.g. the treatment decision depends on the confounder value only when a measurement is available	Differential	Relatively simple to implement Assumptions do not relate to Rubin's taxonomy so may work when standard methods do not	Does not handle missing data on the exposure or outcome Challenging when the number of missingness patterns is large

CC: complete cases; MI: multiple imputation; MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random.

[a] Extensions to accommodate MNAR exist but are challenging to apply in practice.

Figure 1. Causal graphs representing several possible scenarios for missing values. At each time, L_1, L_2 are the two time-varying confounders, A is the time-varying treatment, Y is the outcome, V is an independent risk factor, and R is the missingness indicator. Diagrams represent scenarios in which missingness: A) occurs completely at random (MCAR); B) occurs when there has been no change since the previous measurement (Constant); C) occurs at random given the confounders and treatment (MAR|A,L); D) occurs at random given the confounders, treatment, and outcome (MAR|A,L,Y); E) occurs at random given the confounders, treatment, and independent risk factor F) depends on the treatment and confounders, but the association between the missing value and the subsequent treatment allocation no longer exists (Differential). Under F), the confounder only contributes to the treatment allocation decision when observed.

Figure 2. Absolute bias of the treatment effect estimate at k=0 (A,D,G), k=1 (B,E,H) and k=2 (C, F,I) on full data and following the use of different missing data approach under the missing completely at random (A-C), Constant (D-F) and C Differential (G-I) missingness mechanisms. N=10000. 40% of missing data.

CC: complete cases; LOCF: last observation carried forward; MPA: missing pattern approach; MI: multiple imputation; IPMW: inverse probability of missingness weighting. For multiple imputation, 10 imputed datasets were generated.

Figure 3. Absolute bias of the treatment effect estimate at k=0 (A,D,G), k=1 (B,E,H) and k=2 (C, F,I) on full data and following the use of different missing data approach under 3 scenarios of data missing at random: A-C) missing at random given the covariates and the treatment (MAR|A,L), D-F) given the covariates, the treatment and the outcome (MAR|A,L,Y) and G-I) given the covariates, the treatment and the independent risk factor (MAR|A,L,V). N=10000. 40% of missing data.

CC: complete cases; LOCF: last observation carried forward; MPA: missing pattern approach; MI: multiple imputation; IPMW: inverse probability of missingness weighting. For multiple imputation, 10 imputed datasets were generated.

Figure 4: Results of the illustrative example. Each point represents the difference in sleepiness scores at the end of the study (12 months) between compliers (CPAP use>=4 hours per night) and non-compliers (CPAP use<4 hours per night) at each of 3 time-points (3 months (squares), 6 months (circle) and 12 months (triangles)). The vertical bars are the 95% confidence intervals.