

Published in final edited form as:

Med Care. 2019 March; 57(3): 237-243. doi:10.1097/MLR.000000000001063.

Missing data in marginal structural models: A plasmode simulation study comparing multiple imputation and inverse probability weighting

Shao-Hsien Liu, PhD^{1,2}, Stavroula A. Chrysanthopoulou, PhD³, Qiuzhi Chang, ScD⁴, Jacob N. Hunnicutt, PhD⁵, and Kate L. Lapane, PhD²

¹Clinical and Population Health Research Program, Graduate School of Biomedical Sciences, University of Massachusetts Medical School, Worcester, MA 01605;

²Division of Epidemiology of Chronic Diseases and Vulnerable Populations, Department of Quantitative Health Sciences, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605;

³Department of Biostatistics, Brown University School of Public Health, 121 South Main Street, Providence, RI 02903;

⁴Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115;

⁵Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA 15240

Abstract

Background: The use of marginal structural models (MSMs) to adjust for time-varying confounding has increased in epidemiologic studies. However, in the setting of MSMs, recommendations for how best to handle missing data are contradictory. We present a plasmode simulation study to compare the validity and precision of MSMs estimates using complete case analysis (CC), multiple imputation (MI), and inverse probability weighting (IPW) in the presence of missing data on time-independent and time-varying confounders.

Methods: Simulations were based on a cohort sub-study using data from the Osteoarthritis Initiative which estimated the marginal causal effect of intra-articular injection use on yearly changes in knee pain. We simulated 81 scenarios with parameter values varied on missing mechanisms (MCAR, MAR, and MNAR), percentages of missing (10%, 20%, and 30%), type of confounders (time-independent, time-varying, either or both), and analytical approaches (CC, IPW, and MI). The performance of CC, IPW, and MI methods was compared using relative bias, mean squared error (MSE) of the estimates of interest, and empirical power.

Corresponding Author: Shao-Hsien Liu, PhD, MPH, Division of Epidemiology of Chronic Diseases and Vulnerable Populations, Department of Quantitative Health Sciences, University of Massachusetts Medical School, 368 Plantation street, Worcester, MA 01605, Telephone: 774-455-3616; Fax: 508-856-8993, shaohsien.liu@umassmed.edu).

Conflicts of interest: We have no conflicts of interest to declare.

The abstract of this manuscript has been presented at the International Conference on Pharmacoepidemiology in Montreal, Canada in August, 2017.

Results: Across scenarios defined by missing data mechanism, extent of missing data, and confounder type, MI generally produced less biased estimates (range: 1.2% to 6.7%) with better precision (range: 0.17 to 0.18) compared to IPW (relative bias: –5.3% to 8.0%; precision: 0.19 to 0.53). Empirical power was constant across the scenarios using MI.

Conclusions: Under simple yet realistically constructed scenarios, MI appears to confer an advantage over IPW in MSMs applications.

Keywords

Missing data; Marginal structural models; Plasmode simulation, Multiple imputation; Inverse probability weighting

Introduction

Marginal structural models (MSMs) using inverse-probability-of-treatment—weighted estimation (IPTW) have been proposed to estimate unbiased causal effects when time-varying confounding is a concern. ^{1–3} Briefly, this technique creates a pseudo-population in which bias has been eliminated by simultaneously adjusting for time-varying confounding (without blocking indirect effects from former exposures² and avoiding collider-stratification bias⁴) and selection bias owing to informative censoring. ³ Methodologic research has provided guidance on appropriate weight construction⁵, how best to build the outcome models, ^{6,7} and what assumptions are needed to identify causal effects. ^{8,9} Little guidance, however, exists regarding how to handle missing data when MSMs are applied to longitudinal data.

To our knowledge, two studies compared methods to handle missing confounder information in the setting of MSMs. ^{10,11} Both of these studies limited their evaluation to situations with missing data in time-varying confounders. The guidance provided from these studies appears to be contradictory. While one study recommends that multiple imputation (MI) is superior to inverse probability weighting (IPW) where missing data are strongly predicted by the available data, ¹⁰ the other suggests that IPW performs better than MI. ¹¹ Whether the differences in the study findings could be due to the effect of the confounder on the outcome ¹¹ or some artifact of the simulations is unknown. Specific guidance regarding which missing data techniques should be used for scenarios of associations between confounders and exposure-outcome relationship remains unclear. Further, the extent of biases resulting from applying commonly used missing data techniques for time-independent confounding in MSMs settings has not (to our knowledge) been explored.

Given the increasing use of MSMs in epidemiologic studies, ¹² we sought to compare the validity and precision of commonly used missing data approaches for time-independent or time-varying confounders (i.e., MI and IPW) in a simulated cohort study using MSMs. We used plasmode simulation to generate data (a pseudo-sample) which preserved the underlying associations among observed covariates from an empirical cohort study. ¹³

Materials and Methods

The University of Massachusetts Institutional Review Board considered this study exempt since we used publicly available data to construct the cohort for simulation.

Empirical data

We used data from a previously published retrospective cohort study using publicly available data from the Osteoarthritis Initiative (OAI). ¹³ The OAI is a multi-center (i.e., Baltimore, MD; Columbus, OH; Pittsburgh, PA; and Pawtucket, RI), longitudinal, prospective observational study examining the development and progression of knee osteoarthritis (OA) and the effectiveness of disease-modifying therapies. The OAI cohort includes 4,796 men and women ages 45-79 enrolled between February 2004 and May 2006 and followed for 9 years (http://oai.epi-ucsf.org/). The OAI collects clinical assessments such as symptoms and function of the knee, quality of life, physical performance, health behaviors, medications and supplements, and biologic specimens including blood and urine for up to 9 years of follow up. Data on clinical, joint status, and risk factors for the progression and development of knee OA were collected at baseline and the yearly follow-up clinic visits. We based simulations on a previously published study using a new-user cohort design to compare the initiation of intra-articular injection versus non-use among participants with knee OA.¹⁴ This setting is particularly well-suited for using a plasmode simulation framework. 13 Using the empirical sample derived in the study, ¹⁴ we used participants who newly received a corticosteroid injection to construct the cohort. Only complete cases - participants with no missing data on the variables of interest - were used to generate the simulated datasets. We required no missing data at this stage so that we could impose various missing data mechanisms and vary the extent of missing data. The cohort included 646 participants (213 participants who newly received a corticosteroid injection and 433 participants who did not).

Data generation: plasmode simulation

We simulated datasets using the plasmode simulation framework. 13 The causal diagrams in Figure 1 depict the causal relationship between the exposure, outcome, covariates, and missing data mechanisms. Figure 1 shows three discrete time points given $t=0,\,1,\,$ and 2 with two time points of treatment (Injection1 and Injection2). L_0 indicates a set of prespecified confounders measured at baseline (t=0). L_1 represents time-varying confounders affected by baseline covariates and measured at the same time the exposure was measured (t=1). L_2 represents time-varying confounders affected by previously measured confounders (L_1) and the use of treatment measured at the same time the exposure was measured (t=2). C' indicates a set of additional potential confounders. To simplify the presentation of the causal diagram, we omitted some arrows from Figure 1. The missing data mechanism is represented by M. Figure 1, Panel A shows missing data in the confounder measured at baseline (t=0). Figure 1, Panel B shows missing data in the time-varying confounder measured at t=1 and t=2. Figure 1, Panel C shows missing data in both the confounder measured at t=0, 1, and 2.

To generate data, we used the plasmode simulation framework based on the parameters for data generation shown in Table 1. In the first step, we estimated a linear model with the

observed study outcome (i.e., one-year change in knee pain) as a function of the exposure status, baseline covariates, and a subset of the potential confounders using data from the constructed cohort. 14,15 We then sampled with replacement among those exposed and unexposed participants from the constructed cohort to achieve the desired sample size (N =500). We compared our findings with previous studies. ^{10,11} Because the information on the covariates and the exposure for each participant was preserved without modification, the associations among these variables remained intact in the sampled populations. Using the linear model described in the first step above, we generated outcome values by building outcome-generating models in which we substituted a pre-identified treatment effect as the coefficient for the exposure term. We used 1.2 as the pre-identified average treatment effect since this value was considered as the threshold for achieving a minimal clinically important change in knee pain in patients with knee OA. 16-19 The values of the other model coefficients remained unchanged from the linear model described in the first step above. The outcome Y₁ and Y₂ (i.e., one-year change in knee pain from baseline to year 1 and one-year change in knee pain from year 1 to year 2) were then generated with a normally distributed error term resulting from the outcome-generating model:

$$Y_1 = \beta_0 + \beta_1 (Injection_1) + \beta_2 L_0 + \beta_3 L_1 + \beta_4 C' + \varepsilon$$
 (1)

$$Y_2 = \beta_0 + \beta_1(Injection_1) + \beta_2(Injection_2) + \beta_3 L_0 + \beta_4 L_1 + \beta_5 C' + \varepsilon$$
 (2)

Where $\beta 1$ and β_2 were set to the pre-identified treatment effect, and β_0 , β_2 , β_3 , β_4 and β_5 were determined in the first step. We then repeated this process to generate 1,000 simulated datasets.

Missing data mechanisms and missing information

Based on our experience OAI data, 14,15 we developed scenarios where missing information on the time-independent (L₀) (Figure 1, Panel A), time-varying (L₁, L₂) (Figure 1, Panel B), and both time-independent and time-varying confounders (Figure 1, Panel C). We selected the severity of knee OA (Kellgren-Lawrence grade, 3 levels) measured at baseline as the time-independent confounder of interest. Knee pain measured at the point of injection use was used as a time-varying confounder (continuous, ranging from 0-20).

Separate simulated datasets were generated for each missing data mechanism including MCAR, MAR, and MNAR. 20,21 Data are considered MCAR when the probability of missingness does not depend on the values of observed covariates. To impose MCAR in the simulated data sets, we randomly selected participants and forced the information on the time-independent (i.e., Figure 1.A), time-varying (i.e., Figure 1.B), or both (i.e., Figure 1.C) confounders to be missing. Data are considered MAR if the probability of missingness depends on values of observed covariates. 20,21 To simulate this situation, we assumed the probability of missingness on the time-independent confounder (i.e., Kellgren-Lawrence grade measured at time t=0) to be jointly related to a set of observed covariates (i.e., age,

sex, visit number of initial injection). For the time-varying confounders (i.e., knee pain score concurrently measured with exposure at time t=1 and t=2), we assumed missingness to be jointly dependent on observed covariates using information from age, sex, household income, and race/ethnicity. For this process, we used the R package "simstudy".²²

Data are considered to be MNAR when the probability of missing depends on values of unobserved covariates. 20,21 Using the empirical information from the constructed cohort, we assumed that participants who were more obese were more likely to have missing data on the time-independent confounder. Among all participants missing information on time-independent confounder (i.e., Kellgren-Lawrence grade measured at time t=0), 70% were randomly selected from the obese, 20% from overweight, and 10% from normal weight group. For the time-varying confounder (i.e., knee pain score concurrently measured with exposure status at time t=1 and t=2), we assumed that participants with more severe pain were more likely to be missing. Among all participants missing information on knee pain scores concurrently measured with the injection use, 70% were randomly selected from the knee pain score 10, 20% from the knee pain score between 5–9, and 10% from knee pain score <5. For each missing data mechanism (MCAR, MAR, MNAR), we altered the extent of missing data such that 10%, 20%, and 30% of the data were imposed as missing.

Inverse probability weighting

IPW is a particularly straightforward approach to use in MSMs settings.^{23,24} It shares the similarity of weight building process for the inverse probability of observed treatment or censoring weights that are performed when estimating parameters of MSMs.² IPW proceeds by calculating the probability of having complete data for each individual in the study. Using logistic regression models, contributions of each individual are weighted by the inverse probability of having complete data conditional on other relevant covariates.

In the missing data mechanisms of MCAR and MAR on the time-independent confounder, we created a binary variable indicating a missing status (1=yes, 0=no). We then modeled the weights of missingness proportional to the inverse probability of the value being observed using the logistic regression conditional on all other available information at baseline from L_0 (Figure 1). For MNAR, information at baseline from C' was used to construct the weights. For missing information on the time-varying confounders, a similar approach was used (e.g., missing data at t=1, yes/no). We generated the weights of missingness proportional to the inverse probability of the value being observed at time t = 1 using a logistic model conditional on injection use and observed values from L_0 , L_1 , and/or C'.

Multiple imputation

In general, the MI approach generates *m* complete datasets where missing values in the incomplete observed data are imputed. Each of the *m* datasets is then analyzed using the same model and estimation method. The estimates from each of the *m* datasets are then combined to produce a single estimate that incorporates the usual sampling variability as well as the variability of the missing data.²⁵ We implemented MI using MICE (Multiple Imputation by Chained Equations) package in R.²⁶ Unlike joint modeling approach (JM) (e.g., Markov chain Monte Carlo (MCMC) technique) which assumed joint multivariate

normality of all variables, this method is based on Fully Conditional Specification (FCS), where each incomplete variable is imputed by a separate modeling process.²⁷ As such, FCS MI generally allows some flexibilities since an appropriate imputing model can be selected on a variable-by-variable basis, especially when no suitable multivariate distribution can be drawn from for the missing values using JM approach.^{28,29} For the imputations models, we used all available information (including the outcome) in the data to predict missing values for the time-independent and time-varying confounders.³⁰ In addition, some baseline variables such as age and sex were specified and included in the prediction model.²⁶ For each simulated dataset, five imputed datasets were generated for the missing information on the time-independent and time-varying confounders.³¹

Analytical approaches and evaluation of methods performance

Overall, the analytic approach was carried out in two steps. Each step was applied to scenarios based on different extent of missing data (10%, 20%, 30%), different missing data mechanisms (MCAR, MAR, MNAR), and different variables affected (time-independent confounder, time-varying, and both). First, we applied IPW and MI to deal with missing data in the analysis and compared results to those from complete case analysis (CC). Then, using a different model than the true outcome generating model shown in equation (1) above, we fit the weighted outcome models using generalized estimating equations (GEE) to estimate the average causal effect in MSMs. We did this for each approach for handling missing data (for the CC, IPW, and MI). In all analyses, we used stabilized weights to yield estimates with greater precision compared to the unstabilized weight.^{2,5} The numerator was the conditional probability of receiving observed treatment given baseline confounders. The denominator was the conditional probability of receiving observed treatment given timevarying confounders in addition to baseline confounders. For the application of IPW, the final weights incorporated in the outcome models were calculated as the product of stabilized treatment weights and censoring (missing) weights developed using a similar approach as described above. Since multiple imputed datasets were used, we used the mi.meld function in R package "Amelia". 32 Results generated from the function reflected the average estimates with standard errors that accounted for average uncertainty and disagreement in the estimated values across the models.²⁵

The performance of the CC, IPW, and MI methods in each scenario was compared using relative bias, mean squared error (MSE), and empirical power of the estimates of interest.

Relative bias was calculated as $\left(\frac{\hat{\beta} - \beta_{truth}}{\beta_{truth}}\right)$ *100%. MSE was calculated combining bias and

true variance (bias² + standard error($\hat{\beta}$)²), where standard error ($\hat{\beta}$) was calculated as $\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\widehat{\beta}_i-\overline{\beta})^2}$. Empirical power was defined as (1 – empirical type II error). The empirical type II error was calculated as the total number P values > 0.05 divided by the total number of simulations. The empirical power is the percentage of times that we will reject a false null hypothesis.

Results

We simulated 81 scenarios in total with parameter values varied on missing mechanisms (MCAR, MAR, and MNAR), percentages of missing (10%, 30%, and 50%), type of confounders (time-independent, time-varying, either or both), and analytical approaches (CC, IPW, and MI).

Table 2 shows results from missing values in a time-independent confounder (L_0). While estimates of relative bias showed similar results for CC, and IPW, with a slightly larger range for MI under MCAR and MAR (range: CC, -2.3% to 3.5%; MI, 3.6% to 5.5%; IPW, 2.3% to 3.5%), CC and IPW displayed a trend of underestimating the effects given the increasing proportion of missingness under MNAR. Regardless of missing data mechanisms, the MI procedure produced consistent and smaller MSE across all missing data scenarios. The CC and IPW showed a trend of increased estimates given the increasing proportion of missingness. Similarly, while the empirical power using MI was constant across scenarios, CC and IPW showed a trend of decreased power given the increasing proportion of missingness.

Table 3 shows results from missing values in time-varying confounders. While MI showed a smaller range of estimates for relative bias regardless of scenarios (range: 3.3% to 5.0%), CC and IPW showed an increased relative bias under MCAR and MAR (range: CC, 2.9% to 8.4%; IPW, 3.1% to 8.0%) and underestimated the true effects under MNAR. Relative to CC and IPW, the MI yielded a smaller MSE and maintained empirical power across all scenarios. The CC and IPW both showed a trend of increased MSE and decreased power given the increasing proportion of missingness.

Results from missing values in either a time-independent, time-varying, or both confounders are displayed in Table 4. The MI method showed a relatively consistent bias across scenarios (range: 4.1% to 6.7%) as compared to CC and IPW (range: CC, -1.9% to 11.7%; IPW -5.3% to 7.8%). The IPW method performed slightly better for the estimates of relative bias under MCAR and MAR despite larger MSE and lower empirical power under MNAR. While MI provided a consistently smaller MSE and maintained empirical power across all scenarios, the CC and IPW methods both showed a trend of increased MSE and decreased power given the increasing proportion of missingness. For relative bias, MSE, and empirical power, the worst scenario occurred when missing information on the confounders reached 30% regardless of missing data mechanism (MCAR, MAR, MNAR) using CC and IPW and both also underestimated the true effects under the MNAR.

Discussion

Using the plasmode simulation framework in which we imposed missing data on time-independent and/or time-varying confounders, our simulation study demonstrated the performance of commonly used missing data approaches including CC, IPW, and MI in the context of MSMs analyses. Compared to CC and IPW, MI consistently produced less biased marginal estimates with better precision regardless of the missing data mechanisms and the extent of missingness. In addition, while the empirical power for the MI procedure was

constant across scenarios, CC and IPW both displayed a trend of decreased power given the increasing proportion of missingness.

Our findings are consistent with the previous study¹⁰ but different from the other¹¹ with respect to the results from the MI procedure which demonstrated less biased marginal estimates and noticeably less variation for different missing data mechanisms and extent of missingness. For the implementation of the MI technique, we used a similar approach as the previous study¹⁰ which included baseline information of the confounding variables to impute the missing values. We also included the outcome variable in the prediction model for the time-independent and time-varying confounders.³⁰ Since the purpose of the MI method is to model the missing values, our approaches may provide additional advantages of MI over IPW if the information provided was predictive of the missing variables. However, it is not clear if similar approaches were used in the other study's setting and thus results the discrepancy.¹¹

The IPW technique focuses on predicting missing data mechanisms.³³ As such, there may well be some situations where IPW outperforms MI which was demonstrated in this previous study.¹¹ One plausible example includes situations lacking strong predictors of the missing values or situations in which the missing data mechanism is well-understood. In our study, we only selected predictors that were fully observed and associated with missing values to model the missingness. Using this approach, we had all the data needed to fit the missingness model and to estimate individuals' weights but this may hamper the comparison.³⁴ Another scenario where the IPW may yield more satisfactory performance relative to MI is because the large numbers missing values are due to missed visits (considered as monotone missing). In the previous study, missed visits were also discussed for one of the missing-data scenarios.¹¹. Yet, our study is limited to further explore this particular situation given that we had just two time intervals. Therefore, whether this scenario explains the observed differences across the simulation studies remains unclear.

In addition to comparing method performance using bias and precision, our study demonstrated that use of the MI approach maintained empirical power consistently across scenarios. The empirical power decreased using CC and IPW given increasing proportion of missingness across different types of missing data. MSMs analyses partially mimic a sequentially randomized trial design and thus allows estimation of the marginal treatment effect through the application of IPTWs. Thowever, similar to the findings from the CC approach, we noticed that the statistical power to detect a pre-identified non-null treatment effect using IPW was decreased compared to MI in our simulated scenarios. This is an important issue since statistical power allows both investigators and readers with information to help interpret potentially null conclusions. Our findings provide additional perspectives regarding the choice of analytic methods when dealing with missing data in the context of MSMs. However, despite that the trend of decreasing power is shown, it is possible that the current setting is overpowered given the use of clinically meaningful treatment effect and therefore the magnitude of changes in empirical power for the missing data approaches remains unknown.

Several limitations must be acknowledged. First, we considered a simplified context which only used a two-time interval setting for data-generating scenarios related to treatment use and outcome. For situations involving more time intervals, the mechanisms regarding continued treatment use becomes more complicated. Information on time-varying confounders affecting treatment use is needed to correctly model the complex mechanism of treatment use.³⁶ Second, the model used to generate outcome values was based on a set of pre-defined covariates from a cohort sub-study using data from OAI. It is possible that outcomes generated from a much larger set of factors (e.g., using data from claims datasets) may be different due to the influence of both measured and unmeasured covariates.¹³ Therefore, the performance of methods observed based on this cohort sub-study may not extend to other studies simulated from claims data. Third, we did not have scenarios where 50% of data were missing and thus may hamper the comparison with previous research, 10,11 we considered our scenarios are less extreme and maybe more suitable for the consideration of missing data approach in the setting. In addition, while missing data can also occur in the exposure of interest,³⁷ our simulation study only introduced missingness on confounders. Whether our findings extend to different types of variables, including the outcome and exposure of interest, needs to be explored.

Despite these limitations, the strengths of our study included the use of plasmode simulation which keeps the data and associations among covariates unchanged with the advantage of manipulating other parameters such as strength of confounders and exposure of interest. ¹³ By using the cohort sub-study from OAI to construct the cohort for simulated datasets, we provided measurements not only on clinical assessments (e.g. symptoms and joint status) but also on risk factors and concurrent medication use for the progression of knee OA. Given the pre-identified potential confounders and treatment effect, ¹⁵ data-generating scenarios in our study may perform better than approaches using ordinary methods or healthcare claims which may not capture important features of this population. While previous studies focused on time-varying confounders, ^{10,11} our study also assessed the methods performance using baseline confounders and mixed scenarios which provided a more comprehensive and realistic evaluation under MSMs analyses.

In conclusion, with a range of simulated missing data scenarios under MSMs analyses, our simulation study demonstrated that the MI approach generally produced less biased estimates with better precision over a range of missing data mechanisms and extent of missingness. Moreover, using MI procedure provided a constant empirical power across scenarios. The power decreased using CC and IPW given increasing proportion of missingness across different types of miss data. Under simple yet realistically constructed scenarios, the MI approach may confer an advantage over IPW in MSMs applications.

Acknowledgement:

The OAI is a public-private partnership comprised of five contracts (N01-AR-2–2258; N01-AR-2–2259; N01-AR-2–2260; N01-AR-2–2261; N01-AR-2–2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Pfizer, Inc.; Novartis Pharmaceuticals Corporation; Merck Research Laboratories; and GlaxoSmithKline. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health.

Sources of financial support: This work was supported by the National Institute on Nursing Research (grant number 1R56NR015498–01A1 to Dr. Kate Lapane); the National Cancer Institute (Grant number 1R21CA198172 to Dr. Kate Lapane]; and the National Heart, Lung and Blood Institute (Contract number: HHSN268201000020C, Reference Number: BAA-NHLBI-AR1006).

References

- Robins J A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. Math Model. 1986;7(9–12):1393–1512. doi:10.1016/0270-0255(86)90088-6.
- 2. Robins JM, Hernán M a, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550–560. doi:10.1097/00001648-200009000-00011. [PubMed: 10955408]
- 3. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615–625. doi:10.1097/01.ede.0000135174.63482.43. [PubMed: 15308962]
- 4. Greenland S Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14(3):300–306. doi:10.1097/01.EDE.0000042804.12056.6C. [PubMed: 12859030]
- 5. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168(6):656–664. doi:10.1093/aje/kwn164. [PubMed: 18682488]
- Mortimer KM, Neugebauer R, Van Der Laan M, Tager IB. An application of model-fitting procedures for marginal structural models. Am J Epidemiol. 2005;162(4):382–388. doi:10.1093/aje/kwi208. [PubMed: 16014771]
- Lefebvre G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a marginal structural model. Stat Med. 2008;27(18):3629–3642. doi:10.1002/sim. 3200. [PubMed: 18254127]
- 8. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012;21(1):31–54. doi: 10.1177/0962280210386207. [PubMed: 21030422]
- 9. Brumback BA, Hernán MA, Haneuse SJPA, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med. 2004;23(5): 749–767. doi:10.1002/sim.1657. [PubMed: 14981673]
- 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. 2008;4(1):1–23. doi:10.2202/1557-4679.1106.
- 11. Vourli G, Touloumi G. Performance of the marginal structural models under various scenarios of incomplete marker's values: A simulation study. Biometrical J. 2015;57(2):254–270. doi:10.1002/bimj.201300159.
- Yang S, Eaton CB, Lu J, Lapane KL. Application of marginal structural models in pharmacoepidemiologic studies: A systematic review. Pharmacoepidemiol Drug Saf. 2014;23(6): 560–571. doi:10.1002/pds.3569. [PubMed: 24458364]
- Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. Comput Stat Data Anal. 2014;72:219–226. doi:10.1016/j.csda.2013.10.018. [PubMed: 24587587]
- 14. Liu S-H, Dubé CE, Driban JB, McAlindon TE, Eaton CB, Lapane KL. Patterns of intra-articular injection use after initiation of treatment in patients with knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthr Cartil. 2017;25(10):1607–1614. doi:10.1016/j.joca. 2017.05.023. [PubMed: 28627466]
- Lapane KL, Liu S-H, Dubé CE, Driban JB, McAlindon TE, Eaton CB. Factors Associated with the Use of Hyaluronic Acid and Corticosteroid Injections among Patients with Radiographically Confirmed Knee Osteoarthritis: A Retrospective Data Analysis. Clin Ther. 2017;39(2):347–358. doi:10.1016/j.clinthera.2017.01.006. [PubMed: 28139290]
- 16. Vignon E, Piperno M, Le Graverand MPH, et al. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: Comparison of standing anteroposterior

- and Lyon schuss views. Arthritis Rheum. 2003;48(2):378–384. doi:10.1002/art.10773. [PubMed: 12571846]
- 17. Dougados M, Hawker G, Lohmander S, et al. OARSI/OMERACT criteria of being considered a candidate for total joint replacement in knee/hip osteoarthritis as an endpoint in clinical trials evaluating potential disease modifying osteoarthritic drugs. J Rheumatol. 2009;36(9):2097–2099. doi:10.3899/jrheum.090365. [PubMed: 19738220]
- 18. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower ex. Arthritis Rheum. 2001;45(4):384–391. doi:10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0.. [PubMed: 11501727]
- 19. Greco NJ, Anderson AF, Mann BJ, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal art. Am J Sports Med. 2010;38(5):891–902. doi:10.1177/0363546509354163. [PubMed: 20044494]
- 20. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–592. doi:10.1093/biomet/63.3.581.
- Little RJA, Rublin D Statistical Analysis with Missing Data. New York, NY: John Wiley & Sons Inc.; 1987. doi:10.1002/9781119013563.
- 22. Goldfeld K simstudy: Simulation of Study Data. 2016 https://cran.r-project.org/package=simstudy.
- Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: AIDS Epidemiology, Methodological Issues.; 1992:297–331. doi: 10.1007/978-1-4757-1229-2_14.
- Robins JM, Rotnitzky A, Zhao LP. Estimation of regression-coefficients when some regressors are not always observed. J Am Stat Assoc. 1994;89:846–866.
- Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons Inc.; 2004.
- 26. van Buuren, Stef; Groothuis-Oudshoorn K. {mice}: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45(3):1–67.
- 27. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul. 2006. doi:10.1080/10629360600810434.
- 28. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Software; Vol 1, Issue 3 12 2011 https://www.jstatsoft.org/v045/i03.
- 29. Lee KJ, Carlin JB. Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. Am J Epidemiol. 2010. doi:10.1093/aje/kwp425.
- 30. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol. 2006;59(10):1092–1101. doi:10.1016/j.jclinepi. 2006.01.009. [PubMed: 16980150]
- 31. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res. 1999;8(1):3–15. doi: 10.1191/096228099671525676. [PubMed: 10347857]
- 32. Honaker James; King Gary; Blackwell M. {Amelia II}: A Program for Missing Data. J Stat Softw. 2011;45(7):1–47.
- 33. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychol Methods. 2002;7(2):147–177. doi:10.1037//1082-989X.7.2.147. [PubMed: 12090408]
- 34. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013;22(3):278–295. doi:10.1177/0962280210395740. [PubMed: 21220355]
- 35. Joffe MM, Pistilli M, Kempen JH. Marginal structural models for comparing alternative treatment strategies in ophthalmology using observational data. Ophthalmic Epidemiol. 2013;20(4):197–200. doi:10.3109/09286586.2013.792939. [PubMed: 23819465]
- 36. Velentgas P, Dreyer N, Nourjah P, Smith S, Torchia M. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Vol 12(13)-EHC.; 2013.

37. 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;29(4):431–443. doi: 10.1002/sim.3801. [PubMed: 20025082]

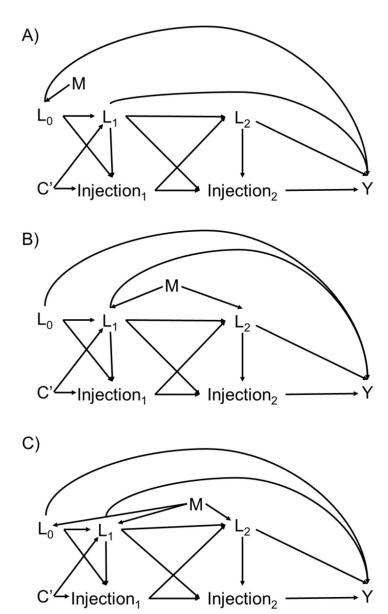


Figure 1 - Causal Diagrams Depicting Relationship in Simulated Datasets.

LEGEND: Data generation and missing mechanisms for simulation studies. L_0 indicates a set of confounders (e.g., Kellgren-Lawrence grade) measured at time t=0. L_1 and L_2 represent time-varying confounders (e.g., knee pain score) of the Injection-Y outcome association concurrently measured with exposure status at time t=1 and t=2. C' indicates an additional set of potential confounders. M indicates missing data (Panel A: The baseline confounder L_0 measured at time t=0 has missing data; Panel B: The time-varying confounders L_1 and L_2 measured at time t=1 and t=2 have missing data; and Panel C: Both baseline and time-varying confounders have missing data).

Table 1.

Values of Parameters for Data Generation Using Plasmode Simulation Framework and Osteoarthritis Initiative

Parameter	Meaning	Value
N	Sample size	500
В	Total simulations	1,000
m	Number of imputed data sets	5
Missing data mechanism for a time-independent or a time-varying confounder (M) *		
MCAR	Missing complete at random	10%, 20%, 30%
MAR	Missing at random	10%, 20%, 30%
MNAR §	Missing not at random	10%, 20%, 30%
$Pr(Injection_1 = 1)$	Probability of receiving intra-articular corticosteroid injection at time $t=1$	Empirical distribution from the constructed cohort: ~33.0%
β	Average true simulated effect: the average difference in one-year change in knee pain (Y) of receiving injection use for two years	1.2
Y	Predicted values of one-year change (from baseline to year 1) in knee pain depend on observed values of exposure status and confounders Predicted values of one-year change (from year 1 to year 2) in knee pain depend on observed values of exposure status and confounders	$Y_{I} = \beta_{0} + \beta_{1}(Injection_{I}) + \beta_{2}L_{0} + \beta_{3}L_{I}$ $+ \beta_{4}C' + \varepsilon$ $Y_{2} = \beta_{0} + \beta_{1}(Injection_{I}) +$ $\beta_{2}(Injection_{2}) + \beta_{3}L_{0} + \beta_{4}L_{I} + \beta_{5}C' + \varepsilon$
Е	Error term	~ <i>N</i> (0,1)
L	Confounders including predefined important time-independent (time $t=0$) confounders (L_0), time-varying (time $t=1$ and $t=2$) confounders (L_1 , L_2), and other pre-specified confounders (C ')	Empirical distribution from the constructed cohort

We introduced missing data within the context of a data source given the complete information from the measured covariates. The severity of knee OA status (Kellgren-Lawrence grade) measured at baseline (time t=0) was used as the time-independent confounder. Knee pain score measured at the time t=1 and t=2 of receiving injection use was used as time-varying confounders.

The probability of missing data for the time-independent confounder (Kellgren-Lawrence grade) was a joint function of observed covariates (i.e. age, sex, and time of initiating injection use) associated with each variable. For the time-varying confounder (knee pain measured concurrently with exposure status at time t=1 and t=2), we assumed missingness to jointly dependent on observed covariates using information from age, sex, household income, and race/ethnicity.

For the <u>time-independent confounder</u> (Kellgren-Lawrence grade) measured at time t = 0, we imposed the following missing data distributions. Among all participants missing information on Kellgren-Lawrence grade, 70% were randomly selected from the obese, 20% from overweight, and 10% from normal weight group. As such, participants with more obese were more likely to be missing given the original distribution from the constructed cohort.

For the <u>time-varying confounder</u> (knee pain score) measured concurrently with exposure status at time t = 1 and t = 2, we imposed the following missing data distributions. Among all participants missing information on knee pain scores concurrently measured with the injection use, 70% were randomly selected from the knee pain score 10, 20% from the knee pain score between 5–9, and 10% from knee pain score <5. As such, participants with more severe pain were more likely to be missing given the original distribution from the constructed cohort.

Author Manuscript

Liu et al. Page 15

Table 2.

Missing Data in a Time-independent Confounder*: Comparison of Percent Bias, Mean Squared Error, and Empirical Power of Methods for Handling Missing Data (Complete Case, Multiple Imputation, Inverse Probability Weighting) Under Various Mechanisms and Extent of Missingness.

Missing		Bias (%)		M	Mean Squared Error	ror		Empirical power	er
mechansm and % missing data	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting
Missing completely at random (MCAR)									
10%	2.3	3.6	2.3	0.19	0.17	0.19	94.6	96.3	94.6
20%	5.9	4.3	2.9	0.21	0.18	0.21	94.3	6.96	94.3
30%	2.5	3.4	2.4	0.24	0.18	0.24	92.5	92.6	92.5
Missing at random (MAR)									
10%	2.9	3.8	2.9	0.20	0.18	0.20	95.1	96.1	95.1
20%	2.5	4.6	2.5	0.21	0.18	0.21	93.9	96.9	93.9
30%	3.5	5.5	3.5	0.25	0.18	0.25	92.1	97.0	92.0
Missing not at random (MNAR)									
10%	-0.4	1.2	1.3	0.19	0.17	0.19	93.7	95.8	93.9
20%	-1.2	3.1	0.5	0.21	0.18	0.20	92.2	96.4	92.9
30%	-1.5	2.0	0.3	0.23	0.18	0.23	0.06	95.4	91.1

 $\stackrel{*}{*}$ Time-independent confounder (Kellgren-Lawrence grade) measured at time t=0.

Table 3.

Missing Data in a Time-varying Confounder*: Comparison of Percent Bias, Mean Squared Error, and Empirical Power of Methods for Handling Missing Data (Complete Case, Multiple Imputation, Inverse Probability Weighting) Under Various Mechanisms and Extent of Missingness.

Missing		Bias (%)		M	Mean Squared Error	rror	[Empirical power	er
mechansm and % missing data	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting
Missing completely at random (MCAR)									
10%	3.1	3.6	3.4	0.22	0.17	0.22	93.6	2.96	93.6
20%	3.7	3.4	4.1	0:30	0.17	0:30	5.88	6.3	88.3
30%	6.2	3.3	2.9	0.45	0.17	0.45	4.67	9.96	5.08
Missing at random (MAR)									
10%	2.9	3.9	3.1	0.22	0.18	0.23	<i>1</i> .26	6.3	92.0
20%	4.5	3.7	5.0	0.31	0.18	0.31	0.68	6.96	2.88
30%	8.4	3.6	8.0	0.40	0.18	0.41	9:88	96.2	82.1
Missing not at random (MNAR)									
%01	-3.8	5.0	-1.2	0.23	0.18	0.27	<i>L</i> .68	9.96	88.3
20%	-5.4	4.6	-1.3	0.25	0.18	0:30	88.3	5.79	87.4
30%	7.4-	4.1	-2.2	0.28	0.18	0.34	87.0	8.96	85.0

 $_{\rm r}^*$ Time-varying confounder; knee pain score measured concurrently as the exposure status at time t=1 and 2.

Table 4.

Missing Data in Either a Time-independent, Time-varying, or Confounders: Comparison of Percent bias, Mean Squared Error, and Empirical Power of Methods for Handling Missing Data (Complete Case, Multiple Imputation, Inverse Probability Weighting) Under Various Mechanisms and Extent of Missingness.

Missing		Bias (%)		Me	Mean Squared Error	rror		Empirical power	ıa
mecnamsm and % missing data	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting	Complete Case	Multiple Imputation	Inverse Probability Weighting
Missing completely at random (MCAR)									
10%	4.6	4.9	3.8	0.26	0.18	0.26	91.9	96.4	92.2
20%	8.2	4.1	6.2	0.40	0.18	0.40	75.7	97.1	83.1
30%	12.5	4.7	7.4	0.58	0.18	0.51	72.4	6.96	74.6
Missing at random (MAR)									
10%	3.6	4.3	3.7	0.26	0.18	0.26	91.3	95.5	91.4
20%	8.1	4.1	7.2	0.38	0.18	0.38	85.0	95.9	84.3
30%	11.7	5.0	7.8	0.56	0.18	0.53	75.2	9.96	73.1
Missing not at random (MNAR)									
10%	-0.8	4.7	0.1	0.24	0.18	0.28	90.4	96.8	87.8
20%	-1.3	4,4	-2.1	0.31	0.18	0.37	85.0	97.0	81.9
30%	-1.9	6.7	-5.3	0.35	0.18	0.43	82.0	97.6	75.3

Page 17