



Practice of Epidemiology

The Impact of Sparse Follow-up on Marginal Structural Models for Time-to-Event Data

Nassim Mojaverian, Erica E. M. Moodie*, Alex Bliu, and Marina B. Klein

* Correspondence to Dr. Erica E. M. Moodie, Department of Epidemiology, Biostatistics, and Occupational Health, Faculty of Medicine, McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, QC H4B 2V2, Canada (e-mail: erica.moodie@mcgill.ca).

Initially submitted September 3, 2014; accepted for publication June 3, 2015.

The impact of risk factors on the amount of time taken to reach an endpoint is a common parameter of interest. Hazard ratios are often estimated using a discrete-time approximation, which works well when the by-interval event rate is low. However, if the intervals are made more frequent than the observation times, missing values will arise. We investigated common analytical approaches, including available-case (AC) analysis, last observation carried forward (LOCF), and multiple imputation (MI), in a setting where time-dependent covariates also act as mediators. We generated complete data to obtain monthly information for all individuals, and from the complete data, we selected “observed” data by assuming that follow-up visits occurred every 6 months. MI proved superior to LOCF and AC analyses when only data on confounding variables were missing; AC analysis also performed well when data for additional variables were missing completely at random. We applied the 3 approaches to data from the Canadian HIV–Hepatitis C Co-infection Cohort Study (2003–2014) to estimate the association of alcohol abuse with liver fibrosis. The AC and LOCF estimates were larger but less precise than those obtained from the analysis that employed MI.

available-case analysis; last observation carried forward; marginal structural models; missing data; multiple imputation; survival analysis

Abbreviations: AC, available-case; APRI, aspartate aminotransferase:platelet ratio index; CCC, Canadian HIV–Hepatitis C Co-infection Cohort; HCV, hepatitis C virus; HIV, human immunodeficiency virus; LOCF, last observation carried forward; MCAR, missing completely at random; MI, multiple imputation; MSM, marginal structural model.

Many researchers are interested in studying influences on the amount of time taken to reach a certain endpoint, often death. When studying time-dependent exposures, additional challenges are posed by time-dependent covariates that act simultaneously as confounders and mediators. Three types of bias can arise in this context: bias due to confounding, if one doesn’t control for a variable that is a common cause of the exposure and outcome (1); blocking of the total effect by conditioning on an intermediate variable (2); and collider stratification bias, which can occur when one conditions on an intermediate variable in situations where there is an unmeasured variable that causes both the outcome and that intermediate variable (3). The most commonly used method of addressing these challenges is the marginal structural model (MSM) (4), frequently fitted via inverse probability weighting.

When fitting MSMs for time-to-event data, a discrete-time approximation is often used (4–8); the resulting odds ratio

will be a good approximation of the hazard ratio if the per-interval event rate is low (9). Using shorter time intervals in order to have fewer events per interval is a means of improving the discrete-time approximation. However, this poses a key challenge if the desired interval frequency exceeds that of the actual observation times, by inducing a considerable missing-data problem. It is not unusual to see analyses in which the last observation carried forward (LOCF) method is used to impute data for between-visit covariates in a discrete-time setting. While it alleviates the issue of missing data, this approach may induce bias.

Andersen and Liestøl (10) noted attenuation of regression coefficients in Cox proportional hazards models when time-varying covariates were measured infrequently. In fitting a marginal structural Cox model, the possibly outdated covariates are not included in the outcome model but rather are used in the weighting models, where the impact of measurement error is unpredictable (11).

In this paper, we examine the impact of the LOCF technique and compare it with 2 alternative approaches: the available-case (AC) approach and multiple imputation (MI). We demonstrate these approaches using data from the Canadian HIV–Hepatitis C Co-infection Cohort (CCC) Study to examine the impact of alcohol abuse on time to development of liver fibrosis, as measured by the aspartate aminotransferase:platelet ratio index (APRI), in a population coinfecting with human immunodeficiency virus (HIV) and hepatitis C virus (HCV).

MISSING DATA: TYPES AND APPROACHES TO MITIGATE ITS EFFECTS

Missing data are common in studies whose subjects are followed over time, but they can also occur in other settings—for example, if participants refuse to answer a subset of questions on a questionnaire. There are 3 classes of missing data: missing completely at random (MCAR), missing at random, and missing not at random (12). In MCAR data, the probability of a measurement's being missing is independent of both observed and unobserved measurements. Under MCAR, estimates from most standard analysis approaches are unbiased. However, there will almost surely be a loss of statistical power. In the “missing at random” situation, the probability of a measurement's being missing depends on observed data but not on the unobserved data. In the last case, missing not at random, the value of the missing measurement itself predicts the probability of data being missing; in general, this can only be corrected by making strong, untestable assumptions about the distribution of the missing values or the missingness mechanism.

There are several approaches to missing data. Here, we consider 3: AC analysis (i.e., analyzing only the observed data), LOCF, and MI.

AC analysis

It is not uncommon for analysts to remove records with missing values; for many analytical software programs, this is the default approach. Unless data are MCAR, AC analyses can yield seriously biased estimators. When the data are MCAR, the estimators will not exhibit bias but can suffer from low precision.

Last observation carried forward

The LOCF method simply uses the last recorded value to impute the missing value(s). This approach preserves the sample size but can yield biased estimators, even under MCAR; the direction of the bias varies across settings (13–15).

Multiple imputation

MI is a Monte Carlo method in which missing values are replaced in m repeated simulations (16). MI is commonly used in standard regression settings but is less frequently applied in MSMs, despite some evidence to suggest its utility (17).

To fully describe MI, we must first introduce some terminology. Let Q be the estimator of interest, such as the causal log odds ratio from the MSM. Y denotes the outcome and is

partitioned into 2 parts, Y_{mis} and Y_{obs} . $\hat{Q} = \hat{Q}(Y_{\text{mis}}, Y_{\text{obs}})$ will estimate Q if complete data are available. $U = U(Y_{\text{mis}}, Y_{\text{obs}})$ is the squared standard error of the estimator. Assume that in the presence of complete data, $(\hat{Q} - Q)/\sqrt{U} \sim N(0, 1)$.

In MI, $m > 1$ independent imputations are used to fill in Y_{mis} , indexed by l . The imputation-specific estimates are $\hat{Q}^{(l)} = \hat{Q}(Y_{\text{obs}}, Y_{\text{mis}}^{(l)})$ and $U^{(l)} = U(Y_{\text{obs}}, Y_{\text{mis}}^{(l)})$, $l = 1, \dots, m$. The overall estimate of Q is given by $\bar{Q} = \sum \hat{Q}^{(l)}/m$ (i.e., the simple average of the estimates resulting from each of the m analyses of the completed data sets), and the standard error for \bar{Q} is $T = (1 + m^{-1})B + \bar{U}$, where $B = \sum (\hat{Q}^{(l)} - \bar{Q})^2 / (m - 1)$ is the between-imputation variance and $\bar{U} = \sum U^{(l)}/m$ is the within-imputation variance. An overview of MI by chained equations is provided in the Web Appendix, available at <http://aje.oxfordjournals.org/>.

Unlike the complete-case approach, MI preserves the sample size by ensuring that no individuals are dropped from the study due to incomplete measurement, a feature also shared by the LOCF method. However, unlike LOCF (or an AC analysis), MI can yield an unbiased treatment effect estimator provided that data are missing at random and the models used to perform the imputation are correctly specified.

SIMULATION STUDY

We conducted a simulation study to examine the performance of 3 different analytical approaches to missing data in MSMs for time-to-event data performed via weighted pooled logistic regression. We used the “gold standard” analysis of complete data, AC analysis (observed data), and LOCF and MI to impute missing values in our simulated data sets. The bias, standard error, and root mean squared error of the treatment effect estimator were used as metrics to compare the 3 analytical approaches.

Methods

Data generation. We used the data-generating algorithm proposed by Young et al. (18). Let n be the number of subjects and M the maximum possible number of observation times. T is the failure time, and $Y_{m'}$ is an indicator for failure by time m' . Treatment during the interval $[m', m' + 1)$ is denoted $A_{m'}$; $L_{m'}$ is the binary time-varying confounder measured at the start of interval $[m', m' + 1)$; and T_0 is the counterfactual survival time if an individual is never exposed. Survival data consistent with a Cox MSM can then be generated according to the following procedure.

Step 1: T_0 is drawn from an exponential distribution with rate $\lambda_0 = 0.01$. That is, λ_0 represents the rate of events (e.g., liver fibrosis) in the absence of exposure. Set $L_{-1} = A_{-1} = Y_0 = 0$. For each m' , we follow steps 2–4.

Step 2: Generate confounder $L_{m'}$ from a binomial distribution with

$$\begin{aligned} \text{logit}[P(L_{m'} = 1 | \bar{L}_{m'-1}, \bar{A}_{m'-1}, Y_{m'} = 0)] \\ = \beta_0 + \beta_1 I[T_0 < c] + \beta_2 A_{m'-1} + \beta_3 L_{m'-1}, \end{aligned}$$

where $c = 30$ as in the study by Young et al. (18).

Step 3: Generate the treatment $A_{m'}$ from a binomial distribution with

$$\begin{aligned} \text{logit}[P(A_{m'} = 1 | \bar{L}_{m'-1}, \bar{A}_{m'-1}, Y_{m'} = 0)] \\ = \alpha_0 + \alpha_1 L_{m'} + \alpha_2 L_{m'-1} + \alpha_3 A_{m'-1}. \end{aligned}$$

Step 4: For generating the event indicator $Y_{m'}$ and survival time T , we have

- If $T_0 > \int_0^{m'+1} \exp(\varphi A_j) dj$, then $Y_{m'+1} = 0$
- If $T_0 \leq \int_0^{m'+1} \exp(\varphi A_j) dj$, then $Y_{m'+1} = 1$ and $T \in (m', m' + 1]$ with

$$T = m' + (T_0 - \int_0^{m'} \exp(\varphi A_j) dj \exp(-\varphi A_{m'})),$$

where φ is the (marginal) causal treatment effect. Thus, we modify the counterfactual treatment-free survival to account for observed treatment and see whether the resulting survival time lies within the interval $[m', m' + 1)$.

Parameter values were chosen based on the example provided by Young et al. (18). Specifically, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (\log(3/7), 2, \log(0.5), \log(1.5))$, and $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (\log(2/7), 0.5, 0.5, \log(4))$; 500 data sets were considered for each scenario.

Three different values were considered for the treatment effect ($\varphi = 0, -\log(3), \log(2)$), so as to consider the null case, a situation where survival time is decreased by the exposure, and a situation where it is increased by the exposure. See Web Figures 1 and 2 for a causal diagram of the data-generating procedure.

Conventionally, it has been seen as desirable to have short intervals with fewer events, as this enables the analyst to apply pooled logistic regression to approximate the hazard ratio in an MSM instead of using a Cox regression model. Thus, complete data were generated using the above algorithm to obtain monthly information for all individuals. However, we assumed that participants were followed up only every 6 months, and therefore we “pretended” to see only every sixth observation of the time-dependent confounder throughout the simulation study. Thus, we kept every sixth observation in a vector of length of the total number of intervals and deleted the between-visit observations; data were MCAR. In initial simulations, only data on confounders were missing. This assumption is not unreasonable, since for many exposures (particularly pharmacological exposures), records are available to confirm treatment between follow-up visits. Additionally, we considered situations where data on both the confounding variables and the exposure variables were missing and where, in addition to missing 5 out of 6 visits, outcome data were missing for a random 15% of the sample.

Analysis. Four analyses were undertaken. First, we analyzed the complete data. This type of analysis provides results in the ideal setting in which observations are made at the same intervals as those used in the analysis, that is, at a monthly frequency. Second, we analyzed available data only. With this approach, we analyzed the data from every sixth observation and ignored the visits at which confounders were not measured. Third, we analyzed the data using LOCF to impute the missing values. Finally, in the fourth analysis we used MI to fill in the

missing values: MI was carried out with the use of the *mice* function in R (R Foundation for Statistical Computing, Vienna, Austria) (19), with logistic regression used to model the missing data, since all variables were binary. The imputation was carried out with the data in long format, where each row of the data set represented a person-visit; the subject identification number and interval number were included in the model to account for the clustering in the data. All available information at the current visit was used in the imputation model. Thus, when only confounding information was missing, the current and previous-interval treatment, previous-interval confounder, interval number, and subject identification number were all included as linear terms in the model. Note, however, that the indicator variable $I[T_0 < c]$ was not included in the imputation model; while this indicator is part of the data-generating mechanism, the treatment-free survival, T_0 , is an unmeasured quantity. In the simulations where treatment information was missing, the logistic imputation model for treatment depended linearly on the current and previous-interval confounding variable, previous-interval treatment, interval number, and subject identification number. Missing outcomes were imputed using a logistic imputation model that was linear in terms of the current- and previous-interval confounding variables, current- and previous-interval treatment, interval number, and subject identification number.

MSM parameters were estimated via inverse probability weighting, thereby creating a pseudopopulation in which exposures were not predicted by the time-dependent confounders included in the weighting models.

Let $\bar{A}_k = (A_0, A_1, \dots, A_k)$ denote exposure history over intervals 0– k , and similarly let L denote all relevant confounding preexposure variables in each interval. Further, let $Y_k = 1$ if a subject experiences the event in interval k and $Y_k = 0$ otherwise; similarly, $C_k = 1$ if the subject was lost to follow-up by interval k and $C_k = 0$ otherwise. In our simulations, stabilized weights,

$$\begin{aligned} sw_{ik} \\ = \frac{\prod_{t=0}^k pr(A_t = a_{it} | \bar{A}_{t-1} = \bar{a}_{i,t-1}, C_{t-1} = Y_{t-1} = 0)}{\prod_{t=0}^k pr(A_t = a_{it} | \bar{A}_{t-1} = \bar{a}_{i,t-1}, \bar{L}_{t-1} = \bar{l}_{i,t-1}, C_{t-1} = Y_{t-1} = 0)}, \end{aligned}$$

are used, where we take $\bar{A}_{-1} = 0$. In the absence of any censoring or loss to follow-up, an unbiased estimator of the marginal effect of the exposure on the outcome can be obtained by regressing the binary outcome on (some function of) the exposure history and any baseline covariates used in the numerator of the stabilized weights, weighting each person-observation by sw_{ik} .

If censoring is purely administrative, as in our simulation study, no changes to the modeling procedure outlined above are required. If, however, censoring depends on measured covariates (as in the analysis presented in the following section), we may treat censoring like another time-varying exposure, provided there are no unmeasured confounders for both treatment and censoring. We must update the weights in the following fashion, taking the subject-specific weight to be $sw_{ik} \times sw'_{ik}$, where

$$\begin{aligned} sw'_{ik} \\ = \frac{\prod_{t=0}^k pr(C_t = 0 | \bar{A}_{t-1} = \bar{a}_{i,t-1}, C_{t-1} = Y_{t-1} = 0)}{\prod_{t=0}^k pr(C_t = 0 | \bar{A}_{t-1} = \bar{a}_{i,t-1}, \bar{L}_{t-1} = \bar{l}_{i,t-1}, C_{t-1} = Y_{t-1} = 0)}. \end{aligned}$$

For each of the 4 analyses, we report and compare bias, variability, and the combined measure of the root mean squared error.

Results

From Tables 1–3, it is evident that variability tends to dominate bias in all analyses. This is perhaps unsurprising, since the data are MCAR and the assumed follow-up scheme is independent of all individual-level covariates. In the setting where only confounding variables are unobserved, MI performs very well. This pattern does not hold across the other

2 scenarios. When the exposure was also missing, none of the 3 approaches was uniformly the best. MI appeared to suffer when data on the outcome were missing, particularly in comparison with the AC analysis; however, this seemingly surprising finding can be explained: Because the data were MCAR, the AC analysis would be expected to be unbiased but inefficient. While the MI analysis imputes, on average, the

Table 1. Simulation Results Based on Complete Data Generated Using Young et al.'s (18) Algorithm, Assuming That Every Sixth Confounder Measurement Is Observed and the Remaining 5 Measurements Are Missing

φ , Sample Size (n), and Statistic	Analytical Approach			
	Complete Data	Available-Case ^a	LOCF	Multiple Imputation
0				
100				
Bias	0.012	0.067	0.069	0.089
SD	0.302	0.659	0.282	0.258
rMSE	0.302	0.663	0.290	0.273
500				
Bias	-0.035	-0.309	0.033	0.059
SD	0.332	2.235	0.318	0.294
rMSE	0.334	2.256	0.320	0.300
-log(3)				
100				
Bias	-0.018	0.061	0.055	0.068
SD	0.165	0.276	0.151	0.139
rMSE	0.166	0.282	0.161	0.155
500				
Bias	0.012	0.067	0.069	0.089
SD	0.302	0.659	0.282	0.258
rMSE	0.302	0.663	0.290	0.273
log(2)				
100				
Bias	0.000	0.056	0.068	0.083
SD	0.141	0.230	0.126	0.117
rMSE	0.141	0.237	0.144	0.144
500				
Bias	0.031	0.086	0.093	0.108
SD	0.129	0.262	0.136	0.116
rMSE	0.132	0.275	0.164	0.159

Abbreviations: LOCF, last observation carried forward; rMSE, root mean squared error; SD, standard deviation.

^a Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 0; (0, 500): 0; (-log(3), 100): 8; (-log(3), 500): 0; (log(2), 100): 1; (log(2), 100): 0.

Table 2. Simulation Results Based on Complete Data Generated Using Young et al.'s (18) Algorithm, in Which Every Sixth Measurement of Both the Confounder and the Exposure Are Observed and the Remaining 5 Measurements Are Missing

φ , Sample Size (n), and Statistic	Analytical Approach			
	Complete Data	Available-Case ^a	LOCF	Multiple Imputation ^b
0				
100				
Bias	0.013	0.090	0.046	0.002
SD	0.305	0.637	0.528	0.225
rMSE	0.305	0.643	0.530	0.225
500				
Bias	-0.005	0.050	0.062	0.003
SD	0.139	0.238	0.231	0.131
rMSE	0.139	0.243	0.239	0.131
-log(3)				
100				
Bias	-0.037	-0.043	0.885	0.898
SD	0.354	0.699	0.639	0.265
rMSE	0.356	0.700	1.092	0.936
500				
Bias	-0.023	0.017	0.852	0.912
SD	0.170	0.287	0.280	0.163
rMSE	0.172	0.288	0.897	0.926
log(2)				
100				
Bias	0.043	0.161	-0.422	-0.574
SD	0.284	0.666	0.509	0.208
rMSE	0.287	0.685	0.662	0.611
500				
Bias	0.039	0.097	-0.437	-0.575
SD	0.123	0.250	0.173	0.140
rMSE	0.129	0.268	0.470	0.592

Abbreviations: LOCF, last observation carried forward; rMSE, root mean squared error; SD, standard deviation.

^a Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 0; (0, 500): 0; (-log(3), 100): 13; (-log(3), 500): 0; (log(2), 100): 1; (log(2), 100): 0.

^b Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 8; (0, 500): 2; (-log(3), 100): 6; (-log(3), 500): 5; (log(2), 100): 6; (log(2), 100): 1.

Table 3. Simulation Results Based on Complete Data Generated Using Young et al.'s (18) Algorithm, in Which Every Sixth Measurement of Both the Confounder and the Exposure Are Observed, the Remaining 5 Measurements Are Missing, and Data on 15% of Outcomes Are Missing

φ , Sample Size (n), and Statistic	Analytical Approach			
	Complete Data	Available-Case ^a	LOCF ^b	Multiple Imputation ^c
0				
100				
Bias	0.009	0.072	0.057	0.129
SD	0.293	0.579	0.746	0.662
rMSE	0.293	0.584	0.748	0.674
500				
Bias	0.004	0.076	0.048	0.118
SD	0.135	0.250	0.356	0.338
rMSE	0.135	0.261	0.359	0.358
−log(3)				
100				
Bias	−0.030	−0.025	0.225	0.023
SD	0.360	0.730	0.953	0.925
rMSE	0.361	0.730	0.979	0.925
500				
Bias	−0.019	0.046	0.281	0.199
SD	0.163	0.277	0.399	0.458
rMSE	0.164	0.281	0.488	0.500
log(2)				
100				
Bias	0.054	0.145	−0.163	0.209
SD	0.286	0.638	0.640	0.789
rMSE	0.291	0.654	0.660	0.816
500				
Bias	0.029	0.087	−0.203	0.188
SD	0.123	0.267	0.250	0.493
rMSE	0.126	0.281	0.322	0.527

Abbreviations: LOCF, last observation carried forward; rMSE, root mean squared error; SD, standard deviation.

^a Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 1; (0, 500): 0; (−log(3), 100): 13; (−log(3), 500): 0; (log(2), 100): 0; (log(2), 100): 0.

^b Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 0; (0, 500): 0; (−log(3), 100): 2; (−log(3), 500): 0; (log(2), 100): 1; (log(2), 100): 0.

^c Some estimates were excluded when calculating the performance statistics (bias, SD, and rMSE) because of an excessively high (≥ 10) odds ratio; for each (φ , n) pair, the following numbers of estimates were excluded: (0, 100): 1; (0, 500): 1; (−log(3), 100): 13; (−log(3), 500): 3; (log(2), 100): 1; (log(2), 100): 1.

correct number of outcomes, they are frequently imputed “too early” in a simulated participant’s lifetime. In all analyses, the confounder imputation models are imperfectly specified,

since the latent variable, $I[T_0 < c]$, is not available to the analyst. Finally, the data were imputed in “long format,” which does not take full advantage of the longitudinal nature of the data; because of the exceptionally high rate of missingness and the very small number of variables (the simulated data contained only 1 confounder, an exposure, and an outcome), the imputation procedure failed with the data in “wide format” (a common problem when predictors in the imputation model are collinear). There is therefore no method that emerges as a clear winner: When the imputation procedure is not taxed by too high a rate of missing information, it performs well. When data are MCAR, results from AC analysis are unbiased but inefficient. Of course, in practice, if the event rate in an AC analysis is high, this analytical approach will be undesirable, since the odds ratio will serve as a poor approximation of the hazard ratio.

THE ASSOCIATION OF ALCOHOL ABUSE WITH LIVER FIBROSIS: AN ANALYSIS OF DATA FROM THE CCC STUDY

The CCC Study

Data were obtained from the CCC Study, a cohort study of a Canadian population coinfectd with HIV and HCV. Recruitment began in 2003; to date (2014), 1,153 patients have been enrolled from 17 sites across Canada. Eligible patients were at least 16 years of age with documented HIV infection and chronic HCV infection or evidence of HCV exposure. At each of the follow-up visits, which are scheduled to take place every 6 months, participants fill out a questionnaire, supplementary information is extracted from their medical records, and blood tests are performed by the research personnel (20). The primary objective of this cohort study is to investigate the association between antiretroviral therapy and progression to end-stage liver disease among persons coinfectd with HIV and HCV; the researchers have also examined the contributions of social factors, toxicities, and immunological factors that may modify the progression of liver fibrosis (20–22).

Outcome, exposure, and confounding variables

We examined the marginal association between alcohol abuse and the development of liver fibrosis as measured by an APRI score of at least 1.5, using the methods described above to address information from missed visits. APRI is a non-invasive surrogate for liver fibrosis and is defined as $100 \times (\text{aspartate aminotransferase/upper limit of normal})/\text{platelet count}$ ($10^9/\text{L}$) (23, 24). $\text{APRI} \geq 1.5$ has been validated as a marker of significant liver fibrosis in coinfectd patients (23, 25).

Alcohol abuse is highly associated with liver fibrosis, and therefore it was chosen as the exposure to demonstrate the potential impact of the various approaches for handling missing data. Alcohol abuse was defined as self-reported alcohol intake of more than 2 drinks per day or binge drinking (>6 drinks at any one time). Alcohol abuse may be predicted by factors such as previous injection drug use and smoking, while also being a cause of future injection drug use and smoking; injection drug use and smoking may themselves also predict liver

fibrosis. Similarly, CD4-positive T lymphocyte (CD4 cell) count and HIV viral load may also act as both time-dependent confounders and mediators. MSMs are thus an appropriate modeling choice for accounting for such factors, which act as both confounding variables and mediating variables. See Web Figure 3 for a simplified causal diagram of the CCC data.

Statistical analysis

Investigation of the association between alcohol abuse in the past 6 months and liver fibrosis was undertaken using MSMs, fitted via pooled logistic regression with inverse probability weighting to account for covariate imbalances in both the exposure and the censoring. The following variables were used to create a model with which to estimate the denominator of the stabilized weights for both the treatment and censoring models: baseline age, baseline ln(APRI), sex, an indicator for aboriginal ethnicity, an indicator for being HCV RNA-positive at baseline, duration of HCV infection, an indicator for being hepatitis B surface antigen-positive at baseline, lagged alcohol abuse, lagged CD4 cell count (per 100 cells/ μ L), lagged HIV viral load (log copies/mL), an indicator for lagged injection drug use, an indicator for lagged receipt of antiretroviral treatment, and an indicator for lagged smoking. The numerator model included only the variables that were time-invariant. The exposure model was fitted using binomial logistic regression, whereas the censoring model was fitted using a multinomial logistic regression model to account for censoring via 1) HCV treatment, 2) death, or 3) other causes. All variables were included as linear terms; see Web Table 1 for additional details.

MI was carried out in such a way as to recognize the longitudinal nature of the data. In particular, the data were imputed in a “wide format,” with 1 row of data per participant (in contrast to the “long format,” where each row corresponds to 1 person-visit). The imputation models were built forward in time, imputing from the first missing interval to the last, to acknowledge the temporally ordered structure of the data.

Predictors in the imputation models included all relevant variables from the previous intervals as well as all time-invariant covariates. For example, CD4 cell count in interval j was imputed on the basis of previous CD4 cell counts, HIV viral load, use of/interruptions in antiretroviral treatment, injection drug use, smoking status, alcohol abuse, and all time-invariant variables. The APRI scores were also used in the imputation model, and APRI scores were themselves imputed. The imputation model also included an indicator of whether APRI was greater than or equal to 1.5 in the observed data. Following imputation, the event time was recalculated and was taken to be the first instance in which either the imputed APRI or the observed APRI was greater than or equal to 1.5. Continuous variables were imputed via predictive mean matching, and categorical variables were imputed via logistic regression, using all other covariates as linear terms without interactions in the logistic models.

We considered the impact of the number of imputations on the stability of the resulting estimates, varying the number of imputations from 5 to 25 in increments of 5. We found that estimates were similar whether we used 20 imputations or 25, so we chose the lower of these numbers to reduce the computational burden of the analysis (which was considerable

because of the large sample size and the approach used to compute confidence intervals; see below).

We used a nonparametric bootstrap to derive confidence intervals for the treatment effect estimate which fully accounted for the variability of the missing-data procedure and the estimation of the weights. The resampling process was performed on the individual, rather than on the visit.

Results

A total of 1,107 participants in the CCC Study had not received HCV treatment at baseline, among whom 224 had baseline APRI ≥ 1.5 and were not included in any of the analyses. A total of 843 participants had observed baseline APRI

Table 4. Risk of Liver Fibrosis Associated With Alcohol Abuse in the Canadian HIV–Hepatitis C Co-infection Cohort Study, 2003–2014^a

Analytical Approach and Variable	Odds Ratio	95% Confidence Interval
Available-case		
Alcohol abuse	1.55	0.91, 2.38
Age at baseline ^b	1.03	0.93, 1.19
Baseline ln(APRI)	7.29	4.15, 19.61
Female sex	1.85	0.98, 2.84
Aboriginal ethnicity	0.64	0.21, 1.07
HCV RNA-positive at baseline	1.42	0.86, 4.79
Duration of HCV infection ^b	1.00	0.89, 1.25
HBsAg-positive at baseline	0.99	0.00, 1.91
LOCF		
Alcohol abuse	1.74	0.65, 2.77
Age at baseline ^b	0.99	0.84, 1.18
Baseline ln(APRI)	6.20	1.98, 12.02
Female sex	1.30	0.90, 2.97
Aboriginal ethnicity	0.86	0.33, 1.36
HCV RNA-positive at baseline	1.56	0.73, 2.26
Duration of HCV infection ^b	1.00	0.91, 1.16
HBsAg-positive at baseline	0.88	0.00, 1.18
Multiple imputation		
Alcohol abuse	1.32	1.11, 1.69
Age at baseline ^b	0.99	0.85, 1.09
Baseline ln(APRI)	4.86	3.02, 8.12
Female sex	1.26	0.89, 1.57
Aboriginal ethnicity	1.02	0.63, 1.48
HCV RNA-positive at baseline	1.12	0.94, 2.09
Duration of HCV infection ^b	1.00	0.97, 1.09
HBsAg-positive at baseline	1.15	0.68, 2.09

Abbreviations: APRI, aspartate aminotransferase:platelet ratio index; HBsAg, hepatitis B surface antigen; HCV, hepatitis C virus; LOCF, last observation carried forward.

^a The hazard ratio is approximated by the odds ratio obtained from a weighted pooled logistic regression.

^b Age and duration of HCV infection were scaled so that the coefficient was associated with a change of 5 years.

scores less than 1.5 and were used in the LOCF analysis, while the MI analysis also included those participants with missing baseline APRI, bringing the sample size up to 883. The AC analysis included only 750 participants. There were 951 missed visits, and 4,216 records were observed in total.

The average estimated coefficients from the inverse weighting models in the MI analysis can be found in Web Tables 1–4. Previous alcohol abuse and previous smoking proved to be the strongest predictors of current alcohol abuse, while previous alcohol abuse and aboriginal ethnicity were the strongest predictors of censoring for reasons other than death or HCV treatment initiation; unsurprisingly, APRI score was the strongest predictor of initiating HCV treatment. These broad patterns were also observed in the AC and LOCF analyses.

Table 4 gives hazard ratios, approximated by the odds ratios from pooled logistic regression, for the association of alcohol abuse (and baseline covariates) with liver fibrosis, accompanied by 95% bootstrap confidence intervals. The odds ratio associated with alcohol abuse was largest under both AC and LOCF analysis, though it was not significantly associated with liver fibrosis in either case. The odds ratio associated with alcohol abuse was smallest in the MI analysis; however, its confidence interval was also the narrowest and excluded 1. Results obtained following truncation of the weights are provided in Web Table 5.

DISCUSSION

MI has been suggested in numerous contexts to be a useful tool for addressing missing data (17, 26, 27). It can be implemented with greater ease than many likelihood-based approaches such as the expectation-maximization algorithm, while offering substantial reductions in bias over naive methods such as AC or LOCF analysis. Our work suggests that this may be the case for MSM analyses of time-to-event outcomes; however, AC analyses may also prove a reasonable choice when data are sparse due to infrequently scheduled follow-up visits (and thus MCAR).

MI was demonstrated to be a reliable approach to addressing missing confounder information for longitudinal data in situations where analyses are performed using MSMs fitted by inverse probability weighting in a pooled logistic regression. Whereas AC methods omit a potentially large number of person-visits and LOCF will omit anyone with missing baseline information, imputation can lead to increased precision by preserving all person-visits. MI proved less reliable in simulations where the exposure, confounders, and outcome were all subject to missingness. Care must be taken to evaluate the imputation models and to ensure that a sufficient number of imputations are used.

Our simulation results can be compared with those of Vourli and Touloumi (28), who also recently studied the impact of missing confounder information on Cox MSMs under a variety of settings. As in our study, the authors found that MI did not perform particularly well when there was a high proportion of missed visits, even when the missingness mechanism was completely at random (28). Viourli and Touloumi also considered LOCF and an approach based on inverse-probability-of (not being)-missing weighting. LOCF performed poorly in all scenarios, whereas the inverse weighting approach performed

better than MI in some circumstances (28). The authors did not investigate settings where, in addition to missing confounding information, missingness was present in both the exposure and the outcome.

In our analysis of the impact of alcohol abuse on the development of liver fibrosis, only the MI approach yielded a confidence interval that excluded 1. In this example, exposure was determined by the cohort participants, likely on the basis of their current circumstances rather than on information collected by researchers. However, there are situations in which the exposure is determined based on mismeasured or outdated covariates, such as the last available measurement. In this very particular case, the analyst *should* use LOCF so that the variables used in the treatment allocation process are those used to construct the inverse probability weights (R. P. Kyle, McGill University, personal communication, 2015 (unpublished manuscript)).

Unlike our simulations, where missingness was completely at random, our analysis of the CCC Study filled in information from missing *scheduled* visits, a more complex form of missingness than that in the simulation study. We performed careful diagnostics of our imputations which suggested that the models were good. Unfortunately, the data in the CCC did not closely match the scenarios considered in the simulation study. Young et al.'s algorithm (18) generates an outcome that is distinct from covariates, which is not reflective of the CCC setting, where the continuous variable APRI was used to define event status. There are currently no data-generation protocols that mimic the CCC setting and permit direct calculation of the true marginal hazard ratio in closed form. Proposing an algorithm that more closely represents realistic cohort settings is an important direction for future research. Finally, we note that the discrete-time approximation (i.e., pooled logistic regression), often used for computational reasons, is no longer strictly required, since many software programs now support the use of weights in an extended Cox model (9, 29).

ACKNOWLEDGMENTS

Author affiliations: Ottawa Hospital Research Institute, Institute for Clinical Evaluative Sciences, Ottawa, Ontario, Canada (Nassim Mojaverian); Department of Epidemiology, Biostatistics, and Occupational Health, Faculty of Medicine, McGill University, Montreal, Quebec, Canada (Erica E. M. Moodie, Alex Bliu); Biologics and Genetic Therapies Directorate, Health Canada, Ottawa, Ontario, Canada (Alex Bliu); Division of Infectious Diseases/Chronic Viral Illness Service, Royal Victoria Hospital, McGill University Health Centre, Montreal, Quebec, Canada (Marina B. Klein); and Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, Montreal, Quebec, Canada (Marina B. Klein).

E.E.M.M. was supported by a “Chercheurs boursiers” career award from the Fonds de recherche en santé du Québec. M.B.K. was supported by a “Chercheurs nationaux” career award from the Réseau SIDA/Maladies infectieuses. The Canadian HIV–Hepatitis C Co-infection Cohort (CCC) Study

was funded by the Fonds de recherche en santé du Québec, the Réseau SIDA/Maladies infectieuses, the Canadian Institutes of Health Research (CIHR) (grant MOP-79529), and the CIHR Canadian HIV Trials Network (grant CTN222). The research presented in this paper was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

We thank the study coordinators and nurses for their assistance with study coordination, participant recruitment, and care.

The CCC Study investigators: Drs. Jeff Cohen, Windsor Regional Hospital Metropolitan Campus, Windsor, Ontario; Brian Conway, Vancouver Infectious Diseases Research and Care Centre, Vancouver, British Columbia; Curtis Cooper, The Ottawa Hospital Research Institute, Ottawa, Ontario; Pierre Côté, Clinique du Quartier Latin, Montréal, Quebec; Joseph Cox, Montréal General Hospital, Montréal, Quebec; John Gill, Southern Alberta HIV Clinic, Calgary, Alberta; Shariq Haider, McMaster University, Hamilton, Ontario; Marianne Harris, St. Paul's Hospital, Vancouver, British Columbia; David Haase, Capital District Health Authority, Halifax, Nova Scotia; Mark Hull, British Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia; Julio Montaner, St. Paul's Hospital, Vancouver, British Columbia; Erica Moodie, McGill University, Montreal, Quebec; Neora Pick, Oak Tree Clinic, Children's and Women's Health Centre of British Columbia, University of British Columbia, Vancouver, British Columbia; Anita Rachlis, Sunnybrook & Women's College Health Sciences Centre, Toronto, Ontario; Danielle Rouleau, Centre Hospitalier de l'Université de Montréal, Montréal, Quebec; Roger Sandre, HAVEN Program, Sudbury, Ontario; Joseph Mark Tyndall, Department of Medicine, Infectious Diseases Division, University of Ottawa, Ottawa, Ontario; Marie-Louise Vachon, Centre Hospitalier Universitaire de Québec, Québec, Québec; Sharon Walmsley, University Health Network, Toronto, Ontario; and David Wong, University Health Network, Toronto, Ontario.

Conflict of interest: none declared.

REFERENCES

- Hernán MA, Robins JM. *Causal Inference* [unpublished e-book]. <http://www.hsph.harvard.edu/miguel-Hernán/causal-inference-book/>. Accessed April 15, 2013.
- Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*. 2009;20(4):488–495.
- Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14(3):300–306.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
- Zhang Y, Thamer M, Kaufman JS, et al. High doses of epoetin do not lower mortality and cardiovascular risk among elderly hemodialysis patients with diabetes. *Kidney Int*. 2011;80(6):663–669.
- Delaney JA, Daskalopoulou SS, Suissa S. Traditional versus marginal structural models to estimate the effectiveness of beta-blocker use on mortality after myocardial infarction. *Pharmacoepidemiol Drug Saf*. 2009;18(1):1–6.
- Barnato AE, Chang CC, Farrell MH, et al. Is survival better at hospitals with higher “end-of-life” treatment intensity? *Med Care*. 2010;48(2):125–132.
- Edmonds A, Yotebieng M, Lusiana J, et al. The effect of highly active antiretroviral therapy on the survival of HIV-infected children in a resource-deprived setting: a cohort study. *PLoS Med*. 2011;8(6):e1001044.
- Xiao Y, Abrahamowicz M, Moodie EE. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *Int J Biostat*. 2010;6(2):Article 13.
- Andersen PK, Liestøl K. Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*. 2003;4(4):633–649.
- Regier MD, Moodie EE, Platt RW. The effect of error-in-confounders on the estimation of the causal parameter when using marginal structural models and inverse probability-of-treatment weights: a simulation study. *Int J Biostat*. 2014;10(1):1–15.
- Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2nd ed. Hoboken, NJ: Wiley-Interscience; 2002.
- Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*. 2004;5(3):445–464.
- Heyting A, Tolboom JT, Essers JG. Statistical handling of drop-outs in longitudinal clinical trials. *Stat Med*. 1992;11(16):2043–2061.
- Myers WR. Handling missing data in clinical trials: an overview. *Drug Inf J*. 2000;34(2):525–533.
- Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8(1):3–15.
- Moodie EEM, Delaney JA, Lefebvre G, et al. Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation. *Int J Biostat*. 2008;4(1):Article 13.
- Young JG, Hernán MA, Picciotto S, et al. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Anal*. 2010;16(1):71–84.
- Van-Buuren S, Oudshoorn CGM. *Multivariate Imputation by Chained Equations (MICE v1.0 User's Manual)*. Amsterdam, the Netherlands: TNO Prevention and Health; 2000.
- Klein MB, Saeed S, Yang H, et al. Cohort profile: the Canadian HIV–Hepatitis C Co-infection Cohort Study. *Int J Epidemiol*. 2010;39(5):1162–1169.
- Klein MB, Rollet KC, Saeed S, et al. HIV and hepatitis C virus coinfection in Canada: challenges and opportunities for reducing preventable morbidity and mortality. *HIV Med*. 2013;14(1):10–20.
- Brunet L, Moodie EE, Rollet K, et al. Marijuana smoking does not accelerate progression of liver disease in HIV-hepatitis C coinfection: a longitudinal cohort analysis. *Clin Infect Dis*. 2013;57(5):663–670.
- Wai CT, Greenson JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518–526.
- Al-Mohri H, Cooper C, Murphy T, et al. Validation of a simple model for predicting liver fibrosis in HIV/hepatitis C virus-coinfected patients. *HIV Med*. 2005;6(6):375–378.
- Nunes D, Fleming C, Offner G, et al. HIV infection does not affect the performance of noninvasive markers of fibrosis for

- the diagnosis of hepatitis C virus-related liver disease. *J Acquir Immune Defic Syndr*. 2005;40(5):538–544.
26. Peyre H, Leplège A, Coste J. Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. *Qual Life Res*. 2011;20(2):287–300.
 27. Ware JH, Harrington D, Hunter DJ, et al. Missing data. *N Engl J Med*. 2012;367(14):1353–1354.
 28. Vourli G, Touloumi G. Performance of the marginal structural models under various scenarios of incomplete marker's values: a simulation study. *Biom J*. 2015;57(2):254–270.
 29. Howe CJ, Cole SR, Mehta SH, et al. Estimating the effects of multiple time-varying exposures using joint marginal structural models: alcohol consumption, injection drug use, and HIV acquisition. *Epidemiology*. 2012;23(4):574–582.