ORIGINAL RESEARCH

A Principled Approach to Characterize and Analyze Partially Observed Confounder Data from Electronic Health Records

Janick Weberpals [ID][1], Sudha R Raman[2], Pamela A Shaw[3], Hana Lee[4], Massimiliano Russo[1], Bradley G Hammill[2], Sengwee Toh [ID][5], John G Connolly[5], Kimberly J Dandreo [ID][6], Fang Tian[7], Wei Liu[7], Jie Li[7], José J Hernández-Muñoz [ID][7], Robert J Glynn[1], Rishi J Desai [ID][1]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [2]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, USA; [3]Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; [4]Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA; [5]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; [6]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA; [7]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

Correspondence: Janick Weberpals, Instructor in Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont Street, Suite 3030-R, Boston, MA, 02120, USA, Tel +1 617-278-0932, Fax +1 617-232-8602, Email jweberpals@bwh.harvard.edu

Objective: Partially observed confounder data pose challenges to the statistical analysis of electronic health records (EHR) and systematic assessments of potentially underlying missingness mechanisms are lacking. We aimed to provide a principled approach to empirically characterize missing data processes and investigate performance of analytic methods.

Methods: Three empirical sub-cohorts of diabetic SGLT2 or DPP4-inhibitor initiators with complete information on HbA1c, BMI and smoking as confounders of interest (COI) formed the basis of data simulation under a plasmode framework. A true null treatment effect, including the COI in the outcome generation model, and four missingness mechanisms for the COI were simulated: completely at random (MCAR), at random (MAR), and two not at random (MNAR) mechanisms, where missingness was dependent on an unmeasured confounder and on the value of the COI itself. We evaluated the ability of three groups of diagnostics to differentiate between mechanisms: 1)-differences in characteristics between patients with or without the observed COI (using averaged standardized mean differences [ASMD]), 2)-predictive ability of the missingness indicator based on observed covariates, and 3)-association of the missingness indicator with the outcome. We then compared analytic methods including "complete case", inverse probability weighting, single and multiple imputation in their ability to recover true treatment effects.

Results: The diagnostics successfully identified characteristic patterns of simulated missingness mechanisms. For MAR, but not MCAR, the patient characteristics showed substantial differences (median ASMD 0.20 vs 0.05) and consequently, discrimination of the prediction models for missingness was also higher (0.59 vs 0.50). For MNAR, but not MAR or MCAR, missingness was significantly associated with the outcome even in models adjusting for other observed covariates. Comparing analytic methods, multiple imputation using a random forest algorithm resulted in the lowest root-mean-squared-error.

Conclusion: Principled diagnostics provided reliable insights into missingness mechanisms. When assumptions allow, multiple imputation with nonparametric models could help reduce bias.

Keywords: electronic health records, missing data, diagnostics, imputation, analytics

Introduction

Administrative data from health insurance claims have traditionally been the backbone for the majority of pharmacoepidemiologic research. A key limitation of claims data is their inability to capture granular clinical information. To address this limitation, large scale efforts to link claims databases to electronic health records (EHR) are underway, for instance, in the FDA Sentinel Initiative. These linked data capture more detailed information related to clinical parameters such as laboratory

329

tests, vital signs and physician assessments, which can enhance confounding capture and adjustment in estimation of treatment effects.[1] However, in EHR such covariates are often observed only in selected patients included in an analytic cohort which challenges the statistical analysis of the data.

Common missing data taxonomies such as missingness patterns[2] and mechanisms[3,4] have been proposed to describe missing data. Although the true underlying missingness generating mechanism cannot be inferred from the observed data, general frameworks[5–7] have given recommendations on how missing data may be characterized empirically. However, limited attention has been paid within existing frameworks to factors that are salient to partially observed confounder information in inferential studies of treatment effects using routinely collected healthcare data. Examples of such factors include the assumed relationship among a partially observed confounder, its missingness generating process and a given exposure-outcome effect,[8] as well as the presence of heterogeneous treatment effects (HTE) across levels of the partially observed confounder,[9] and the availability of auxiliary covariates correlated with the partially observed confounder.[10]

In this study, we sought to (i) develop a principled approach to empirically characterize missing data processes for partially observed EHR confounders and (ii) investigate the comparative performance of several analytic methods under various missingness scenarios using the plasmode simulation framework.

Methods

The data generation is illustrated in Figure 1.

Data Sources and Study Population for Plasmode Data Generation

As the basis for our plasmode data generation,[11] we used an *empirical cohort* of patients with type 2 diabetes who were new-users of either a sodium-glucose-cotransporter-2 (SGLT2i) or dipeptidyl-peptidase-4 (DPP4i) inhibitor (exposure). The empirical cohort was identified using data from the Mass General Brigham Research Patient Data Registry EHR in Boston[12] linked with Medicare fee-for-service claims data from 2012–2019. The EHR data contain information on patient
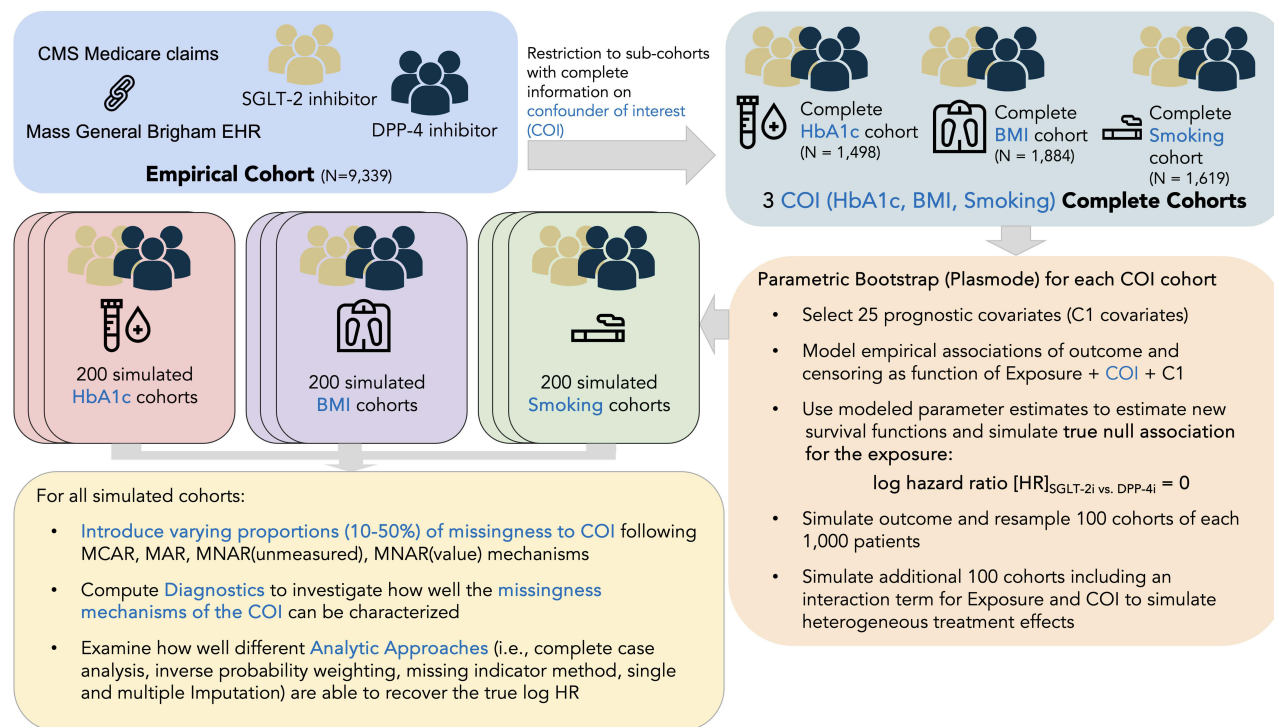

Figure 1 Illustration of plasmode data generating process.
Abbreviations: COI, Confounder of interest; DPP-4, dipeptidyl-peptidase-4 (DPP4i) inhibitor; MCAR, Missing completely at random; MAR, Missing at random; MNAR (unmeasured), Missing not at random; missingness depends on an unmeasured confounder; MNAR (value), Missing not at random; missingness depends on the value of the confounder of interest itself; SGLT-2, sodium-glucose-cotransporter-2 (SGLT2i).

demographics, medications, vital signs, smoking status, body mass index (BMI) and laboratory data. The Medicare claims data include enrollment files, inpatient claims, outpatient claims, and Part D (prescription drug coverage) files.

We defined the date of the first dispensing of either a SGLT2i or a DPP4i using Part D claims as the index date on which follow-up started. Covariates were measured in a 180-day window prior to the index date. The outcome was defined as the time to major adverse cardiovascular events (MACE), a composite endpoint that included myocardial infarction, stroke, hospitalization for heart failure or death. The follow-up time was calculated from the index date until the occurrence of the event or censoring (ie, disenrollment, end of follow-up at 365 days after the index date or end of study period on Dec 31st, 2019), whichever was first (details see Supplementary Figure 1).

Based on this empirical cohort, containing information on all study variables of interest, we derived three sub-cohorts with complete information on hemoglobin a1c (HbA1c) lab values, BMI, and smoking, respectively (*complete cohorts*). These three covariates served as the confounder of interest (COI) for the simulation.

Plasmode Data Generation

The plasmode simulation setup followed the framework described in detail by Franklin et al.[11] In brief, we first identified 25 prognostic covariates, other than the COI, covering information on each patient's demographics, comorbidities, co-medications, and healthcare utilization. Henceforth, these covariates are referred to as *C1* covariates (Table 1 and Supplementary Table 1) while all remaining covariates are referred to as *C0* or *auxiliary* covariates (Supplementary Table 2). The C1 covariates were fitted with the exposure (SGLT2i vs DPP4i) and the COI using a multivariable Cox proportional hazards regression to model the empirical outcome and censoring events in the respective complete cohorts. Next, we used the resulting parameter estimates to define the event-free and censoring survival functions. To predict event and censoring times for all patients given their C1 covariate values, we extracted the Breslow estimates of the baseline event-free survival and baseline censoring survival functions and additionally recorded the vector of estimated coefficients from these models. We then injected the desired investigator-defined null treatment effect estimate for the SGLT2 vs DPP4 exposure by replacing the estimated coefficient for the treatment effect with the desired null treatment effect (log hazard ratio $[HR]_{true} = 0$).[13] In the final step, the baseline event-free survival function was adjusted to get the desired event-rate, and a censored time-to-event outcome was generated for each patient as the minimum between the predicted event time and censoring time.[14] C0 covariates had no direct association with the outcome;, but as the plasmode framework preserves empirical correlations between variables, they may be indirectly associated with the outcome depending on their strength of correlations with C1 covariates.

For each COI complete cohort, we performed a resampling with replacement to arrive at 100 simulated plasmode datasets of 1000 patients each. For the BMI and smoking plasmode datasets, the resampling was stratified on levels of exposure and COI to retain the same distribution among these strata as in the original complete cohorts.

We repeated this procedure for each COI complete cohort to additionally simulate HTE by including an interaction term between the COI and the exposure.

Missingness Generation

To simulate missingness across all generated plasmode datasets, we followed the methodology described by Schouten and Vink.[17,18] To illustrate our assumed causal relationships between missingness with exposure, outcome, C1 covariates and COI, we provide M-graphs for each simulated scenario (Figure 2).[8,9,19,20]

- Missing completely at random – MCAR: Under a MCAR mechanism, the missingness is independent of any observed or unobserved covariates. Hence, in our simulation, all patients had the same probability of having an unobserved COI.
- Missing at random – MAR: A MAR mechanism is characterized by the missingness of the COI being dependent on one or multiple observed covariates. To simulate MAR, we calculated a weighted sum score for each patient. The weighted sum score was derived using a multivariable linear regression using all C1 covariates as predictors with each covariate contributing the same weight. Patients with a higher weighted sum score were more likely to have an unobserved value for the COI.

Table 1 Overview of All Plasmode Simulation Parameters. In a First Step, Complete Plasmode Datasets with a Known Underlying Outcome-Generating Model Were Simulated Using an Empirical Exposure, Some Fully Observed Prognostic Factors (C1) and a Confounder of Interest (COI) That Missingness Was Subsequently Imposed on. For Each COI, 200 Plasmode Datasets Were Generated. The Simulated Missingness-Generating Models for the COIs Followed Four Different Missingness Mechanisms in Which the Missing Probability Was the Same for Each Patient (MCAR), Dependent on Observed C1 Covariates (MAR), Dependent on an Unmeasured Confounder (MNAR [Unmeasured]) or Dependent on the COI Itself (MNAR [Value]), Respectively. Besides the COI and the Underlying Missingness Mechanisms, Other Simulation Parameters That Were Altered Involved the Proportion of Missingness That Was Imposed on the COI, the Covariates Included in the Imputation Model and the Presence of Exposure Treatment Effect Modification by the COI

Simulation Element	Simulation Parameter
Confounder of interest (COI) and variable type	• HbA1c value in % (continuous) • Body mass index (ordinal: underweight/normal, overweight, obese) • Smoking (binary: current/former vs never)
C1 covariates (used for outcome generation together with exposure and COI)	• **Demographics**: age, sex, race, Medicare subsidy, year of index date • **Comorbidities at baseline**: Combined comorbidity score,[15] hyperglycemia, cardiomyopathy, atrial fibrillation, anemia, stroke/ischemia, diabetic neuropathy, diabetes with peripheral circulatory disorders, diabetic nephropathy • **Use of concomitant drugs at baseline**: statin, sulfonylureas (current use), anticoagulants, antiplatelets, antihypertensive • **Healthcare utilization**: number of internal physician visits, hospital admission, generics, # diabetes medications, # HbA1c tests during baseline period, # emergency room visits
Missingness mechanisms simulated	• Missing completely at random (MCAR) • Missing at random (MAR) • Missing not at random - MNAR(unmeasured) • Missing not at random - MNAR(value)
# resampled datasets and sampled cohort size	200 resampled datasets per COI (100 datasets with and 100 without HTE), 1000 patients each
Missingness proportion	10 –50% (incremental increases by 10%)

Analytic approaches to address missing data	• Complete case analysis (baseline method)
	• Inverse probability weighting (IPW)
	• Missingness indicator
	• *missForest*[16] (single imputation random forest algorithm)
	Multiple imputation (m = 5 imputed datasets each):
	• MICE defaults for specific variable types:
	  ○ Predictive mean matching (HbA1c, continuous)
	  ○ Proportional odds model (BMI, ordinal)
	  ○ Logistic regression (smoking, binary)
	• Classification and regression tree (CART)
	• Random forest (RF)

Covariates included in diagnostics and imputation models	• C1 covariates only • C1 covariates + C0 covariates as auxiliary variables • + Exposure, Outcome (ie, time-to-event and event indicator)
Modifications to causal effect estimation	± treatment effect modification by EHR COI

Abbreviations: #, Number of; COI, Confounder of interest; EHR, Electronic health records; HbA1c, Hemoglobin A1C; HTE, Heterogeneous treatment effects; MAR, Missing at random mechanism in which the missingness probability depends on observed covariates; MCAR, Missing completely at random mechanism in which each patients has the same missingness probability; MNAR(unmeasured), Missing not at random mechanism in which the missingness can only be explained by a covariate which is not observed in the underlying dataset; MNAR(value), Missing not at random mechanism in which the missingness just depends on the actual value of the partially observed confounder of interest itself; MICE, Multiple imputation by chained equations.
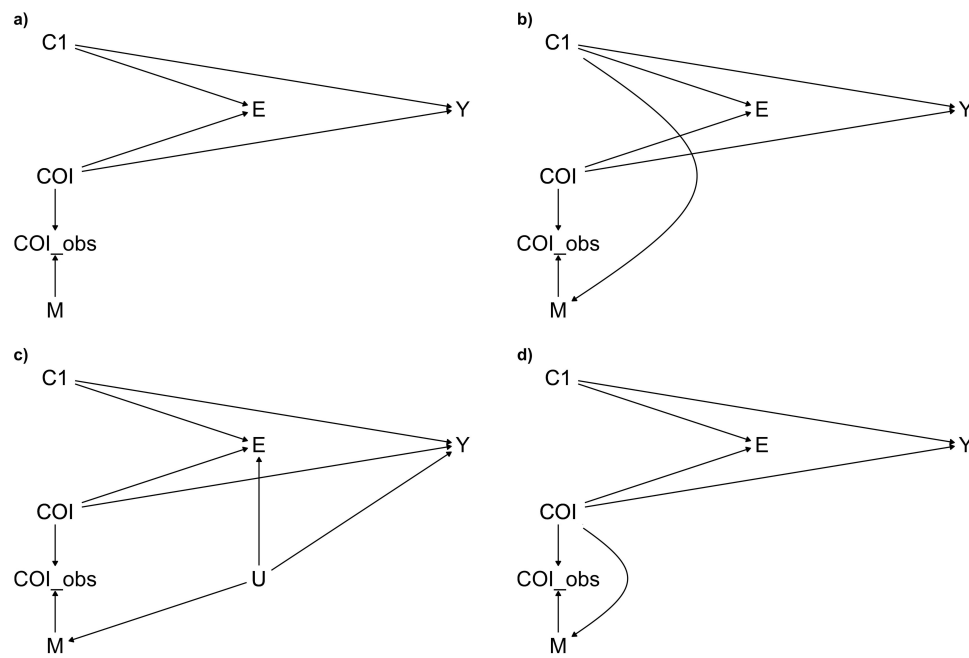
Figure 2 M-graphs depicting structural missing data assumptions and simulated mechanisms. (**a**) Missing completely at random: MCAR, (**b**) Missing at random: MAR), (**c**) Missing not at random, missingness depends on an unmeasured confounder: MNAR(unmeasured) and d) Missing not at random, missingness depends on the value of the confounder of interest itself: MNAR(value). (Notation: C1 = Fully observed C1 confounders used in outcome generation, COI = Confounder of interest [HbA1c, BMI, Smoking], COI_obs = observed portion of COI, E = Exposure, M = Missingness of COI with M=0 fully observed and M=1 fully missing, Y = Outcome).

- Missing not at random – MNAR (unmeasured): The first MNAR scenario assumed that the missingness of the COI is dependent on an unmeasured confounder. To simulate MNAR (unmeasured), missingness was generated to be solely dependent on a single continuous C1 covariate (exemplified by *age* in our study), ie, patients with a higher age were more likely to be missing. In contrast to a MAR mechanism, however, age was treated as "unmeasured" by subsequently dropping it from all following analyses.
- Missing not at random – MNAR (value): The second MNAR mechanism assumed that the missingness of the COI is dependent on the value of the COI itself. That is, for example, patients with a higher level of HbA1c were more like to have HbA1c unobserved.

For all assumed missingness mechanisms, we altered the missingness proportion between 10% and 50%. Further details on the missingness simulation are provided in the Supplementary Methods and Supplementary Table 3.

Missingness Diagnostics to Characterize Missing Data Processes

We employed three group diagnostics and evaluated their ability to differentiate and characterize different missingness mechanisms (Table 2).

Group 1 diagnostics investigated differences in the distribution of characteristics between patients with and without an observed value of the COI. We hypothesized that if missingness can be explained by observed covariates such as in the MAR mechanism, patient characteristics will significantly differ. To evaluate such differences, we computed the median absolute standardized mean difference (ASMD) across all observed variables between patients with and without an observed COI value. Analogous to balance measures in propensity score analysis[23] or randomized trials,[24] we considered an ASMD > 0.1 as indicative of a meaningful difference between cohorts. In addition, we applied Hoteling's[21] and Little's[22] tests, which have been proposed as formal hypothesis tests for MCAR. For both tests, high test statistics and a rejection of the null hypothesis would provide evidence for differences in the distribution of patient characteristics and suggest the underlying mechanism is not MCAR or MNAR.

Group 2 diagnostics assessed the ability to predict missingness based on observed covariates by fitting a classification model to predict the missingness indicator M of the COI with $M=1$ being missing and $M=0$ being observed.[7] To that end,

Table 2 Diagnostics to Empirically Differentiate and Characterize Missing Data Mechanisms. The Three Group Diagnostics are Composed of Analytic Models and Tests That Contextualize and Provide Information to Differentiate and Characterize Potentially Underlying Missingness Mechanisms

	Group 1 Diagnostics		Group 2 Diagnostics	Group 3 Diagnostics
Diagnostic metric	Absolute Standardized Mean Difference (ASMD)	P-value Hoteling[21]/ Little[22]	Area Under the Receiver Operating Curve (AUC)	Log HR (Missingness Indicator)
Purpose	Comparison of distributions between patients with vs without observed value of the partially observed covariate.		Assessing the ability to predict missingness based on observed covariates.	Check whether missingness of a covariate is associated with the outcome (differential missingness).
Example value	ASMD = 0.1	p-value < 0.001	AUC = 0.5	log HR = 0.1 (0.05 to 0.2)
Interpretation	<0.1[a]: no imbalances in observed patient characteristics; missingness may be likely completely at random or not at random (~MCAR, ~MNAR). >0.1[a]: imbalances in observed patient characteristics; missingness may be likely at random (~MAR).	High test statistics and low p-values indicate differences in baseline covariate distributions and null hypothesis would be rejected (~MAR).	AUC values ~ 0.5 indicate completely random or not at random prediction (~MCAR, ~MNAR). Values meaningfully above 0.5 indicate stronger relationships between covariates and missingness (~MAR).	No association in either univariate or adjusted model and no meaningful difference in the log HR after full adjustment (~MCAR). Association in univariate but not fully adjusted model (~MAR). Meaningful difference in the log HR also after full adjustment (~MNAR).

Note: [a]Analogous to propensity score-based balance measures.[23]
Abbreviations: ASMD, Median absolute standardized mean difference across all covariates; AUC, Area under the curve; CI, Confidence interval; MAR, Missing at random mechanism in which the missingness probability depends on observed covariates; MCAR, Missing completely at random mechanism in which each patients has the same missingness probability; MNAR(unmeasured), Missing not at random mechanism in which the missingness can only be explained by a covariate which is not observed in the underlying dataset; MNAR(value), Missing not at random mechanism in which the missingness just depends on the actual value of the partially observed confounder of interest itself.

we fit a random forest (RF) classification model using observed covariates with a 70/30 train-test split of the complete cohort. A sufficiently high area under the receiver operating characteristic curve (AUC) metric of the test dataset may demonstrate that M can be predicted well and could point towards MAR as a likely mechanism as opposed to an AUC~0.5 which would suggest MCAR or MNAR. It is generally important to check for potential monotone missing data patterns when missingness is multivariate,[25] as otherwise the group 2 diagnostics could lead to inflated AUC values when the missingness indicator of one partially observed covariate is a perfect predictor for another. In such a case, it would be rather advisable to perform the missingness diagnostics in the absence of the correlated partially observed covariates. However, since we aimed to investigate the influence of different COI variable types on the performance of the missingness diagnostics, this was not applicable in this simulation.

Group 3 diagnostics evaluated the association between M and the outcome under study. If the missingness of a confounder cannot be explained or approximated by observed covariates and a difference in the outcome is observed depending on M (eg, $logHR_M \neq 0$), this may be indicative of an underlying MNAR mechanism. We evaluated the possibility of such a differential missingness by fitting an outcome model including M. Since the outcome of this study was a time-to-event outcome, we fitted a univariate and an adjusted Cox proportional hazards regression. We hypothesized that the contrast of a univariate and an adjusted model conditional on observed C1 and C0 covariates might be able to reveal important characteristics between different mechanisms. Specifically, we would expect no meaningful differences between the two resulting log HRs for MCAR, while this may not be true for a MAR or MNAR mechanism.

For all diagnostic models, we altered the inclusion of auxiliary (C0) covariates as an additional simulation parameter and summarized results by averaging the group diagnostic metrics across all simulated plasmode datasets and missingness mechanisms. We computed confidence intervals (CI) for all diagnostics using the empirical Monte-Carlo standard

error, which was derived as the standard deviation of the estimate over the square root of the number of stratum-specific simulated plasmode datasets. We additionally visualized the variability of the diagnostic results across all iterations and scenarios using boxplots. More details on the specifications of the three group diagnostics can be found in the Supplementary Methods (sections 1.1.3 and 1.1.5).

Analytic Approaches to Address Missingness

Finally, to examine the performance of different analytic approaches across the simulated missingness scenarios, we compared methods including complete case analysis (CCA), missing indicator method, inverse probability weighting (IPW) with weights based on probabilities derived from a logistic regression model predicting complete cases given C1, missing indicator method, single imputation with RF and multiple imputation (MI) approaches using different parametric and non-parametric models as summarized in Table 1 and Supplementary Methods, section 1.1.4. To evaluate the performance of these approaches, we derived the root mean squared error (RMSE), bias and variance as described by Morris et al[26,27] to compare differences in the estimated treatment effect estimates (logHR) of each method and the fixed true treatment effect estimate ($logHR_{true} = 0$). Further details on model specifications and evaluation metrics are given in the Supplementary Methods (section 1.1.5).

Empirical Case Example

To augment learnings from our simulation studies, we applied the proposed missingness diagnostics to a case-example study using an empirical cohort independent of the one we used for the simulation. Briefly, a cohort of initiators of either an opioid (exposure) or non-steroidal anti-inflammatory drug (NSAID, comparator) was identified using Medicare claims-EHR linked data to investigate the comparative risk of acute kidney injury (AKI) (Supplementary Figure 2).[28] Since chronic use of NSAIDs is a well-established risk factor for the onset of AKI, we hypothesized that physicians are more likely to prescribe opioids to those patients who are susceptible to AKI based on their baseline risk factors which would lead to strong confounding bias with a spuriously increased risk of AKI among opioid initiators (HR > 1, Supplementary Figure 3).[29] To adjust for confounding, we derived multiple, fully observed confounding factors from Medicare claims including demographics, comorbid conditions, and comedications, and *three partially observed EHR labs* that are prognostic biomarkers of a patient's baseline kidney function:[30] blood urea nitrogen (BUN), estimated glomerular filtration rate (eGFR) and serum creatinine (Supplementary Table 4). To adjust for these three EHR labs, we investigated potential missing data patterns,[25] applied the missingness diagnostics as described earlier and used the results to guide the most appropriate analytic approach.

Statistical Software

All patient cohorts used in this study were created using SAS version 9.4 and all analyses were conducted in R version 4.1.2. Detailed information on used packages and versions can be found in the Supplementary Methods and code used in this study is available at https://gitlab-scm.partners.org/drugepi/missingehr. The proportional hazards assumption for the complete cohorts and base cohort was checked using using the cox.zph() function of the "survival" package and plotted using the ggcoxzph() function of the "survminer" package (Supplementary Figures 4 and 5) without indication of a significant violation.

Results

Diagnostics

The diagnostic results, averaged across all simulated datasets, are summarized in Table 3 and the distributions of the different diagnostic results by mechanism across all scenarios and iterations are illustrated in Supplementary Figure 6.

The three group diagnostics successfully identified characteristic patterns of the simulated missingness mechanisms. For Group 1 diagnostics, the patient characteristics showed substantial differences for MAR, but not for MCAR and both types of MNAR (median ASMD 0.20 vs 0.05, 0.09, and 0.06). In group 2 diagnostics, the discrimination of the prediction models for missingness was highest for MAR (0.59) and lowest for MCAR (0.50). For MNAR, but not for MAR or MCAR, missingness was significantly associated with the outcome even in models adjusting for other observed

Table 3 Averaged Simulation Results of Missingness diagnostics Across All Simulated Plasmode Datasets

Mechanism	ASMD (95% CI)[a]	p(Hotelling)[a]	p(Little)[a]	AUC (95% CI)[b]	log HR(univariate) (95% CI)[c]	log HR (95% CI)[c]
MCAR	0.05 (0.05, 0.05)	0.50	0.50	0.50 (0.50, 0.50)	0.00 (−0.01, 0.00)	0.00 (−0.01, 0.00)
MAR	0.20 (0.20, 0.20)	<0.001	<0.001	0.59 (0.58, 0.59)	0.53 (0.53, 0.54)	0.00 (−0.01, 0.00)
MNAR(unmeasured)	0.09 (0.09, 0.09)	0.02	0.02	0.54 (0.54, 0.54)	0.43 (0.43, 0.44)	0.31 (0.31, 0.32)
MNAR(value)	0.06 (0.06, 0.06)	0.10	0.10	0.53 (0.53, 0.53)	0.05 (0.04, 0.05)	0.10 (0.09, 0.11)

Notes: [a]Group 1 diagnostic: Differences in patient characteristics between patients with and without covariate. [b]Group 2 diagnostic: Ability to predict missingness. [c]Group 3 diagnostic: Assessment if missingness is associated with the outcome (univariate, adjusted).
Abbreviations: ASMD, Median absolute standardized mean difference across all covariates; AUC, Area under the curve; CI, Confidence interval; Confidence intervals are based on the empirical Monte-Carlo standard error.

patient characteristics. Investigating the potential for differential missingness in group 3 diagnostics, both univariate and fully adjusted models indicated no association under MCAR (logHR 0.00, [95% confidence interval −0.01–0.00]), while MAR showed a spurious association in the univariate model (logHR(univariate) 0.53, [0.53, 0.54]) which disappeared after adjusting for other fully observed covariates (logHR 0.00 [−0.01, 0.00]). In contrast, a significant difference for the association between the missingness and the outcome was found in both MNAR mechanisms even after comprehensive adjustment.

Analyzing the diagnostics results by proportion missing, COI, and inclusion of auxiliary covariates did not meaningfully change these characteristic patterns (Table 4, Supplementary Tables 5 and 6). A minor exception was that there was a noticeable increase in the ability to predict missingness (group 2 diagnostics) with increasing proportions of missingness under MAR (AUC 0.53 with 10% to 0.63 with 50% missingness, Table 4).

Table 4 Averaged Simulation Results of Missingness Diagnostics by Proportion Missingness

Mechanism	ASMD (95% CI)[a]	p(Hotelling)[a]	p(Little)[a]	AUC (95% CI)[b]	log HR(univariate) (95% CI)[c]	log HR (95% CI)[c]
Proportion missing: 10%						
MCAR	0.07 (0.07, 0.07)	0.52	0.52	0.50 (0.50, 0.50)	−0.01 (−0.02, 0.01)	0.00 (−0.02, 0.01)
MAR	0.22 (0.22, 0.23)	<0.001	<0.001	0.53 (0.53, 0.53)	0.61 (0.60, 0.62)	0.00 (−0.01, 0.02)
MNAR(unmeasured)	0.11 (0.11, 0.11)	0.06	0.06	0.52 (0.51, 0.52)	0.48 (0.47, 0.49)	0.35 (0.34, 0.37)
MNAR(value)	0.08 (0.08, 0.09)	0.18	0.18	0.52 (0.51, 0.52)	0.06 (0.05, 0.08)	0.12 (0.11, 0.14)
Proportion missing: 20%						
MCAR	0.05 (0.05, 0.05)	0.50	0.50	0.50 (0.50, 0.50)	0.00 (−0.01, 0.01)	0.00 (−0.01, 0.01)
MAR	0.20 (0.20, 0.21)	<0.001	<0.001	0.56 (0.56, 0.56)	0.55 (0.54, 0.56)	0.00 (−0.02, 0.01)
MNAR(unmeasured)	0.10 (0.10, 0.10)	0.02	0.02	0.53 (0.52, 0.53)	0.45 (0.44, 0.46)	0.33 (0.32, 0.34)
MNAR(value)	0.07 (0.07, 0.07)	0.11	0.11	0.53 (0.53, 0.53)	0.05 (0.04, 0.06)	0.11 (0.10, 0.13)
Proportion missing: 30%						
MCAR	0.05 (0.05, 0.05)	0.50	0.49	0.50 (0.50, 0.50)	0.00 (−0.01, 0.01)	0.00 (−0.01, 0.01)
MAR	0.19 (0.19, 0.20)	<0.001	<0.001	0.59 (0.59, 0.60)	0.52 (0.51, 0.53)	0.00 (−0.01, 0.01)
MNAR(unmeasured)	0.09 (0.09, 0.09)	0.01	0.01	0.54 (0.54, 0.54)	0.43 (0.42, 0.44)	0.32 (0.31, 0.33)
MNAR(value)	0.06 (0.06, 0.06)	0.07	0.07	0.54 (0.53, 0.54)	0.04 (0.04, 0.05)	0.09 (0.08, 0.10)

(Continued)

Table 4 (Continued).

Mechanism	ASMD (95% CI)[a]	p(Hotelling)[a]	p(Little)[a]	AUC (95% CI)[b]	log HR(univariate) (95% CI)[c]	log HR (95% CI)[c]
Proportion missing: 40%						
MCAR	0.04 (0.04, 0.04)	0.49	0.49	0.50 (0.50, 0.50)	0.00 (−0.01, 0.00)	0.00 (−0.01, 0.01)
MAR	0.19 (0.19, 0.19)	<0.001	<0.001	0.62 (0.62, 0.62)	0.49 (0.49, 0.50)	−0.01 (−0.02, 0.01)
MNAR(unmeasured)	0.08 (0.08, 0.08)	0.01	0.01	0.55 (0.55, 0.55)	0.40 (0.39, 0.41)	0.29 (0.28, 0.30)
MNAR(value)	0.06 (0.06, 0.06)	0.07	0.07	0.54 (0.54, 0.54)	0.04 (0.03, 0.05)	0.09 (0.08, 0.10)
Proportion missing: 50%						
MCAR	0.04 (0.04, 0.04)	0.50	0.50	0.51 (0.51, 0.51)	0.00 (−0.01, 0.01)	0.00 (−0.01, 0.01)
MAR	0.19 (0.18, 0.19)	<0.001	<0.001	0.63 (0.63, 0.63)	0.48 (0.47, 0.49)	0.00 (−0.01, 0.01)
MNAR(unmeasured)	0.08 (0.08, 0.08)	0.02	0.02	0.56 (0.56, 0.56)	0.39 (0.38, 0.40)	0.29 (0.28, 0.30)
MNAR(value)	0.05 (0.05, 0.05)	0.08	0.08	0.55 (0.54, 0.55)	0.04 (0.03, 0.04)	0.08 (0.07, 0.09)

Notes: [a]Group 1 diagnostic: Differences in patient characteristics between patients with and without covariate. [b]Group 2 diagnostic: Ability to predict missingness. [c]Group 3 diagnostic: Assessment if missingness is associated with the outcome (univariate, adjusted).
Abbreviation: ASMD, Median absolute standardized mean difference across all covariates; AUC, Area under the curve; CI, Confidence interval.

Analytic Approaches

MI approaches resulted in the lowest RMSE, bias and variance, with a RF-based algorithm showing the best overall performance, closely followed by the default MICE implementation for the specific variable types (Figure 3). The single imputation *missForest* algorithm tended to show a higher bias and RMSE as compared to the MI counterparts. CCA and IPW performed equally poor in terms of bias, but IPW was observed to have higher variance than CCA and resulted in the highest RMSE (the corresponding IP weights by proportion of imposed missingness are displayed in Supplementary Figure 7). Comparing the different methods by missingness mechanism, MNAR (value) led to the highest RMSE across analytic methods (Figure 4). Including auxiliary covariates in imputation models led to slightly lower RMSEs, independent of the imputation method (Supplementary Figure 8, the corresponding correlation strengths between covariates and COI are displayed in Supplementary Figures 9–11).
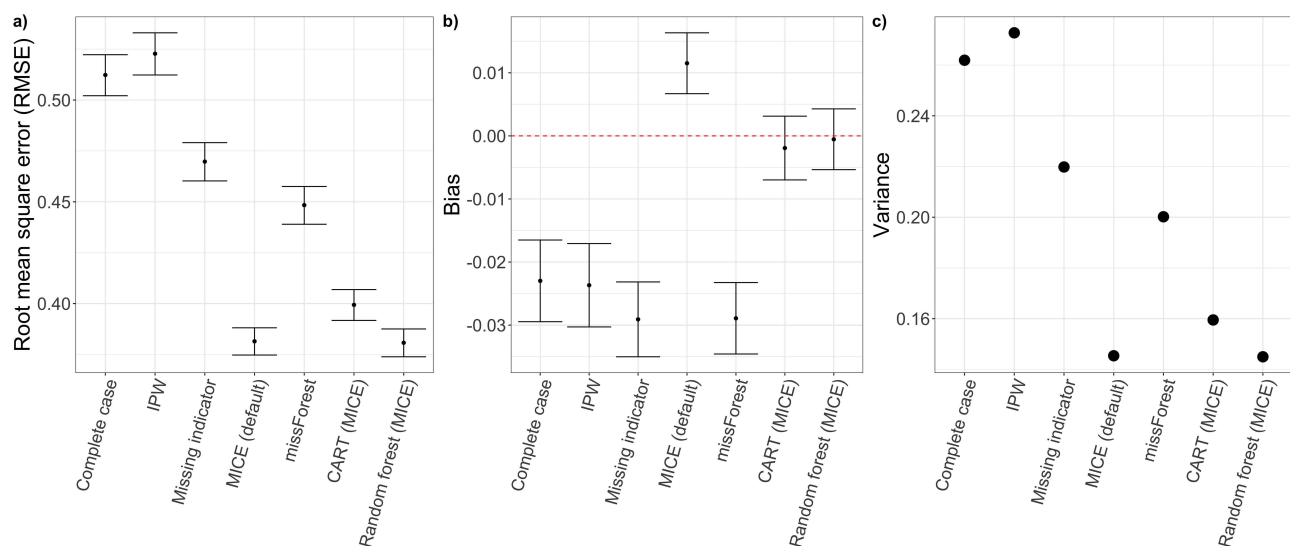

Figure 3 Panel illustrating (**a**) root mean squared error (RMSE), (**b**) % bias and (**c**) Variance for each analytic missing data method averaged across all simulation scenarios.
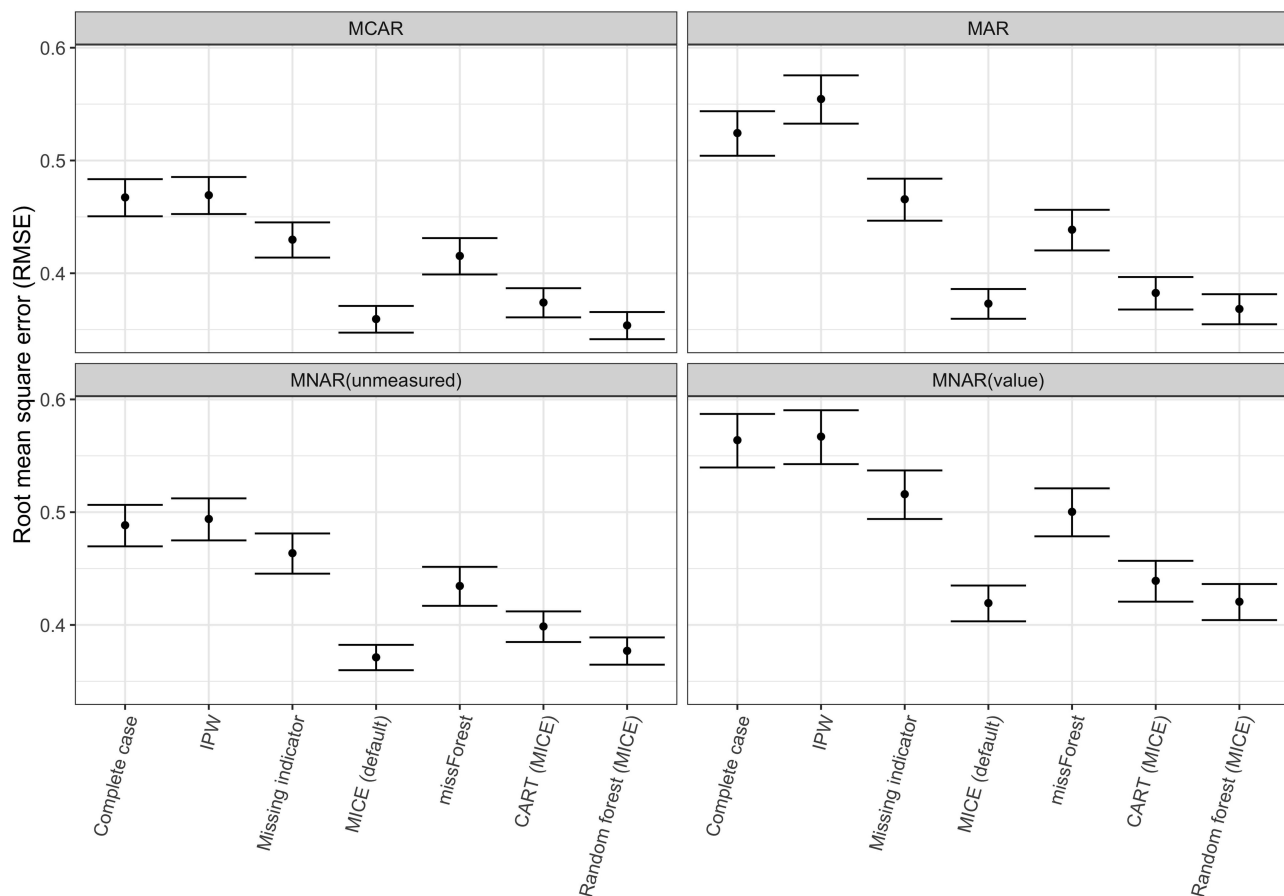
Figure 4 Root mean square error (RMSE) metrics by simulated missingness mechanism.

Empirical Case-Example

After applying all in- and exclusion criteria, the final cohort comprised 13,340 patients. The BUN, eGFR and serum creatinine labs were missing in 52.07%, 55.51% and 51.51% of all patients, respectively (Supplementary Figure 12).

Of 7458 patients with at least one unobserved value for either of the three labs, 6862 (92%) patients were observed with missing values for all three labs concurrently, suggesting an almost monotone missing data pattern (Figure 5).[31] To avoid inflated AUCs in the group 2 diagnostics by the missingness of one lab being a perfect predictor for another, we applied the missingness diagnostics for each lab independent of the other two labs as potential predictors (except for Little's test). The patterns of the missingness diagnostics were similar across all three EHR labs and matched the ones observed for MAR in the simulation study (Table 5). We therefore concluded that MI was an appropriate analytic method and imputed 50 datasets including the three EHR labs, exposure, outcome and fully observed covariates using a RF-based MI model, as this was the best performing algorithm in the simulation study.

As expected, the unadjusted outcome model suggested a spuriously increased risk among opioid users (hazard ratio [HR] 1.36, 95% 1.22–1.52, Figure 6). Adjusting for all fully observed covariates decreased the HR meaningfully towards the null and became insignificant (1.07, 0.94–1.22). Adding the three imputed EHR labs led to a further decrease of the pooled HR in the expected direction without appreciable increase in variance (HR 1.05, 0.92–1.20).

Discussion

In this study, we developed, validated, and illustrated a principled approach to empirically characterize missing data processes for partially observed EHR confounders and investigated the performance of analytic methods under different missingness scenarios using the plasmode simulation framework. The three proposed group diagnostics successfully
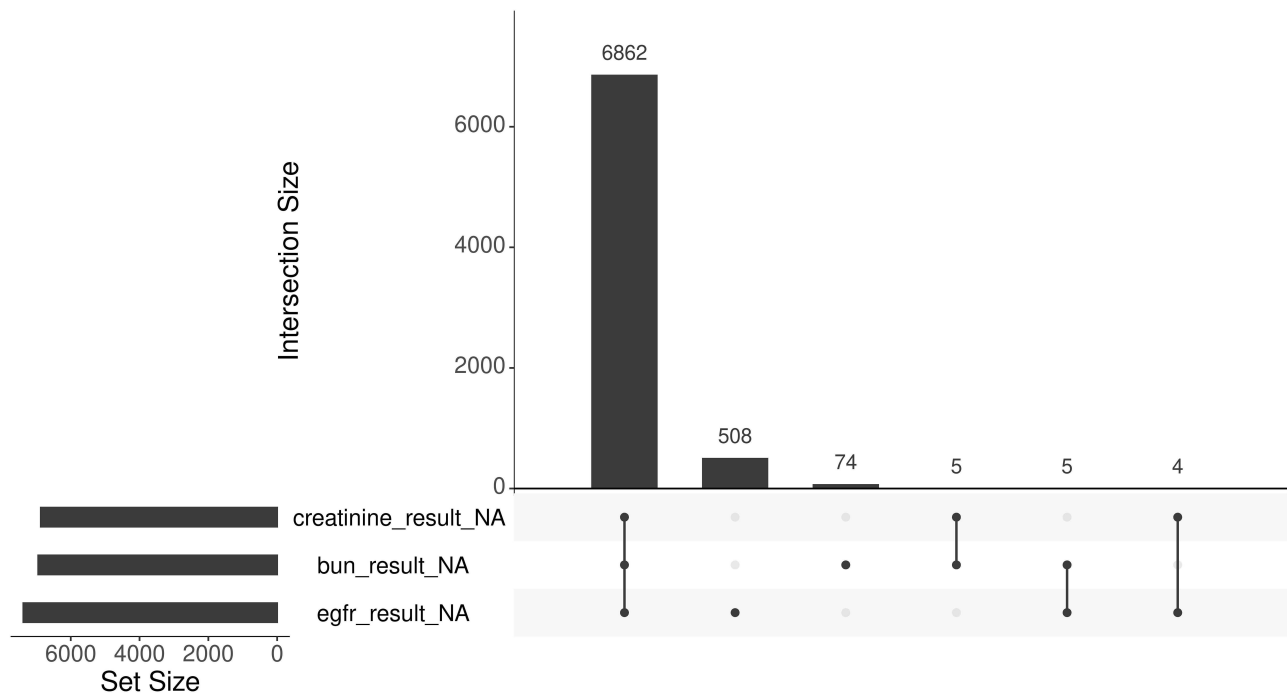
Figure 5 Missingness pattern[25] of Blood urea nitrogen (*bun_results_NA*), estimated glomerular filtration rate (*egfr_results_NA*) and serum creatinine (*creatinine_result_NA*) in the base case cohort. The set size displays the count of missing observations for each lab result individually while the intersection size illustrates the count of intersecting missing observations across the three labs.

identified patterns that matched assumptions for a set of simulated missing data structures. We additionally found that of all the analytic approaches to handle missing data we considered, MI approaches were most efficient and led to the lowest bias when compared to other methods.

Given that typically only few studies outline their assumptions on missing data mechanisms and fail to provide reasoning as to why certain analytic choices are made,[32] systematic approaches to handle partially observed covariates are essential to mitigate bias stemming from inappropriate methods and to increase transparency. In combination with M-graphs to outline missingness assumptions and to aid decisions on the recoverability of the studied estimand,[8,20] our proposed missingness diagnostics can be a valuable tool to routinely characterize missing data processes. Particularly when supplemented with substantive expert knowledge (eg, clinicians), this can help researchers understand how covariates are measured in routine care processes and contextualize potential reasons for missingness. In our simulations, the three group diagnostics differentiated well between MCAR and MAR. Distinguishing between MCAR and MNAR (value) may be difficult as both these mechanisms could lead to no real detectable difference in patient characteristics

Table 5 Missingness Diagnostics of the Exemplary Base Case Study for Three Blood Laboratory Measurements Representing a Patient's Baseline Kidney Function. Due to the Expected Monotone Missingness Pattern, the Missingness Diagnostics for One EHR Lab Were Performed in Absence of the Other Two EHR Labs, Except for Little's Test Which Globally Tests for Presence of MCAR Considering All EHR Labs Jointly

Covariate	ASMD (Min/Max)[a]	p(Hotelling)[a]	AUC[b]	Log HR(Univariate) (95% CI)[c]	log HR (95% CI)[c]
Blood urea nitrogen (BUN)	0.056 (0.001, 0.389)	<0.001	0.621	−0.13 (95% CI −0.23, −0.02)	0.03 (95% CI −0.08, 0.14)
Serum creatinine	0.056 (0.000, 0.380)	<0.001	0.622	−0.13 (95% CI −0.24, −0.02)	0.03 (95% CI −0.09, 0.14)
Estimated glomerular filtration rate (eGFR)	0.047 (0.001, 0.359)	<0.001	0.611	−0.13 (95% CI −0.24, −0.03)	0.01 (95% CI −0.11, 0.12)

Notes: p little: <0.001, [a]Group 1 diagnostic: Differences in patient characteristics between patients with and without covariate. [b]Group 2 diagnostic: Ability to predict missingness. [c]Group 3 diagnostic: Assessment if missingness is associated with the outcome (univariate, adjusted).
Abbreviations: ASMD, Median absolute standardized mean difference across all covariates; AUC, Area under the curve; beta, beta coefficient; CI, Confidence interval; max, Maximum; min, Minimum.
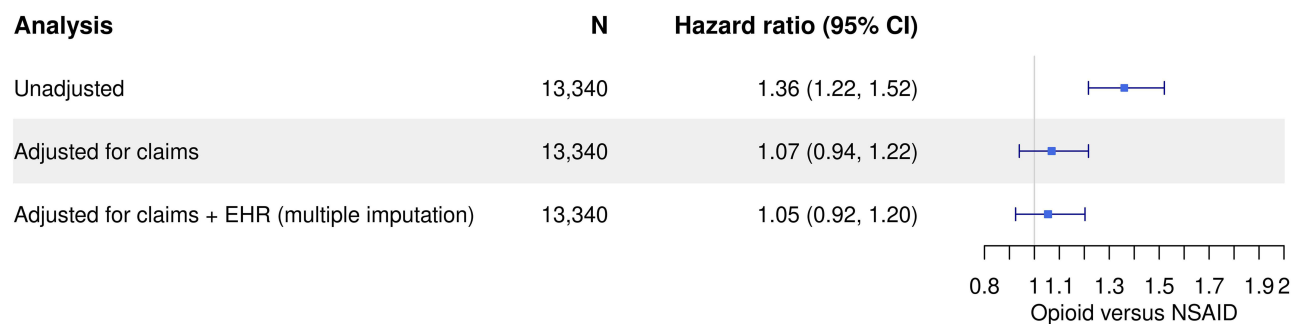
Analysis	N	Hazard ratio (95% CI)	
Unadjusted	13,340	1.36 (1.22, 1.52)	
Adjusted for claims	13,340	1.07 (0.94, 1.22)	
Adjusted for claims + EHR (multiple imputation)	13,340	1.05 (0.92, 1.20)	

0.8 1 1.1 1.3 1.5 1.7 1.9 2
Opioid versus NSAID

Figure 6 Comparison of results from the empirical case example assessing opioid versus non-steroidal anti-inflammatory drug use on the time to acute kidney injury dependent on level of adjustment.

between those with and without missingness in group 1 diagnostics, which in turn would lead to low discrimination in prediction model for missingness indicator in group 2 diagnostics. In such situations, a strong association between missingness indicator and the outcome of interest in both univariate and adjusted models, as suggested in our group 3 diagnostics, could indicate differential MNAR mechanisms.

Analytically, we found that MI algorithms addressed missingness well under MCAR and MAR, which has been also shown in previous studies.[9,33,34] When the missingness mechanisms followed either of the two MNAR mechanisms, higher RMSEs were observed across all methods, although in relative comparison, MI approaches still performed best. Although for this simulation we operated under the assumption of the COIs being critical confounders, it may also be a reasonable approach to ignore a partially observed covariate when the diagnostics indicate likely MNAR and if the confounding strength is negligible.[35] IPW models showed similar bias compared to a CCA but consistently had higher variance. This was also observed under a true MAR mechanism, where we would have expected IPW to show less biased results than the CCA. One potential explanation could be unstable weights resulting from a violation of the positivity assumption in small samples with low prevalence COIs, which could lead the probability of being a complete case very close to 0 (Supplementary Table 7 and Supplementary Figure 7).[36,37] Further, the IPW models in this study were rather simple and model misspecification may have also contributed to this observation. More sophisticated approaches such as the augmented IPW (AIPW) may improve the performance due to their more flexible properties as doubly robust models.[36]

Given that the missingness under a MNAR (unmeasured) scenario depends on an unobserved confounder, the availability of highly correlated auxiliary covariatesfor instance, recorded diagnosis of chronic obstructive pulmonary disease as a proxy for partially observed smoking, as proxies may help shift an MNAR (unmeasured) to a MAR mechanism and thereby enable the imputation of the data.[10,38] Such covariate correlations can be easily inspected, eg, utilizing correlation heatmaps. Given the presence of a few moderate correlations between auxiliary covariates and COIs in our simulation (Supplementary Figures 9–11), the inclusion of auxiliary covariates in imputation models led to lower RMSEs (Supplementary Figure 8). This is not applicable to MNAR (value) mechanisms since values between patients with and without an observed COI systematically differ and are not inferable from observed data. One analytic method that has been proposed for this mechanism are fully conditional specified imputation models that include a sensitivity parameter which accounts for this systematic difference conditional on all remaining variables of the data and their respective missingness indicators (*delta adjustment*).[39] The challenge among these models is the determination of the sensitivity parameter, as this quantity is typically unknown.[40] Hence, one popular application of this method is the use as sensitivity analysis by modeling multiple sensitivity parameters over a range of plausible values to elucidate 'tipping points', ie, systematic differences, which would yield a qualitative change of the primary analysis results.

Applying the group diagnostics in the empirical case example demonstrated how these diagnostics can be routinely used in real-world evidence studies involving partially observed confounders. The diagnostics suggested a near monotone missingness patterns and MAR mechanism between the three EHR labs, which is clinically plausible given that serum creatinine is needed to calculate eGFR and hence these labs are typically measured together as part of a kidney labs

panel. The corresponding multiple imputation analysis combining claims and EHR confounders led to a decreased HR which moved the treatment effect estimate closer in the expected direction. However, the difference to the claims only analysis was only marginal and a stronger decrease to a HR < 1 in a fully adjusted model would have been expected. This may be explained by residual confounding of unmeasured prognostic factors and the modeling of the exposure in an intention-to-treat analysis, which may have biassed the estimates towards the null as patients were not censored upon treatment switch or discontinuation.

Strengths and Limitations

One weakness of this study is that we based our assumptions on four different missingness mechanisms, which may be an oversimplification. Mixed mechanisms may also be possible, which would increase the number of potential combinations of scenarios by several orders of magnitude. Nevertheless, the four considered mechanisms are plausible and common in EHR, as also corroborated by observations in the empirical case example. In addition, our study solely focused on partially observed confounders and did not address missingness in exposure or outcome variables, which would be important future avenues for research.

To our knowledge, this is the first study that systematically investigated an empirical approach to characterize missingness processes in EHR confounder data. We used plasmode simulations as more realistic alternatives to de novo simulations since it retains more complex covariate correlations inherent to real-world data. We further considered a wide range of different simulation parameters including two different MNAR mechanisms that so far has received limited attention in the literature. Finally, although the simulations focused on one partially observed COI at a time, the empirical case-example study illustrated that these missingness diagnostics can be readily extended to multivariate missing data. Finally, a major contribution of the current paper is that we provide an R package for implementation (https://janickweberpals.gitlab-pages.partners.org/smdi) of the proposed diagnostics.[41]

Conclusions

The proposed principled diagnostics successfully differentiated between four different simulated missingness mechanisms, provided insights into underlying missingness structures and may be routinely used in real-world evidence studies involving partially observed confounders. MI approaches performed well in this study under MCAR and MAR mechanisms and, combined with M-graphs and sensitivity analyses, the proposed missingness diagnostics can help elucidate if such analytic approaches are viable options and thereby improve the robustness of studies involving partially observed confounders.

Data Sharing Statement

Patient-level data cannot be shared due to restriction under our data use agreement with CMS. However, to test the missing data diagnostics and imputation analyses, the *smdi* R package includes a simulated dataset (*smdi_data*) with similar missingness structures as the data used in this manuscript. Patient cohorts used in this study were queried using SAS version 9.4 and all analyses were conducted in R version 4.1.2. Detailed information on used packages and versions can be found in the Supplementary Methods and code used in this study is available at https://gitlab-scm.partners.org/drugepi/missingehr. The missing data diagnostics presented in this manuscript can be implemented using the *smdi* R package, available from https://janickweberpals.gitlab-pages.partners.org/smdi/ and CRAN via install.packages("smdi").

Ethics Statement

This study does not meet the criteria for human subject research as defined by Mass General Brigham Human Research Office policies, Health and Human Services (HHS) regulations set forth in 45 CFR 46, and Food and Drug Administration (FDA) regulations set forth in 21 CFR 56. The study involves public health surveillance activity as defined by HHS regulations at 45 CFR 46.102(l)(2).

Acknowledgments

Funding

Disclosure

The FDA approved the study protocol, statistical analysis plan and reviewed and approved this manuscript. Coauthors from the FDA participated in the results interpretation and in the preparation and decision to submit the manuscript for publication. The FDA had no role in data collection, management, or analysis. The views expressed are those of the authors and not necessarily those of the US FDA. Janick Weberpals reports prior employment by Hoffmann-La Roche and previously held shares in Hoffmann-La Roche. Pamela Shaw is a named inventor on a patent licensed to Novartis by the University of Pennsylvania for an unrelated project. Sengwee Toh serves as a consultant for Pfizer, Inc. and TriNetX, LLC. Robert J Glynn has received research funding through his employer from Amarin, Kowa, Novartis, and Pfizer. Dr. Desai reports serving as Principal Investigator on investigator-initiated grants to the Brigham and Women's Hospital from Novartis, Vertex, and Bristol-Myers-Squibb on unrelated projects. All remaining authors report no disclosures or conflicts of interest.

References

1. Desai RJ, Matheny ME, Johnson K, et al. Broadening the reach of the FDA Sentinel system: a roadmap for integrating electronic health record data in a causal analysis framework. *Npj Digit Med*. 2021;4(1):170. doi:10.1038/s41746-021-00542-0
2. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. Vol. 793. John Wiley & Sons; 2019.
3. Heymans MW, Twisk JWR. Handling missing data in clinical research. *J Clin Epidemiol*. 2022;151:185–188. doi:10.1016/j.jclinepi.2022.08.016
4. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–592. doi:10.1093/biomet/63.3.581
5. Lee KJ, Tilling KM, Cornish RP, et al. Framework for the treatment and reporting of missing data in observational studies: the treatment and reporting of missing data in observational studies framework. *J Clin Epidemiol*. 2021;134:79–88. doi:10.1016/j.jclinepi.2021.01.008
6. Carpenter JR, Smuk M. Missing data: a statistical framework for practice. *Biom J*. 2021;63(5):915–947. doi:10.1002/bimj.202000196
7. Sondhi A, Weberpals J, Yerram P, et al. A systematic approach towards missing lab data in electronic health records: a case study in non-small cell lung cancer and multiple myeloma. *CPT Pharmacomet Syst Pharmacol*. 2023;12:1201–1212. doi:10.1002/psp4.12998
8. Lee KJ, Carlin JB, Simpson JA, Moreno-Betancur M. Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *Int J Epidemiol*. 2023;dyad008. doi:10.1093/ije/dyad008
9. Choi J, Dekkers OM, le Cessie S. A comparison of different methods to handle missing data in the context of propensity score analysis. *Eur J Epidemiol*. 2019;34(1):23–36. doi:10.1007/s10654-018-0447-z
10. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol*. 2019;110:63–73. doi:10.1016/j.jclinepi.2019.02.016
11. 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
12. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the Benefits of a Research Patient Data Repository. *AMIA Annu Symp Proc*. 2006;2006:1044.
13. Huitfeldt A, Stensrud MJ, Suzuki E. On the collapsibility of measures of effect in the counterfactual causal framework. *Emerg Themes Epidemiol*. 2019;16(1):1. doi:10.1186/s12982-018-0083-9
14. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med*. 2005;24(11):1713–1723. doi:10.1002/sim.2059
15. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol*. 2011;64(7):749–759. doi:10.1016/j.jclinepi.2010.10.004
16. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–118. doi:10.1093/bioinformatics/btr597
17. Buuren van S. Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67. doi:10.18637/jss.v045.i03
18. Schouten RM, Lugtig P, Vink G. Generating missing values for simulation purposes: a multivariate amputation procedure. *J Stat Comput Simul*. 2018;88(15):2909–2930. doi:10.1080/00949655.2018.1491577
19. Mohan K, Pearl J. Graphical models for processing missing data. *J Am Stat Assoc*. 2021;116(534):1023–1037. doi:10.1080/01621459.2021.1874961
20. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. *Am J Epidemiol*. 2018;187(12):2705–2715. doi:10.1093/aje/kwy173
21. Hotelling H. The Generalization of Student's Ratio. *Ann Math Stat*. 1931;2(3):360–378. doi:10.1214/aoms/1177732979
22. Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc*. 1988;83(404):1198–1202. doi:10.1080/01621459.1988.10478722

23. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res*. 2011;46(3):399–424. doi:10.1080/00273171.2011.568786

24. Schober P, Vetter TR. Correct baseline comparisons in a randomized trial. *Anesth Analg*. 2019;129(3):639. doi:10.1213/ANE.0000000000004211

25. Ruddle RA, Adnan M, Hall M. Using set visualisation to find and explain patterns of missing values: a case study with NHS hospital episode statistics data. *BMJ Open*. 2022;12(11):e064887. doi:10.1136/bmjopen-2022-064887

26. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–2102. doi:10.1002/sim.8086

27. Gasparini A. Rsimsum: summarise results from Monte Carlo simulation studies. *J Open Source Softw*. 2018;3(26):739. doi:10.21105/joss.00739

28. Patel UD, Hardy NC, Smith DH, et al. Validation of acute kidney injury cases in the mini-sentinel distributed database; 2013.

29. Weinstein RB, Ryan PB, Berlin JA, Schuemie MJ, Swerdel J, Fife D. Channeling bias in the analysis of risk of myocardial infarction, stroke, gastrointestinal bleeding, and acute renal failure with the use of paracetamol compared with ibuprofen. *Drug Saf*. 2020;43(9):927–942. doi:10.1007/s40264-020-00950-3

30. Gounden V, Bhatt H, Jialal I. Renal Function Tests. In: *StatPearls*. StatPearls Publishing; 2023. Available from. http://www.ncbi.nlm.nih.gov/books/NBK507821/. Accessed June 1, 2023.

31. Van Buuren S. *Flexible Imputation of Missing Data*. CRC press; 2018.

32. Carroll OU, Morris TP, Keogh RH. How are missing data in covariates handled in observational time-to-event studies in oncology? A systematic review. *BMC Med Res Methodol*. 2020;20(1):134. doi:10.1186/s12874-020-01018-7

33. Getz K, Hubbard RA, Linn KA. Performance of multiple imputation using modern machine learning methods in electronic health records data. *Epidemiol Camb Mass*. 2023;34(2):206–215. doi:10.1097/EDE.0000000000001578

34. Vader DT, Mamtani R, Li Y, Griffith SD, Calip GS, Hubbard RA. Inverse probability of treatment weighting and confounder missingness in electronic health record-based analyses: a comparison of approaches using plasmode simulation. *Epidemiol Camb Mass*. 2023;34:520–530. doi:10.1097/EDE.0000000000001618

35. Toh S, Rodríguez LAG, Hernán MA. Analyzing partially missing confounder information in comparative effectiveness and safety research of therapeutics. *Pharmacoepidemiol Drug Saf*. 2012;21(0 2):13–20. doi:10.1002/pds.3248

36. 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

37. Sun B, Tchetgen Tchetgen EJ. On inverse probability weighting for nonmonotone missing at random data. *J Am Stat Assoc*. 2018;113 (521):369–379. doi:10.1080/01621459.2016.1256814

38. Mustillo S, Kwon S. Auxiliary variables in multiple imputation when data are missing not at random. *J Math Sociol*. 2015;39(2):73–91. doi:10.1080/0022250X.2013.877898

39. Leacy FP. *Multiple Imputation under Missing Not at Random Assumptions via Fully Conditional Specification* [Dissertation. Ph.D. Thesis]; 2018.

40. Tompsett DM, Leacy F, Moreno-Betancur M, Heron J, White IR. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. *Stat Med*. 2018;37(15):2338–2353. doi:10.1002/sim.7643

41. Weberpals J, Raman SR, Shaw PA, et al. smdi: an R package to perform structural missing data investigations on partially observed confounders in real-world evidence studies. *JAMIA Open*. 2024;7(1):ooae008. doi:10.1093/jamiaopen/ooae008