

# The Impact of Low EHR-Continuity on Effect Estimates: Evidence from Two EHR-Medicare Linked Databases

Yinzhu Jin <sup>1</sup>, Angela Y Tong <sup>1</sup>, Richard Wyss <sup>1</sup>, Jie Yang <sup>1</sup>, Kueiyu Joshua Lin <sup>1,2</sup>

<sup>1</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; <sup>2</sup>Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Correspondence: Kueiyu Joshua Lin, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont St. Suite 3030, Boston, MA, 02120, USA, Tel +1 617 278 0930, Fax +1 617 232 8602, Email jkclin@bwh.harvard.edu

**Purpose:** To compare the effect estimates obtained from patients with different levels of electronic health record (EHR)-continuity in four empirical studies: risk of pneumonia among 1) new users of proton pump inhibitors (PPI) vs H2 receptor antagonists, or 2) new users of PPI vs non-PPI users; risk of major bleeding among 3) new users of warfarin vs direct-acting oral anticoagulants, or 4) new users of oral anti-coagulants (OAC) vs non-OAC users.

**Patients and Methods:** Patients were identified in 2 US EHR systems (MA system, NC system) linked with Medicare claims data (2007/1/1 – 2014/12/31) separately. We calculated incidence rates (IR), incidence rate differences (IRD), and hazard ratios (HR) in the total linked study population and after excluding patients with the lowest 25%, 50%, or 75% of EHR-continuity scores. We quantified bias in IRD and propensity score (PS) decile-adjusted HR.

**Results:** In the MA system, IRs based on EHR-only data underestimated true rates by 44.1% to 76.2%, reduced to 12.9% to 46.5%. After excluding the lowest 75% of EHR-continuity patients, underestimation was more pronounced in non-user comparator designs. Absolute IRD bias was small for PPI vs H2RA (0.4%) and warfarin vs DOAC (0.7%), but larger for PPI vs non-PPI (19.1%) and OAC vs non-OAC (7.8%). Relative HR bias was 13.3% (PPI vs H2RA), 18.9% (PPI vs non-PPI), 3.0% (warfarin vs DOAC), and 31.5% (OAC vs non-OAC). Excluding lower-continuity patients and PS adjustment reduced IRD bias, while restricting to higher-continuity patients modestly improved HR bias.

**Conclusion:** Limiting analyses to patients with higher EHR-continuity can reduce IR underestimation and bias in effect estimates, particularly on the IRD scale. While PS adjustment mitigates some bias, EHR-discontinuity remains a source of bias, especially in studies using non-user comparators. These findings underscore the importance of balancing EHR-continuity with sample size considerations in pharmacoepidemiologic research.

**Keywords:** data leakage, care continuum, patient connectedness, data completeness, bias, effect estimates, electronic health records

## Introduction

Over the past decade, the utilization of electronic health records (EHR) data in clinical research has expanded rapidly.<sup>1</sup> EHRs capture rich clinical information essential for observational studies, such as medical diagnoses, prescribed medications, ordered procedures, vital signs, laboratory results, lifestyle factors, and clinical documents.

However, several inherent limitations of typical US EHR data can threaten the validity of the study findings. A key limitation is EHR-discontinuity (also referred to as lack of EHR-continuity), defined as clinical encounters occurring outside the study network and therefore unobserved in the data (eg, medication prescription from one hospital, but later had a hospitalization for pneumonia at an unaffiliated hospital). We use EHR-continuity as the standard term (synonym: data completeness) throughout this article, and its inverse EHR-discontinuity to denote care occurring outside the study network (previously also described as out-of-network care or data leakage).

Low EHR-continuity may result in missing or misclassified health events, leading to biased risk estimates. This could affect treatment safety evaluations and potentially misinform clinical decision-making or public health policy. To address this challenge, we previously developed and validated a prediction model to quantify the level of EHR-continuity and identify patients with high EHR-continuity.<sup>2,3</sup> In subsequent work,<sup>4</sup> we demonstrated that patients in the lowest quartile of EHR-continuity introduced substantial bias in relative effect estimates: 6–99% for short-term outcomes and 62–112% for long-term outcomes when using hazard ratios.<sup>4</sup>

However, little is known about how EHR continuity affects the underestimation of incidence rate (IR) and incidence rate differences (IRD) between treatment groups, that is, effect estimates on the absolute scale. Furthermore, while previous studies have compared effect estimates across four mutually exclusive sub-cohorts, which maybe infeasible in smaller EHR systems with limited sample sizes. A more practical alternative is to sequentially exclude patients with lower EHR-continuity—for example, by removing those in the lowest 25% of predicted EHR-continuity. By excluding patients with the lowest level of EHR-continuity, we may retain more patients than when dividing the population into mutually exclusive sub-cohorts. In light of this, the study aimed to evaluate the bias introduced by low EHR-continuity on both relative and absolute scale of effect estimates, by sequentially excluding patients with the lowest 25%, 50%, and 75% of EHR-continuity.

## Materials and Methods

### Data Source

We linked EHR-Medicare data from Massachusetts (MA) and North Carolina (NC) EHR systems. Specifically, the longitudinal claims data from fee-for-service Medicare Parts A, B, and D databases was linked with MA EHR system (2007/1/1 – 2014/12/31) and NC EHR system (2007/1/1 – 2014/12/31) separately. Detailed descriptions of each EHR system are provided elsewhere.<sup>5</sup> Both EHR systems provide structured data on demographics, medical diagnoses, procedures, medication prescriptions, and rich clinical information such as laboratory test results and lifestyle measures. The Medicare claims data provide longitudinal information on demographics, claims records for inpatient and outpatient medical diagnoses, procedures, dispensed medications. All data were accessed and analyzed in accordance with applicable data protection and privacy regulations. The study protocol was reviewed and approved by the Mass General Brigham Institutional Review Board.

### Base Study Population

The base study population included individuals who: 1) were aged  $\geq 65$  years, 2) had  $\geq 365$  days of continuous Medicare enrollment in Parts A, B, and D, and 3) had  $\geq 1$  study EHR encounter that is overlapping with the Medicare continuous enrollment period.

From the base population, we derived four empirical study cohorts. Two of them employed an active comparator, new-user design, while the other two used a non-user comparator design. The first two study examples compared the risk of pneumonia among 1) new users of proton pump inhibitors (PPI) vs H2 receptor antagonists (H2RA), and 2) new users of PPI vs non-PPI users. The other two study examples compared the major bleeding risk among 3) new users of warfarin vs direct-acting oral anticoagulants (DOAC, dabigatran, rivaroxaban, and apixaban), or 4) new users of oral anti-coagulants (OAC: warfarin, dabigatran, rivaroxaban, and apixaban) vs non-OAC users. For studies 3 and 4, we additionally required individuals to have  $\geq 1$  diagnosis of non-valvular atrial fibrillation (AF) 180 days prior to the index date (defined in the following section).

### Exposure Measurement

Exposures were measured using EHR data alone to ensure that the comparisons between EHR-based vs. EHR-claims-based estimates were conducted within the same population. Since the cohorts were fixed for each study, any observed could be attributed solely to outcome misclassification or confounder measurement error resulting from low EHR-continuity.

New users of PPI and H2RA were defined as individuals without use of the study medication in the 365 days prior to the index date, which was defined as the first prescription record in the EHR. For non-PPI users, index dates were randomly assigned within their Medicare enrollment eligibility period, excluding individuals with any PPI use in the preceding 365 days. Non-users who later initiated PPI treatment remained eligible for the non-PPI group. Non-PPI users were randomly selected from the base population and 1:1 matched to PPI users.

For the warfarin and DOAC comparisons, a 180-day washout period was applied to define new use. For the OAC vs non-OAC cohort, each new OAC user was matched to five non-OAC users based on the number of days elapsed since their most recent AF diagnosis prior to the OAC prescription date. The index date for each non-OAC user was defined as their last AF diagnosis date plus the corresponding interval between AF diagnosis and OAC initiation observed in the matched OAC user. Additional details on the OAC vs non-OAC user cohort definition are provided in the Supplementary Data ([Figure S1](#)).

## Follow-up and Outcome Ascertainment

The follow-up started one day after the index date until the earliest of 1) death, 2) 365 days after the index date (maximum follow-up), 3) end of data availability, and 4) loss of Medicare enrollment (for EHR-claims linked data only). Outcomes were measured separately based on EHR data alone and EHR-claims linked data. We used International Classification of Diseases, Ninth Revision (ICD-9) diagnostic codes to define pneumonia and major bleeding. The specific codes used to define the outcome are listed in the Supplementary Data ([Table S1](#)). To evaluate whether the impact of EHR continuity varies for more acute outcome events, we performed sensitivity analyses restricting the follow-up period to 180 days.

## Covariates Ascertainment and Algorithm Predicted EHR Continuity

Potential confounders were assessed during the 365-day pre-index period. Similar to outcome assessment, covariates were measured separately using EHR data alone and using the linked EHR-claims data. A total of 70 baseline covariates were measured, including demographics, comorbidities, prior medication use, and healthcare utilization. A full list of pre-index covariates is listed in the Supplementary Data ([Tables S2](#) and [S3](#)).

We applied the previously validated algorithm to calculate predicted EHR-continuity.<sup>2,4</sup> A full list of predictors their corresponding coefficients is provided in the Supplementary Data ([Table S4](#)). Briefly, the model predictors of the EHR-continuity were measured based on EHR data alone during a 365-day pre-index period. The model predictors are mainly indicators related to primary care follow-up in the study EHR, demographics (age, sex, and race), as well as healthcare utilization factors. The predicted EHR-continuity has been compared to the observed EHR-continuity (mean proportion of encounters captured) in our previous work, and the area under the curve for predicting the measured EHR-continuity of  $\geq 60\%$  was 0.86 in the training and 0.88 in the testing set. Furthermore, the predicted EHR-continuity value was highly correlated with the measured EHR-continuity with a Spearman coefficient = 0.78.<sup>2</sup>

The full list of ICD codes and matching algorithm (for non-user cohorts) for cohort formation and the covariates definition are available in the supplementary materials ([Table S1](#) and [Figure S1](#)) and can be provided upon reasonable request.

## Estimation of Treatment Effects

For each of the four empirical study examples described above, we calculated the incidence rate (IR) stratified by exposure status, incidence rate difference (IRD), hazard ratio (HR), and 95% confidence interval (CI) of each estimate. For hazard ratio, we checked Kaplan–Meier plots to visually check the proportional hazards assumption, and we used robust standard errors to estimate the variance. Confounding was addressed by propensity score (PS) deciles, and both IRD and HR adjusted by PS deciles were reported. Since Medicare claims capture encounters that occur outside of the study EHR systems, estimates derived from EHR-claims linked data were treated as the “reference-standard” for evaluating the impact of EHR-continuity on effect estimates. As sensitivity analysis, we also applied inverse probability weights (IPW) as an alternative confounding adjustment strategy, using the same propensity score model.

## Quantifying the Bias Due to Misclassification of Outcomes and Covariates

We quantified discrepancies between IR estimates between EHR alone and EHR-claims linked data using the “underestimation proportion”, defined as the relative difference in IR between two data sources:  $\% IR \text{ underestimation} = \left[ \frac{IR_{EHR+claims} - IR_{EHR}}{IR_{EHR+claims}} \right] \times 100\%$ .

The underestimation proportion is expected to be higher among non-user comparator groups, as these individuals are more likely to experience discontinuity of care within the study EHR system. To assess whether IR underestimation differed between non-user and active comparator designs, we calculated the relative difference in IR underestimation within each cohort as:  $\frac{[\%IR \text{ underestimation}_{exp} - \%IR \text{ underestimation}_{ref}]}{\%IR \text{ underestimation}_{exp}}$ , where  $\%IR \text{ underestimation}_{exp}$  and  $\%IR \text{ underestimation}_{ref}$  are  $\% IR$  underestimation in the exposure and reference group, respectively.

Bias in the IRD were assessed as the standardized difference between  $IRD_{EHR+claims}$  and  $IRD_{EHR}$ . Bias in HR was calculated on the logarithmic scale as follows:

$$bias (\%) = \exp\{[\ln(HR_{EHR+claims}) - \ln(HR_{EHR})]\} * 100\% - 100\%$$

To evaluate the impact of predicted EHR-continuity on bias in IR, IRD, and HR, we compared bias measured in the overall cohort and cohorts excluding the lowest 25%, 50%, or 75% of predicted EHR-continuity values.

Furthermore, we assessed the proportion of overlap in the 95% CIs for IRD and HR derived from EHR data alone and EHR+claims linked data. For example, the overlap in the 95% CIs for the crude IRD from the EHR and EHR+claims data was calculated by dividing length of overlapping segment by the total range spanned by both 95% CIs:

$$overlap = \frac{\max(0, \min(U_{EHR}, U_{EHR+claims}) - \max(L_{EHR}, L_{EHR+claims}))}{\max(U_{EHR}, U_{EHR+claims}) - \min(L_{EHR}, L_{EHR+claims})}$$

U and L are upper and lower limits of 95% CI, respectively.

## Evaluation of Bias–Variance Trade-off via Mean Squared Error

To select a practical exclusion threshold for low EHR-continuity, we quantified the bias–variance trade-off using mean squared error (MSE) at each threshold (total, excluding the lowest 25%, 50%, 75% of EHR-continuity). MSE combines systematic error (bias) and statistical imprecision (variance), with lower values indicating better overall accuracy.

For IRD, the MSE was obtained using the following formula, where  $Var(IRD_{EHR})$  is the variance of IRD from EHR data only:

$$MSE_{IRD} = (IRD_{EHR+claims} - IRD_{EHR}) + Var(IRD_{EHR})$$

For HR, we computed MSE on log scale, where  $Var(\ln(HR_{EHR}))$  is the variance of log form of HR from EHR data only:

$$MSE_{\log(HR)} = (\ln(HR_{EHR+claims}) - \ln(HR_{EHR})) + Var(\ln(HR_{EHR}))$$

## Results

### Study Population

From the MA EHR system, we identified a total of 51,099 PPI users, 14,447 H2RA users, and 51,330 non-PPI users. There were 10,590 warfarin, 1,562 DOAC, 12,152 OAC, and 44,252 non-OAC users who had non-valvular AF (Tables 1 and S2). In the NC system, we identified 5,620 PPI, 1,421 H2RA, 5,694 non-PPI, 859 warfarin, 266 DOAC, 1,125 OAC, and 5,103 non-OAC users (Table S3).

The mean age across cohorts ranged from 75.4 to 78.5 years in the MA system, and from 74.9 to 77.8 years in the NC system. The proportion of female ranged from 42.5% to 62.9% in the MA system cohorts and from 45.5% to 60.7% in the NC system cohorts. Full characteristics of the study populations are presented in Tables S2 and S3.

**Table 1** Patients' Characteristics and Absolute Standardized Mean Difference (%) at Baseline, MA System

Variable	Study Cohort: Medication Exposures and Outcomes											
	PPI vs H2RA and Pneumonia			PPI vs non-PPI and Pneumonia			Warfarin vs DOAC and Bleeding			OAC vs Non-OAC and Bleeding		
	PPI	H2RA	ASMD	PPI	Non-PPI	ASMD	Warfarin	DOAC	ASMD	OAC	Non-OAC	ASMD
<b>Total</b>	51099	14447		51099	51330		10590	1562		12152	44252	
<b>Age in years, mean (SD)</b>	76.0 (7.4)	75.6 (7.3)	5.4	76.0 (7.4)	75.4 (7.1)	8.1	78.4 (7.3)	76.3 (6.8)	30.1	78.2 (7.3)	78.5 (8)	4.9
<b>Female sex, %</b>	60.6	62.9	4.6	60.6	61.7	2.2	47.3	42.8	9.1	46.7	55.0	<b>16.6</b>
<b>Race, %</b>												
<b>Caucasian</b>	85.0	83.1	5.3	85.0	85.5	1.4	88.6	91.3	8.9	89.0	86.4	7.9
<b>African American</b>	3.5	4.4	4.4	3.5	3.1	2.5	2.9	2.1	5.2	2.8	2.7	0.7
<b>Others</b>	11.5	12.6	3.3	11.5	11.4	0.2	8.4	6.6	7.0	8.2	10.9	9.2
<b>Recent hospitalization<sup>a</sup>, %</b>	9.2	8.7	1.7	9.2	3.0	<b>26.3</b>	9.8	9.5	0.7	9.7	22.0	<b>34.1</b>
<b>Number of distinct prescriptions, %</b>												
<b>0-2</b>	5.0	5.1	0.2	5.0	16.8	<b>38.4</b>	1.7	1.3	3.8	1.7	3.6	<b>12.2</b>
<b>3-4</b>	9.5	8.9	2.3	9.5	20.3	<b>30.6</b>	6.9	6.9	0.2	6.9	7.8	3.5
<b>5-7</b>	21.0	18.8	5.6	21.0	28.1	<b>16.7</b>	21.9	19.0	7.2	21.5	18.7	7.0
<b>&gt;7</b>	64.4	67.3	6.0	64.4	34.8	<b>62.1</b>	69.5	72.9	7.3	70.0	69.9	0.1
<b>Number of hospitalizations, %</b>												
<b>0-1</b>	31.3	30.8	1.2	31.3	15.7	<b>37.4</b>	32.3	35.0	5.8	32.6	38.4	<b>12.1</b>
<b>&gt;1</b>	21.3	23.1	4.4	21.3	7.4	<b>40.4</b>	24.8	23.4	3.3	24.6	32.0	<b>16.5</b>
<b>Length of stay in hospital in days, %</b>												
<b>1-3</b>	8.3	7.6	2.6	8.3	5.8	10.0	10.0	15.3	<b>15.9</b>	10.7	9.0	5.6
<b>&gt;3</b>	44.3	46.2	3.9	44.3	17.3	<b>60.9</b>	47.0	43.1	7.9	46.5	61.4	<b>30.2</b>
<b>Number of office visits, %</b>												
<b>0-1</b>	4.0	4.0	0.4	4.0	4.9	4.6	3.0	1.9	6.7	2.8	4.9	<b>10.9</b>

(Continued)

Table I (Continued).

Variable	Study Cohort: Medication Exposures and Outcomes											
	PPI vs H2RA and Pneumonia			PPI vs non-PPI and Pneumonia			Warfarin vs DOAC and Bleeding			OAC vs Non-OAC and Bleeding		
	PPI	H2RA	ASMD	PPI	Non-PPI	ASMD	Warfarin	DOAC	ASMD	OAC	Non-OAC	ASMD
2-3	6.3	6.6	1.2	6.3	10.5	<b>15.1</b>	5.6	3.4	<b>10.7</b>	5.3	6.5	5.2
4-5	9.3	9.5	0.5	9.3	13.6	<b>13.5</b>	9.4	8.6	2.5	9.3	9.1	0.6
>5	80.4	79.9	1.4	80.4	71.0	<b>22.1</b>	82.1	86.0	<b>10.8</b>	82.6	79.4	8.1
Alcohol abuse, %	1.9	1.9	0.1	1.9	0.8	9.0	1.7	2.2	3.8	1.7	2.7	6.8
Drug abuse, %	0.8	1.1	2.5	0.8	0.3	6.9	0.6	1.0	4.2	0.7	1.3	6.1
Hypertension, %	88.9	88.7	0.8	88.9	81.5	<b>21.0</b>	98.3	98.7	3.4	98.4	96.3	<b>13.0</b>
Hyperlipidemia, %	75.3	74.5	1.7	75.3	71.0	9.5	77.5	80.5	7.3	77.9	75.6	5.4
Coronary artery disease, %	40.8	41.1	0.5	40.8	27.2	<b>29.1</b>	55.9	51.4	9.0	55.3	54.4	2.0
Systemic embolism, %	1.5	1.8	1.9	1.5	0.8	7.3	2.1	1.7	3.0	2.1	2.1	0.4
DVT, %	6.7	7.3	2.2	6.7	3.9	<b>12.8</b>	9.6	5.4	<b>15.9</b>	9.1	8.2	3.0
Pulmonary embolism, %	3.0	3.1	0.5	3.0	1.5	<b>10.4</b>	4.6	2.4	<b>11.6</b>	4.3	3.5	4.2
Atrial fibrillation/flutter, %	25.0	23.0	4.7	25.0	16.6	<b>20.8</b>	97.3	99.4	<b>16.4</b>	97.5	100.0	<b>22.5</b>
Peripheral vascular disease, %	13.1	13.7	1.8	13.1	9.0	<b>13.3</b>	18.1	11.1	<b>19.7</b>	17.2	17.0	0.5
Heart failure, %	18.8	18.4	0.9	18.8	8.5	<b>30.4</b>	37.7	29.4	<b>17.6</b>	36.7	31.4	<b>11.1</b>
Ischemic stroke, %	10.4	12.5	6.5	10.4	6.2	<b>15.1</b>	16.1	14.5	4.6	15.9	16.4	1.4
Transient ischemic attack, %	5.0	5.6	2.6	5	3.4	8.2	7.7	7.6	0.4	7.7	8.3	2.2
Old myocardial infarction, %	10.2	11.4	4.0	10.2	5.5	<b>17.5</b>	14.2	12.9	3.8	14.1	14.7	1.7
Acute myocardial infarction, %	7.5	8.8	5.0	7.5	2.5	<b>22.8</b>	9.9	9.0	2.9	9.8	12.8	9.6
Diabetes, %	27.9	28.6	1.4	27.9	21.1	<b>16.1</b>	29.6	23.8	<b>13.3</b>	28.9	29.3	1.1
COPD, %	22.3	22.2	0.1	22.3	13.0	<b>24.4</b>	21.5	18.1	8.5	21.0	26.7	<b>13.5</b>
Asthma, %	14.1	14.5	1.0	14.1	8.9	<b>16.2</b>	11.4	12.2	2.7	11.5	13.3	5.5
Pneumonia, %	16.2	17.4	3.2	16.2	7.4	<b>27.4</b>	17.3	13.2	<b>11.5</b>	16.8	24.9	<b>20.0</b>

Upper GI bleeding, %	2.9	2.0	5.8	2.9	0.5	<b>18.4</b>	1.7	1.8	0.6	1.7	3.1	8.9
Lower GI bleeding, %	12.8	10.5	7.2	12.8	5.7	<b>24.5</b>	10.2	10.0	0.9	10.2	14.0	<b>11.6</b>
Liver disease, %	12.4	11.5	2.6	12.4	6.8	<b>19.1</b>	8.6	7.3	5.0	8.5	11.5	<b>10.1</b>
Renal dysfunction, %	32.4	32.2	0.5	32.4	19.6	<b>29.5</b>	36.6	28.7	<b>16.7</b>	35.6	39.1	7.3
Acute renal disease, %	12.6	13.0	1.2	12.6	4.6	<b>29.1</b>	14.8	10.8	<b>11.9</b>	14.3	20.1	<b>15.5</b>
End-stage renal disease, %	3.0	3.4	2.1	3.0	1.3	<b>11.6</b>	4.4	0.8	<b>22.3</b>	3.9	4.1	1.2
Osteoarthritis, %	38.5	37.2	2.8	38.5	32.6	<b>12.4</b>	36.0	37.2	2.4	36.2	37.4	2.5
Rheumatoid arthritis, %	4.5	4.0	2.6	4.5	2.9	8.4	3.1	3.7	3.5	3.2	4.0	4.3

**Notes:** <sup>a</sup>Recent hospitalization within 30 days before the cohort entry date. Bolded values represent ASMD > 10%, indicating imbalance between exposure groups.

**Abbreviations:** ASMD, absolute standardized mean difference; COPD, chronic obstructive pulmonary disease; DOAC, direct oral anticoagulants; DVT, deep vein thromboembolism; GI, gastrointestinal; H2RA, histamine type-2 receptor antagonists; OAC, oral anticoagulant; PPI, proton pump inhibitor; SD, standard deviation.

## Underestimated Incidence Rates

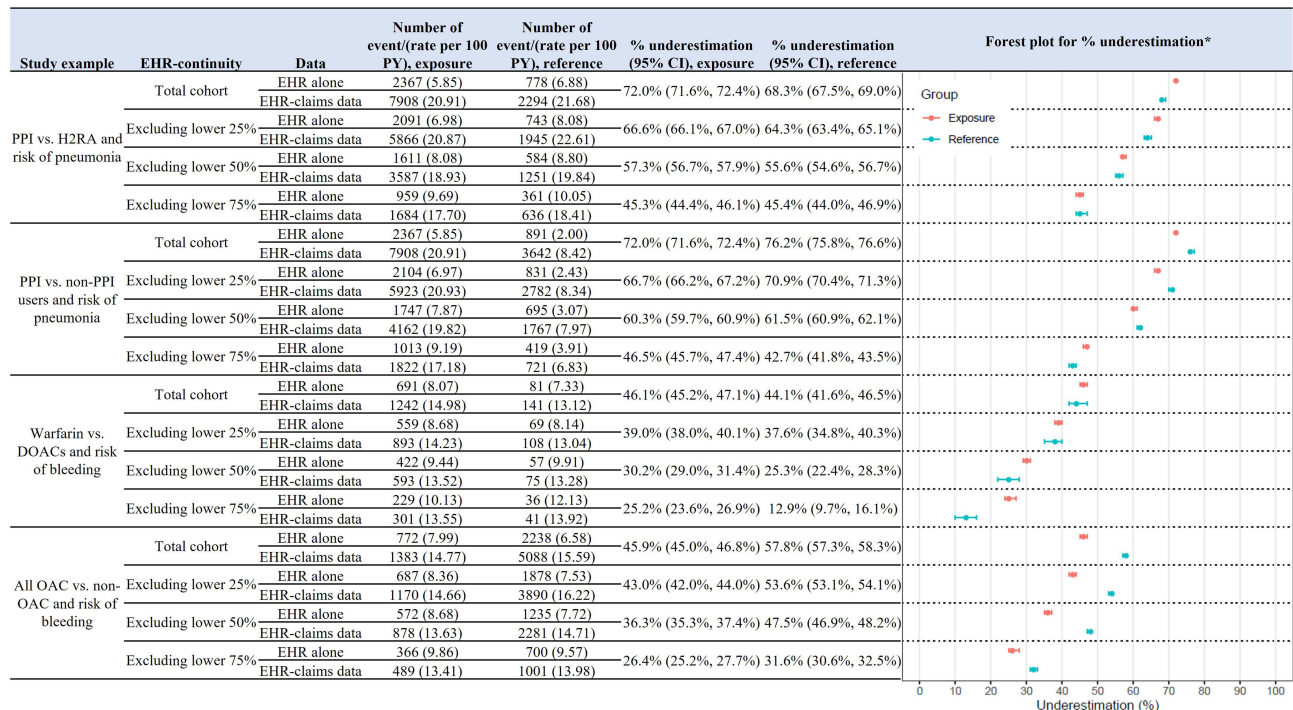
In the MA EHR system, IR underestimation was highest in the total cohort and progressively decreased with the exclusion of the lowest 25%, 50%, 75% of predicted EHR-continuity patients (Figure 1). The decreasing trend in IR underestimation was consistently observed across all study examples in both EHR systems (Figure 2). The number of events and follow-up time are listed in Table S5.

In MA total cohort, the IR underestimation among PPI users was 72.0%. Non-PPI users had higher underestimation compared to the active comparator H2RA users (76.2% vs 68.3%). However, after excluding the lowest 75% EHR continuity, the level of IR underestimation became comparable across the PPI, non-PPI, and H2RA users (45.3%, 45.4%, and 42.7%, respectively). Similarly, in the NC EHR system, we observed that the non-PPI comparator group exhibited greater IR underestimation than both the PPI exposure and H2RA comparator groups (Figure 2). The relative difference in IR underestimation in MA total cohort was 9.5% for PPI vs H2RA and 28.6% for PPI vs non-PPI cohort.

The difference in IR underestimation between the exposure and non-user comparator group was more pronounced in study examples 3 and 4 (Figure 1). While active comparators groups showed similar levels of underestimation (46.1% for warfarin and 44.1% for DOACs in MA cohort), the non-user comparator group consistently showed higher underestimation compared to the exposure groups (45.9% for OAC and 57.8% for non-OAC users in total cohorts). The relative difference in IR underestimation in MA total cohort was 4.4% for warfarin vs DOACs cohort and 25.9% for OAC vs non-OAC cohort. The difference was greater in the NC EHR system (27.2% for OAC vs 52.6% for non-OAC with a relative difference of 93.5%; Figure 2).

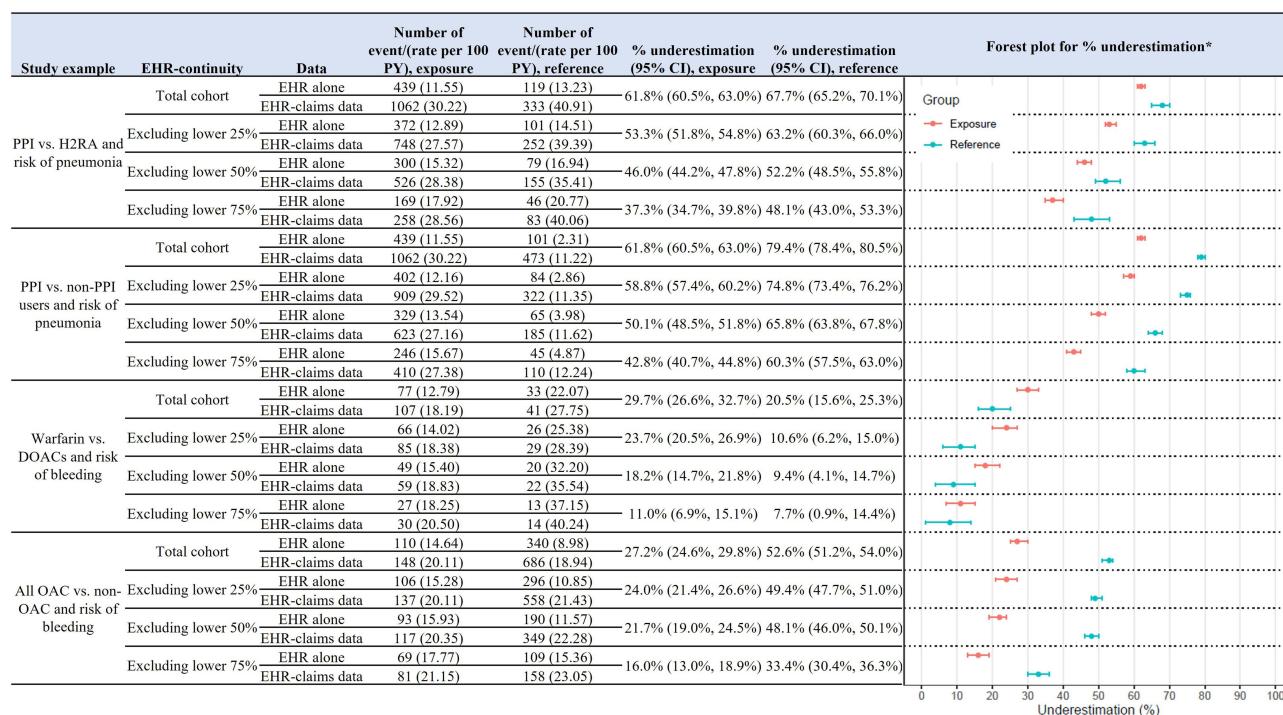
## Comparing Bias and CI Overlap in IRD Estimates

Overall, we observed relatively greater bias in IRDs among the non-user comparator cohorts compared to the active comparator cohorts. For example, the crude IRD bias was 0.4% among PPI vs H2RA %, 19.1% among PPI vs non-PPI cohort (Table 2). Similarly, the bias was 0.7% in the warfarin vs DOAC cohort, whereas it reached 7.8% in the OAC vs non-OAC cohort. After adjusting for PS deciles, we observed a reduction in bias among non-user comparator cohorts –



**Figure 1** Incidence rates and percentage of underestimated incidence rates by exposure group: MA EHR system. \*Underestimation of incidence rates was calculated as difference in the incidence rate from EHR+claims data and EHR only, divided by the incidence rate from EHR+claims data:  $((IR_{EHR+claims} - IR_{EHR})/IR_{EHR+claims})$ .

**Abbreviations:** CI, confidence interval; DOAC, direct-acting oral anticoagulant; EHR, electronic health records; H2RA, H2 receptor antagonists; OAC, oral anticoagulant; PPI, proton pump inhibitors; PY, person-years.



**Figure 2** Incidence rates and percentage of underestimated incidence rates\* by exposure group: NC EHR system. \*Underestimation of incidence rates was calculated as difference in the incidence rate from EHR+claims data and EHR only, divided by the incidence rate from EHR+claims data:  $((IR_{EHR+claims} - IR_{EHR})/IR_{EHR+claims})$ .

**Abbreviations:** CI, confidence interval; DOAC, direct-acting oral anticoagulant; EHR, electronic health records; H2RA, H2 receptor antagonists; OAC, oral anticoagulant; PPI, proton pump inhibitors; PY, person-years.

for example, in the PPI vs non-PPI cohort, the bias decreased from 19.1% to 8.4%. Excluding patients with lower EHR-continuity also reduced bias, for example, bias decreased from 19.1% to 11.5% after excluding the lowest 75% of EHR-continuity patients. The overlap in the 95% CI for crude IRD was 59% in the PPI vs H2RA and 54% warfarin vs DOACs cohorts. We did not observe any overlap in the 95% CI for the non-user comparator cohorts. There was an improvement in the overlap in the OAC vs non-OAC cohort after excluding the lowest 75% of EHR-continuity and following confounding adjustment (Table 2 and Figure S2). We observed similar trend in IRD bias and 95% CI overlap from the NC system results (Table S6).

### Comparing Bias and CI Overlap in HR Estimates

The HR bias in the PPI vs H2RA and PPI vs non-PPI from the MA system remained relatively small, ranging from 0% to 19% for the crude HR from 1 to 11% for the PS decile-adjusted HR (Table 3). Excluding lower EHR-continuity subsets improved the bias in crude HR modestly. However, after confounding adjustment using PS deciles, the HR bias was similar in the total cohort and in the cohorts excluding patients with lower EHR-continuity, ranging from 1% to 3%. We observed similar pattern in PPI vs non-PPI study example (Table 3).

The OAC vs non-OAC study cohort showed smaller bias in the crude HR than the warfarin vs DOAC cohort (3% vs 31%, respectively). After adjusting for PS deciles, HR bias was similar across different EHR-continuity levels in both cohorts. Consistent with the 95% CI overlaps observed in the IRD estimates, we found that CI overlap was poor in the cohorts with non-user comparators. Excluding patients with lower EHR-continuity or adjusting for PS deciles improved the 95% CI overlap in these study examples (Table 3 and Figure S3). We observed similar trend in HR bias and 95% CI overlap from the NC system results (Table S7). Visual inspection of Kaplan–Meier plots indicated that the proportional hazards assumption was reasonably satisfied (Figures S4–S7).

Results for bias and CI overlap in IRD and HR were similar when we restricted the follow-up time to 180 days (Tables S8–S10) or used IPW as confounding adjustment approach (Tables S11 and S12).

**Table 2** Metrics for Crude and Adjusted Incidence Rate Difference per 100 PY and by Level of EHR-Continuity: MA EHR System

Medication/ outcome	Level of EHR-continuity	Crude IRD				Adjusted* IRD			
		IRD <sub>EHR</sub> (95% CI)	IRD <sub>EHR+claims</sub> (95% CI)	Bias	95% CI overlap	IRD <sub>EHR</sub> (95% CI)	IRD <sub>EHR+claims</sub> (95% CI)	Bias	95% CI overlap
PPI vs H2RA and risk of pneumonia	Total cohort	-1.03 (-1.55, -0.51)	-0.78 (-1.66, 0.11)	0.4%	59%	0.00 (-0.34, 0.34)	-0.21 (-0.96, 0.54)	0.4%	46%
	Excluding lower 25%	-1.10 (-1.73, -0.48)	-1.74 (-2.75, -0.74)	1.0%	44%	-0.07 (-0.50, 0.36)	-0.59 (-1.43, 0.25)	1.0%	42%
	Excluding lower 50%	-0.72 (-1.50, 0.06)	-0.91 (-2.04, 0.22)	0.3%	69%	0.22 (-0.34, 0.77)	0.29 (-0.63, 1.20)	0.1%	60%
	Excluding lower 75%	-0.36 (-1.50, 0.78)	-0.71 (-2.22, 0.79)	0.6%	76%	0.47 (-0.38, 1.32)	0.29 (-0.92, 1.50)	0.4%	70%
PPI vs non-PPI users and risk of pneumonia	Total cohort	3.84 (3.58, 4.11)	12.48 (12.00, 12.97)	19.1%	0%	1.49 (1.28, 1.70)	4.39 (4.02, 4.76)	8.4%	0%
	Excluding lower 25%	4.54 (4.21, 4.87)	12.59 (12.03, 13.15)	17.5%	0%	1.35 (1.10, 1.59)	4.39 (3.97, 4.82)	8.8%	0%
	Excluding lower 50%	4.80 (4.38, 5.22)	11.85 (11.21, 12.50)	15.5%	0%	1.44 (1.13, 1.75)	4.39 (3.90, 4.89)	8.6%	0%
	Excluding lower 75%	5.27 (4.62, 5.93)	10.35 (9.49, 11.21)	11.5%	0%	1.62 (1.14, 2.10)	3.80 (3.15, 4.46)	6.6%	0%
Warfarin vs DOACs and risk of bleeding	Total cohort	0.74 (-0.91, 2.38)	1.86 (-0.29, 4.02)	0.7%	54%	-0.70 (-2.30, 0.90)	0.45 (-1.64, 2.54)	0.7%	52%
	Excluding lower 25%	0.54 (-1.43, 2.50)	1.20 (-1.25, 3.65)	0.4%	74%	-1.02 (-2.95, 0.90)	-0.83 (-3.26, 1.59)	0.1%	79%
	Excluding lower 50%	-0.48 (-3.07, 2.11)	0.24 (-2.73, 3.22)	0.4%	77%	-2.22 (-4.73, 0.29)	-1.54 (-4.41, 1.33)	0.4%	77%
	Excluding lower 75%	-2.00 (-5.91, 1.92)	-0.38 (-4.58, 3.83)	0.9%	67%	-3.85 (-7.81, 0.12)	-2.33 (-6.47, 1.82)	0.9%	68%
All OAC vs non-OAC and risk of bleeding	Total cohort	1.40 (0.80, 2.01)	-0.83 (-1.64, -0.01)	7.8%	0%	1.25 (0.44, 2.06)	1.94 (0.84, 3.04)	1.8%	47%
	Excluding lower 25%	0.83 (0.15, 1.51)	-1.56 (-2.47, -0.66)	8.7%	0%	1.49 (0.57, 2.40)	1.91 (0.65, 3.16)	1.1%	68%
	Excluding lower 50%	0.96 (0.16, 1.75)	-1.08 (-2.09, -0.07)	7.9%	0%	2.28 (1.20, 3.37)	2.61 (1.21, 4.01)	0.9%	77%
	Excluding lower 75%	0.30 (-0.88, 1.47)	-0.57 (-1.94, 0.80)	3.4%	49%	2.45 (0.84, 4.06)	3.40 (1.51, 5.29)	2.7%	57%

**Notes:** \*Adjusted by propensity score decile.

**Abbreviations:** CI, confidence interval; DOAC, direct-acting oral anticoagulant; HER, electronic health records; H2RA, H2 receptor antagonists; IRD, incidence rate difference; OAC, oral anticoagulant; PPI, proton pump inhibitors.

**Table 3** Metrics for Crude and Adjusted Hazard Ratio by Level of EHR-Continuity: MA EHR System

Medication/ Outcome	Level of EHR-Continuity	Crude				Adjusted*			
		HR <sub>EHR</sub> (95% CI)	HR <sub>EHR+claims</sub> (95% CI)	Bias	95% CI overlap	Adjusted* HR <sub>EHR</sub> (95% CI)	Adjusted HR <sub>EHR+claims</sub> (95% CI)	Bias	95% CI overlap
PPI vs H2RA and risk of pneumonia	Total cohort	0.85 (0.79, 0.92)	0.97 (0.92, 1.01)	13.3%	0%	1.00 (0.93, 1.08)	0.99 (0.95, 1.03)	1.2%	58%
	Excluding lower 25%	0.87 (0.80, 0.94)	0.93 (0.89, 0.97)	7.0%	30%	0.99 (0.92, 1.07)	0.97 (0.93, 1.02)	1.8%	61%
	Excluding lower 50%	0.92 (0.85, 1.01)	0.96 (0.90, 1.02)	3.9%	60%	1.03 (0.94, 1.12)	1.02 (0.96, 1.08)	1.3%	67%
	Excluding lower 75%	0.97 (0.87, 1.09)	0.97 (0.89, 1.05)	0.2%	75%	1.05 (0.94, 1.18)	1.02 (0.93, 1.11)	3.3%	70%
PPI vs non-PPI users and risk of pneumonia	Total cohort	2.87 (2.66, 3.10)	2.42 (2.32, 2.51)	18.9%	0%	1.58 (1.44, 1.73)	1.46 (1.40, 1.52)	8.4%	23%
	Excluding lower 25%	2.83 (2.61, 3.07)	2.45 (2.34, 2.56)	15.6%	0%	1.44 (1.32, 1.58)	1.46 (1.39, 1.53)	1.4%	54%
	Excluding lower 50%	2.54 (2.33, 2.77)	2.44 (2.31, 2.58)	4.1%	54%	1.39 (1.26, 1.52)	1.49 (1.40, 1.58)	7.2%	39%
	Excluding lower 75%	2.33 (2.08, 2.61)	2.48 (2.27, 2.71)	6.5%	54%	1.35 (1.20, 1.53)	1.50 (1.37, 1.65)	10.9%	35%
Warfarin vs DOACs and risk of bleeding	Total cohort	1.15 (0.89, 1.49)	1.19 (0.98, 1.44)	3.0%	78%	0.93 (0.71, 1.22)	1.04 (0.85, 1.28)	12.3%	64%
	Excluding lower 25%	1.12 (0.84, 1.48)	1.14 (0.91, 1.42)	1.9%	79%	0.90 (0.67, 1.21)	0.95 (0.76, 1.20)	5.8%	84%
	Excluding lower 50%	1.00 (0.74, 1.35)	1.06 (0.82, 1.38)	6.6%	82%	0.82 (0.59, 1.12)	0.91 (0.69, 1.20)	11.5%	71%
	Excluding lower 75%	0.88 (0.59, 1.29)	1.02 (0.71, 1.46)	16.3%	67%	0.73 (0.48, 1.12)	0.87 (0.59, 1.26)	18.1%	67%
All OAC vs non-OAC and risk of bleeding	Total cohort	1.33 (1.19, 1.48)	1.01 (0.93, 1.09)	31.5%	0%	1.27 (1.11, 1.46)	1.17 (1.07, 1.28)	8.8%	43%
	Excluding lower 25%	1.20 (1.07, 1.35)	0.97 (0.89, 1.06)	24.1%	0%	1.24 (1.08, 1.43)	1.16 (1.05, 1.28)	6.8%	54%
	Excluding lower 50%	1.33 (1.16, 1.52)	1.04 (0.93, 1.15)	28.5%	0%	1.47 (1.25, 1.73)	1.27 (1.12, 1.43)	16.0%	29%
	Excluding lower 75%	1.22 (1.02, 1.47)	1.14 (0.98, 1.32)	7.7%	62%	1.53 (1.24, 1.88)	1.42 (1.19, 1.70)	7.4%	67%

**Notes:** \*Adjusted by propensity score decile.

**Abbreviations:** CI, confidence interval; DOAC, direct-acting oral anticoagulant; HER, electronic health records; H2RA, H2 receptor antagonists; IRD, incidence rate difference; OAC, oral anticoagulant; PPI, proton pump inhibitors.

## MSE for IRD and HR Across EHR-Continuity Levels

PS-adjusted IRD MSE was low (0–2%) across thresholds and insensitive to exclusions ([Table S13](#) and [Figure S8](#)). Crude IRD MSE was low (MA 4–14%, NC 8–32%) and tended to rise with 75% exclusion. Crude HR MSE was highest and generally increased with exclusion (warfarin vs DOACs: MA 36 to 47%, NC 48 to 61%), except NC PPI vs non-PPI where it declined (73 to 51–53%). PS adjustment sharply reduced HR MSE (eg, MA 36 to 13%; NC 48 to 22%) and remained stable. Overall, MSEs were low; crude HR had the highest MSE but dropped substantially after PS-decile adjustment.

## Discussion

The growing use of EHR data in recent epidemiologic research has enabled the identification of key study variables that are often not available in insurance claims databases.<sup>6</sup> However, EHR-discontinuity, or incomplete capture of care across health care systems, can lead to misclassification of key variables, potentially compromising the validity of effect estimates derived from EHR data alone. Our previous work has provided solid evidence that patients with low predicted EHR-continuity had a greater magnitude of misclassification of study variables.<sup>2,3,5,7,8</sup> In our prior work, we assessed bias only on the relative scale and across sub-cohorts defined by EHR-continuity quartiles. In this study, we build on that foundation by demonstrating the impact of EHR-continuity on both absolute and relative effect estimates. We also evaluate practical strategies to balance the trade-off between minimizing bias and preserving precision. Using four empirical CER examples drawn from two US EHR systems in MA and NC, we demonstrated that lower EHR-continuity is consistently associated with greater underestimation in IR. This effect was more pronounced in non-user comparator designs than in active user comparator designs.

## Underestimation Risk and Rates in the Descriptive Epidemiology

We observed that IRs estimated using EHR data alone consistently underestimated the “true” IRs derived from EHR-claims linked data. This finding aligns with the earlier study,<sup>4</sup> in which we reported reduced outcome event capture rates in certain empirical CER examples. These findings suggest that in EHR-based descriptive epidemiologic studies, researchers should exercise caution when reporting IRs. The risk of underestimation is especially pronounced in cohorts that include non-user comparator groups. Therefore, EHR-continuity should be considered when reporting risks or incidence rates using EHR data, and sensitivity analyses excluding patients with lower EHR-continuity subset should be conducted to assess the robustness of risk/rate estimates. This issue is not limited to CER studies, but also applies equally to descriptive and surveillance studies. Results from EHR-based analyses should be reported with a clear acknowledgement of the potential of underestimation of risk and rates due to incomplete data capture.

## Differential vs Non-Differential Misclassification of Outcome

If outcome misclassification differs by exposure status, both absolute measures such as risk/rate difference and relative measures such as risk/rate/hazard ratio could be biased in unpredictable directions, either toward or away from the null. Our findings demonstrate a greater likelihood of differential outcome misclassification in the non-user comparator cohorts compared to active comparator cohorts. We observed in the PPI vs H2RA and warfarin vs DOAC examples, IR underestimation was similar between the exposure and reference groups. In contrast, the corresponding non-user comparisons – PPI vs non-PPI and OAC vs non-OAC – exhibited greater differences in IR underestimation between groups. To mitigate bias from differential misclassification, an active-comparator design is preferred. This design not only helps reduce outcome misclassification but is also widely recommended to minimize confounding and other potential sources bias.<sup>9–11</sup> In situations where a non-user comparator is unavoidable, the study cohort could be restricted to higher EHR-continuity (eg, excluding the lowest 25%) to mitigate potential differential misclassification between the exposure and comparator groups. We recommend reporting both absolute and relative measures and presenting CI-overlap diagnostics across different EHR-continuity levels. Future research is warranted to examine the magnitude of differential misclassification by considering negative-control outcomes or exposures.

## Other Factors Related to the Effect Estimation

Outcome misclassification is not the only source of bias nor the determinant of the bias direction in epidemiologic studies. The magnitude of both outcome misclassification and confounder misclassification can significantly affect the effect estimate validity. We noted that the impact of EHR-continuity on bias is more pronounced in the crude estimates than in the adjusted IRD or HR. This may be explained by partial adjustment for EHR-continuity levels when other confounders were accounted for in the PS model. In other words, adjusting for comorbidity or healthcare utilization as confounders may indirectly adjust for EHR-continuity, since certain comorbidity or healthcare utilization patterns are correlated with patients' EHR-continuity. We also observed that the bias related to EHR-continuity appears to be greater when effect estimates were measured on the absolute scale than on the relative scale. However, given the limited number of examples investigated in our study, these findings should be interpreted as descriptive. Future studies encompassing a broader range of research questions and setting is needed to assess the generalizability of these results.<sup>12</sup>

Another factor related to bias is the study sample size and the number of observed outcome events. Smaller sample sizes are more susceptible to random variation (“chance-finding”) and are therefore associated with greater bias.<sup>13</sup> In our study examples, the population size was smaller in the NC system, and the warfarin/DOAC/OAC/non-OAC study examples had relatively fewer outcome events. As a result, effect estimates in these cohorts became more imprecise, leading to greater fluctuation in bias.

MSE did not reveal a clear “optimal” exclusion threshold — patterns were flat or worsened with aggressive exclusions—so there is no practical MSE-based “sweet spot”. Given the consistent bias reduction with minimal impact on variance, a pragmatic threshold is to exclude the lowest 25% of EHR-continuity. This yields lower bias (and, in NC PPI vs non-PPI, lower HR MSE) while avoiding the diminishing returns and precision loss seen with  $\geq 50$ –75% exclusions.

## Recommendations for Future EHR-Based CER Studies

We recommend reporting both absolute and relative measures' effect estimates, along with CI-overlap diagnostics. Furthermore, EHR-based CER should routinely quantify and report predicted EHR-continuity distributions and present sensitivity analyses by exclusion thresholds. These recommendations align with current RWE guidance emphasizing data relevance and reliability, and transparent study conduct. In particular, FDA (US Food and Drug Administration)'s 2024 guidance on using EHR/claims data calls for documentation of data provenance, linkage quality, missingness, validation of key variables, and fitness-for-use assessments.<sup>14</sup> The recommendations also align with the ISPE-ISPOR (International Society of Pharmacoepidemiology – International Society for Pharmacoeconomics and Outcomes Research) initiatives to improve transparency and reliability of RWE.<sup>15,16</sup>

## Limitations

The study results have several limitations. First, EHR systems across the US vary substantially in documentation practices and data availability. Although we observed relatively consistent findings across two US systems in different states, both are based in academic institutions; therefore, generalizability to non-academic or community-based health-care networks may be limited. Second, we treated effect estimates derived from EHR-claims data as the “reference-standard”, though these do not represent the true causal effect. The objective of this study was to assess the relative bias resulting from data leakage in EHR-only analyses, compared with estimates from EHR-claims data. Third, we did not include all commonly used analytical methods, such as as-treated analysis or alternative confounding adjustment strategies such as propensity score matching or stratification. Lastly, while restricting study cohorts to individuals with higher EHR-continuity can reduce bias due to data leakage, it inevitably reduces sample size and lead to less precise effect estimates.

## Conclusion

Using data from two US academic network in different states, we found that restricting study cohorts to individuals with higher EHR-continuity, such as by excluding patients in the lowest 25% of predicted EHR-continuity, can reduce the

underestimation of IRs when relying on EHR data alone. Bias in effect estimates due to EHR-discontinuity was more pronounced in crude than in the adjusted estimates. We also demonstrated that, in EHR-based CER studies, effect estimates in relative scale are less susceptible to bias from data discontinuity than those on the absolute scale. Researchers should carefully weigh the trade-off between optimizing EHR-continuity and preserving adequate sample size to ensure both validity and precision of effect estimates. We recommend that future EHR-based CER routinely quantify and report predicted EHR-continuity distributions and present sensitivity analyses by exclusion thresholds. Our findings derived from two academic EHR networks and four exemplar pharmacoepidemiologic scenarios, and these may not fully generalize to non-academic systems or other therapeutic areas, and replication in diverse settings is warranted.

## Funding

This project was funded by NIH 1R01LM013204 and 1R01LM012594.

## Disclosure

All authors declared no competing interests for this work.

## References

- Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Method.* 2021;21(1):234. doi:10.1186/s12874-021-01416-5
- Lin KJ, Singer DE, Glynn RJ, Murphy SN, Lii J, Schneeweiss S. Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. *Clin Pharmacol Ther.* 2018;103(5):899–905. doi:10.1002/cpt.861
- Merola D, Schneeweiss S, Jin Y, Lii J, Lin KJ. Advancing an algorithm for the identification of patients with high data-continuity in electronic health records. *Clin Epidemiol.* 2022;14:1339–1349. doi:10.2147/CLEP.S370031
- Joshua Lin K, Jin Y, Gagne J, et al. Longitudinal data discontinuity in electronic health records and consequences for medication effectiveness studies. *Clin Pharmacol Ther.* 2022;111(1):243–251. doi:10.1002/cpt.2400
- Lin KJ, Rosenthal GE, Murphy SN, et al. External validation of an algorithm to identify patients with high data-completeness in electronic health records for comparative effectiveness research. *Clin Epidemiol.* 2020;12:133–141. doi:10.2147/clep.S232540
- Cowie MR, Blomster JL, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol.* 2017;106(1):1–9. doi:10.1007/s00392-016-1025-6
- Kar S, Besette LG, Wyss R, Kesselheim AS, Lin KJ. The impact of electronic health record discontinuity on prediction modeling. *PLoS One.* 2023;18(7):e0287985. doi:10.1371/journal.pone.0287985
- Jin Y, Schneeweiss S, Merola D, Lin KJ. Impact of longitudinal data-completeness of electronic health record data on risk score misclassification. *J Am Med Inform Assoc.* 2022;29(7):1225–1232. doi:10.1093/jamia/ocac043
- Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol.* 2015;11(7):437–441. doi:10.1038/nrrheum.2015.30
- Wintzell V, Svanström H, Pasternak B. Selection of comparator group in observational drug safety studies: alternatives to the active comparator new user design. *Epidemiology.* 2022;33(5):707–714. doi:10.1097/ede.0000000000001521
- Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep.* 2015;2(4):221–228. doi:10.1007/s40471-015-0053-5
- Nab L, Groenwold RHH, van Smeden M, Keogh RH. Quantitative bias analysis for a misclassified confounder: a comparison between marginal structural models and conditional models for point treatments. *Epidemiology.* 2020;31(6):796–805. doi:10.1097/ede.0000000000001239
- Springate SD. The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. *Eur J Orthodont.* 2011;34(2):158–163. doi:10.1093/ejo/cjr010
- Food and Drug Administration. Real-world evidence: considerations regarding non-interventional studies for drug and biological products; 2023.
- Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ.* 2021;372:m4856. doi:10.1136/bmj.m4856
- Wang SV, Pottegård A. Building transparency and reproducibility into the practice of pharmacoepidemiology and outcomes research. *Am J Epidemiol.* 2024;193(11):1625–1631. doi:10.1093/aje/kwae087

Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

**Dovepress**  
Taylor & Francis Group