



Published in final edited form as:

Epidemiology. 2018 May ; 29(3): 356–363. doi:10.1097/EDE.0000000000000794.

Out-of-system care and recording of patient characteristics critical for comparative effectiveness research

Kueiyu Joshua Lin, MD, ScD, MPH^{1,2,3}, Robert J. Glynn, PhD, ScD^{1,3}, Daniel E. Singer, MD^{1,2}, Shawn N. Murphy, MD, PhD⁴, Joyce Lii, MS, MA¹, and Sebastian Schneeweiss, MD, ScD^{1,3}

¹Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School

²Department of Medicine, Massachusetts General Hospital, Harvard Medical School

³Department of Epidemiology, Harvard School of Public Health

⁴Department of Neurology, Massachusetts General Hospital, Harvard Medical School

Abstract

BACKGROUND—It is unclear how out-of-system care or electronic health record (EHR) discontinuity (i.e., receiving care outside of an EHR system) may affect validity of comparative effectiveness research using these data. We aimed to compare the misclassification of key variables in patients with high vs. low EHR-continuity.

METHODS—The study cohort comprised patients aged ≥ 65 identified in electronic health records from two US provider networks linked with Medicare insurance claims data from 2007–2014. By comparing electronic health records and claims data, we quantified EHR-continuity by the proportion of encounters captured by the EHRs (i.e., “capture proportion”). Within levels of EHR-continuity, for 40 key variables, we quantified misclassification by mean standardized differences between coding based on EHRs alone vs. linked claims and EHR data.

RESULTS—Based on 183,739 patients, we found mean capture proportion in a single electronic health record system was 16%–27% across two provider networks. Patients with highest level of EHR-continuity (capture proportion $\geq 80\%$) had 11.4- to 17.4-fold less variable misclassification, when compared to those with lowest level of EHR-continuity (capture proportion $< 10\%$). Capturing at least 60% of the encounters in an EHR system was required to have reasonable variable classification (mean standardized difference < 0.1). We found modest differences in comorbidity profiles between patients with high and low EHR-continuity.

Corresponding author: 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, Phone: (617) 278-0930; Fax: (617) 232-8602, jklyn@mgh.harvard.edu.

Conflict of interest: None declared

Data for replication: Data used to generate the study findings are not available due to restriction of the data use agreement with Centers for Medicare & Medicaid Services (CMS). The detailed definition of all variables in the study can be found in the supplemental materials (eAppendix 1).

CONCLUSIONS—EHR-discontinuity may lead to substantial misclassification in key variables. Restricting comparative effectiveness research to patients with high EHR-continuity may confer a favorable benefit (reducing information bias) to risk (losing generalizability) ratio.

Keywords

Electronic medical records; claims database; data linkage; comparative effectiveness research; data leakage; information bias

Introduction

Research on comparative effectiveness and drug safety with sufficient statistical power is often needed in a timely fashion as new medications are marketed with limited information about their effectiveness in routine care.^{1,2} There has been a remarkable growth in drug effectiveness research using electronic health record (EHR) databases as the primary data source in the last decade and there are currently more than 50 EHR-based research networks in the US. EHRs contain rich clinical information essential for many drug effectiveness studies (e.g., smoking status, body mass index, blood pressure levels, laboratory test results, etc.) that are not available in other administrative databases, such as insurance claims data. It is thus critical to understand how we can best conduct valid comparative effectiveness research using EHRs.

Most EHR systems in the US, with the exception of highly integrated plans, do not comprehensively capture medical encounters across all care settings (e.g. ambulatory office, emergency room, hospitalizations, etc.) and may miss substantial amounts of information that characterize the health state of its patient population. Medical information recorded at a facility outside of a given EHR system is “invisible” to the investigators and therefore often assumed to be absent in the study. Without linkage to an additional data source, however, researchers typically could not assess completeness of the records captured by the EHR system and their study findings are subject to likely misclassification of the exposure, outcome, and covariates used for confounding adjustment³. In contrast, insurance claims data have defined enrollment (start and end) dates and recording of all covered healthcare encounters, although the level of clinical detail is less than in an EHR system⁴.

Given that EHRs and data networks have become increasingly available for effectiveness research combined with the fact that linking EHRs to claims data is rarely done because of compliance and privacy concerns (e.g., sensitive identifiers are often required for reliable linkage), it is critical to understand the magnitude of out-of-system care or EHR-discontinuity, defined as “*receiving care outside of an EHR system*”, and its impact on data completeness in EHR databases. It is also important to provide a generalizable framework on how to best assess and remedy the potential biases due to data incompleteness and EHR-discontinuity.

We sought to evaluate the completeness of comparative effectiveness research-relevant information in an EHR system by comparing records of encounters in an EHR with claims data. Within levels of EHR data completeness, for a list of key variables in comparative effectiveness research, we quantified information bias due to EHR-discontinuity by

comparing classification based on EHRs alone with that based on EHRs linked with claims data. We wanted to identify a reasonable level of EHR-continuity for comparative effectiveness research and evaluated whether high EHR-continuity patients would have representative comorbidity profiles when compared to the remaining population in the EHRs.

Methods

Data sets

We linked longitudinal claims data from Medicare to electronic health records data for two medical care networks deterministically by insurance policy number, date of birth, and sex. The first network (EHR system 1) consists of one tertiary hospital, two community hospitals, and 17 primary care centers. The second network (EHR system 2) includes one tertiary hospital, one community hospital, and 16 primary care centers. The EHR database contains information on patient demographics, diagnosis/procedure codes, medications, lifestyle factors, laboratory data, and various clinical notes and reports. The Medicare claims data contain information on demographics, enrollment start and end dates, diagnosis/procedure codes, and dispensed medications.⁴

Study population

Among patients aged 65 and older with at least 180 days of continuous enrollment in Medicare (including inpatient, outpatient, and prescription coverage) from 2007/1/1 to 2014/12/31, we identified those with at least one encounter recorded in the study EHR system during their active Medicare enrollment period. The date when these criteria were met was assigned as the index (cohort entry) date after which we started the evaluation of their EHR completeness and classification of key variables. Those with private commercial insurance and Medicare as secondary payor were excluded to ensure we have comprehensive claims data for the study population.

Study design

Whether an EHR system holds adequate data for a particular individual (so called “EHR continuity status”) may change over time because patients may seek medical care in different provider systems over time. Therefore, we allowed the EHR continuity status to change every 365 days (Figure 1). The assumption was that most active patients aged 65 and older would present for a regular follow-up with records in the claims data at least annually. A short assessment period may lead to unstable estimates of the capture rates and a long period would make the continuity status less time-flexible. We followed patients until the earliest of the following: 1) loss of Medicare coverage; 2) death; 3) 2014/12/31, the end of the study period.

Measurement of EHR-continuity in an EHR system

To assess EHR-continuity, we calculated mean proportions of encounters captured by the EHRs (mean capture proportion) for each person:

$$\text{mean capture proportion} = \left(\frac{\text{Outpatient encounters recorded in EHR}}{\text{Outpatient encounters recorded in claims data}} + \frac{\text{Inpatient encounters recorded in EHR}}{\text{Inpatient encounters recorded in claims data}} \right) / 2$$

Patients generally have substantially more outpatient than inpatient visits. This formula purposefully gives higher weight to inpatient than to outpatient visits. (This is consistent with usual data considerations in comparative effectiveness research where recording of inpatient diagnosis is considered more complete and accurate than in outpatient settings.⁵⁻⁷) A match was said to exist if the type of encounter (inpatient vs. outpatient), admission and discharge dates (the same date for outpatient encounters) in the EHRs and claims databases match for the same patient. The incomplete terminal year during follow-up (with length less than 365 days) was not used to calculate mean capture proportion to avoid unstable estimates. We stratified the study population by bins of 10-point mean capture proportion reaching from 0% to 100% with the intent to stratify subsequent analysis by levels of mean capture proportion (because patients with mean capture proportion >80% comprised < 5% of the study population, those with mean capture proportion 90%–100% and 80%–90% were merged).

Quantifying misclassification of key variables

Our key variables for misclassification evaluation were, first of all, a combined comorbidity score (based on 20 comorbidity variables), which outperformed two widely used comorbidity scores, Charlson comorbidity index and Elixhauser system, in predicting 1-year mortality.⁸ The combined comorbidity score ranges between -2 to 26 with a higher score associated with higher mortality. The second group were 40 selected variables commonly used as drug exposure (n=15), outcome (n=10), or confounders (n=15) in comparative effectiveness research (see list of variables in Table 1 and definitions in eAppendix 1). The 10 outcome variables were based on previously validated algorithms.^{6,9,10,11,12,13,14,7} For each year following the index date, we assessed mean capture proportion and the classification of all the listed variables during the same 365-day period.

Metrics of misclassifications

For each individual, we calculated combined comorbidity scores based on (1) EHRs alone and (2) linked claims-EHR data. The difference between these two terms represents how much a patient's combined comorbidity score would be underestimated if relying on only EHRs than with access to both claims and EHRs. We then computed group mean of this difference for patients with different levels of mean capture proportion. In its derivation, each 1 point increase (decrease) in combined comorbidity score corresponded to 35% increase (decrease) in the odds of dying in one year.⁸

Next, we quantified misclassification of the 40 selected variables by two methods: (a) Sensitivity of positive coding in EHRs when compared to coding in the linked claims-EHRs data:

$$\text{Sensitivity}_{\text{variable}(i)} = \frac{\# \text{ of patients with variable}(i) = 1 \text{ based on EHR alone}}{\# \text{ of patients with variable}(i) = 1 \text{ based on the linked claims - EHR data}} \quad i = 1 - 40$$

Because we used all available data as the gold standard, for a given variable, those classified as negative by the gold standard (having no codes of interest in either EHRs or claims data) would be certainly coded as negative by EHRs alone (i.e., specificity should be 100% by design for all variables) but if the study EHR system did not capture medical information recorded in other systems, the sensitivity could be low; (b) Standardized difference comparing the classification based on only EHRs vs. that based on the linked claims-EHR data: Standardized difference is a measure of distance between two group means standardized by their standard deviations. This metric is often used to assess balance of covariates for exposure groups under comparison.¹⁵ Small mean standardized differences between proportions based on EHRs alone vs linked claims-EHR data would indicate that the EHR system has sufficient data that lead to similar classification compared to the gold standard. Within levels of mean capture proportion, we computed mean sensitivity and standardized difference over the 40 variables. We used a formula derived by Becker¹⁶ to construct 95% confidence intervals (CIs) of the calculated standardized differences, accounting for the correlation between the repeated variable classifications in the same population. It was suggested that standardized differences of less than 0.1 indicate satisfactory balance of covariates in the context of achieving adequate confounding adjustment.¹⁷ Because reducing misclassification of confounders is one of the major pathways to improve study validity, this cut-off is relevant for our study.

Evaluation of the representativeness of the cohort with high EHR-continuity

The minimum mean capture proportion needed to achieve acceptable classification (standardized differences < 0.1) of the key variables was used as the cut-off and patients with mean capture proportion greater than this cut-point were defined as the “EHR continuity cohort”. We compared the distribution of combined comorbidity score in those within vs. outside of EHR continuity cohort to understand if the comorbidity profiles differ in those with high vs. low EHR-continuity. We used claims data for the representativeness assessment, assuming similar completeness in patients with different levels of EHR-continuity.

Sensitivity analyses

(1) We presented the results for the first year following cohort entry and evaluated if a similar pattern can be observed in the subsequent years. (2) Some encounters tend to be routine and repetitive with the similar purposes, such as physical therapy, rehabilitation, psychotherapy/psychiatric visits, or ophthalmology visits. In a quantitative bias analysis, we excluded these types of encounters and repeat the analyses. (3) We evaluated if our results were sensitive to the length of EHR continuity assessment period. We compared results when assessing EHR continuity status every 180, 545, and 730 days instead of every 365 days. The statistical analyses were conducted with SAS 9.4 (SAS Institute Inc., Cary, NC).

Results

Measurement of EHR-continuity

There were a total of 183,739 patients in our study cohort (104,403 in EHR system 1 and 79,336 in EHR system 2). The mean follow-up time for the study cohort was 3.1 years based on the length of Medicare enrollment. Figure 2 shows number and proportion of patients by different levels of EHR-continuity after the initial records. The mean capture proportion was 27% in system 1 and 22% in system 2 for the first year, which remained consistently low (ranges from 16% to 26%) in all subsequent years across two EHR systems.

Quantifying misclassification by combined comorbidity score

Based on data in the first year, Figure 3 demonstrates a trend of decreasing discrepancy associated with increasing mean capture proportion. In EHR system 1, the mean difference in combined comorbidity score in those with mean capture proportion less than 10 % (1.95, 95% CI: 1.92–1.98) was 17.0 fold greater than that for mean capture proportion 80% or higher (0.11, 95% CI: 0.10–0.13). Similar findings were found in system 2.

Quantifying misclassification by 40 selected variables

(a) Sensitivity—Based on data in the first year, Figure 4 shows a clear trend of increasing mean sensitivity of EHRs capturing the codes for 40 selected variables when compared to the gold standard (Sensitivity_{40_variables}) associated with increasing mean capture proportion. In system 1, the Sensitivity_{40_variables} in those with mean capture proportion 80% (0.91, 95% CI: 0.89–0.93) was six-fold greater than that for mean capture proportion <10 % (0.15, 95% CI: 0.13–0.18). Similar findings were found in system 2. A similar trend was observed when the analysis was done for 25 co-morbidity and 15 medication use variables separately.

(b) Standardized difference—Based on data in the first year, Figure 5 demonstrates a trend of decreasing mean standardized differences between the proportions of 40 selected variables based on EHRs alone vs. the linked claims-EHR data (mean standardized difference_{40_variables}) associated with increasing mean capture proportion. In system 1, the mean standardized difference_{40_variables} in those with mean capture proportion <10 % (0.53, 95% CI: 0.51–0.54) was 11.4-fold greater than that for mean capture proportion 80% (0.05, 95% CI: 0.04–0.06). Similar findings were found in system 2. For the same level of EHR-continuity, mean standardized difference for medication use variables tended to be larger than that for comorbidity variables (diagnosis-code based variables, see Table 1). For example, among patients with mean capture proportion 80% in EHR system 1, mean standardized difference was 0.09 (95% CI, 0.08–0.10) for 15 medication use variables and 0.02 (95% CI, 0.01–0.03) for 25 co-morbidity variables. A similar pattern was observed for other EHR-continuity levels and in EHR system 2. Using mean standardized difference_{40_variables} of 0.1 as the cut-off¹⁷, a mean capture proportion of 60% was associated with acceptable classification of 40 selected variables in both system 1 and 2. Accordingly, we defined those with mean capture proportion 60% as the “EHR continuity cohort”.

EHR continuity cohort

(1) Variable classification—In the EHR continuity cohort (mean capture proportion 60%), our EHR systems had sensitivity greater than 0.7 for all 40 selected variables and sensitivity greater than 0.8 for 34 out of 40 variables (85%). In contrast, in non-EHR continuity cohort (mean capture proportion <60%), our EHR systems had sensitivity less than 0.5 for 37 out of 40 variables (93%). The mean standardized difference_{40_variables} was 6 fold greater in non-EHR continuity cohort (mean standardized difference_{40_variables} =0.36) than that in the EHR continuity cohort (mean standardized difference_{40_variables} =0.06, eTable 1).

(2) Representativeness—We found small to modest differences in the distribution of combined comorbidity score in those with mean capture proportion 60% vs. <60%. The mean standardized difference between proportions of all combined comorbidity score categories across the two populations was 0.04 (Table 2).

Sensitivity analyses

(1) We observed a similar pattern for both systems in all the years after cohort entry (Table 3). (2) After excluding encounters with higher repetitive tendency, we observed a very similar pattern in variable misclassification in all years (eTable 3). (3) When assessing EHR continuity status every 180, 545, and 730 days rather than every 365 days, the resulting mean capture proportion was highly correlated with the mean capture proportion generated by the primary analysis (Spearman coefficient =0.90, 0.94, and 0.93, respectively).

Discussion

We found that the mean proportion of encounters captured by a single EHR system was consistently lower than 30% in two metropolitan EHR systems and across years after the initial records. There was a clear trend of increasing misclassification of variables based on only EHRs when the completeness of EHRs decreases (up to 17-fold greater misclassification in those with record capture rate <10% vs. 80%). Calculating a comorbidity score to quantify the information bias showed that one may underestimate one-year odds of death by 68%-76% in the lowest level of EHR-continuity but only 4%–5% in the highest level of EHR-continuity in EHRs. Capturing at least 60% of the encounters in an EHR system was required to have a mean standardized difference of <0.1 when compared with the gold standard, one possible reference point to indicate acceptable classification. We found small to modest differences in comorbidity profiles between EHRs continuity and non-continuity cohorts, which was in a much smaller magnitude (about one-ninth) than the amount of variable misclassification among non-continuity patients when quantified by the same metric (mean standardized difference).

Our findings are useful for comparative effectiveness research in the following ways: (a) Researchers could use our findings to quantify the potential magnitude of bias due to EHR-discontinuity in their study even in the setting where linking EHRs with additional data is not possible (e.g. due to lack of necessary identifiers for data linkage). They can check whether this bias can explain away their findings through a simple quantitative bias analysis.

¹⁸ This will help researchers achieve more valid qualitative conclusions based on EHRs alone; (b) Our findings suggest that restricting to those with higher record capture rates by EHRs could substantially reduce misclassification resulting from EHR-discontinuity. Although assessing EHR data completeness for the entire cohort may not always be feasible, we have demonstrated how investigators could obtain a sample of data within which information on the completeness of their records could be ascertained through linkage to additional data sources like an insurance claims database. In this linked sample, they can quantify information bias by the metrics proposed in our study and use that data for quantitative bias analysis¹⁸ or external adjustment (e.g. propensity score calibration)¹⁹.

There are several limitations to our work. First, the ultimate impact of information bias on study estimates is research question and context specific. Therefore, future investigations on how EHR-discontinuity influences relative risks in a wide range of research questions are needed. Second, we used mean standardized differences less than 0.1 as a reference point to better understand the magnitude of misclassification for selected variables. This cut-off was suggested in the context of assessing covariate balance to achieve adequate confounding adjustment. As tolerance of misclassification may differ by research contexts, the minimum capture rate needed may also depend on research purposes. We additionally provided data on sensitivity of EHRs capturing relevant codes or medication records when compared to the gold standard as a more generic reference tool. Next, we have only EHR data from two metropolitan provider networks, and it is not clear if these findings are applicable to other EHR systems. While the two EHR networks had different mean EHR capture rates, the amount of misclassification of key variables turned out to be very similar within the same levels of the EHR completeness in the two EHR systems. These findings suggest that the proportion of those with high EHR-continuity may vary by EHR systems, but once restricted to the EHR continuity cohort, the misclassification would be reduced in a similar fashion. Moreover, our study cohort consisted of only those aged 65 and older. The older adults are the most critical population to investigate the impact of EHR-continuity on study validity using EHRs because they often need more complex care which may not be fulfilled in one system due to resource limitations. Nonetheless, it is important to note that our findings may not be applicable to the younger populations as their medical seeking behaviors and general health state are quite different. Lastly, limiting to patients with high EHR-continuity will inevitably reduce study sizes and statistical power.

In summary, our findings support the strategy to restrict comparative effectiveness research to patients with high EHR-continuity, as the risk of losing generalizability is relatively small in comparison to the benefit of substantial misclassification reduction. These results are relevant for the majority of healthcare systems in the US that are not integrated with a payor/insurer and where EHR-discontinuity in the EHR system is likely.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: This project was supported by NIH Grant 1R01LM012594-01. Kueiyu Joshua Lin received a stipend from the Pharmacoepidemiology program in the Department of Epidemiology, Harvard School of Public Health and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School. Dr. Singer was supported by the Eliot B. and Edith C. Shoolman Fund of Massachusetts General Hospital. Drs. Schneeweiss and Murphy were supported by PCORI grant 282364.5077585.0007 (ARCH/SCILHS).

References

1. Strom, BL., Kimmel, SE., Hennessy, S. Textbook of pharmacoepidemiology. Wiley Blackwell; Chichester, West Sussex England; Hoboken, NJ: 2013.
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005; 58:323–337. [PubMed: 15862718]
3. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform.* 2013; 46:830–836. [PubMed: 23820016]
4. Hennessy S. Use of health care databases in pharmacoepidemiology. *Basic & clinical pharmacology & toxicology.* 2006; 98:311–313. [PubMed: 16611207]
5. Fang MC, et al. Validity of Using Inpatient and Outpatient Administrative Codes to Identify Acute Venous Thromboembolism: The CVRN VTE Study. *Medical care.* 2016
6. Wahl PM, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiology and drug safety.* 2010; 19:596–603. [PubMed: 20140892]
7. Myers RP, Leung Y, Shaheen AA, Li B. Validation of ICD-9-CM/ICD-10 coding algorithms for the identification of patients with acetaminophen overdose and hepatotoxicity using administrative data. *BMC Health Serv Res.* 2007; 7:159. [PubMed: 17910762]
8. 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; 64:749–759. [PubMed: 21208778]
9. Birman-Deych E, et al. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Medical care.* 2005; 43:480–485. [PubMed: 15838413]
10. Andrade SE, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiology and drug safety.* 2012; 21(Suppl 1):100–128. [PubMed: 22262598]
11. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiology and drug safety.* 2012; 21(Suppl 1):154–162. [PubMed: 22262602]
12. Waikar SS, et al. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification Codes for Acute Renal Failure. *Journal of the American Society of Nephrology : JASN.* 2006; 17:1688–1694. [PubMed: 16641149]
13. Cushman M, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med.* 2004; 117:19–25. [PubMed: 15210384]
14. Cunningham A, et al. An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiology and drug safety.* 2011; 20:560–566. [PubMed: 21387461]
15. Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Statistics in medicine.* 2014; 33:1685–1699. [PubMed: 24323618]
16. Becker BJ. Synthesizing standardized mean change measures. *British Journal of Mathematical and Statistical Psychology.* 1988; 41:257–278.
17. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine.* 2009; 28:3083–3107. [PubMed: 19757444]

18. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and drug safety*. 2006; 15:291–303. [PubMed: 16447304]
19. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005; 162:279–289. [PubMed: 15987725]

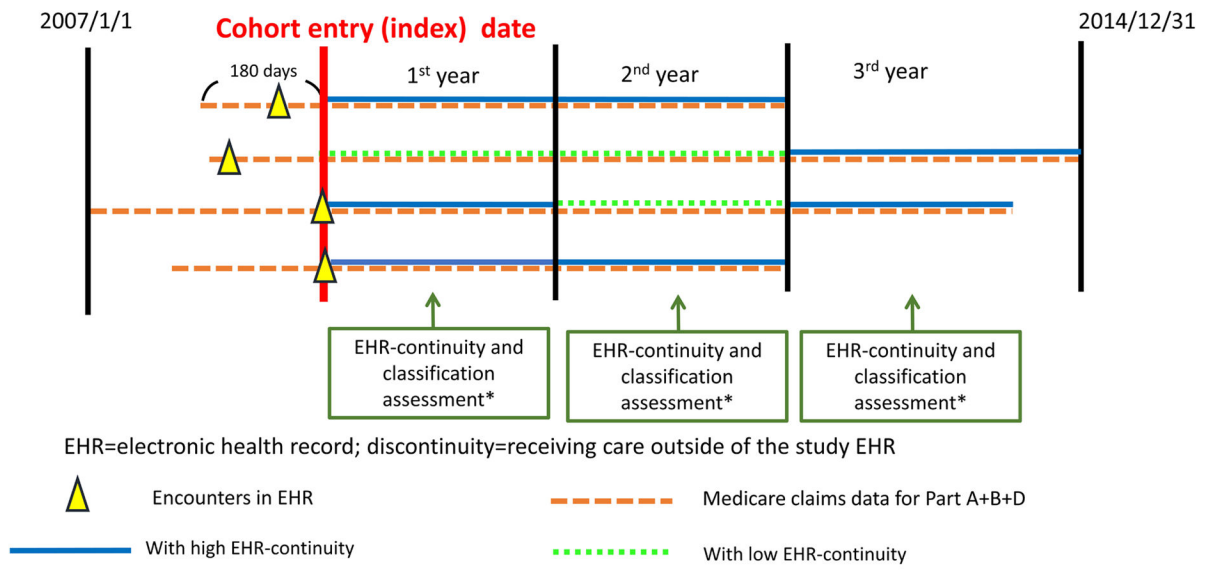


Figure 1. Assessment of Electronic Health Record continuity and proxy indicators

*Assess the completeness of the records captured by the EHR and classification of key variables during the same period

Author Manuscript

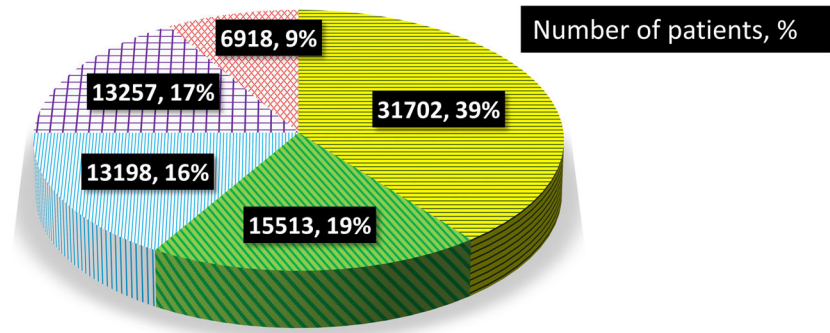
Author Manuscript

Author Manuscript

Author Manuscript

A. Patients with different capture proportions (CP) in EHR system 1*

■ CP=0 to <0.1 ■ CP=0.1 to <0.25 ▨ CP=0.25 to <0.5 ▩ CP=0.5 to <0.75 ✖ CP>=0.75



B. Mean capture proportions by EHR system and year after cohort entry

Year after cohort entry	1	2	3	4	5	6	7
Mean capture proportion in EHR system 1	0.27	0.22	0.22	0.22	0.23	0.24	0.26
Mean capture proportion in EHR system 2	0.22	0.16	0.16	0.16	0.16	0.18	0.20

Figure 2.

Proportion of encounters captured by single electronic healthrecord system

EHR= electronic health record, CP=capture proportion, * Proportions were based on data in the EHRsystem 1 in the first year, but the pattern was similar in the subsequent years and in EHR system 2.

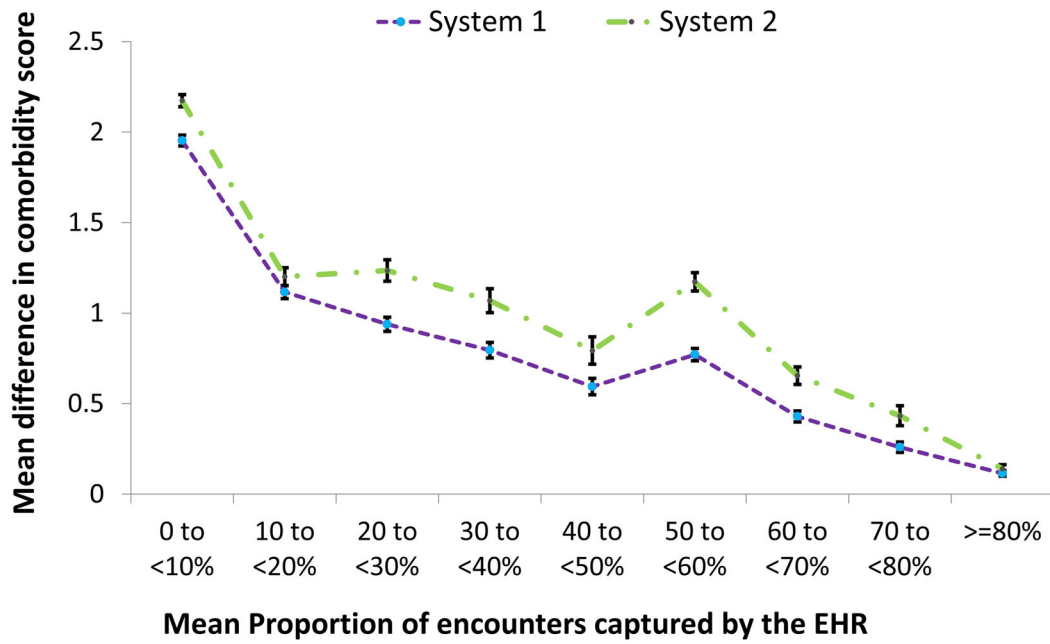


Figure 3. Underestimation of combined comorbidity score based on EHR alone in relation to EHR-continuity
 EHR=electronic health record; Combined comorbidity score⁸ ranges between -2 to 26 with a higher score associated with higher mortality. The error bars represent the 95% confidence intervals around the point estimates of the means difference in comorbidity scores.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

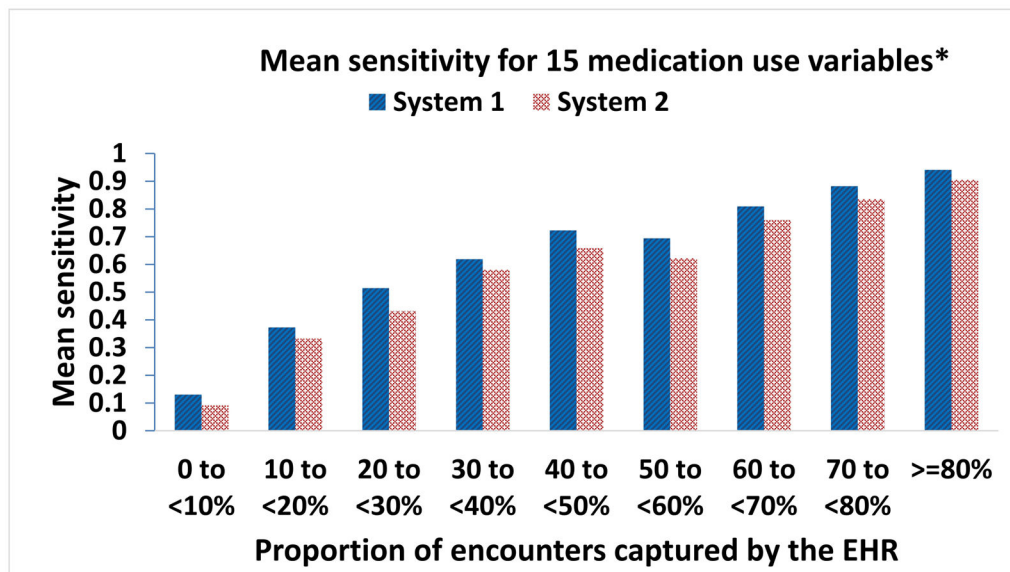
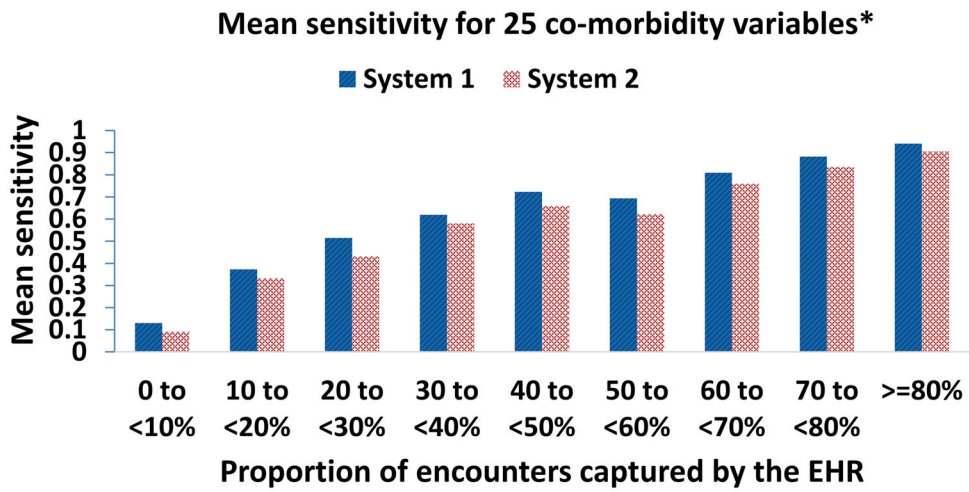


Figure 4. Sensitivity of capturing medical information in relation to EHR-continuity
 EHR=electronic health record *See list of variables in Table 1

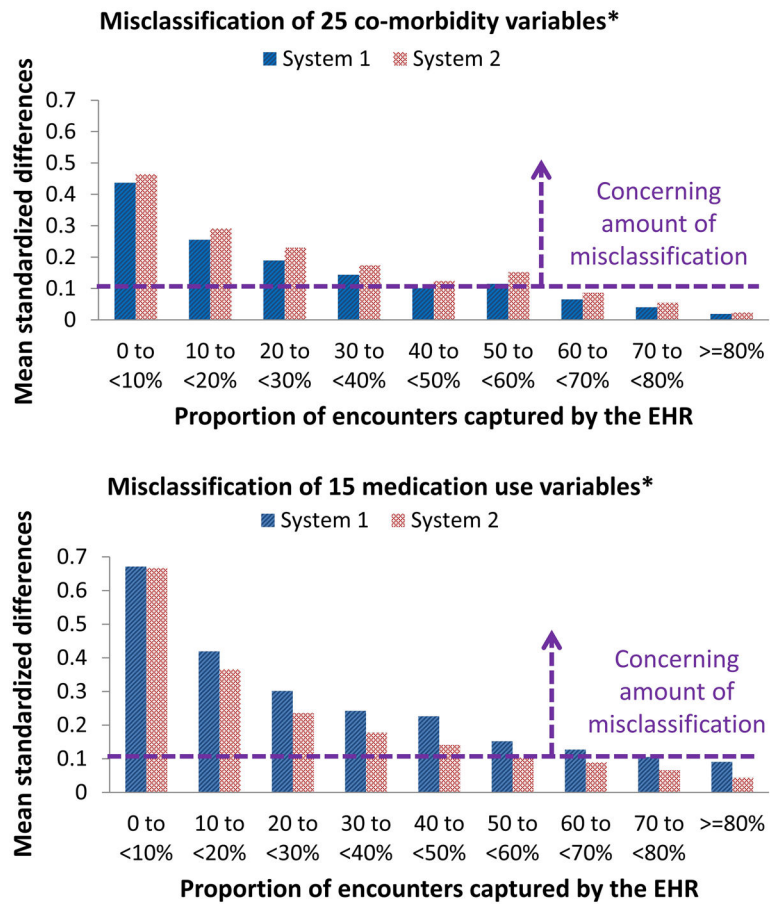


Figure 5. Misclassification in relation to EHR-continuity
 EHR=electronic health record. *See list of variables in Table 1

Table 1

Patient characteristics to be assessed for accuracy of classification

25 co-morbidity variables	<p>a. 15 variables commonly used as covariates for confounding adjustment: Dementia, atrial fibrillation, chronic lung disease, chronic liver disease, chronic kidney disease, cancer, diabetes, hypertension, anemia, psychosis, depression, pneumonia, HIV, fracture, and RA</p> <p>b. 10 variables with validated algorithm commonly used as outcome variables: ischemic stroke, ICH, MI, AKI, hepatotoxicity, GI Bleeding, major bleeding, DVT, PE, and CHF ^{6,7,9-14}</p>
15 medication use variables	antiplatelet agents, antidiabetics, antihypertensives, NSAIDs, opioids, antidepressants, antipsychotics, anticonvulsants, PPIs, antiarrhythmics, statins, dementia, hormone therapy, antibiotics, and oral anticoagulants

CHF= congestive heart failure, HIV= human immunodeficiency virus, RA=rheumatoid arthritis, AKI=acute kidney injury, ICH= intracranial hemorrhage, MI= myocardial infarction, PE= pulmonary embolism, DVT= deep vein thrombosis, NSAIDs= nonsteroidal anti-inflammatory drugs, PPI= Proton pump inhibitors

Table 2

Representativeness: comparison of combined comorbidity score in those with high vs. low EHR continuity

Comorbidity score categories	Low EHR-continuity N (%) ^a	High EHR-continuity N (%) ^b	Stand. Diff. ^c
-1	15453 (13)	3774 (18)	0.15
0	28637 (24)	5604 (27)	0.08
1	21677 (18)	3500 (17)	0.03
2	13971 (12)	2142 (11)	0.04
3	9546 (8)	1382 (7)	0.05
4	7017 (6)	964 (5)	0.05
5	5555 (5)	769 (4)	0.04
6	4918 (4)	657 (3)	0.05
7	3846 (3)	543 (3)	0.03
8	2967 (3)	407 (2)	0.03
9	2128 (2)	274 (1)	0.04
10	1457 (1)	186 (1)	0.03
11	1001 (1)	119 (1)	0.03
12	647 (1)	73 (0.4)	0.03
13	410 (0.3)	46 (0.2)	0.02
14	243 (0.2)	28 (0.1)	0.02
15	197 (0.2)	12 (0.1)	0.03
16	88 (0.1)	5 (0.02)	0.02
Total N/mean stand. diff.	119851	20497	0.04

^aThe proportion of encounters captured by the electronic health record system <60%.

^bThe proportion of encounters captured by the electronic health record system ≥60%.

^cStand diff= Standardized difference. Combined comorbidity score⁸ ranges between -2 to 26 with a higher score associated with higher mortality; cell size <5 were not presented here.

Table 3
 Mean standardized difference of 40 selected variables, comparing electronic health records with claims-electronic health records in all years after cohort entry.^a

	Year after cohort entry						
	1	2	3	4	5	6	7
0 to <10%	0.53	0.55	0.56	0.56	0.56	0.55	0.54
10 to <20%	0.32	0.29	0.28	0.28	0.28	0.25	0.23
20 to <30%	0.23	0.21	0.21	0.21	0.19	0.18	0.16
30 to <40%	0.18	0.16	0.16	0.15	0.15	0.13	0.13
40 to <50%	0.14	0.13	0.13	0.13	0.11	0.10	0.09
50 to <60%	0.13	0.11	0.11	0.11	0.10	0.08	0.08
60 to <70%	0.09	0.08	0.07	0.07	0.07	0.06	0.06
70 to <80%	0.06	0.06	0.06	0.06	0.06	0.04	0.04
>=80%	0.04	0.04	0.04	0.04	0.04	0.03	0.03

Proportion of encounters captured by the electronic health records

^a A mean standardized difference <0.1 indicates acceptable discrepancy¹⁷; estimates < 0.1 were marked in boldface.