

Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients

Erik Youngson, MMath^{a,b}, Robert C. Welsh, MD^{a,c,*}, Padma Kaul, PhD^a, Finlay McAlister, MD^a, Hude Quan, PhD^d, Jeffrey Bakal, PhD^{a,b}

Abstract

Administrative health databases are used in research to define comorbid conditions, diagnosis, and procedures. Our objectives were to validate a diagnosis of ST-elevation myocardial infarction (STEMI) and invasive cardiac procedure coding against a comprehensive registry of STEMI patients and determine an optimal algorithm for defining comorbidities using administrative hospitalization and ambulatory databases, but without using a physician claims database, which is unavailable for use in many jurisdictions.

A registry of consecutive STEMI patients was used to define a reference cohort and linked to the hospitalization and ambulatory databases. Four administrative case definitions for defining comorbidities, as well as STEMI diagnosis and in-hospital procedures using the International Classification of Diseases, 10th Revision (ICD-10) and the Canadian Classification of Health Interventions (CCI) were evaluated. Metrics were used to evaluate algorithm performance and compare discriminative ability using the C statistic.

The 3236 patients had median age of 60 years (interquartile range 52–71) and 75.7% were male. A diagnosis of STEMI was correctly identified in the administrative records for 3043 (94.0%) patients. In-hospital procedures (coronary artery bypass grafting, percutaneous coronary intervention, and angiogram) were well identified using administrative definitions (Kappa statistic 0.83–1.00). Validation of comorbidities varied by condition but an algorithm using 2 inpatient/ambulatory visits in the previous 2 years maximized PPV, ranging from 28.6% for previous heart failure to 95.7% for previous MI. The c statistic was similar for each of the methods, ranging from 0.76 to 0.80.

ICD-10 and CCI codes can identify hospitalized STEMI patients with high sensitivity and accurately define in-hospital cardiac procedures. Comorbidities can be defined with high PPV using a definition of 2 inpatient/ambulatory visits in the previous 2 years.

Abbreviations: CABG = coronary artery bypass grafting, CCI = Canadian Classification of Health Interventions, CI = confidence interval, DAD = Discharge Abstract Database, ECG = electrocardiogram, EPICORE = epidemiology coordinating and research, ICD-10 = International Classification of Diseases, 10th revision, ICD-10-CA = ICD-10, Canadian enhancement, ICD-9 = International Classification of Diseases, 9th revision, IQR = interquartile range, MI = myocardial infarction, NACRS = National Ambulatory Care Reporting System, NPV = negative predictive value, NSTEMI = non-ST elevation myocardial infarction, PCI = percutaneous coronary intervention, PPV = positive predictive value, STEMI = ST elevation myocardial infarction, VHR = vital heart response.

Keywords: comorbidities, ICD-10, procedures, STEMI, validation

1. Introduction

In contrast to many chronic diseases, ST-elevation myocardial infarction (STEMI) is acute and often has outcomes related to comorbidities and in-hospital procedures at the index

hospitalization. Because administrative health databases are increasingly being used for epidemiological and population health research studies, it is important to be able to identify comorbidities and procedures appropriately for the purpose of creating patient comorbidity profiles and risk-adjusted regression models, and indices such as the Charlson Comorbidity Index.^[1] Administrative case definitions using the *International Classification of Diseases 9th (ICD-9)* and *10th (ICD-10)* revision have been well validated for many common conditions.^[2–6] For STEMI and non-STEMI myocardial infarction, the case definitions have only been described using ICD-9 codes, which have not been used for inpatient data since 2002 in Canada. A previous study compared ICD-10 codes with their ICD-9 counterparts for 9 common comorbidities in cardiovascular patients and found a high level of agreement between the 2 versions.^[7]

In addition to the validation of the codes themselves, a key area of study is to determine the optimal strategy for defining comorbid conditions. Defining comorbid conditions using only codes from an index hospitalization often results in under-reporting of conditions^[2,8,9] and therefore it has become common practice to utilize algorithms that search for the occurrence of the particular diagnosis codes in the previous 1 to 3 years, in inpatient and outpatient records, to increase the identification rate of these conditions. These algorithms typically include data

Editor: Davide Piraino.

The Vital Heart Response registry was funded by Alberta Health Services and Astra Zeneca Canada with support from Boehringer Ingelheim Canada.

None of the authors have conflicts of interest related to this manuscript.

Supplemental Digital Content is available for this article.

^a University of Alberta, Edmonton, ^b Alberta Health Services, ^c Mazankowski Alberta Heart Institute, ^d University of Calgary, Calgary, Alberta, Canada.

* Correspondence: Robert C. Welsh, 2C2 Cardiology, WMC, 8440 112 St NW, Edmonton, Alberta T6G 2B7, Canada (e-mail: Robert.Welsh@ahs.ca).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

Medicine (2016) 95:32(e4554)

Received: 31 March 2016 / Received in final form: 20 June 2016 / Accepted: 18 July 2016

<http://dx.doi.org/10.1097/MD.0000000000004554>

from hospitalizations, ambulatory encounters, and/or physician billing claims,^[10] which enable capture of diseases that may be managed primarily in an outpatient setting. Owing to privacy regulations, the use of physician billing claims databases is often prohibited for research purposes; additionally, their use may generate false-positive records owing to rule-out diagnoses performed in a primary care office (e.g., if a patient was seen or tests were done to rule out a condition). Therefore, establishing algorithms to accurately identify key comorbid conditions in the absence of physician claims data (i.e., based entirely on inpatient and ambulatory databases) is of high value.

The 3 objectives of this study are to: validate the *ICD-10* STEMI diagnosis in the Administrative Data holdings; validate the procedure codes associated with reperfusion namely coronary artery bypass grafting (CABG), percutaneous coronary intervention (PCI), and angiogram; determine the optimal algorithm for 7 common comorbidities using only hospitalization and outpatient ambulatory databases.

2. Methods

2.1. VHR registry

The Vital Heart Response (VHR) registry captured consecutive STEMI hospitalizations at 5 hospitals in Edmonton, Alberta, Canada, between October 2006 and March 2011. Patients were considered to have a STEMI and thus captured in VHR if they had positive Troponin and elevated ST waves based on an adjudication of the electrocardiogram (ECG) by experienced clinician investigators. Clinical information and medical histories, including age, sex, mode of arrival, reperfusion therapy, in-hospital procedures and events, and medications, were acquired by independent review of the patient's chart by the Epidemiology Coordinating and Research (EPICORE) Centre—a dedicated research organization located at the University of Alberta. Seven comorbidities (previous angina, previous MI, hypercholesterolemia, hypertension, diabetes, previous heart failure, atrial fibrillation) were documented in VHR only if the condition was known before admission. Data collection for atrial fibrillation and previous heart failure were added in April 2010 and thus were only completed for approximately one-quarter of the patients. Angiogram was recorded when it was done without any additional intervention, and thus we considered patients who had a PCI to also have had an angiogram. PCI (including primary, rescue, urgent, or elective) and CABG were recorded separately in the VHR database. VHR has been utilized previously for research purposes^[11–13] and has undergone detailed quality review and validation, thus making it an ideal candidate as a reference standard for validating administrative definitions.

2.2. Administrative databases

The Discharge Abstract Database (DAD) contains 1 record for each acute care hospitalization in the province including dates of admission and discharge, a most responsible diagnosis and up to 24 other diagnoses, and up to 20 in-hospital interventions/procedures. The National Ambulatory Care Reporting System (NACRS) contains all patient visits to emergency departments and in-hospital clinics in Alberta and includes date of visit, up to 10 diagnoses, and up to 10 interventions/procedures—of particular relevance to this study, this dataset does not capture

visits to a primary care physician or specialist offices outside of hospital settings. The DAD and NACRS databases used the Canadian enhancement of the International Statistical Classification of Diseases and Health Related Interventions, 10th revision (*ICD-10-CA*) and the Canadian Classification of Health Interventions (CCI) for defining diagnoses and procedures, respectively, during the study time period and are coded by trained nosologists within each hospital according to Canadian Institute for Health Information and Provincial guidelines.

2.3. Administrative definitions

A diagnosis of STEMI was defined by searching each patient's index hospitalization for the presence of *ICD-10* codes I21.0*, I21.1*, I21.2*, or I21.3* in the most responsible diagnosis field. To define CABG, PCI, and angiogram procedures, the index hospitalization records and outpatient records that occurred during, or 1 day before, the index admission were searched for the presence of relevant CCI codes (CABG: 1.IJ.76*; PCI: 1.IJ.50*; angiogram: 3.IP.10*) in any of the intervention fields.

The 7 comorbidities were defined based on the presence of *ICD-10* codes in any diagnosis field: previous angina: I20.*, I25.11, I25.7; previous MI: I21.*, I22.*, I25.2; hypercholesterolemia: E78.*; hypertension: I10.*, I11.*, I12.*, I13.*, I15.*; diabetes: E10.*, E11.*, E12.*, E13.*, E14.*; previous heart failure: I50.*; and atrial fibrillation: I48.*. We considered 4 case definitions to define comorbidities using the administrative databases: the index hospitalization DAD record(s), the Index hospitalization or any DAD/NACRS records in the previous 1 year, the index hospitalization or any DAD/NACRS records in the previous 2 years, and at least 2 occurrences in any DAD/NACRS records in the previous 2 years (i.e., index plus at least 1 record in the previous 2 years, or at least 2 non-index occurrences in the previous 2 years).

2.4. Data linking

Patients in the VHR registry were linked to the 2 administrative databases using unique patient identifiers, and records that could not be linked (out of province residence or invalid patient identifiers recorded in VHR) were removed. For patients with >1 STEMI hospitalization in VHR, only the first instance was retained in the cohort. To define the index STEMI hospitalizations in the administrative data, the patients in VHR were linked to corresponding hospitalizations in the inpatient DAD, as under the VHR protocol most patients are admitted directly and skip the Emergency Department, and so do not have an index record in NACRS. Because VHR includes patients transferred from other hospitals, we accounted for transfers in the administrative data by combining concurrent hospitalization records, into an episode of care. Therefore, the index hospitalization defined by the administrative DAD data sometimes contained >1 DAD record (but with dates overlapping with those recorded in VHR).

2.5. Statistical analysis

Baseline patient characteristics were summarized using frequencies and percentages for discrete variables and medians and interquartile ranges (IQR) for continuous variables. VHR was considered to be the reference standard and the performance of the administrative comorbidity and procedure definitions was

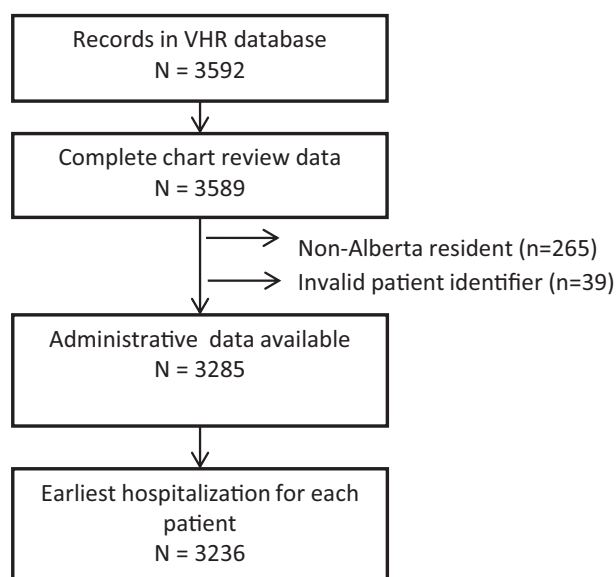


Figure 1. Study cohort derivation. The study cohort consisted of 3592 STEMI patients of which 3589 had completed registry data. This registry cohort was matched to administrative data successfully in 3285 with the final cohort being 3236 after repeat admissions were removed. VHR=Vital Heart Response STEMI registry.

evaluated by calculating the positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, and Cohen Kappa statistic, all with corresponding 95% Wald confidence intervals (CIs). As VHR contains exclusively STEMI patients, only sensitivity was calculated for evaluating the administrative STEMI definition. To assess and compare the discrimination of the 4 sets of administrative comorbidity case definitions, we modeled in-hospital mortality using logistic regression models that included the 7 comorbidities defined by each algorithm as well as age and sex, and reported the corresponding C statistics for each of the 4 models. Patients with missing values for comorbidities in the VHR registry were excluded from each comorbidity specific analysis but included in analyses for which their values were known. All analyses were conducted using SAS v9.4 (SAS Institute, Cary, NC). This study received approval from the University of Alberta Health Research Ethics Board.

3. Results

3.1. Cohort demographics

Between October 2006 and March 2011, there were 3592 hospitalizations for STEMI captured in the VHR database. After excluding patients with incomplete VHR chart review data (n=3), non-Alberta residents (n=265), patients with invalid patient identifiers (n=39), and repeat visits by the same patient (n=49), the resulting cohort consisted of 3236 unique patients (Fig. 1). The median age of the cohort was 60 years (IQR 52–71), 75.7% of patients were male, and the median length of stay for the index hospitalization was 4 days (IQR 3–7). 43.9% of patients arrived by ground ambulance. Nearly half (48.6%) were transferred from another hospital as only 2 of the hospitals have PCI capability. Primary PCI and fibrinolysis were used for similar proportions of patients (43.7% vs. 40.5%), whereas 15.9% of patients did not receive any reperfusion therapy (Table 1).

Table 1

Patient characteristics.

	STEMI Cohort
No. of patients	N=3236
Age, y	60 (52, 71)
Male	2450 (75.7)
Index hospitalization characteristics	
Arrived by ambulance	1419 (43.9)
Transferred	1572 (48.6)
Died in hospital	187 (5.8)
Length of stay, days	4 (3, 7)
Reperfusion	
Primary PCI	1414 (43.7)
Fibrinolytic	1309 (40.5)
None	513 (15.9)

Values are displayed as median (IQR) for continuous variables and n (%) for binary/categorical. PCI=percutaneous coronary intervention, STEMI=ST elevation myocardial infarction

3.2. Validity of STEMI diagnosis

Among the index hospitalization DAD records, a most responsible diagnosis of STEMI was correctly recorded for 3043 patients (94.0%). Of the 193 patients that were incorrectly coded in DAD (i.e., did not have a most responsible diagnosis of STEMI), the most frequent diagnoses were Non-ST elevation (NSTEMI) myocardial infarction (62.2%), AMI-unspecified (21.2%), and atherosclerotic heart disease of native coronary artery (7.8%).

3.3. Validity of cardiac procedures

Of the 3 in-hospital procedures examined, all yielded excellent results using the administrative CCI codes compared to chart review, with Kappas of 0.83 (PCI), 0.87 (angiogram), and 1.00 (CABG). Administrative coding for CABG correctly identified all but 1 procedure yielding sensitivity, specificity, PPV, and NPV between 99.2% and 100.0%, whereas PCI and angiogram coding resulted in sensitivity, specificity, PPV, and NPV between 79.9% and 99.1%, respectively (Table 2). All administrative procedure codes for CABG were identified in the DAD; however, the DAD data captured only 78.6% of PCI codes and 80.5% angiogram codes, with the rest coming from the NACRS database.

3.4. Validity of comorbidities

In evaluating our primary objective, we found that although the performance measures of the different algorithms for the identification of comorbid conditions varied considerably depending on condition under consideration, the use of 2 visits in either the DAD or the NACRS in the previous 2 years consistently yielded the highest PPV (Table 3) (appendix Figure 1, <http://links.lww.com/MD/B180>: Comparison of algorithms for defining comorbid conditions). For defining diabetes, all algorithms performed very well and although the algorithm that required 2 visits in either the DAD or the NACRS in the previous 2 years yielded the best results (PPV 90.1%, NPV 96.3%, Kappa 0.84), even the simplest algorithm using only the index hospitalization was nearly as good (PPV 79.9%, NPV 98.6%, Kappa 0.83). For hypertension, the 2 visits in either dataset in 2 years algorithm resulted in PPV of 86.3% (NPV 68.5%, Kappa 0.50), whereas the other algorithms resulted in higher Kappa

Table 2

Validity of administrative databases for in-hospital procedures.

Procedure	Chart prevalence %	Administrative prevalence %	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Kappa (95% CI)
CABG	3.8	3.8	100.0 (100.0, 100.0)	100.0 (99.9, 100.0)	99.2 (97.6, 100.0)	100.0 (100.0, 100.0)	1.00 (0.99, 1.00)
PCI	81.0	77.4	98.8 (98.4, 99.2)	79.9 (77.0, 82.8)	94.4 (93.5, 95.3)	95.1 (93.4, 96.8)	0.83 (0.81, 0.86)
Angiogram	93.2	93.2	99.1 (98.8, 99.5)	87.8 (83.5, 92.1)	99.1 (98.8, 99.4)	88.2 (83.9, 92.4)	0.87 (0.84, 0.91)

CABG=coronary artery bypass grafting, PCI=percutaneous coronary intervention, PPV=positive predictive value, NPV=negative predictive value.

values (range 0.62–0.65) but with lower PPV, ranging from 74.8% for 1 inpatient/outpatient visit in the previous year to 78.6% for the index hospitalization. In this population of patients with acute coronary syndrome, administrative definitions for other chronic conditions performed less well overall, with optimal Kappa values for each condition ranging from 0.26 to 0.46. However, with the exception of previous heart failure (PPV 28.6%, NPV 98.0%, Kappa 0.26), the 2 visits in either dataset in 2 years algorithm yielded PPV of >70% for all conditions and NPV of >80% for all but 1 condition (Table 3).

Discriminative ability for in-hospital death was similar for the 4 algorithms, with the first 3 algorithms yielding C statistics of 0.80, whereas the algorithm using 2 visits in either dataset in 2 years had a C statistic of 0.76.

4. Discussion

This study utilized an existing registry of STEMI patients in Alberta and was linked to administrative inpatient and ambulatory databases to provide a platform for validating

Table 3

Validity of algorithms for defining comorbid conditions.

Algorithm	Algorithm prevalence, %	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Kappa (95% CI)
Previous angina (valid n=2997; chart prevalence=19.7%)						
Index hospitalization DAD record(s)	2.3	83.8 (75.1, 92.6)	81.8 (80.4, 83.2)	9.7 (7.3, 12.1)	99.5 (99.3, 99.8)	0.14 (0.10, 0.17)
Index hosp. or 1 DAD/NACRS visit in previous year	9.2	57.8 (52.0, 63.7)	84.2 (82.8, 85.6)	27.0 (23.4, 30.6)	95.2 (94.3, 96.0)	0.28 (0.23, 0.32)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	10.1	60.9 (55.4, 66.3)	85.0 (83.6, 86.3)	31.4 (27.7, 35.2)	95.1 (94.2, 95.9)	0.32 (0.28, 0.37)
2 DAD/NACRS visits in previous 2 years	3.1	94.6 (89.9, 99.2)	82.7 (81.3, 84.1)	14.8 (11.9, 17.6)	99.8 (99.6, 100.0)	0.21 (0.17, 0.25)
Previous MI (valid n=3007; chart prevalence=15.5%)						
Index hospitalization DAD record (s)	9.2	88.4 (84.7, 92.2)	91.9 (90.8, 92.9)	52.5 (47.9, 57.0)	98.7 (98.3, 99.2)	0.61 (0.57, 0.66)
Index hosp. or 1 DAD/NACRS visit in previous year	13.4	83.9 (80.3, 87.5)	95.0 (94.2, 95.9)	72.4 (68.3, 76.4)	97.4 (96.8, 98.1)	0.74 (0.70, 0.77)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	13.9	83.5 (79.9, 87.1)	95.4 (94.6, 96.2)	74.7 (70.8, 78.7)	97.3 (96.7, 97.9)	0.75 (0.72, 0.79)
2 DAD/NACRS visits in previous 2 years	6.2	95.7 (92.8, 98.6)	89.8 (88.6, 90.9)	38.1 (33.7, 42.5)	99.7 (99.5, 99.9)	0.50 (0.45, 0.55)
Hypercholesterolemia (valid n=2919; chart prevalence=47.2%)						
Index hospitalization DAD record (s)	60.6	62.4 (60.1, 64.6)	76.2 (73.7, 78.6)	80.1 (78.0, 82.2)	56.8 (54.3, 59.3)	0.36 (0.33, 0.40)
Index hosp. or 1 DAD/NACRS visit in previous year	64.4	62.0 (59.8, 64.2)	79.7 (77.2, 82.1)	84.7 (82.8, 86.6)	53.6 (51.1, 56.1)	0.38 (0.34, 0.41)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	64.8	62.0 (59.8, 64.2)	80.1 (77.7, 82.6)	85.2 (83.3, 87.1)	53.4 (50.9, 55.9)	0.38 (0.35, 0.41)
2 DAD/NACRS visits in previous 2 years	15.0	76.5 (72.6, 80.5)	58.0 (56.1, 60.0)	24.4 (22.1, 26.7)	93.3 (92.1, 94.6)	0.18 (0.16, 0.21)
Hypertension (valid n=3082; chart prevalence=51.2%)						
Index hospitalization DAD record (s)	59.1	78.6 (76.7, 80.5)	88.4 (86.6, 90.2)	90.8 (89.3, 92.2)	74.1 (71.8, 76.3)	0.65 (0.62, 0.68)
Index hosp. or 1 DAD/NACRS visit in previous year	64.8	74.9 (73.0, 76.9)	92.4 (90.8, 93.9)	94.7 (93.6, 95.8)	66.7 (64.4, 69.1)	0.62 (0.59, 0.65)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	65.0	74.8 (72.9, 76.7)	92.5 (90.9, 94.1)	94.9 (93.8, 96.0)	66.5 (64.1, 68.9)	0.62 (0.59, 0.64)
2 DAD/NACRS visits in previous 2 years	36.0	86.3 (84.3, 88.3)	68.5 (66.5, 70.6)	60.7 (58.3, 63.1)	89.9 (88.4, 91.4)	0.50 (0.47, 0.53)
Diabetes (valid n=3072; chart prevalence=19.0%)						
Index hospitalization DAD record (s)	22.4	79.9 (76.9, 82.9)	98.6 (98.1, 99.1)	94.2 (92.3, 96.1)	94.5 (93.6, 95.4)	0.83 (0.81, 0.85)
Index hosp. or 1 DAD/NACRS visit in previous year	24.0	76.0 (72.9, 79.1)	99.0 (98.6, 99.4)	96.1 (94.5, 97.6)	92.9 (91.9, 93.9)	0.81 (0.78, 0.83)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	24.3	75.2 (72.1, 78.3)	99.0 (98.6, 99.4)	96.1 (94.5, 97.6)	92.6 (91.5, 93.6)	0.80 (0.78, 0.83)
2 DAD/NACRS visits in previous 2 years	17.7	90.1 (87.5, 92.6)	96.3 (95.5, 97.0)	83.9 (80.9, 86.9)	97.8 (97.3, 98.4)	0.84 (0.81, 0.86)
Prior heart failure (valid n=803; chart prevalence=2.7%)						
Index hospitalization DAD record (s)	9.1	19.2 (10.1, 28.2)	98.9 (98.1, 99.7)	63.6 (43.5, 83.7)	92.4 (90.6, 94.3)	0.26 (0.15, 0.38)
Index hosp. or 1 DAD/NACRS visit in previous year	10.1	18.5 (10.1, 27.0)	99.0 (98.3, 99.7)	68.2 (48.7, 87.6)	91.5 (89.6, 93.5)	0.26 (0.15, 0.37)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	10.5	17.9 (9.7, 26.0)	99.0 (98.3, 99.7)	68.2 (48.7, 87.6)	91.2 (89.2, 93.2)	0.25 (0.14, 0.36)
2 DAD/NACRS visits in previous 2 years	2.6	28.6 (9.2, 47.9)	98.0 (97.0, 98.9)	27.3 (8.7, 45.9)	98.1 (97.1, 99.0)	0.26 (0.08, 0.44)
Atrial Fibrillation (valid n=794; chart prevalence=3.5%)						
Index hospitalization DAD record (s)	9.8	28.2 (18.2, 38.2)	99.2 (98.5, 99.8)	78.6 (63.4, 93.8)	92.7 (90.8, 94.5)	0.38 (0.26, 0.50)
Index hosp. or 1 DAD/NACRS visit in previous year	10.6	27.4 (17.8, 36.9)	99.3 (98.7, 99.9)	82.1 (68.0, 96.3)	92.0 (90.1, 94.0)	0.38 (0.26, 0.49)
Index hosp. or 1 DAD/NACRS visit in previous 2 years	10.6	27.4 (17.8, 36.9)	99.3 (98.7, 99.9)	82.1 (68.0, 96.3)	92.0 (90.1, 94.0)	0.38 (0.26, 0.49)
2 DAD/NACRS visits in previous 2 years	1.8	71.4 (47.8, 95.1)	97.7 (96.6, 98.7)	35.7 (18.0, 53.5)	99.5 (99.0, 100.0)	0.46 (0.28, 0.65)

Valid n represents the number of records with nonmissing values included in validation for each comorbidity. CI=confidence interval, DAD=Discharge Abstract Database, MI=myocardial infarction, NACRS=National Ambulatory Care Reporting System, NPV=negative predictive value, PPV=positive predictive value.

STEMI diagnosis and cardiac procedures as well as common comorbidity definitions that do not rely on physician claims databases.

We found STEMI diagnosis and procedure validation yielded excellent agreement compared to chart review with Kappa values ranging from 0.84 to 1.00 for procedures defined using CCI codes. It is important to utilize ambulatory as well as inpatient data for defining PCI and angiogram, as we found that inpatient data alone captured only 80% of these procedures. A previous study showed prevalence of administrative coding for CABG remained constant in an Alberta city in the years immediately following implementation of CCI coding,^[14] suggesting comparability across time periods, whereas another study in Alberta showed good agreement between administrative CCI codes compared to chart review for coronary angiography, with Kappa 0.74 (sensitivity 61%, specificity 99.9%, PPV 98%, NPV 98%),^[15] results that are similar to ours.

Comorbidity validation varied depending on the conditions, and these results are important because they provide a basis for defining comorbidities when physician billing claims data are not available. Using DAD inpatient data alone to define comorbidities is feasible, but not ideal as algorithms that included the NACRS ambulatory database enhanced validity. In general, researchers can use the algorithm based on 2 DAD/NACRS visits in the previous 2 years to optimize PPV, or select an alternate algorithm if determined to be more appropriate for specific research objectives (e.g., to maximize sensitivity). Despite variation in each of the comorbidity algorithms, there was little difference when comparing discrimination for in-hospital mortality. Our results are consistent with other studies that considered case definition algorithms for defining comorbid conditions. Two recent studies that validated administrative definitions for diabetes compared to chart review had a Kappa value of 0.85 using hospital discharge data with a 2-year lookback period,^[16] whereas another achieved Kappa of 0.79 using an algorithm of 2 physician claims within 2 years or 1 hospitalization.^[17] For defining hypertension, 1 study obtained Kappa of 0.71 using a definition of 2 physician claims within 2 years or 1 hospitalization,^[18] whereas another achieved Kappa of >0.70 using 2 physician claims within 3 years.^[19]

4.1. Strengths and limitations

Our study has several strengths, including availability of a detailed clinical registry of STEMI patients and the ability to link to comprehensive administrative datasets. However, there are also some potential limitations. First, the chart review data that were used as our reference standard were collected by trained study personnel, but not for the purpose of validating comorbidities. Therefore, we cannot calculate inter-rater reliability for the coding of comorbidities in VHR. Second, the VHR registry recorded true comorbidities (i.e., conditions present on arrival to hospital), whereas the administrative data include conditions that developed during the episode of care, which present a possible source of discrepancy for our comparisons. Third, as our cohort includes only STEMI patients, we were unable to determine specificity, PPV, or NPV for the administrative STEMI definition. Fourth, we were unable to determine underlying causes of discrepancies between VHR and the administrative datasets. For example, PPV was higher than NPV in the validation of cardiac procedures, which suggest the administrative definitions tended to miss some procedures, but the underlying reasons for this are not known. Fifth, our cohort

represents patients hospitalized for STEMI at 5 hospitals in Edmonton, Alberta, and may not be representative of a larger, more general population within different healthcare systems; however, our demographics and comorbidity profiles closely mirror those of larger, multisite studies for patients hospitalized with MI.^[20–23] Finally, although we utilized the Canadian enhancement to *ICD-10* for defining comorbidities, our definitions relied mostly on 3-digit codes, which are consistent among countries that use *ICD-10* morbidity coding.

5. Conclusions

In general, Canadian *ICD-10* codes can be used to accurately define comorbidities using administrative algorithms that rely on DAD (inpatient) and NACRS (ambulatory) data when it is not feasible to include physician billing claims as part of the comorbidity algorithms. Although definitions for some conditions were less than optimal, this had little effect on the discriminative ability of the comorbidities for predicting in-hospital mortality. Comorbidity algorithms that use 2 DAD/NACRS visits in the previous 2 years should be used to maximize PPV; however, other algorithms can be used if interest instead lies in maximizing sensitivity, specificity, or NPV. Additionally, *ICD-10* and CCI codes can be used to accurately identify patients hospitalized for STEMI and define cardiac procedures, respectively.

References

- [1] Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
- [2] Quan H, Li B, Saunders LD, et al. Assessing validity of *ICD-9-CM* and *ICD-10* administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424–41.
- [3] Lee DS, Donovan L, Austin PC, et al. Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. *Med Care* 2005;43:182–8.
- [4] Humphries KH, Rankin JM, Carere RG, et al. Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol* 2000;53:343–9.
- [5] Henderson T, Shephard J, Sundararajan V. Quality of diagnosis and procedure coding in *ICD-10* administrative data. *Med Care* 2006;44:1011–9.
- [6] Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in *ICD-9-CM* and *ICD-10* administrative data. *Med Care* 2005;43:1130–9.
- [7] So L, Evans D, Quan H. *ICD-10* coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res* 2006;6:161.
- [8] Januel JM, Luthi JC, Quan H, et al. Improved accuracy of co-morbidity coding over time after the introduction of *ICD-10* administrative data. *BMC Health Serv Res* 2011;11:194.
- [9] Leal JR, Laupland KB. Validity of ascertainment of co-morbid illness using administrative databases: a systematic review. *Clin Microbiol Infect* 2010;16:715–21.
- [10] Tonelli M, Wiebe N, Fortin M, et al. Methods for identifying 30 chronic conditions: application to administrative data. *BMC Med Inform Decis Mak* 2015;15:31.
- [11] Shavadia J, Ibrahim Q, Sookram S, et al. Bridging the gap for nonmetropolitan STEMI patients through implementation of a pharmacoinvasive reperfusion strategy. *Can J Cardiol* 2013;29:951–9.
- [12] Bainey KR, Ferguson C, Ibrahim QI, et al. Vital Heart Response Registry Investigators Impact of reperfusion strategy on aborted myocardial infarction: insights from a large Canadian ST-Elevation Myocardial Infarction Clinical Registry. *Can J Cardiol* 2014;30:1570–5.
- [13] Toleva O, Ibrahim Q, Brass N, et al. Treatment choices in elderly patients with ST: elevation myocardial infarction—insights from the Vital Heart Response registry. *Open Heart* 2015;2:e000235.

- [14] Quan H. ICD-10-CA/CCI coding algorithms for defining clinical variables to assess outcome after aortic and mitral valve replacement surgery. *Can J Cardiol* 2006;22:153–4.
- [15] De Coster C, Li B, Quan H. Comparison and validity of procedures coded With ICD-9-CM and ICD-10-CA/CCI. *Med Care* 2008;46:627–34.
- [16] Nedkoff L, Knuiman M, Hung J, et al. Concordance between administrative health data and medical records for diabetes status in coronary heart disease patients: a retrospective linked data study. *BMC Med Res Methodol* 2013;13:121.
- [17] Chen G, Khan N, Walker R, et al. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010;89:189–95.
- [18] Quan H, Khan N, Hemmelgarn BR, et al. Validation of a case definition to define hypertension using administrative data. *Hypertension* 2009;54:1423–8.
- [19] Tu K, Campbell NR, Chen ZL, et al. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18–26.
- [20] Abuzeid W, Bennell M, Qiu F, et al. Clinical outcomes of early repatriation for patients with ST-segment elevation myocardial infarction: a propensity-matched analysis. *Can J Cardiol* 2015;31:1225–31.
- [21] Madan M, Halvorsen S, Di Mario C, et al. Relationship between time to invasive assessment and clinical outcomes of patients undergoing an early invasive strategy after fibrinolysis for st-segment elevation myocardial infarction: a patient-level analysis of the randomized early routine invasive clinical trials. *JACC Cardiovasc Interv* 2015;8:166–74.
- [22] Canto JG, Rogers WJ, Goldberg RJ, et al. Association of age and sex with myocardial infarction symptom presentation and in-hospital mortality. *JAMA* 2012;307:813–22.
- [23] Katanuki H, Honda T, Haze K, et al. A large-scale prospective cohort study on the current status of therapeutic modalities for acute myocardial infarction in Japan: rationale and initial results of the HIJAMI Registry. *Am Heart J* 2005;150:411–8.