

# Optimal look back period and summary method for Elixhauser comorbidity measures in a US population-based electronic health record database

Yannick Fortin<sup>1,2</sup>James AG Crispo<sup>1-4</sup>Deborah Cohen<sup>2,5,6</sup>Douglas S McNair<sup>7</sup>Donald R Mattison<sup>1,8</sup>Daniel Krewski<sup>1,2,8</sup>

<sup>1</sup>McLaughlin Centre for Population Health Risk Assessment, <sup>2</sup>School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>3</sup>Fulbright Canada, <sup>4</sup>Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; <sup>5</sup>Canadian Population Health Initiative, Canadian Institute for Health Information, Ottawa, <sup>6</sup>Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; <sup>7</sup>Cerner Corporation, Kansas City, MO, USA; <sup>8</sup>Risk Sciences International, Ottawa, ON, Canada

**Background:** Comorbidity risk-adjustment tools are widely used in health database research to control for clinical differences between individuals, but they need to be validated a priori. This study aimed to identify the optimal parameters for predicting all-cause in-hospital mortality using Quan's enhanced Elixhauser comorbidity measures (ECMs) in the US-based Cerner Health Facts® (HF) electronic health record database.

**Methods:** Health care recipients aged 18–89 years between 2002 and 2011 were included. Prevalent comorbidities recorded, 1) during the index encounter; 2) in the prior year; and 3) in the prior 2 years were identified using the ECMs. Multiple logistic regression models, with in-hospital mortality at index and at 1 year as the predicted outcomes, were fitted with comorbidities summarized as binary indicators, total counts, or weighted scores for the three look back periods. Baseline variables included sex and age. The receiver operating characteristic (ROC) curves of the competing models were compared with a non-parametric Mann–Whitney *U* test to identify the optimal parameters.

**Results:** A sample of 3,273,298 unique health care recipients were included, of whom 31,298 (1.0%) and 50,215 (1.5%) died during the index encounter and within the 1-year follow-up, respectively. Models of comorbidity based on binary and weighted indicators had near-identical performance and were statistically better than the models based on total counts ( $p < 0.0001$ ). Discrimination of in-hospital mortality was highest with a look back period limited to the index encounter, while in-hospital mortality at 1 year was best predicted with 1 year of look back ( $p < 0.0001$ ).

**Conclusion:** In Cerner HF, the binary and weighted methods for summarizing the Quan ECM were the best predictors of all-cause in-hospital mortality at index and at 1 year. Observed differences in predictive performance between models with diagnostic ascertainment periods of up to 2 years of look back were statistically significant but not practically important.

**Keywords:** comorbidity, ICD-9, electronic health records, risk adjustment, mortality, statistical modeling

## Introduction

Risk-adjustment measures of patient comorbidity are commonly used in health database research and are associated with short- and long-term mortality, hospital costs, inpatient length of stay (LOS), physician visits, and hospital readmissions.<sup>1–9</sup>

The strong discriminatory performance of two Elixhauser comorbidity measure (ECM) variants, by Quan et al<sup>10</sup> and the Agency for Healthcare Research and Quality (AHRQ),<sup>11</sup> for predicting in-hospital mortality was recently confirmed in the

Correspondence: Yannick Fortin  
McLaughlin Centre for Population Health Risk Assessment, University of Ottawa,  
850 Peter Morand Crescent, Room 119,  
Ottawa, ON, K1G 3Z7 Canada  
Tel +1 613 261 8713  
Fax +1 613 562 5380  
Email yfort006@uottawa.ca

Cerner Health Facts® (HF) database.<sup>12</sup> HF is a longitudinal electronic health record (EHR) data source populated by health care centers located across the continental US in compliance with the US Health Insurance Portability and Accountability Act (HIPAA). HIPAA-compliant data repositories must adhere to strict patient de-identification procedures to ensure the privacy and protection of personal information.<sup>13</sup> These operational constraints preclude any possibility of subsequent data linkage to national registries such as the National Death Index or to other health provider data sources that are not covered under a common HIPAA network. EHR databases such as HF are therefore limited in their ability to fully capture the utilization of health services of their constitutive patients, possibly leading to missing information about health outcomes and health service use happening outside their privacy network. How this potential missing information affects the performance of comorbidity risk assessment tools in HF and similar health data sources deserves investigation, if only for the interpretation of studies derived from them.

Selecting an appropriate ascertainment look back period (henceforth, look back) is a standard decision researchers have to make prior to relying on comorbidity risk-adjustment tools. A look back period is the length of a patient's medical history available for the assessment of prevalent comorbidities. When look back is expanded to include health encounters preceding an index encounter, additional diagnoses, possibly unreported during the index, are more likely to be detected. One reason cited for expanding the look back period beyond a single health encounter is to limit discharge coding bias, whereby existing and less serious chronic health conditions are less likely to be recorded within EHRs during acute health encounters and vice versa.<sup>1,14</sup> In a recent systematic review, Yurkovich et al<sup>15</sup> concluded that combining health data from outpatient and inpatient encounters yields comorbidity indices with better predictive performance for mortality outcomes than inpatient data alone. Evidence also suggests that longer look back is associated with comorbidity indices with better predictive performance for a wide array of health outcomes.<sup>4,16</sup> However, performance increases beyond 1 year of look back may be insignificant.<sup>17</sup> In HF and similar sources of health data in which patient follow-up cannot be confirmed with complete accuracy, the selection of an optimal look back period should be confirmed through empirical investigation.

The comorbidity status of patients, inferred from the health conditions recorded in EHRs, can be summarized in

several ways. Unlike the Charlson<sup>8</sup> comorbidity index and its variants,<sup>18–21</sup> which summarize overall comorbidity in a single score on a continuous scale, ECMs consist of 30 binary indicators.<sup>1,10</sup> This makes ECMs impractical when conducting association studies or when matching cases to controls based on health status. Fortunately, several approaches are available for improving the manageability of ECMs. The simplest method is to count the total number of prevalent comorbidities identified during the look back period (henceforth, the total method). Comorbidities can also be assigned relative weights based on the strength of their association with an outcome of interest, e.g., in-hospital mortality. Once these relative weights are derived, they can be added to produce a single comorbidity score for each patient. The ECM point system developed and validated by van Walraven et al<sup>22</sup> (henceforth, the weighted method) is based on this approach but has yet to be validated in a US health data source. The principal and arguably questionable assumption of the total method is that every ECM health condition has the same strength of association with the outcome of interest. The weighted method addresses this limitation by producing a weighted comorbidity score that accounts for the relative ability of each ECM condition to predict an outcome.

Alternative methods for summarizing overall comorbidity in patients result in comorbidity risk-adjustment tools with divergent predictive performances.<sup>16</sup> In a recent meta-analysis, Sharabiani et al<sup>16</sup> tested the discrimination performance of comorbidity measures for predicting short- and long-term outcomes with different comorbidity summary methods. When predicting inpatient and short-term mortality (<30 days), model performance ranked in increasing order from the binary method to the total method and then to the weighted method. For predicting long-term outpatient mortality (>30 days), the performance of comorbidity summary methods ranked differently. They improved from the total method to the weighted method and then to the binary method. The predictive performance of competing ECM summary methods has not been compared in the HF database, and it is unclear if one approach outperforms the others in this type of data source.

The primary objective of this study was to identify the optimal look back period (index encounter alone versus 1 year versus 2 years) for the Quan ECM when predicting all-cause in-hospital mortality and in-hospital mortality at 1 year in HF. The secondary objective was to ascertain which comorbidity summary method, the binary, total, or weighted, best predicts the mortality outcomes.

## Methods

### Source of health data

The study is based on EHRs from the Cerner HF (Kansas City, MO, USA) data repository and relies on inpatient, outpatient, and emergency care encounters recorded between January 2000 and December 2012. Cerner is a global provider of health informatics solutions, and HF is managed in accordance with the HIPAA. To date, >500 health care facilities have contributed patient-level data to HF, including details on hospital admissions, diagnoses, medication orders, laboratory tests, medical procedures, and surgical interventions.<sup>23</sup> Data-contributing health care facilities are located in all US census regions and are categorized by teaching status, urban/rural setting, and bed size. Participating data sources are primarily located in the Northeast census region (36%) and ~60% of HF health encounters are recorded by university-affiliated teaching hospitals. Permission to access and analyze HF data was granted by Cerner Corporation, Kansas City, MO, USA.

HIPAA compliance by the Cerner Corporation ensures complete de-identification and anonymization of HF patients. Ethical approval for this study was granted by the University of Ottawa's Office for Research Ethics and Integrity who deemed patient consent unnecessary due to the de-identification process.

### Study population

Individuals aged 18–89 years with one or more inpatient or emergency department (ED) visits between January 2002 and December 2011 were admissible for inclusion in the study. For each admissible patient, a single health care encounter was selected at random and assigned as the index encounter. Random selection of the index encounter was completed to limit systematic information bias while allowing for the possibility of look back and follow-up time. The study inclusion dates were selected to ensure index encounters had 2 years of available look back time and 1 year of follow-up. Individuals younger than 18 years were excluded due to the low prevalence of ECM conditions and mortality in this age group. Individuals aged 90 years or older were excluded because age is treated as a continuous variable in this study and age values are pooled in a single category beyond 89 years in HF due to HIPAA requirements. To avoid bias resulting from missing information, we excluded index encounters in which individuals were admitted from or transferred to an outside health care facility.<sup>22</sup> Individual characteristics included sex (male and female), race limited to the four highest frequency categories (Caucasian, African American, Hispanic, and Asian),

and age captured during the index encounter. Primary health insurance status at the index encounter was classified based on recommendations from the AHRQ<sup>24</sup> as private, Medicaid, Medicare, uninsured/self-pay, other Tricare (formerly known as the Civilian Health and Medical Program of the Uniformed Services [CHAMPUS]), (international plan, research funded, Title V, worker's compensation), or missing. An important proportion of health insurance status values were expected to be missing since data contributors often elect to retain this type of information from being transferred to external data repositories such as HF.

### ECMs

Quan et al's<sup>10</sup> enhanced International Classification of Diseases, Ninth Edition (ICD-9), ECM was selected to identify the prevalence of 30 health conditions. This choice was based on prior evidence of superior performance by this ECM for predicting inhospital mortality in HF compared to version 3.7 of the AHRQ ECM.<sup>25</sup> The Quan ECM relies on specific combinations of ICD-9 diagnostic codes to identify cases of congestive heart failure, cardiac arrhythmia, valvular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension (un/complicated), paralysis, neurological disorders, chronic pulmonary disease, uncomplicated diabetes, complicated diabetes, hypothyroidism, renal failure, liver disease, peptic ulcer disease without bleeding, AIDS/HIV, lymphoma, metastatic cancer, solid tumor without metastasis, rheumatoid arthritis/collagen vascular diseases, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anemia, deficiency anemia, alcohol abuse, drug abuse, psychoses, and depression, from electronic health data.<sup>26,27</sup>

### Look back periods

To test the consequences of varying the diagnostic ascertainment look back period on the predictive performance of the ECM, prevalent health conditions were identified during the index encounter alone (index only), during the index encounter and any encounter recorded in the preceding year (1-year look back), and during the index encounter and any encounter recorded during the previous 2 years (2-year look back). Prior studies have reported that when the objective is to predict mortality, disease prevalence captured from multiple health care settings generally performs better (higher c-statistic values) than disease prevalence assessments limited to a single health care setting.<sup>4,15,28,29</sup> For this reason, diagnostic ascertainment in this study included information recorded in inpatient, outpatient, and emergency care settings.

## ECM summary methods

The overall comorbidity status of patients was summarized using three approaches. The binary method reflected the original Elixhauser format, and the 30 health conditions assessed were coded as binary indicators (present or absent). In the total method, each health condition identified was assigned a value of 1 and a total count was derived to produce a continuous variable ranging from 0 to 30. For the weighted method, the van Walraven (VW)<sup>22</sup> point system was implemented and relative weights were calculated based on HF data. The VW point system is principally founded on the methodology from the Framingham Heart Study for compiling risk score for developing coronary heart disease over time.<sup>30</sup> To begin, a random number generator based on the Bernoulli distribution was used to divide the HF encounters into two subsamples. The first dataset (derivation) was reserved to generate the VW weights and the second dataset (validation) was retained to conduct the comparative statistical analyses. In the VW approach, backward stepwise logistic regression was used to predict the outcome of interest (e.g., in-hospital mortality) using the ECM binary indicators as parameters. Conditions significantly associated ( $p < 0.05$ ) with the dependent in the best-fitted model were retained and parameter estimates were divided by the absolute value of the parameter with the lowest admissible estimate. ECM conditions not significantly associated with the dependent were assigned a weight of zero. A weighted summary score was then compiled for each patient by adding the relative weight of each positively identified health condition during the diagnostic ascertainment step. Since this study has three look back periods and two outcomes, six sets of VW weights were generated.

## Statistical analysis

The study outcomes were all-cause in-hospital mortality during the index encounter and at 1 year. In-hospital mortality at 1 year was defined as a death recorded during an HF encounter (inpatient stay or emergency visit) in the 365 days that followed the admission date of the index encounter. Deaths occurring during the index encounter are therefore included in the in-hospital mortality at 1-year outcome. Because deaths occurring in outpatient settings are rare relative to those occurring in acute care settings,<sup>31,32</sup> they were excluded from the study.

Patient demographic and index encounter characteristics for the derivation, validation, and complete datasets are reported as counts and percentages for categorical variables and as mean and standard error (SE) values for continuous

variables. Observed differences between the derivation and validation datasets were compared with Pearson's chi-squared test for the categorical variables and with Student's *t*-test for the continuous variables. Multiple logistic regression was used to fit the six combinations of look back and comorbidity summary method and to predict the mortality outcomes. Since the weights required by the VW point system were produced from the derivation dataset, the comparative analyses were conducted with the validation dataset as opposed to the complete sample. The baseline model was limited to sex and age at index to allow for comparisons with prior studies.<sup>4,25,33</sup>

Statistical analyses were completed with SAS version 9.4 (SAS Institute Inc., Cary, NC, USA), which allowed for concordance index (c-statistic) values to be outputted from the logistic regression function using the receiver operating characteristic (ROC) option.<sup>34</sup> Interpretation of the c-statistics follows recommendations by Hosmer Jr, Lemeshow, and Sturdivant,<sup>35</sup> whereby values  $>0.7$  are considered reasonable, those  $>0.8$  are strong, and those  $>0.9$  are exceptional. Differences in the capacity of each model to discriminate between the mortality outcomes were compared using the ROCCONTRAST option, which relies on a variation of the non-parametric Mann–Whitney *U* test developed by DeLong et al.<sup>36</sup>

## Results

A dataset based on the inclusion of 3,273,298 unique health care recipients was randomly parsed to create a derivation and a validation dataset containing 1,637,572 and 1,635,726 patients, respectively (Table 1). In the complete dataset, the majority were females (53.8%) and Caucasians (72.3%). The mean age was 41.1 years (SE = 0.01), and 31,298 (1.0%) deaths of any cause were recorded during the index. At 1 year, 50,215 (1.5%) deaths were recorded. A larger percentage of index encounters originated from health care facilities located in the Northeast (36.1%) census region compared to the Midwest (19.8%), South (32.9%), and West (11.2%). Forty-one percent of health insurance statuses were missing (41.0%), and the privately insured (24.3%) made up the majority of non-missing values. The observed differences between the derivation and validation samples were not significant.

The prevalence of the health conditions ascertained ranked consistently across all look back periods and was led by hypertension (17.5–19.5%), chronic pulmonary disease (7.8–9.2%), diabetes without complications (6.9–7.8%), fluid and electrolyte disorders (5.7–7.0%), and cardiac arrhythmia (5.3–6.2%; Table 2). The greatest increase in prevalence between the index encounter and 2 years of look back were for hypertension (2.0%), chronic pulmonary disease (1.3%),

**Table 1** Patient demographic and encounter characteristics, for the complete, derivation, and validation datasets

Variables	Complete, N = 3,273,298 (%)	Derivation, n = 1,637,572 (%)	Validation, n = 1,635,726 (%)	p value <sup>a</sup>
Sex				0.366 <sup>b</sup>
Female	1,761,525 (53.8)	881,667 (53.8)	879,858 (53.9)	
Age (years)				0.706 <sup>c</sup>
Mean $\pm$ SE	44.1 $\pm$ 0.01	44.0 $\pm$ 0.01	44.1 $\pm$ 0.01	
Race				0.468 <sup>b</sup>
Caucasian	2,366,665 (72.3)	1,183,791 (72.3)	1,182,874 (72.3)	
African American	711,051 (21.7)	355,705 (21.7)	355,346 (21.7)	
Hispanic	146,877 (4.5)	73,765 (4.5)	73,112 (4.5)	
Asian	48,705 (1.5)	24,311 (1.5)	24,394 (1.5)	
Health insurance status				0.685 <sup>b</sup>
Private	795,449 (24.3)	397,520 (24.3)	397,929 (24.3)	
Medicare	370,701 (11.3)	185,497 (11.3)	185,204 (11.3)	
Medicaid	248,009 (7.6)	124,091 (7.6)	123,918 (7.6)	
Uninsured	378,536 (11.6)	189,729 (11.6)	188,807 (11.5)	
Other	139,745 (4.3)	69,905 (4.3)	69,840 (4.3)	
Missing	1,340,858 (41.0)	670,830 (41.0)	670,028 (41.0)	
Census region				0.229 <sup>b</sup>
Northeast	1,180,270 (36.1)	589,807 (36.0)	590,463 (36.1)	
Midwest	648,644 (19.8)	325,141 (19.9)	323,503 (19.8)	
South	1,077,965 (32.9)	539,459 (32.9)	538,506 (32.9)	
West	366,419 (11.2)	183,165 (11.2)	183,254 (11.2)	
Inhospital mortality				0.351 <sup>b</sup>
Deaths	31,298 (1.0)	15,740 (1.0)	15,558 (1.0)	
Inhospital mortality at 1 year				0.185 <sup>b</sup>
Deaths	50,215 (1.5)	25,269 (1.5)	24,946 (1.5)	

**Notes:** <sup>a</sup>p value resulting from the significance tests used to evaluate the difference between the derivation and validation subsamples. <sup>b</sup>Statistical significance derived using Pearson's chi-squared test. <sup>c</sup>Statistical significance derived using Student's t-test.

**Abbreviation:** SE, standard error.

and fluid and electrolyte disorders (1.3%). The adjusted odds and relative weight of each health condition derived during the implementation of the VW point system are reported in Table 3 for inpatient mortality and in Table 4 for inpatient mortality at 1 year. In the models retained for predicting inpatient mortality, uncomplicated diabetes and blood loss anemia were excluded for all look back periods, while the rheumatoid arthritis/collagen disease group was excluded from the models based on 1 and 2 years of look back. The parameters retained in the models predicting inpatient mortality at 1 year differed. Valvular disease was excluded across all look back periods, while diabetes with complications and alcohol abuse were excluded from the 1- and 2-year look back models. Relative points (weights) attributed to health conditions ranged from -6 to 11 across the three look back periods when predicting inpatient mortality, and neurological disorders, metastatic cancer, and fluid and electrolyte disorders had the largest weights. For inpatient mortality at 1 year, health conditions exhibited a broader range of points, from -11 to 27, across the look back periods, with liver disease, lymphoma, and metastatic cancer having the largest weights.

Model performance comparisons based on the validation dataset showed that irrespective of the look back period

and the outcome predicted, the combination of the baseline variables with any of the summary methods improved on the predictive performance of the baseline model alone ( $p < 0.0001$ ) (Table 5). For inpatient mortality, the average performance improvement on the baseline model across the look back periods was 6.6%, 3.4%, and 6.4% for the binary, total, and weighted summary methods, respectively. For inpatient mortality at 1 year, the equivalent improvements in c-statistics were 5.9%, 3.7%, and 5.7%.

For both mortality outcomes, the predictive performance of models improved consistently and significantly in the following order of summary method for every look back period: total < weighted < binary. Irrespective of the look back period and outcome, the predictive performances of the binary and weighted summary methods were nearly identical ( $0.882 \leq \text{c-statistics} \leq 0.887$ ), varying by less than a percent in direct comparisons. The most noticeable performance improvement across the look back periods was the outperformance of the total method by the binary and weighted methods, which averaged approximately 3% for inpatient mortality at index and  $\approx 2\%$  for inpatient mortality at 1 year. The c-statistic confidence intervals of every direct comparison between the binary and the weighted methods

**Table 2** Prevalence of comorbid conditions by look back period, N = 3,273,298

Health condition(s)	Index only, n (%)	1-year look back, n (%)	2-year look back, n (%)
Congestive heart failure	99,280 (3.03)	109,970 (3.36)	116,581 (3.56)
Cardiac arrhythmia	174,656 (5.34)	191,000 (5.84)	202,319 (6.18)
Valvular disease	51,250 (1.57)	58,506 (1.79)	63,974 (1.95)
Pulmonary circulation disorders	21,084 (0.64)	24,790 (0.76)	27,212 (0.83)
Peripheral vascular disorders	41,651 (1.27)	49,317 (1.51)	54,798 (1.67)
Hypertension combined	572,139 (17.48)	611,354 (18.68)	638,363 (19.50)
Hypertension uncomplicated	518,529 (15.84)	558,185 (17.05)	586,568 (17.92)
Hypertension complicated	55,303 (1.69)	62,760 (1.92)	67,135 (2.05)
Paralysis	13,707 (0.42)	15,417 (0.47)	16,675 (0.51)
Other neurological disorders	76,237 (2.33)	84,644 (2.59)	90,360 (2.76)
Chronic pulmonary disease	256,170 (7.83)	281,398 (8.60)	300,230 (9.17)
Diabetes, uncomplicated	226,807 (6.93)	245,625 (7.50)	256,680 (7.84)
Diabetes, complicated	34,016 (1.04)	40,131 (1.23)	44,068 (1.35)
Hypothyroidism	85,493 (2.61)	96,038 (2.93)	103,358 (3.16)
Renal failure	60,238 (1.84)	66,706 (2.04)	70,142 (2.14)
Liver disease	31,307 (0.96)	36,958 (1.13)	41,044 (1.25)
Peptic ulcer disease, excluding bleeding	8,251 (0.25)	10,142 (0.31)	11,719 (0.36)
AIDS/HIV	4,246 (0.13)	4,921 (0.15)	5,235 (0.16)
Lymphoma	6,778 (0.21)	7,506 (0.23)	7,958 (0.24)
Metastatic cancer	22,836 (0.70)	24,646 (0.75)	25,405 (0.78)
Solid tumor without metastasis	56,376 (1.72)	63,399 (1.94)	67,709 (2.07)
Rheumatoid arthritis/collagen	24,143 (0.74)	27,417 (0.84)	29,763 (0.91)
Coagulopathy	27,019 (0.83)	31,378 (0.96)	34,258 (1.05)
Obesity	79,680 (2.43)	90,225 (2.76)	98,747 (3.02)
Weight loss	21,219 (0.65)	26,262 (0.80)	29,644 (0.91)
Fluid and electrolyte disorders	187,321 (5.72)	211,182 (6.45)	228,697 (6.99)
Blood loss anemia	10,884 (0.33)	12,742 (0.39)	14,262 (0.44)
Deficiency anemia	19,745 (0.60)	24,617 (0.75)	28,171 (0.86)
Alcohol abuse	89,460 (2.73)	96,224 (2.94)	101,830 (3.11)
Drug abuse	61,260 (1.87)	68,894 (2.10)	75,069 (2.29)
Psychoses	29,656 (0.91)	33,619 (1.03)	36,329 (1.11)
Depression	113,659 (3.47)	130,320 (3.98)	143,078 (4.37)

overlapped, yet differences in model performance between the methods were statistically significant ( $p < 0.0001$ ).

When predicting in-hospital mortality, the best performing look back period across all summary methods was the diagnostic assessment limited to the index encounter alone. For predictions of in-hospital mortality at 1 year, the best look back period was the index encounter supplemented with 1 year of look back. While the differences in predictive performance across look back periods for each summary method were statistically different, performance improvements were negligible and rarely exceeded more than two-tenths of a percent. ROC curves are available as supplementary materials (Figures S1 and S2).

## Discussion

The goal of this study was to advance methodologies for conducting observational studies founded on privacy-protected data warehouses. More specifically, we aimed to identify the optimal look back period and comorbidity summary

method for the Quan ECM when predicting in-hospital mortality at index and in-hospital mortality at 1 year in Cerner HF. We found that the optimal comorbidity risk-adjustment parameters differed minimally across the mortality outcome. The Quan ECM model based on the binary method with diagnostic ascertainment limited to the index encounter performed the best for predicting in-hospital mortality. When predicting in-hospital mortality at 1 year, the binary method with 1 year of look back had the highest c-statistic. A recent study by Thompson et al<sup>37</sup> conducted using the AHRQ National Inpatient Sample and the Maryland State Inpatient Database showed that the binary method ( $c = 0.809$ , 95% CI: 0.808–0.810) for the original Elixhauser comorbidities performed slightly better than the VW method ( $c = 0.802$ , 95% CI: 0.801–0.804), albeit the cardiac arrhythmia group was excluded from both models. During the validation of the original VW point system, van Walraven et al<sup>22</sup> observed a similar but inverse performance between the binary ( $c = 0.760$ , 95% CI: 0.756–0.764) and weighted ( $c = 0.763$ ,

**Table 3** Adjusted odds of in-hospital mortality and relative points by health condition(s) and look back period, n = 1,637,572

Health condition(s)	Index only, OR (CI)	Points	1-year look back, OR (CI)	Points	2-year look back, OR (CI)	Points
Congestive heart failure	3.33 (3.17–3.50)***	9	3.13 (2.98–3.30)***	8	3.04 (2.89–3.19)***	9
Cardiac arrhythmia	2.87 (2.75–3.00)***	8	2.77 (2.65–2.89)***	7	2.68 (2.57–2.80)***	8
Valvular disease	0.87 (0.81–0.95)***	–1	0.83 (0.77–0.89)***	–1	0.80 (0.74–0.86)***	–2
Pulmonary circulation disorders	1.38 (1.25–1.52)***	2	1.31 (1.19–1.43)***	2	1.29 (1.18–1.41)***	2
Peripheral vascular disorders	1.98 (1.84–2.14)***	5	1.68 (1.56–1.81)***	4	1.55 (1.44–1.66)***	4
Hypertension	0.85 (0.82–0.89)***	–1	0.86 (0.82–0.89)***	–1	0.85 (0.82–0.89)***	–1
Paralysis	3.54 (3.20–3.92)***	9	3.12 (2.82–3.44)***	8	2.86 (2.60–3.16)***	9
Neurological disorders	4.20 (3.99–4.42)***	11	3.73 (3.55–3.93)***	9	3.55 (3.38–3.73)***	11
Chronic pulmonary disease	1.25 (1.19–1.31)***	2	1.16 (1.10–1.21)***	1	1.13 (1.08–1.19)***	1
Diabetes without complications	E	E	E	E	E	E
Diabetes with complications	0.68 (0.61–0.77)***	–3	0.64 (0.57–0.71)***	–3	0.64 (0.58–0.71)***	–4
Hypothyroidism	0.79 (0.73–0.86)***	–2	0.80 (0.74–0.87)***	–2	0.79 (0.73–0.85)***	–2
Renal failure	1.85 (1.74–1.97)***	5	1.67 (1.57–1.78)***	4	1.65 (1.56–1.76)***	4
Liver disease	3.50 (3.24–3.78)***	9	2.99 (2.77–3.22)***	7	2.80 (2.60–3.01)***	9
Peptic ulcer disease, excluding bleeding	0.49 (0.36–0.67)***	–5	0.51 (0.39–0.66)***	–5	0.51 (0.41–0.64)***	–6
AIDS/HIV	1.77 (1.39–2.25)***	4	1.66 (1.32–2.09)***	3	1.66 (1.33–2.08)***	4
Lymphoma	2.73 (2.32–3.21)***	7	2.29 (1.96–2.68)***	6	2.15 (1.84–2.50)***	6
Metastatic cancer	3.62 (3.31–3.97)***	9	3.73 (3.42–4.08)***	9	3.78 (3.46–4.13)***	11
Solid tumor without metastasis	2.20 (2.04–2.38)***	6	1.88 (1.75–2.03)***	4	1.80 (1.67–1.94)***	5
Rheumatoid arthritis/collagen	1.15 (1.00–1.31)*	1	E	E	E	E
Coagulopathy	2.87 (2.67–3.08)***	8	2.52 (2.35–2.69)***	6	2.38 (2.23–2.55)***	7
Obesity	0.60 (0.54–0.66)***	–4	0.59 (0.54–0.65)***	–4	0.60 (0.55–0.65)***	–4
Weight loss	2.84 (2.64–3.05)***	8	2.51 (2.35–2.69)***	6	2.44 (2.28–2.61)***	7
Fluid and electrolyte disorders	3.74 (3.60–3.90)***	10	3.58 (3.44–3.73)***	9	3.45 (3.32–3.59)***	10
Blood loss anemia	E	E	E	E	E	E
Deficiency anemia	0.47 (0.40–0.56)***	–5	0.55 (0.48–0.63)***	–4	0.53 (0.46–0.60)***	–5
Alcohol abuse	0.82 (0.76–0.90)***	–1	0.86 (0.79–0.94)***	–1	0.89 (0.82–0.96)***	–1
Drug abuse	0.57 (0.50–0.65)***	–4	0.52 (0.46–0.60)***	–4	0.51 (0.45–0.58)***	–6
Psychoses	0.73 (0.62–0.87)***	–2	0.72 (0.62–0.84)***	–2	0.73 (0.63–0.84)***	–3
Depression	0.59 (0.54–0.65)***	–4	0.54 (0.50–0.59)***	–4	0.52 (0.47–0.56)***	–6

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

Abbreviations: OR, odds ratio; CI, 95% confidence interval; E, eliminated from the model.

95% CI: 0.759–0.766) methods when predicting in-hospital mortality for a Canadian population. Our findings pertaining to the best method for summarizing Elixhauser conditions fit somewhere between the above studies. While the binary method was statistically superior, when holding the look back period constant, we found that improvements in predictive performance between the binary and weighted methods were not likely clinically important. As for the total method, it performed significantly worse on average than the other methods. Relatively lower performance by the total method was recently demonstrated<sup>37</sup> and was expected because it is unrealistic to assume that every ECM condition has an equivalent relationship to the outcome of interest.

This study asked if the length of the ascertainment look back period affected the capacity of models to discriminate between mortality statuses. Previous studies had shown that longer look back periods are better predictors of long-term mortality and that look back periods limited to the index

encounter are better predictor of in-hospital mortality at index.<sup>4,16,17</sup> In this study, every direct comparison between look back periods proved highly statistically different ( $p < 0.0001$ ), yet varying the look back period appeared practically irrelevant. For in-hospital mortality, we observed statistically significant decreases in predictive performance with increasing look back. For predicting in-hospital mortality at 1 year, the 1-year look back option was statistically the best option. Complementary results from sensitivity analyses based on 3-, 4-, and 5-year look back periods (not reported) failed to demonstrate any improvement in predictive performance and did not challenge the main study findings. Within comorbidity summary methods, the statistically significant difference in predictive performance between the look back periods are not likely to have any clinical relevance since they rarely exceeded half a percent. In this study, mortality outcomes were selected because they are the most commonly used end points in validation studies of comorbidity measures

**Table 4** Adjusted odds of in-hospital mortality at 1 year and relative points by health condition(s) and look back period, n = 1,637,572

Health condition(s)	Index only, OR (CI)	Points	1-year look back, OR (CI)	Points	2-year look back, OR (CI)	Points
Congestive heart failure	3.32 (3.19–3.45)***	24	3.11 (2.98–3.23)***	13	3.01 (2.89–3.13)***	15
Cardiac arrhythmia	2.40 (2.31–2.49)***	17	2.30 (2.22–2.39)***	10	2.24 (2.16–2.32)***	11
Valvular disease	E		E		E	
Pulmonary circulation disorders	1.24 (1.15–1.34)***	4	1.24 (1.15–1.33)***	2	1.22 (1.14–1.31)***	3
Peripheral vascular disorders	2.00 (1.88–2.13)***	14	1.77 (1.67–1.87)***	7	1.65 (1.56–1.74)***	7
Hypertension	1.05 (1.02–1.09)**	1	1.09 (1.06–1.13)***	1	1.10 (1.06–1.13)***	1
Paralysis	3.22 (2.94–3.52)***	23	2.86 (2.62–3.13)***	12	2.67 (2.45–2.90)***	13
Neurological disorders	3.20 (3.06–3.35)***	23	2.93 (2.80–3.06)***	13	2.78 (2.66–2.90)***	14
Chronic pulmonary disease	1.56 (1.51–1.62)***	9	1.45 (1.40–1.50)***	4	1.41 (1.36–1.46)***	5
Diabetes without complications	1.15 (1.10–1.20)***	3	1.10 (1.06–1.15)***	1	1.08 (1.04–1.12)***	1
Diabetes with complications	1.09 (1.01–1.19)*	2	E		E	
Hypothyroidism	0.89 (0.84–0.95)***	–2	0.92 (0.87–0.97)**	–1	0.90 (0.85–0.95)***	–1
Renal failure	2.14 (2.03–2.24)***	15	1.88 (1.79–1.97)***	7	1.83 (1.75–1.92)***	8
Liver disease	3.46 (3.24–3.69)***	25	2.94 (2.77–3.12)***	13	2.74 (2.58–2.90)***	14
Peptic ulcer disease, excluding bleeding	0.76 (0.62–0.93)**	–5	0.77 (0.65–0.91)**	–3	0.76 (0.65–0.89)***	–4
AIDS/HIV	2.24 (1.85–2.70)***	16	2.00 (1.67–2.40)***	8	1.97 (1.65–2.35)***	9
Lymphoma	3.92 (3.47–4.43)***	27	3.46 (3.08–3.89)***	15	3.26 (2.91–3.66)***	16
Metastatic cancer	3.93 (3.65–4.22)***	27	4.04 (3.77–4.34)***	16	4.07 (3.79–4.36)***	19
Solid tumor without metastasis	2.77 (2.61–2.94)***	20	2.41 (2.28–2.56)***	10	2.31 (2.18–2.44)***	11
Rheumatoid arthritis/collagen	1.30 (1.18–1.44)***	5	1.29 (1.17–1.42)***	3	1.22 (1.12–1.34)***	3
Coagulopathy	2.58 (2.43–2.75)***	19	2.33 (2.20–2.47)***	10	2.22 (2.10–2.35)***	11
Obesity	0.58 (0.54–0.63)***	–11	0.59 (0.55–0.64)***	–6	0.60 (0.56–0.65)***	–7
Weight loss	2.68 (2.52–2.86)***	20	2.38 (2.24–2.52)***	10	2.28 (2.15–2.41)***	11
Fluid and electrolyte disorders	2.87 (2.77–2.96)***	21	2.78 (2.70–2.88)***	12	2.71 (2.62–2.80)***	13
Blood loss anemia	1.22 (1.07–1.39)**	4	1.16 (1.03–1.31)*	2	1.12 (1.00–1.26)*	2
Deficiency anemia	0.73 (0.65–0.82)***	–6	0.83 (0.75–0.91)***	–2	0.80 (0.73–0.87)***	–3
Alcohol abuse	0.91 (0.85–0.98)*	–2	E		E	
Drug abuse	0.61 (0.55–0.68)***	–10	0.59 (0.53–0.65)***	–6	0.59 (0.54–0.65)***	–7
Psychoses	0.85 (0.75–0.96)*	–3	0.84 (0.75–0.94)**	–2	0.85 (0.76–0.94)**	–2
Depression	0.70 (0.65–0.75)***	–7	0.68 (0.63–0.72)***	–5	0.65 (0.61–0.69)***	–6

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

Abbreviations: OR, odds ratio; CI, 95% confidence interval; E, eliminated from the model.

**Table 5** Predictive performance by mortality outcome, look back period, and comorbidity summary method, n = 1,635,726

Model	c-statistic (95% CI)		
	Index only	Index + 1-year look back	Index + 2-year look back
<b>In-hospital mortality</b>			
Baseline: age and sex	0.820 (0.818–0.822)	0.820 (0.818–0.822)	0.820 (0.818–0.822)
Baseline and binary method	0.887 (0.885–0.890) <sup>a</sup>	0.886 (0.883–0.888) <sup>a,b</sup>	0.884 (0.882–0.887) <sup>a,c</sup>
Baseline and total method	0.856 (0.853–0.859) <sup>a,d</sup>	0.854 (0.851–0.856) <sup>a,b,d</sup>	0.851 (0.849–0.854) <sup>a,d</sup>
Baseline and weighted method	0.886 (0.884–0.888) <sup>a,d,e</sup>	0.884 (0.882–0.885) <sup>a,b,d,e</sup>	0.883 (0.881–0.885) <sup>a,e</sup>
<b>In-hospital mortality at 1 year</b>			
Baseline: age and sex	0.826 (0.824–0.827)	0.826 (0.824–0.827)	0.826 (0.824–0.827)
Baseline and binary method	0.884 (0.882–0.886) <sup>a</sup>	0.886 (0.884–0.888) <sup>a,b</sup>	0.885 (0.883–0.887) <sup>a,c</sup>
Baseline and total method	0.861 (0.859–0.863) <sup>a,d</sup>	0.864 (0.862–0.866) <sup>a,b,d</sup>	0.863 (0.862–0.866) <sup>a,d</sup>
Baseline and weighted method	0.882 (0.880–0.884) <sup>a,d,e</sup>	0.883 (0.881–0.885) <sup>a,b,d,e</sup>	0.883 (0.881–0.885) <sup>a,e</sup>

Notes: c-statistic differences were evaluated using the non-parametric Mann–Whitney U statistical method for comparing ROC curves developed and described by DeLong et al.<sup>36</sup> <sup>a</sup>Reference = baseline: age and sex, statistically significant,  $p < 0.0001$ . <sup>b</sup>Reference = index only, statistically significant,  $p < 0.0001$ . <sup>c</sup>Reference = index + 1-year look back, statistically significant,  $p < 0.0001$ . <sup>d</sup>Reference = baseline and binary method, statistically significant,  $p < 0.0001$ . <sup>e</sup>Reference = baseline and total method, statistically significant,  $p < 0.0001$ .

Abbreviations: CI, 95% confidence interval; ROC, receiver operating characteristic.

and allow for cross-study comparisons.<sup>15</sup> However, alternative outcomes, such as hospital readmissions, have been shown to be better predicted by longer look back periods and would require additional validation.<sup>17</sup>

While implementing the VW point system, we observed an inverse relationship between several comorbidity groups and the mortality outcomes. This led to ECM conditions having negative weights. Notably, more than one-third of the

Quan ECM, including valvular disease, hypertension, diabetes with complications, hypothyroidism, peptic ulcer disease without bleeding, obesity, depression, psychoses, drug abuse, alcohol abuse, and deficiency anemia, had negative weights when predicting in-hospital mortality at the index with 1 year of look back. According to Elixhauser et al,<sup>1</sup> conditions negatively associated with mortality possibly reflect the fact that during acute encounters, conditions possibly considered irrelevant to the provision of care are less frequently recorded. In other words, conditions negatively associated with mortality may not be protective as much as they are the result of a combination of clinical relevance and inconsistent data collection. It could also be hypothesized that comorbidities such as drug and alcohol abuse, psychoses, and depression are less likely recorded in medical records because they are stigmatized conditions with a lower likelihood of voluntary disclosure by health care recipients.

The strengths of this study include its broad geographic coverage, numerous contributing health care facilities, and the large and diverse population of US health care recipients in the data source. This afforded the selection of a study sample with complete demographic profiles and available diagnostic information without compromising study power. While studies with enormous sample sizes are prone to being overpowered, the resulting resource utilization consequences are somewhat inconsequential and ethical concerns are irrelevant in administrative database research. The ability to ascertain diagnoses across multiple care settings in HF, i.e., outpatient clinics, ED visits, hospital care, is likely to have improved the accuracy and completeness of individual comorbidity profiles. The study design was based on random encounters over episodes of care. Episodes of care were initially defined by Solon et al<sup>38</sup> as the grouping of health services data from one or more health encounters related to the management of a principal health condition or problem. In the context of validating measures of comorbidity without selecting a primary health condition of interest, as was the case for this study and those to which it should be compared,<sup>22,39</sup> the episode of care approach not only increases complexity but also raises methodological concerns. For instance, in highly comorbid individuals, the aggregation of a larger number of health encounters into a single episode of care is likely to lead to increased opportunities for identifying prevalent health conditions, when compared to people with lower overall morbidity and narrower episodes of care. This situation could result in greater exposure misclassification based on the characteristics of the episode of care itself. The reader is referred to the seminal paper by Wingert et al<sup>40</sup> for

further discussion of the theoretical motives supporting the episodes of care methodology.

This study is also the first to validate the performance of the VW approach using the Quan ECM in a US data source. Our findings highlight the relevance of internal validation prior to selecting comorbidity risk-adjustment parameters. In HF, the choice of look back period is not as critical as the choice of comorbidity summary method compared to other data sources. While the total summary method should be avoided, the consequences of limiting ECM diagnostic ascertainment to the index encounter, the least complex and resource-intensive approach available, is not likely to practically affect the performance of the Quan ECM in HF. We also confirmed the practical equivalence of the weighted and binary methods for predicting in-hospital mortality outcomes. This highlights a real advantage for future epidemiological studies since weighted scores are much easier to manipulate, interpret, and could help avoid problems related to overfitting that arise during regression analyses when too many variables are included in a model.<sup>22</sup>

Limitations of this study included our inability to separate conditions present on admission from complications in care typically identified using diagnostic-type indicators. In HF, the diagnostic-type indicator was not consistently available, which prevented further investigation of this parameter. Thus, we could not test the hypothesis by Quan et al<sup>10</sup> that including conditions emerging from complications in care is likely to benefit predictions for long-term mortality but not necessarily in-hospital mortality. In HF, care recipients are tracked as they seek care from Cerner participating institutions. However, not all US health care institutions are Cerner clients and data contributors. Even in cases where patients received care from two Cerner health facilities, it would only be possible to track patients across time if these facilities were covered under the same privacy (HIPAA) network. Likewise, care recipients in HF may have the option of seeking health services from non-Cerner-affiliated health facilities. In such instances, the clinical encounter data would not be captured in HF and details about health services participation, including diagnoses and health outcomes, would be lost. These limitations are characteristic of observational research conducted with HIPAA-compliant administrative databases and necessitate consideration in experimental design. Notably, evidence of mortality had to be limited to deaths captured in Cerner hospitals since mortality status could not be verified against other sources of vital statistics through data linkage. Therefore, deaths were certainly underrepresented, and this likely introduced some biases. Another possible source of

confounding relates to how health care organizations elect to record the discharge outcome of people transferred to hospice care.<sup>41</sup> Findings from Kozar et al<sup>41</sup> based on data from US Trauma Quality Improvement Program centers suggest that transfers to hospice care assign a survivor status to patients and that in some cases, the more appropriate discharge status might be deceased. The authors also found that patients with two or more comorbidities were more likely to be coded as hospice care transfers than in-hospital deaths.

Results of this study must be contrasted against the possibility of lead time and retention biases, which could vary by condition, by condition latency or intermittency, by condition severity, by services rendered, by age, by health insurance status, and other factors.<sup>42,43</sup> People with certain conditions or comorbidities could theoretically expedite or delay seeking care and experience different lead times till their first encounter with an HIPAA-covered entity source institution in HF. Such factors could crucially affect whether one will or will not have, 1) antecedent encounters in the data warehouse, at 1 year or other look back intervals, and 2) subsequent encounters, at 1 year or other follow-up times, wherein mortality or comorbidities can be ascertained.

Often, comparative studies of comorbidity indices or measures are performed in a population defined by a specific primary diagnosis, e.g., cancer,<sup>44-46</sup> from which comorbid conditions are defined. This study is aligned with the seminal Quan et al<sup>10</sup> study and was conducted with an undifferentiated patient general population. For this reason and until evidence to the contrary is available, it is advised that VW-based weights be rederived for each study population based on a different primary diagnosis. We completed secondary database research in a data source primarily designed for documenting clinical practice, health services, and billing.<sup>23</sup> Hence, our results are subject to the typical biases and limitations accompanying administrative database research,<sup>47,48</sup> which include risks of missing clinical and demographic information, the misclassification of diagnoses during data abstraction, linkage errors related to interoperability issues, and record duplication.

Finally, we confirmed the excellent predictive performance of the Quan ECM for predicting in-hospital mortality and in-hospital mortality at 1 year in Cerner HF, a large longitudinal US multi-payer health database that complies with US HIPAA regulations and limitations. The performance of the Quan ECM expressed as binary indicators and the weighted scores were practically equivalent and both surpassed the total method. Differences in diagnostic ascertainment look back period up to 2 years were statistically but not clinically associated with significant differences in predictive performance for in-hospital mortality.

## Acknowledgments

This study was supported by the Fonds de recherche du Québec – Santé (FRQS), the McLaughlin Centre for Population Health Risk Assessment, and the University of Ottawa's admission and excellence scholarship programs. The authors would like to thank the Cerner Corporation for generously allowing their team to conduct research using the HF database. They also want to acknowledge the journal reviewers for their valuable comments and suggestions.

## Disclosure

YF, JAGC, and DC report no competing interests. DSM is the president of Cerner Math Inc. and has ownership interest in the Cerner Corporation. This paper reflects the opinions of the authors and not necessarily those of the Canadian Institute for Health Information or the Cerner Corporation. DRM serves as chief medical officer of Risk Sciences International (RSI), a company formed in partnership with the University of Ottawa in 2006. RSI undertakes risk assessment work for public and private sector clients in Canada and abroad. RSI has not conducted prior work on the subject of the present research paper. DK is chief risk scientist and CEO of RSI. Since 2002, DK has held a Natural Sciences and Engineering Research Council of Canada (NSERC) Industrial Research Chair in Risk Science, through a peer-reviewed university–industry partnerships program administered by NSERC. The scientific opinions and conclusions expressed in research publications by the chair are the responsibility of the chair alone. The authors report no other conflicts of interest in this work.

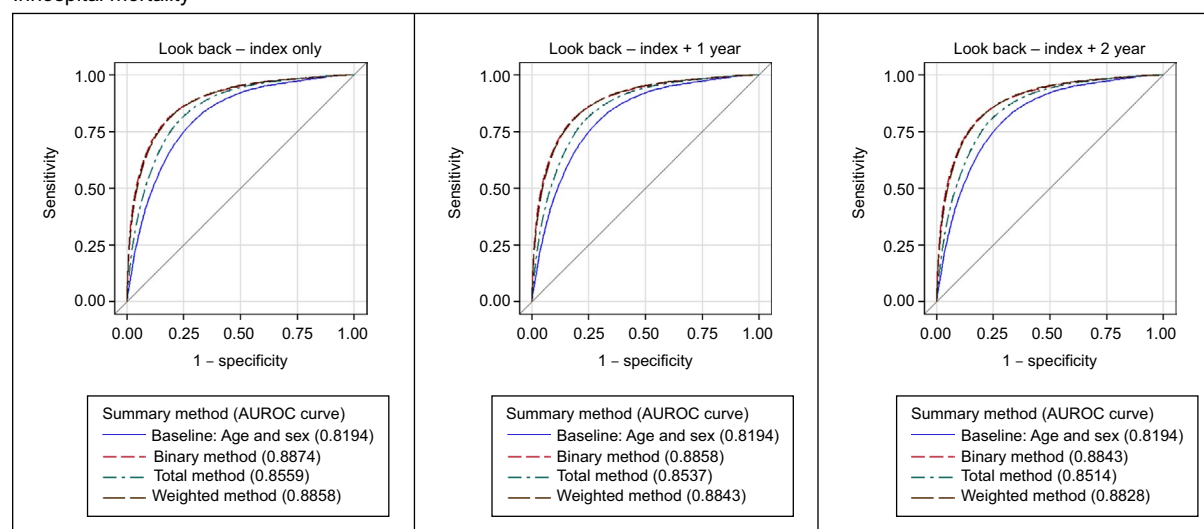
## References

1. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8–27.
2. Tadrous M, Gagne JJ, Stürmer T, Cadarette SM. Disease risk score as a confounder summary method: systematic review and recommendations. *Pharmacoepidemiol Drug Saf*. 2013;22(2):122–129.
3. Dominick KL, Dudley TK, Coffman CJ, Bosworth HB. Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. *Arthritis Care Res (Hoboken)*. 2005;53(5):666–672.
4. Li P, Kim MM, Doshi JA. Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Serv Res*. 2010;10(1):245.
5. Schneeweiss S, Wang PS, Avorn J, Maclure M, Levin R, Glynn RJ. Consistency of performance ranking of comorbidity adjustment scores in Canadian and US utilization data. *J Gen Intern Med*. 2004;19(5p1):444–450.
6. Stukenborg GJ, Wagner DP, Connors AF Jr. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. *Med Care*. 2001;39(7):727–739.
7. Charlson ME, Wells MT, Ullman R, King F, Shmukler C. The Charlson comorbidity index can be used prospectively to identify patients who will incur high future costs. *PLoS One*. 2014;9(12):e112479.

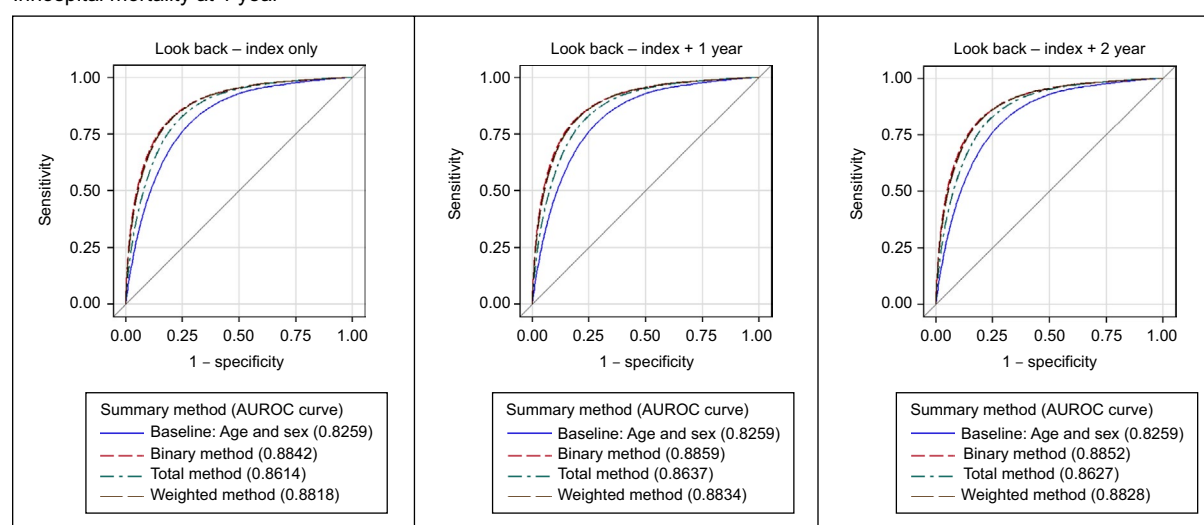
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383.
9. Chu Y-T, Ng Y-Y, Wu S-C. Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMC Health Serv Res*. 2010;10(1):140.
10. 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(11):1130–1139.
11. AHRQ [webpage on the Internet]. *Comorbidity Software, Version 3.7*. 2014. Available from: <http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp>. Accessed October 31, 2014.
12. Fortin Y, Crispo J, Emons M, Mattison D, Krewski D. Comparing the predictive performance of two variants of the Elixhauser comorbidity measures for all-cause in-hospital mortality in a large multi-payer US Administrative database. Paper presented at: Value in Health; 2015; Philadelphia, PA.
13. US Department of Health Human Services. *Summary of the HIPAA Privacy Rule*. U.S. Department of Health & Human Services. Washington; 2013.
14. Hughes JS, Iezzoni LI, Daley J, Greenberg L. How severity measures rate hospitalized patients. *J Gen Intern Med*. 1996;11(5):303–311.
15. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. *J Clin Epidemiol*. 2015;68(1):3–14.
16. Sharabiani MT, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Med Care*. 2012;50(12):1109–1118.
17. Preen DB, Holman CAJ, Spilsbury K, Semmens JB, Brameld KJ. Length of comorbidity lookback period affected regression model performance of administrative health data. *J Clin Epidemiol*. 2006;59(9):940–946.
18. D’Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Methods Inf Med*. 1993;32(5):382–387.
19. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613–619.
20. Romano PS, Roos LL, Jollis JG. Presentation adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol*. 1993;46(10):1075–1079.
21. Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol*. 1996;49(3):273–278.
22. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47(6):626–633.
23. B.R.I.D.G.E. to Data [webpage on the Internet]. *Cerner Health Facts® Database (USA)*. 2014. Available from: <http://www.bridgetodata.org/node/1789>. Accessed October 28, 2014.
24. Barrett M, Lopez-Gonzalez L, Hines A, Andrews R, Jiang J [webpage on the Internet]. An examination of expected payer coding in HCUP databases. In: U.S. Agency for Healthcare Research and Quality, editor. *HCUP Methods Series Report # 2014–03*. 2014. Available from: <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>. Accessed October 7, 2015.
25. Fortin Y, Crispo J, Emons M, Mattison D, Krewski D. PRM20-comparing the predictive performance of two variants of the Elixhauser comorbidity measures for all-cause in-hospital mortality in a large multi-payer US administrative database. *Value Health*. 2015;18(3):A13.
26. CDC. Classification of Diseases, Functioning, and Disability: International Classification of Diseases, Ninth Revision (ICD-9). Atlanta, GA: Centers for Disease Control and Prevention; 2009.
27. Manitoba Centre for Health Policy [webpage on the Internet]. *Concept: Complications and Comorbidities*. 2010. Available from: [http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1099#a\\_sas](http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1099#a_sas). Accessed October 3, 2015.
28. Kurichi JE, Stineman MG, Kwong PL, Bates BE, Reker DM. Assessing and using comorbidity measures in elderly veterans with lower extremity amputations. *Gerontology*. 2007;53(5):255–259.
29. Baldwin L-M, Klabunde CN, Green P, Barlow W, Wright G. In search of the perfect comorbidity measure for use with administrative claims data: does it exist? *Med Care*. 2006;44(8):745–753.
30. Sullivan LM, Massaro JM, D’Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631–1660.
31. Raymond EG, Grossman D, Weaver MA, Toti S, Winikoff B. Mortality of induced abortion, other outpatient surgical procedures and common activities in the United States. *Contraception*. 2014;90(5):476–479.
32. Keyes GR, Singer R, Iverson RE, et al. Mortality in outpatient surgery. *Plast Reconstr Surg*. 2008;122(1):245–250.
33. Zhu H, Hill MD. Stroke the Elixhauser Index for comorbidity adjustment of in-hospital case fatality. *Neurology*. 2008;71(4):283–287.
34. Downer RG, Richardson P. Illustrative Logistic Regression Examples Using PROC LOGISTIC: New Features in SAS/STAT® 9.2. Paper SP03-2009. 2009.
35. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol. 398. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.
37. Thompson NR, Fan Y, Dalton JE, et al. A new Elixhauser-based comorbidity summary measure to predict in-hospital mortality. *Med Care*. 2015;53(4):374–379.
38. Solon JA, Feeney JJ, Jones SH, Rigg RD, Sheps CG. Delineating episodes of medical care. *Am J Public Health Nations Health*. 1967;57(3):401–408.
39. Quan H, Li B, Duncan Saunders L, 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(4):1424–1441.
40. Wingert TD, Kralewski JE, Lindquist TJ, Knutson DJ. Constructing episodes of care from encounter and claims data: some methodological issues. *Inquiry*. 1995;32(4):430–443.
41. Kozar RA, Holcomb JB, Xiong W, Nathens AB. Are all deaths recorded equally? The impact of hospice care on risk-adjusted mortality. *J Trauma Acute Care Surg*. 2014;76(3):634–641.
42. Jensen ET, Cook SF, Allen JK, et al. Enrollment factors and bias of disease prevalence estimates in administrative claims data. *Ann Epidemiol*. 2015;25(7):519–525.e2.
43. Tunnell R, Millar B, Smith G. The effect of lead time bias on severity of illness scoring, mortality prediction and standardised mortality ratio in intensive care – a pilot study. *Anaesthesia*. 1998;53(11):1045–1053.
44. Klabunde CN, Legler JM, Warren JL, Baldwin L-M, Schrag D. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol*. 2007;17(8):584–590.
45. Hollenbeak CS, Stack BC, Daley SM, Piccirillo JF. Using comorbidity indexes to predict costs for head and neck cancer. *Arch Otolaryngol Head Neck Surg*. 2007;133(1):24–27.
46. Khan NF, Perera R, Harper S, Rose PW. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract*. 2010;11(1):1.
47. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol*. 2011;65(2):126–131.
48. Esposito D, Migliaccio-Walle K, Molsen E. Reliability and Validity of Data Sources for Outcomes Research & Disease and Health Management Programs. Lawrenceville, NJ: ISPOR; 2013:467.

## Supplementary materials

### Inhospital mortality



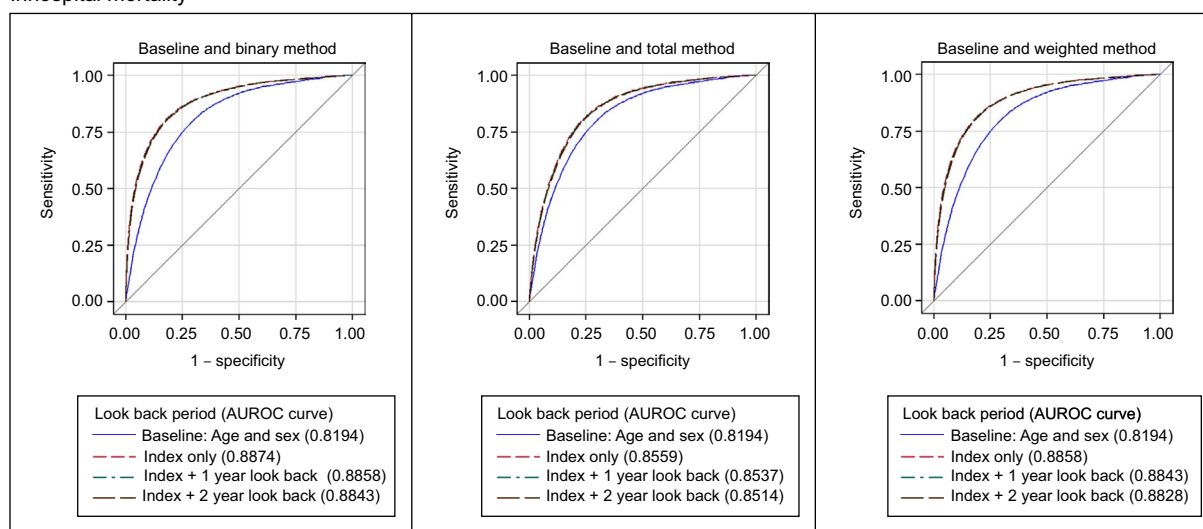
### Inhospital mortality at 1 year



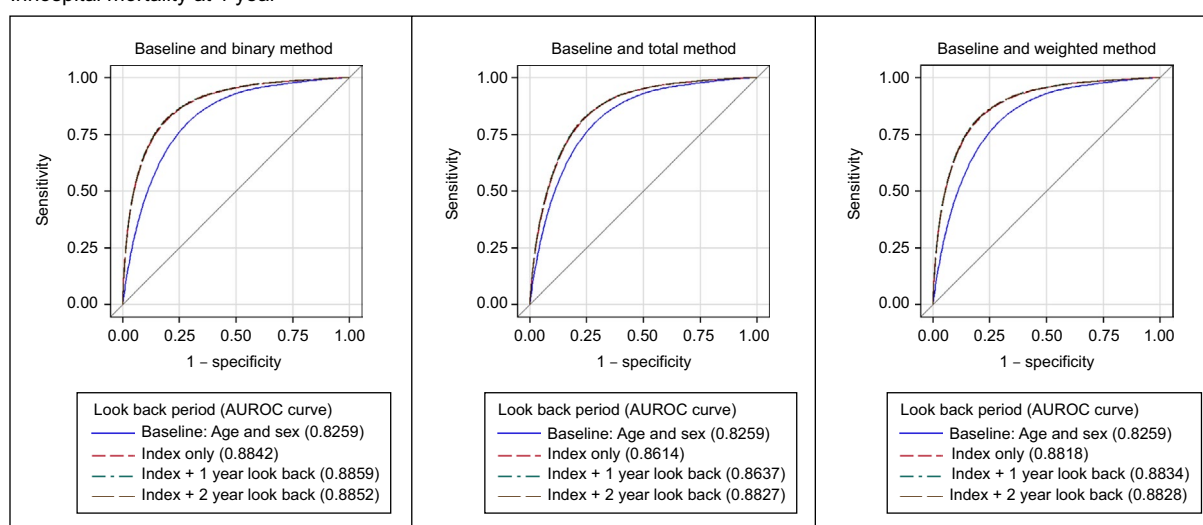
**Figure S1** Comparisons of ROC curves across comorbidity summary methods by mortality outcome and look back period.

**Abbreviations:** ROC, receiver operating characteristic; AUROC, area under the receiver operating characteristic.

## Inhospital mortality



## Inhospital mortality at 1 year



**Figure S2** Comparisons of ROC curves across look back periods by mortality outcome and comorbidity summary method.

**Abbreviations:** ROC, receiver operating characteristic; AUROC, area under the receiver operating characteristic.

## Open Access Medical Statistics

## Publish your work in this journal

Open Access Medical Statistics is an international, peer-reviewed, open access journal publishing original research, reports, reviews and commentaries on all areas of medical statistics. The manuscript management system is completely online and includes a very quick and fair

peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/open-access-medical-statistics-journal>

Dovepress