Validation of the Combined Comorbidity Index of Charlson and Elixhauser to Predict 30-Day Mortality Across ICD-9 and ICD-10

Marc Simard, MSc, Caroline Sirois, PhD,*†‡ and Bernard Candas, PhD†§*

Objectives: To validate and compare performance of an International Classification of Diseases, tenth revision (ICD-10) version of a combined comorbidity index merging conditions of Charlson and Elixhauser measures against individual measures in the prediction of 30-day mortality. To select a weight derivation method providing optimal performance across ICD-9 and ICD-10 coding systems.

Research Design: Using 2 adult population-based cohorts of patients with hospital admissions in ICD-9 (2005, n = 337,367) and ICD-10 (2011, n = 348,820), we validated a combined comorbidity index by predicting 30-day mortality with logistic regression. To appreciate performance of the Combined index and both individual measures, factors impacting indices performance such as population characteristics and weight derivation methods were accounted for. We applied 3 scoring methods (Van Walraven, Schneeweiss, and Charlson) and determined which provides best predictive values.

Results: Combined index [*c*-statistics: 0.853 (95% confidence interval: CI, 0.848–0.856)] performed better than original Charlson [0.841 (95% CI, 0.835–0.844)] or Elixhauser [0.841 (95% CI, 0.837–0.844)] measures on ICD-10 cohort. All weight derivation methods provided close high discrimination results for the Combined index (Van Walraven: 0.852, Schneeweiss: 0.851, Charlson: 0.849). Results were consistent across both coding systems.

Conclusions: The Combined index remains valid with both ICD-9 and ICD-10 coding systems and the 3 weight derivation methods evaluated provided consistent high performance across those coding systems.

Key Words: comorbidity adjustment, administrative data, ICD-9, ICD-10, chronic diseases

(*Med Care* 2018;00: 000–000)

From the *Quebec National Institute of Public Health; †Department of Social and Preventive Medicine, Faculty of Medicine, Laval University; ‡Centre of Excellence on Aging of Quebec, Integrated University Health and Social Services Centres of the Capitale-Nationale; and §National Institute of Excellence in Health and Social Services, Québec, QC, Canada.

The authors declare no conflict of interest.

Reprints: Marc Simard, MSc, Quebec National Institute of Public Health, 945, Wolfe, 5e étage, Québec, QC, Canada G1V 5B3. E-mail: marc.simard@inspq.qc.ca.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

ISSN: 0025-7079/18/000-000

Comorbidity indices using administrative database records are a valid approach to account for individuals' overall burden of illnesses in transversal and longitudinal observational studies in order to adjust for case mix or confounders.[1]

Charlson and Elihxauser are the most widely used comorbidity indices and have long proven usefulness in numerous studies relying on administrative data.[2] More recently, Gagne et al[3] suggested combining the Charlson's index[4] and the Elixhauser's measure of comorbidities[5] into a single index that outperformed both measures individually. Gagne et al's Combined index accounted for an extended list of medical conditions coded with the ninth revision of the International Classification of Diseases (ICD-9). Individuals' scores of comorbidity are made of the summation of condition-specific weights for every chronic condition identified in the person.

Several factors influence an index performance such as the population characteristics and the weights' derivation method. Moreover, the age range, time periods, and source population will lead to differences in weight values and performance.[6,7] Weight derivation methods, that is, the mathematical methods used to derive the impact of a medical condition, are multiple and they attributed different weights to each medical condition. These weights, in turn, influence/determine indices performance.[8]

The Combined index performance assessed by Gagne et al[3] has limitations. First, the 3 set of weights (Charlson, Elixhauser, and Combined) used in the validation study were derived from 3 distinct populations with respect to age, period, and data source. Furthermore, weight derivation methods applied for each index were different. It is therefore difficult to assess the real improvement brought by the combined comorbidity index compared with the other indices because it cannot be isolated from the possible impact of population characteristics and weight derivation methods.

The Combined index, unlike Charlson and Elixhauser measures individually,[9,10] has not yet be defined with the tenth version of the ICD that is growingly implemented. It is imperative to study the ICD-9 and ICD-10-based versions of a Combined index to ensure their validity over periods of time where both coding standards are used.

Hence, in order to provide further extend the validity of a Combined index, the specific objectives of the present study is (1) to validate and assess the performance of a Combined index on a single population, (2) to compare the performance of the Combined index when derived from either ICD-9 or

the ICD-10 coded diagnoses, and (3) to select a weight derivation method that provides high performance across the ICD-9 and ICD-10 coding systems.

METHODS

Study Population

We performed a population-based study of the adult population in the province of Quebec, Canada, using hospital discharge records extracted from the Quebec Integrated Chronic Disease Surveillance System (QICDSS).[11] QICDSS links patient level records of the provincial health system administrative databases that include, among others, the hospital discharge database, the provincial death registry, and the demographic records. Quebec has universal access health care services with free-of-charge hospital stays which allows the QICDSS to cover 98% of the provincial adult population. Diagnoses at admission and established during the stay are coded using an ICD-9 derived classification (ICD-9-QC) up to March 31, 2006, and the ICD-10 Canadian coding standard (ICD-10-CA) thereafter. Discharge abstracts include the admission diagnosis, primary diagnosis, in-hospital cause of death diagnoses (when relevant), as well as up to 15 (ICD-9) and 25 (ICD-10) secondary diagnoses.

Two cohorts were extracted from the hospital discharge database ICD-9 and ICD-10 eras, respectively. We selected the first hospital admission of patients aged 18 years and older between April 1, 2005, and March 31, 2006 (cohort 2005/2006, using ICD-9-QC codes), and between April 1, 2011, and March 31, 2012 (cohort 2011/2012, using ICD-10-CA codes). Maternity-related admissions and same-day surgeries (mostly minor surgeries) were excluded to maintain comparability with other studies.[4,5]

Identification of Medical Conditions

Diagnosis codes for the 17 medical conditions of the Charlson index[4] and the 30 conditions included in the Elixhauser measures of morbidity,[5] were retrieved from Quan work's[12] and adapted to the Quebec-specific context. A 3-step process was performed. First we translated ICD-9 clinical modification (ICD-9-CM) codes from Quan[12] into ICD-9-QC manually to generate our initial list. When no exact match existed, a senior registrar and a medical specialist in the relevant medical field were involved. The medical specialist had 3 options: to approve the code identified by the senior registrar, to select another code with a close definition, or not to retain any code by lack of appropriate correspondence. Second, the list of ICD-9-QC codes resulting from the above-described process was matched to the ICD-10-CA codes from Quan et al.[12] Consistency between ICD-9-QC and ICD-10-CA code definitions was reviewed and minor adjustments performed when needed, thus creating a subsequent list of ICD-9-QC codes with their ICD-10-CA counterparts. Third, we translated the ICD-10-CA codes back into ICD-9-QC codes. Discrepancies between the back-translated and original ICD-9-QC codes were resolved with the assistance of a registrar and a medical specialist as described above. All of the ICD-10-CA codes from the included algorithms are consistent with ICD-10-CM codes, and thus we use ICD-10 term throughout the text. ICD-9-QC and ICD-9-CM differences are rare and illustrated in supplemental digital content (Tables, Supplemental Digital Content 1, http://links.lww.com/MLR/B543, medical conditions definitions, and code). We use ICD-9 term throughout the text.

Combined Index

Both lists of the 17 Charlson[4] and the 30 Elixhauser[5] medical conditions were later combined to create a broader list of 32 conditions. When a code of a medical condition did not match exactly in the 2 lists, the most inclusive was always preferred.[3] As in most recent studies using the Elixhauser measures, hypertension with and without complications have been merged into a single medical condition.[3,8,13] Because we considered hemiplegia and paraplegia as one condition,[4,5] we obtained one fewer comorbidity in our combined list than Gagne et al.[3]

Statistical Analyses

Both cohorts [n = 337,367 in the ICD-9(2005/2006) cohort and n = 348,820 in the ICD-10(2011/2012) cohort] were randomly divided into a derivation and a validation sample of equal size. Derivation samples were used to estimate weights for each medical condition, using 3 lists of medical conditions (Charlson, Elixhauser, Combined) and 3 weight derivation methods. Validation samples were used to compare the predictive performance of each list and of each weight derivation method.

Derivation of Weight

We applied 3 weight derivation methods [Charlson (CH),[4] Schneeweiss (SCH),[14] Van Walraven (VW)[13]] for each list of medical conditions in both derivation samples (ICD-9 and ICD-10). All 3 methods apply different sets of rules to derive the weights from the coefficients of multivariate logistic regressions predicting death within 30 days of hospital admission. Thirty-day mortality is a measure that reduce bias due to length of stay differences across hospitals or time.[7] Independent variables included sex, 5-year age categories, and a binary variable for every medical condition (0 = condition not identified, 1 = condition documented in the hospital discharge record). The CH method rounds up odds ratios >1.2 to the nearest integer for each statistically significant medical condition with a maximal value of 6.[4] Nonsignificant medical conditions receives a weight of 0. The SCH method is based on a regression model including all independent variables. It assigns weights by rounding up, to the nearest integer, the result of the division of the coefficient by 0.30.[14] Medical conditions with coefficients <0.30 in absolute value have a weight of 0. Therefore, a weight of 1 refers to a 35% increase (ie, $\exp^{0.30}$) in the odds of dying within 30 days of the admission. Finally, the VW method uses a backward stepwise elimination process on medical condition that are defined as binary variables to select statistically significant coefficients ($P < 0.05$).[13] Each weight is obtained by dividing the estimated coefficients by the smallest absolute value among all medical condition coefficients, rounded up to the nearest integer. The nonsignificant medical conditions receives a weight of 0.

Performance Assessment

Each set of weights was applied to calculate a comorbidity score for all patients of the validation samples by summing up weights attributed for every existing condition identified in each individual patient. A total of 9 comorbidity scores were available for each patient since 3 lists of medical conditions were available for each patient (Charlson, Elixhauser, Combined), each being associated to 3 set of weights (CH, SCH, VW). The same procedure was applied to the patients form the ICD-9 and ICD-10 validation samples. For each list of medical conditions (from Elixhauser, Charlson, and Combined index), we performed 6 multivariate logistic regression models predicting 30-day mortality in the ICD-9 and ICD-10 validation samples. The baseline model included age and sex. The other 5 models included the following predictors added to the baseline model: (1) CH score; (2) SCH score; (3) VW score; (4) binary variables for each medical condition of the list (BIN); (5) total count of prevalent medical conditions (COUNT). The 2 last predictors (BIN, COUNT) were included in our analyses to provide upper and lower bounds of the predictive capacity of logistic regressions, respectively. Indeed, introducing every medical condition as a binary variable in the equation (BIN method) provides the best possible prediction using logistic regressions on validation samples.[5] In contrast, replacing the score with a simple count of the medical conditions identified in every patient should result in lower predictive capacity (COUNT method).[8]

We calculated 4 performance measures for each logistic regression. Performance measures were also calculated for the 2 reference models (COUNT and BIN methods) to allow for comparisons. Overall performance is provided by the scaled Brier score which value ranges from 0 to 1; higher value indicates better performance.[15] Discrimination, the ability to identify rightfully patients who died within 30 days, is assessed by the area under the receiver operating characteristic curve (AUC).[15] We considered an AUC < 0.70 to represent poor discrimination; 0.70–0.79, acceptable discrimination; 0.80–0.89, excellent discrimination; and ≥ 0.90, outstanding discrimination.[15] Calibration curves provide insight on the level of agreement between observed and predicted probability of the outcome and is appraised visually on plots of observed versus predicted outcomes by quintiles of mortality risk. Data depart from the diagonal as prediction deteriorates.[15] Net reclassification improvement (NRI), evaluates the improvement provided by adding comorbidity index to the baseline logistic regression model that only accounts for age and sex. Higher NRI values indicate better reclassification.[15] Reclassification of subjects with and without 30-day mortality are 2 components of NRI. Higher improvement in proportion of subjects with 30-day mortality who are well classified by adding overall medical condition score (mortality percent improvement) and higher improvement in proportion of subjects without 30-day mortality who are well classified by adding overall medical condition (nonmortality percent improvement) imply a high NRI.

We used bootstrap resampling (200 samples with replacement) to estimate 95% confidence intervals for all performance measures and to compare AUC. Statistical tests were 2-sided with significance level of $P < 0.05$. Analyses were performed using SAS 9.3 (SAS Institute, Cary, NC).

Sensitivity and Additional Analysis

We performed a sensitivity analysis stratifying performance assessment by age (below or above 65 y) and performed an additional analysis using 1-year mortality as the outcome.

Ethics

The use of QICDSS for research and development in population health surveillance has been approved by the custodians of the databases, the provincial Public Health Research Ethics Board and the Quebec Commission protecting privacy and access to information (Commission d'accès à l'information du Québec).

RESULTS

Sociodemographic characteristics of the 2005/2006 and 2011/2012 cohorts are presented in Table 1. As expected from the aging of the population between 2005/2006 and

TABLE 1. Characteristics of ICD-9 (2005/2006) and ICD-10 (2011/2012) Cohorts

	ICD-9 Cohort			ICD-10 Cohort		
Characteristics (%)	**Derivation Sample**	**Validation Sample**	**Total**	**Derivation Sample**	**Validation Sample**	**Total**
N	168,684	168,683	337,367	174,410	174,410	348,820
Age [mean (SD)] (y)	61.2 (18.5)	61.3 (18.5)	61.2 (18.5)	63.4 (18.4)	63.4 (18.4)	63.4 (18.4)
Sex						
Male	48.1	48.1	48.1	48.2	48.2	48.2
Material deprivation						
First quintile—least deprived	15.8	15.7	15.7	15.8	15.8	15.8
Second quintile	17.8	17.9	17.8	18.2	18.4	18.3
Third quintile	20.1	20.2	20.2	20.1	19.8	19.9
Fourth quintile	21.9	21.8	21.9	22.1	22.1	22.1
Fifth quintile—most deprived	24.0	23.9	24.0	23.7	23.8	23.7
30-day mortality	4.5	4.6	4.5	4.8	4.8	4.8
365-day mortality	12.3	12.4	12.3	13.4	13.4	13.4
Length of stay [median (Q1–Q3)] (d)	4 (2–9)	4 (2–9)	4 (2–9)	4 (2–9)	4 (2–9)	4 (2–9)

ICD-10 indicates International Classification of Diseases tenth edition; ICD-9, International Classification of Diseases ninth edition.

2011/2012, patients are older in the latter cohort (mean age, 61.2 vs. 63.4 y). Accordingly, 30-day and 365-day mortality rates are slightly higher in the 2011/2012 cohort (4.5% vs. 4.8% and 12.3% vs. 13.4%, respectively). The median lengths of hospital stay (4 d; interquartile range, 2–9) are identical and differences in the material deprivation quintiles are minor.

Derivation of Weight

The prevalence of the 32 medical conditions accounted for in the Combined index and weights issued from derivation samples are presented in Table 2 for both ICD-9 and ICD-10 cohorts (Tables, Supplemental Digital Content 1, http://links.lww.com/MLR/B543, medical conditions definitions, and code). Weights obtained from CH and SCH derivation methods are very similar, both for the ICD-9 and ICD-10 cohorts. In contrast, the VW weights span over a much larger range of values although the relative weights of medical conditions remain compatible across both cohorts.

Performance Assessment

Performance of the 3 weight derivation methods and of COUNT and BIN logistic regressions applied to the Combined index is presented in Table 3 for the ICD-9 (2005/2006) validation sample and in Table 4 for the ICD-10 (2011/2012) validation sample. Performance results obtained with the Combined index were better than those obtained when using the Elixhauser measure or Charlson index. The Combined index provided higher overall performance (Scaled Brier score) and discrimination (AUC) on all scoring methods from both the ICD-9 and ICD-10 samples (Tables 3, 4). Compared with the baseline model that does not include any medical condition, weight-based methods (VW, SCH, CH) increased predicted probabilities of dying within 30 days in 62%–64% of patients and decreased the predicted probabilities in

TABLE 2. Combined List of Medical Conditions, Prevalence, and Weights According to the VW, SCH, and CH Derivation Weight Methods Obtained in ICD-9 (2005/2006) and ICD-10 (2011/2012) Derivation Sample to Predict 30-day Mortality

	Prevalence (%)		Adjusted Odds Ratio (95% CI)*		CH Weights		SCH Weights		VW Weights	
Comorbidities	**ICD-9**	**ICD-10**	**ICD-9**	**ICD-10**	**ICD-9**	**ICD-10**	**ICD-9**	**ICD-10**	**ICD-9**	**ICD-10**
Hypertension†	36.1	41.7	0.70 (0.68–0.72)	0.75 (0.73–0.77)	0	0	−1	−1	−2	−3
Chronic pulmonary disease‡	16.1	17.4	1.26 (1.23–1.30)	1.24 (1.21–1.28)	1	1	1	1	2	2
Cardiac arrhythmias†	13.4	16.5	1.17 (1.13–1.21)	1.22 (1.18–1.25)	0	1	1	1	1	2
Diabetes, uncomplicated‡	13.9	15.7	1.04 (1.01–1.08)	1.06 (1.03–1.09)	0	0	0	0	0	0
Deficiency anemia†	9.1	13.2	0.85 (0.82–0.88)	0.83 (0.81–0.86)	0	0	−1	−1	−1	−2
Any tumor without metastasis§‖	12.0	12.8	2.44 (2.35–2.53)	2.57 (2.47–2.66)	2	3	3	3	6	11
Hypothyroidism†	10.6	12.5	0.83 (0.79–0.86)	0.92 (0.89–0.95)	0	0	−1	0	−1	−1
Renal disease§¶	8.6	10.3	1.42 (1.37–1.47)	1.30 (1.26–1.35)	1	1	1	1	2	3
Fluid and electrolyte disorders†	5.3	8.6	1.93 (1.86–2.01)	1.75 (1.69–1.81)	2	2	2	2	5	6
Peripheral vascular disorders†	6.6	8.0	1.48 (1.42–1.54)	1.17 (1.13–1.21)	1	0	1	1	3	2
Myocardial infarction§	8.4	7.8	2.03 (1.96–2.10)	1.89 (1.82–1.96)	2	2	2	2	5	7
Congestive heart failure‡	6.5	6.7	2.03 (1.95–2.11)	1.87 (1.81–1.95)	2	2	2	2	5	7
Obesity†	5.2	6.2	0.99 (0.92–1.05)	0.88 (0.83–0.94)	0	0	0	0	0	0
Valvular disease‡	5.1	5.9	0.71 (0.67–0.75)	0.80 (0.76–0.83)	0	0	−1	−1	−2	−3
Metastatic cancer‡	5.0	5.4	5.38 (5.15–5.63)	5.15 (4.93–5.37)	5	5	6	5	12	19
Dementia§	3.7	5.4	1.69 (1.62–1.76)	1.89 (1.82–1.96)	2	2	2	2	4	7
Cerebrovascular disease§	5.9	5.3	2.12 (2.03–2.20)	2.35 (2.26–2.45)	2	2	2	3	5	10
Depression†	5.4	5.2	0.65 (0.60–0.70)	0.66 (0.62–0.71)	0	0	−1	−1	−3	−5
Neurological disorders†	4.3	4.8	2.00 (1.91–2.10)	1.72 (1.64–1.79)	2	2	2	2	5	6
Alcohol abuse†	4.1	4.0	1.18 (1.10–1.27)	1.29 (1.21–1.37)	0	1	1	1	1	3
Liver disease†	3.1	3.9	1.65 (1.54–1.76)	2.28 (2.17–2.39)	2	2	2	3	3	9
Diabetes, complicated‡	2.7	3.8	0.87 (0.81–0.94)	0.83 (0.78–0.88)	0	0	0	−1	−1	−2
Psychoses†	3.3	2.7	1.34 (1.25–1.44)	0.94 (0.84–1.05)	1	0	1	0	2	0
Pulmonary circulation disorders†	1.9	2.6	2.21 (1.99–2.25)	1.47 (1.39–1.55)	2	1	2	1	5	4
Rheumatoid arthritis/collagen vascular disease†	2.1	2.4	1.31 (1.21–1.41)	0.98 (0.91–1.06)	1	0	1	0	2	0
Coagulopathy†	1.6	2.3	2.18 (2.03–2.34)	1.86 (1.76–1.97)	2	2	3	2	5	7
Weight loss†	1.5	2.3	2.19 (2.06–2.33)	1.47 (1.39–1.54)	2	1	3	1	5	4
Drug abuse†	2.3	2.3	0.73 (0.62–0.85)	0.81 (0.70–0.94)	0	0	−1	−1	−2	0
Paralysis‡	1.7	1.5	1.39 (1.29–1.50)	1.40 (1.30–1.51)	1	1	1	1	2	4
Blood loss anemia†	1.6	1.3	0.76 (0.69–0.84)	0.69 (0.92–0.77)	0	0	−1	−1	−2	−4
Ulcer disease§	1.1	1.0	0.85 (0.76–0.96)	1.12 (1.00–1.25)	0	0	−1	0	0	0
AIDS/HIV‡	0.1	0.1	7.79 (5.84–10.37)	2.68 (2.00–3.59)	6	3	7	3	14	11

*In addition to each comorbidities, the multivariate logistic regression model to predict 30-day mortality included age and sex.
†Elixhauser definition used.
‡Charlson and Elixhauser definition are identical.
§Charlson definition used.
‖Includes Elixhauser "solid tumor without metastasis" and "lymphoma."
¶Includes Charlson "mild liver diseases" and "moderate or severe liver disease."
AIDS indicates acquired immunodeficiency syndrome; CH, Charlson; CI, confidence interval; HIV, human immunodeficiency virus; ICD-10, International Classification of Diseases tenth edition; ICD-9, International Classification of Diseases ninth edition; SCH, Schneeweiss; VW, Van Walraven.

TABLE 3. Performance Measures of Elixhauser, Charlson, and Combined Index by Weight Derivation Method Within the ICD-9 (2005/2006) Validation Sample

	Discrimination	Overall Performance	Reclassification Improvement		
Model	**AUC (95% CI)**	**Scaled Brier Score (95% CI)**	**NRI (95% CI)**	**Mortalitiy Percent Improvement**	**Nonmortality Percent Improvement**
Baseline*	0.742 (0.737–0.746)	0.037 (0.035–0.039)			
Elixhauser					
COUNT	0.782 (0.778–0.786)	0.048 (0.045–0.050)	0.44 (0.41–0.48)	57.5	64.8
CH	0.844 (0.844–0.848)	0.097 (0.092–0.103)	0.77 (0.73–0.82)	57.2	81.4
SCH	0.847 (0.842–0.851)	0.100 (0.094–0.105)	0.82 (0.78–0.85)	60.4	80.4
VW	0.848 (0.844–0.852)	0.102 (0.096–0.108)	0.82 (0.78–0.86)	60.8	80.3
BIN	0.849 (0.845–0.853)	0.103 (0.097–0.109)	0.83 (0.79–0.86)	61.4	80.1
Charlson					
COUNT	0.813 (0.809–0.817)	0.059 (0.056–0.062)	0.68 (0.63–0.72)	64.1	70.0
CH	0.845 (0.840–0.848)	0.092 (0.087–0.097)	0.83 (0.78–0.86)	60.7	80.6
SCH	0.845 (0.840–0.848)	0.094 (0.089–0.099)	0.82 (0.78–0.86)	60.9	80.3
VW	0.844 (0.839–0.848)	0.095 (0.089–0.100)	0.83 (0.78–0.87)	60.1	81.2
BIN	0.846 (0.841–0.849)	0.095 (0.089–0.101)	0.82 (0.78–0.86)	59.8	81.1
Combined					
COUNT	0.789 (0.784–0.793)	0.051 (0.048–0.054)	0.52 (0.48–0.55)	60.5	65.3
CH	0.857 (0.853–0.861)	0.099 (0.093–0.104)	0.89 (0.85–0.92)	65.1	79.4
SCH	0.858 (0.854–0.862)	0.107 (0.101–0.112)	0.88 (0.84–0.91)	63.3	80.6
VW	0.860 (0.856–0.864)	0.105 (0.099–0.111)	0.89 (0.85–0.92)	64.4	79.9
BIN	0.861 (0.858–0.865)	0.107 (0.100–0.113)	0.89 (0.85–0.92)	64.4	79.9

*Baseline model includes age and sex.

AUC indicates area under the receiver operating characteristic curve; BIN, one binary variable by comorbidity; CH, Weighted summary score method as used by Charlson; CI, confidence interval; COUNT, number of comorbidities; NRI, net reclassification improvement; SCH, weighted summary score method as used by Schneeweiss; VW, weighted summary score method as used by Van Walraven.

79%–80% of those who did not die, resulting in highest NRIs. Overall, the 3 weight derivation methods provided much improvement over the use of a simple count of medical conditions. However, the VW approach reached a performance that is closest to a logistic regression accounting for every medical condition (BIN).

Calibration was good for each model (Figures, Supplemental Digital Content 2, http://links.lww.com/MLR/B544,

TABLE 4. Performance Measures of Elixhauser, Charlson, and Combined Index by Weight Derivation Method Within the ICD-10 (2011/2012) Validation Sample

	Discrimination	Overall Performance	Reclassification Improvement		
Model	**AUC (95% CI)**	**Scaled Brier Score (95% CI)**	**NRI (95% CI)**	**Mortalitiy Percent Improvement**	**Nonmortality Percent Improvement**
Baseline*	0.735 (0.730–0.739)	0.036 (0.034–0.039)			
Elixhauser					
COUNT	0.773 (0.769–0.777)	0.044 (0.042–0.047)	0.47 (0.43–0.51)	59.2	64.3
CH	0.838 (0.835–0.841)	0.087 (0.082–0.092)	0.76 (0.72–0.80)	58.6	79.4
SCH	0.839 (0.835–0.842)	0.094 (0.089–0.099)	0.80 (0.77–0.84)	59.5	80.5
VW	0.841 (0.837–0.844)	0.095 (0.090–0.100)	0.80 (0.77–0.84)	59.6	80.5
BIN	0.841 (0.837–0.844)	0.096 (0.090–0.101)	0.81 (0.78–0.85)	60.1	80.5
Charlson					
COUNT	0.805 (0.801–0.809)	0.058 (0.054–0.062)	0.64 (0.58–0.68)	61.9	70.0
CH	0.839 (0.834–0.843)	0.097 (0.092–0.102)	0.79 (0.76–0.83)	59.2	80.7
SCH	0.839 (0.835–0.843)	0.098 (0.092–0.102)	0.79 (0.75–0.83)	57.1	82.4
VW	0.841 (0.841–0.844)	0.098 (0.092–0.102)	0.81 (0.77–0.84)	59.8	80.8
BIN	0.841 (0.835–0.844)	0.099 (0.093–0.103)	0.80 (0.76–0.84)	58.6	81.5
Combined					
COUNT	0.780 (0.779–0.784)	0.047 (0.041–0.050)	0.44 (0.40–0.48)	56.6	65.7
CH	0.849 (0.845–0.853)	0.094 (0.088–0.099)	0.81 (0.77–0.85)	61.9	78.6
SCH	0.851 (0.847–0.854)	0.098 (0.092–0.103)	0.84 (0.80–0.87)	62.7	79.1
VW	0.852 (0.847–0.855)	0.100 (0.095–0.106)	0.86 (0.82–0.89)	63.1	79.9
Bin	0.853 (0.848–0.856)	0.102 (0.095–0.107)	0.87 (0.83–0.90)	63.7	79.9

*Baseline model includes age and sex.

AUC indicates area under the receiver operating characteristic curve; BIN, one binary variable by comorbidity; CH, weighted summary score method as used by Charlson; CI, confidence interval; COUNT, number of comorbidities; NRI, net reclassification improvement; SCH, weighted summary score method as used by Schneeweiss; VW, weighted summary score method as used by Van Walraven.

calibration plots). For predicted probabilities <0.3 (ie, 98% of predicted probabilities) the smooth curve superposed the 45-degree dash line which means that predicted probabilities of mortality at 30 days were on average similar to observed probabilities. Miscalibration was essentially limited to about 2% of the patients with predicted probabilities of dying of ≥30% and was more pronounced with the COUNT method.

Sensitivity and Additional Analysis

In sensitivity analysis (Tables, Supplemental Digital Content 3, http://links.lww.com/MLR/B545, sensitivity analysis), the Combined index outperformed Elixhauser measure and Charlson index in predicting 30-day mortality both in younger (<65 y) and older (≥65 y) populations, but provided a better performance in younger population. The improvement in discrimination (AUC) provided by the Combined index was higher when predicting 1-year mortality than when predicting 30-day mortality.

DISCUSSION

In population-based adult cohorts, we validated that combining Elixhauser's and Charlson's indices outperforms either ones to predict 30-day mortality. The superior performance of the Combined index is confirmed by a better discrimination, as reflected by higher AUC. Although the differences may appear small, it has been demonstrated that even slight improvements in AUC can translate into measurable potential benefit in reducing confounding bias.[1] We also demonstrated that the 3 weight derivation methods tested provide consistent performance across ICD-9 and ICD-10. Those results are important because as the ICD-10 is gradually implemented worldwide, it is critical to epidemiologist and health population analysts to have access to an efficient method to adjust for case mix and comorbidities in longitudinal observational studies overlapping ICD-9 and ICD-10 coding systems.

Our results thus confirm the better performance of the Combined index on both Charlson and Elixhauser index across ICD-9 and ICD-10 coding systems. In the Gagne et al[3] study where weights were derived using the SCH method, the AUC of the Combined index when predicting 30-day mortality was 1.7–2.4 point higher than the AUC of Charlson and Elixhauser index. In comparison, the increase in AUC with the Combined index was 1.1–1.3 point in our study. The difference may be explained by the younger population in our study (difference in mean age of ~18 y). Indeed, when we restricted the analyses to the 65 years and older group, the AUC of the Combined index was 1.9–2.0 point higher, similar to the results of Gagne et al.[3] Improvement in performance of the Combined index was also smaller in our study to predict 1-year mortality than in the Gagne et al[3] study. Differences in population characteristics and the fact that different weight derivation methods were used between each index in the Gagne et al[3] study may partly explain this discrepancy. Of note, the performance of all 3 indices remained high to predict 1-year mortality in our study while it decreased in the Gagne et al[3] study.

Our results showed that it is important to evaluate the impact each medical condition plays in predicting mortality over time periods. Comorbidities with relatively high weight in the Charlson/Romano version[16] and VanWalraven version[13] of the Charlson index, such as AIDS, got lower weight in our study. Differences may be explained by a period effect of the availability of novel therapies that have reduced mortality risk.

The 3 tested weight derivation methods based on logistic coefficient offered similar performance to binary approach and outperformed the count of conditions in both ICD-9 and ICD-10 coding systems. Even though weighted and binary approaches offer similar performances, weighted approaches should be preferred. Presence of large number of binary variables may imply collinearity issues in regression models when high correlation exists between ≥2 variables.[13] Furthermore, modeling interactions between covariables and comorbidity is hazardous. Use of weighted approaches prevents such problems. The VW method had a slightly better discriminant capacity than the SCH method. However, the magnitude of the weights in the SCH approach represents a statistically meaningful concept across any population. For example, a weight of 1 refers to a 35% increase in odds of developing the outcome. The CH method has a slightly poorer discriminant capacity and is not mathematically sound.[17] Nonetheless, the method is well-known and easy to interpret: a weight of 2, for example, is equivalent to an odds ratio of 2.[18] Combined Index may be used as a continuous variable or a class variable. On a continuous scale, weights derived through the SCH approach appears an optimal choice because they are mathematically valid and provide better interpretability than the VW method. However, weights derived through the CH method are clinically more intuitive without significantly compromising on the performance and may be preferred when using the Combined index as a class variable.

Our study has limitations. First, a coding algorithm had to be developed to accommodate Quebec-specific ICD coding systems. However, based on validated algorithms from Quan et al,[12] we did not encounter issues that would suggest medical conditions were not identified with sufficient validity. We involved a multidisciplinary team of medical specialists' physician and senior registrar at each step of the process. By the means of our rigorous conversion process, we selected codes that are mainly similar to Quan et al[12] codes. Only a small number of them were different and they had small frequencies. Second, there was no external validation of the results with the ICD-10 coding system. However, results obtained in another population with ICD-9 are similar to our results.[3]

To our knowledge, this is the first study to validate a combined of list of Charlson Index and Elixhauser measure within the ICD-10 coding system. Furthermore, even though some studies have compared score methods,[8,13] to your knowledge none had systematically compared all 5 scores methods using Charlson Index and Elixhauser system. By derivation of weights in one population (same age, same year) and by using comparative weights derivation methods for each 3 indices, we isolated the net improvement of the Combined index on both indices.

Nevertheless, more work will be required to explore the quality and characteristics of such a Combined index.

A priority should be to better understand how cohort characteristics such as age and sex may affect the index weights and overall scores.[19] From a population surveillance perspective, major health determinants such as social, demographic, and economic factors should also be included in such analyses. It is, however, important to note that further studying these properties will not challenge the results presented in this paper but rather refine them when being applied to specific strata of the population or outcomes sensitive to those factors and characteristics.

CONCLUSIONS

A weighted comorbidity index combining conditions included in the Charlson and Elixhauser measures shows significant predictive capacity for 30-day mortality in a general population. The comorbidity index remains valid with both ICD-9 and ICD-10 coding classifications and weight derivation methods provide consistent high performance across those coding classifications. The use of the Combined index, as an approach to control for confounding in health research and surveillance studies, should be easy to implement.

ACKNOWLEDGMENTS

The authors are grateful to the senior archivist, Ginette Therriault, who supported us through the ICD-9 Quebec diagnosis codes adaptation and to each medical specialist who helped us in that work: Dr Claudia Blais, Dr Richard Fachehoun, Dr Linda Perron, Dr Sonia Jean, Dr Alain Lesage, and Dr Edeltraut Kroger. Finally, we thank Myles Gaulin for help in the revision of the final version of the document and all QICDSS team members for their support.

REFERENCES

1. Schneeweiss S, Seeger JD, Maclure M, et al. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154:854–864.
2. Yurkovich M, vina-Zubieta JA, Thomas J, et al. A systematic review identifies valid comorbidity indices derived from administrative health data. *J Clin Epidemiol*. 2015;68:3–14.
3. Gagne JJ, Glynn RJ, Avorn J, et al. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol*. 2011;64:749–759.
4. 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;40:373–383.
5. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.
6. Lix LM, Quail J, Fadahunsi O, et al. Predictive performance of comorbidity measures in administrative databases for diabetes cohorts. *BMC Health Serv Res*. 2013;13:340.
7. Iezzoni L. *Risk Adjustment for Measuring Health-Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press; 2003.
8. Thompson NR, Fan Y, Dalton JE, et al. A new elixhauser-based comorbidity summary measure to predict in-hospital mortality. *Med Care*. 2015;53:374–379.
9. Li B, Evans D, Faris P, et al. Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC Health Serv Res*. 2008;8:12.
10. Sundararajan V, Henderson T, Perry C, et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol*. 2004;57:1288–1294.
11. Blais C, Jean S, Sirois C, et al. Quebec integrated chronic disease surveillance system (QICDSS), an innovative approach. *Chronic Inj Can*. 2014;34:226–235.
12. 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;43:1130–1139.
13. Van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47:626–633.
14. Schneeweiss S, Wang PS, Avorn J, et al. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res*. 2003;38:1103–1120.
15. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138.
16. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol*. 1993;46:1075–1079.
17. Harrell F. Regression coefficients and scoring rules. *J Clin Epidemiol*. 1996;49:819.
18. Tu JV, Naylor CD. Clinical prediction rules. *J Clin Epidemiol*. 1997; 50:743–744.
19. Charlson M, Szatroswski TP, Peterson J. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994;47:1245–1251.