Derivation and Validation of Automated Electronic Search Strategies to Extract Charlson Comorbidities From Electronic Medical Records

Balwinder Singh, MBBS; Amandeep Singh, MBBS; Adil Ahmed, MBBS;
Gregory A. Wilson, BS; Brian W. Pickering, MB, BCh, FFARCSI;
Vitaly Herasevich, MD, PhD; Ognjen Gajic, MD, MSc; and Guangxi Li, MD

Abstract

Objective: To develop and validate automated electronic note search strategies (automated digital algorithm) to identify Charlson comorbidities.

Patients and Methods: The automated digital algorithm was built by a series of programmatic queries applied to an institutional electronic medical record database. The automated digital algorithm was derived from secondary analysis of an observational cohort study of 1447 patients admitted to the intensive care unit from January 1 through December 31, 2006, and validated in an independent cohort of 240 patients. The sensitivity, specificity, and positive and negative predictive values of the automated digital algorithm and *International Classification of Diseases, Ninth Revision (ICD-9)* codes were compared with comprehensive medical record review (reference standard) for the Charlson comorbidities.

Results: In the derivation cohort, the automated digital algorithm achieved a median sensitivity of 100% (range, 99%-100%) and a median specificity of 99.7% (range, 99%-100%). In the validation cohort, the sensitivity of the automated digital algorithm ranged from 91% to 100%, and the specificity ranged from 98% to 100%. The sensitivity of the ICD-9 codes ranged from 8% for dementia to 100% for leukemia, whereas specificity ranged from 86% for congestive heart failure to 100% for leukemia, dementia, and AIDS.

Conclusion: Our results suggest that search strategies that use automated electronic search strategies to extract Charlson comorbidities from the clinical notes contained within the electronic medical record are feasible and reliable. Automated digital algorithm outperformed ICD-9 codes in all the Charlson variables except leukemia, with greater sensitivity, specificity, and positive and negative predictive values.

© 2012 Mayo Foundation for Medical Education and Research ■ Mayo Clin Proc. 2012;87(9):817-824

Comorbidity is defined as any distinct clinical entity that preexists or occurs during a patient's primary disease.¹ Various studies have documented the role of comorbidities in predicting a patient's outcome.²⁻⁶ The Charlson Comorbidity Index (CCI) was developed to estimate the long-term (1-year) mortality of patients admitted to the hospital or enrolled in research studies on the basis of the comorbid conditions.⁷ The CCI consists of 19 comorbid conditions, and each comorbidity is assigned a score of 1, 2, 3, or 6 based on the relative risk of 1-year mortality.⁷ The CCI has been validated in several different populations and is widely used in various health services research and critical care studies.^{3,8-10}

The use of electronic medical records (EMRs) is increasing, and these records are used not only in clinical practice but also in most epidemiologic and health care research. Recent mail survey findings from the National Ambulatory Medical Care Survey conducted by the Centers for Disease Control and Prevention reported an increase in the adoption of EMR systems by US office-based physicians from

18% in 2001 to 57% in preliminary 2011 results.¹¹ As a part of the current health system reform in the United States, the government has invested large sums of money to support and promote the adoption of the EMR system in the country. The 2009 Health Information Technology for Economic and Clinical Health Act provides possible incentives for hospitals to implement EMR systems; hence, most physicians and hospitals intend to implement EMR systems within the next few years.^{12,13} Traditionally, the CCI was identified and calculated solely using the manual medical record review. In 1992, Deyo et al¹⁴ developed an electronic application tool based on *International Classification of Diseases, Ninth Revision (ICD-9)* codes to automatically calculate the CCI.¹⁴ Currently, this method is applied in most research projects for baseline comorbidities adjustments,^{5,10} although the literature has reported various concerns about the accuracy and underreporting of comorbidities using ICD-9 codes.¹⁵⁻²¹

With the growing notion of EMRs as a tool to reduce cost and improve safety,²² adoption of EMR systems in US hospitals is steadily increasing. The

For editorial
comment, see
page 811

From Multidisciplinary Epidemiology and Translational Research in Intensive Care (B.S., A.S., A.A., G.A.W., B.W.P., V.H., O.G., G.L.), Division of Pulmonary and Critical Care Medicine (B.S., A.A., O.G., G.L.), and Division of Critical Care Medicine (A.S., G.A.W., B.W.P., V.H.), Mayo Clinic, Rochester, MN; and Guang An Men Hospital, China Academy of Chinese Medical Science, Beijing (G.L.).

information overload becomes a hindrance to the effective use of EMRs, which may lead to reduction in performance and alter patient safety.²³ Thus, not only is implementation of an automated electronic note search strategy to identify a patient's baseline comorbidities useful in medical research, but early identification of the comorbidities from the EMR might also be helpful for the efficient treatment of patients.

We therefore aimed to develop and validate an automatic note search strategy (automated digital algorithm) based on EMR notes to identify CCI comorbidities. Our secondary aim was to compare the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of automatic note search strategy and ICD-9 code search with comprehensive medical record review (reference standard) in detecting CCI comorbidities from the EMR.

PATIENTS AND METHODS

The study was approved by the Mayo Clinic Institutional Review Board, for the use of medical records for research.

Study Population

The derivation cohort was from a community-based, retrospective study of 1447 eligible patients (patients ≥ 18 years of age who gave research authorization) admitted to the intensive care unit (ICU) during 2006 from Olmsted County, Minnesota.²⁴ The automated digital algorithm was further validated against a randomly selected cohort of 240 patients who gave research authorization from a retrospective cohort of 651 patients with severe sepsis admitted to the ICU from January 2007 to December 2009,²⁵ using JMP statistical software (version 9.0, SAS Institute Inc, Cary, NC). All those who denied research authorization were excluded.

Data Extraction Strategies

All the CCI comorbidities in the 5-year interval before the time of admission were manually collected by trained research fellows according to the published definitions by Charlson et al.⁷ Each record was reviewed by multiple reviewers. We compared the 2 data extraction strategies: (1) CCI scores extracted by the automated digital algorithm and (2) CCI scores extracted by using ICD-9 codes.

Automated Electronic Note Search Strategies (Automated Digital Algorithm)

The development of an automated electronic note search strategy (automated digital algorithm) requires considerable investment in both time and ex-

pertise. This retrospective study used data from the Mayo Clinic Life Sciences System (MCLSS). The MCLSS database is the single centralized database for all the Mayo clinic hospital data. The MCLSS is an exhaustive clinical data warehouse that stores patient demographic characteristics, diagnoses, and hospital, laboratory, flow sheet, clinical, and pathologic data gathered from various clinical and hospital source systems within the institution. The MCLSS encompasses a near real-time model of Mayo's EMR system.²⁶ We used a Data Discovery and Query Builder (DDQB) tool set to access the data contained within the MCLSS database. The DDQB can search for demographic characteristics, clinical data, hospital admissions information, diagnosis codes, procedure codes, laboratory test results, flow sheet data, pathology reports, and genetic data. A valid institutional review board number is needed to retrieve patient data, which can be used for research using the MCLSS. The DDQB provides a unique text search strategy by which researchers can rapidly search for distinct words or entities in the EMR system.

The DDQB is based on Boolean logic to create free text searches.²⁶ All the free text searches were run independently by 2 physician investigators (A.S. and B.S.). To initiate a query, we entered all the synonyms, abbreviations, and the most common symptoms associated with the comorbidity. In addition, we excluded the negative terms mentioned in the clinical notes to make the automated digital algorithms more specific (see Supplemental Appendix 1 for the list of excluded terms; available online at <http://www.mayoclinicproceedings.org>). For the extraction of all the CCI comorbidities, the automated digital algorithm explored the EMR of each eligible patient during the 5 years before the date of admission in the medical and surgical history section of the EMR. For better understanding, an automated digital algorithm for peptic ulcer disease is shown in the Figure. The automated digital algorithm for CCI comorbidities were continuously refined by adding or excluding terms to improve the sensitivity and specificity to 95% or more. The final search terms used for building the automated digital algorithm are shown in Supplemental Appendix 2 (available online at <http://www.mayoclinicproceedings.org>). To validate the automated digital algorithm, the sensitivity and specificity were calculated against the reference standard of comprehensive medical record review.

ICD-9-Based Data Extraction

The MCLSS administrative database was used to calculate the ICD-9 coded CCI comorbidities according to the widely used algorithm of Deyo et al.¹⁴

FIGURE. The automated digital algorithm for searching peptic ulcer disease (PUD).

Manual Data Extraction (Reference Standard)

Manual data extraction is the traditional method of ascertaining comorbidity data from clinical notes. Trained research fellows manually collected comorbidity data according to the definitions published by Charlson et al.⁷ Comorbid conditions are mostly recorded in the medical and surgical history section of the clinical notes. So to maintain uniformity and efficiently identify specific comorbid conditions, only the medical and surgical history sections of the clinical notes were ascertained. If the comorbid condition was not identified in this particular section of

the EMR, it was assumed to be negative. Research fellows involved in manual data extraction were masked to the automated electronic note search strategy results.

Statistical Analyses

Sensitivity and specificity of both the automated digital algorithm and ICD-9 codes search were calculated based on comparisons of the test results and the reference standard in the 2 cohorts. The PPV and NPV were calculated based on the formula:

$$NPV = \frac{(\text{specificity} \times 1 - \text{prevalence})}{(\text{specificity} \times 1 - \text{prevalence}) + (1 - \text{sensitivity} \times \text{prevalence})}$$

$$PPV = \frac{(\text{sensitivity} \times \text{prevalence})}{(\text{sensitivity} \times \text{prevalence}) + (1 - \text{specificity} \times 1 - \text{prevalence})}$$

TABLE 1. Demographic Characteristics and Baseline Comorbidity Status of the Derivation and Validation Cohorts

Demographic characteristic	Derivation cohort (n=1447)	Validation cohort (n=240)
Median age (y) (IQR)	66 (49-80)	69 (58-80)
Male, No. (%)	756 (52.2)	129 (53.8)
White, No. (%)	1312 (90.7)	217 (90.4)
Any comorbidity, No. (%)	982 (67.9)	204 (85.0)

IQR = interquartile range.

The 95% confidence intervals were calculated using an exact test for proportions. JMP statistical software (version 9.0, SAS Institute Inc) was used for all data analysis.

RESULTS

During the study, 1447 consecutive eligible patients admitted to the ICU from January 1 through December 31, 2006, at Mayo Clinic, Rochester, MN, were included in the derivation cohort. The demographic characteristics and baseline comorbidity status of the derivation and validation cohort are summarized in Table 1. The most prevalent comorbidities were chronic lung diseases (24.3%) and diabetes mellitus (24.3%) in the derivation cohort and malignant tumor (32.1%) in the validation cohort. In the derivation cohort, the automated digital algorithm achieved a median sensitivity of 100% (range, 99%-100%) and a median specificity of 99.7% (range, 99%-100%). A summary of the prevalence of CCI comorbidities, sensitivity, and specificity of automated digital algorithm in the derivation and validation cohorts compared with the reference standard is given in Table 2. In addition, the concordance and discordance (true-positive, true-negative, false-positive, and false-negative results) between the automated digital algorithm and reference standard in the derivation and validation cohorts are summarized in Table 2.

Table 3 summarizes the sensitivity, specificity, PPV, and NPV of the 2 data extraction strategies (automated digital algorithm and ICD-9 codes search) for all CCI comorbidities in the validation cohort. The sensitivity for identifying CCI comorbidities using the automated digital algorithm ranged from a minimum of 91% for lymphoma to a maximum of 100% for congestive heart failure (median, 98.5%; interquartile range [IQR], 94%-100%). Sensitivities for extracting comorbidities using ICD-9 codes ranged from a minimum of 8% for dementia to 100% for leukemia (median, 66%; IQR, 51%-76%). The automated digital algorithm achieved a median specificity of 99.6% (IQR, 99%-100%) compared with 97% (IQR, 91%-98%) for ICD-9 codes.

In different CCI comorbidity domains, the PPV ranged from 80% to 100% for the automated digital algorithm, with the lowest being hemiplegia and the highest connective tissue disease, leukemia, lymphoma, metastasis, mild liver disease, peptic ulcer disease, and peripheral vascular disease (100%). The PPV for ICD-9 codes ranged from 0% to 100%, with the highest being leukemia and lowest being AIDS (the ICD-9 codes from Deyo et al for AIDS, ie, 042.x-044.x take into account patients with human immunodeficiency virus and patients with AIDS). The CCI index weight of 6 is given only to AIDS patients and the only patient detected as having AIDS using the ICD-9 code in the validation cohort was a false-positive result, hence the PPV of zero. The NPV ranged from 97% to 100% for the automated digital algorithm compared with 79% to 100% for ICD-9-Clinical Modification codes. The median PPV and NPV were 97.2% (IQR, 94%-100%) and 99.6% (IQR, 99%-100%) for the automated digital algorithm and 62.5% (IQR, 42%-78%) and 97.2% (IQR, 91%-99%) for the ICD-9 codes, respectively.

DISCUSSION

Results of this study suggested the feasibility and validity of the automatic note search strategy in identifying CCI comorbidities in the EMR. Our results indicate that the sensitivities of the automatic note search strategy were considerably better than ICD-9 codes to search for all the CCI variables except one, leukemia, for which the sensitivity of both search strategies was 100%. The specificity and the NPV of the automatic note search strategy were also equal or superior to the ICD-9 codes search in all the CCI comorbidities. In addition, our results confirmed the findings of previous studies on the reliability and accuracy of electronic search; for example, Alsara et al²⁶ also reported that electronic query resulted in accurate and highly efficient data extraction.

The CCI is being widely used by health care researchers to predict short-term (30 days) and long-term (1 year) mortality in ICU patients.²⁷⁻³⁰ To compare the meaningful differences in patients' outcomes, it is essential to balance the baseline comorbidity conditions. The CCI is one of the most commonly used tools to measure the baseline comorbidities before ICU admission. A recent study performed by Christensen et al³¹ discussed the important role of the CCI combined with administrative data in predicting short- and long-term mortality for ICU patients. Although D'Hoore et al³² described the CCI index as a resourceful way to perform risk adjustment from administrative databases, Poses et al⁴ reported enhanced discrimination of inpatient mortality using the CCI index. Currently, an ICD-9 code search is frequently used to automati-

TABLE 2. Prevalence of Charlson Comorbidities, Sensitivity, Specificity, and the Concordance and Discordance Between the Automated Digital Algorithm and Reference Standard in the Derivation Cohort and Validation Cohort

Charlson variable	Cohort	No. of patients	Prevalence (%)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	True-positive result (No.)	True-negative result (No.)	False-positive result (No.)	False-negative result (No.)
AIDS	DC	2	0.1	100 (20-100)	100 (99-100)	2	1445	0	0
	VC	0	0.0	NA	100	0	240	0	0
Cerebrovascular disease	DC	193	13.3	100 (97-100)	99 (98-99)	193	1242	12	0
	VC	40	16.7	95 (82-100)	99 (96-100)	38	198	2	2
Chronic pulmonary disease	DC	352	24.3	99 (97-100)	99 (98-100)	349	1086	9	3
	VC	63	26.3	98 (90-100)	99 (96-100)	62	175	2	1
Congestive heart failure	DC	185	12.8	99 (97-100)	100 (99-100)	184	1261	0	1
	VC	47	19.6	100 (90-100)	99 (96-100)	47	191	2	0
Connective tissue disease	DC	63	4.4	100 (93-100)	100 (99-100)	63	1384	0	0
	VC	15	6.3	93 (67-100)	100 (98-100)	14	225	0	1
Dementia	DC	54	3.7	96 (86-99)	100 (99-100)	52	1389	4	2
	VC	13	5.4	92 (62-100)	100 (97-100)	12	226	1	1
Diabetes mellitus	DC	352	24.3	100 (99-100)	100 (99-100)	352	1095	0	0
	VC	59	24.6	92 (81-97)	99 (96-100)	54	179	2	5
Diabetes with end-organ damage	DC	101	7.0	100 (56-100)	97 (90-99)	101	1303	43	0
	VC	24	10.0	100 (82-100)	99 (96-100)	24	213	3	0
Hemiplegia	DC	14	1.0	100 (73-100)	100 (99-100)	14	1433	0	0
	VC	4	1.7	100 (40-100)	100 (97-100)	4	235	1	0
Leukemia	DC	12	0.8	100 (68-100)	100 (99-100)	12	1433	2	0
	VC	9	3.8	100 (63-100)	100 (98-100)	9	231	0	0
Lymphoma	DC	29	2.0	100 (85-100)	100 (99-100)	29	1418	0	0
	VC	11	4.6	91 (57-100)	100 (98-100)	10	229	0	1
Malignant tumor	DC	289	20.3	98 (96-99)	96 (95-97)	289	1109	44	5
	VC	77	32.1	96 (89-99)	99 (95-100)	74	161	2	3
Metastatic solid cancer	DC	35	2.4	97 (83-100)	99 (98-100)	34	1402	10	1
	VC	11	4.6	100 (68-100)	100 (98-100)	11	229	0	0
Mild liver disease	DC	38	2.6	100 (89-100)	100 (99-100)	38	1406	3	0
	VC	9	3.8	100 (63-100)	100 (98-100)	9	231	0	0
Moderate-severe liver disease	DC	16	1.1	100 (76-100)	100 (99-100)	16	1431	0	0
	VC	10	4.2	100 (67-100)	99 (97-100)	10	228	2	0
Moderate-severe renal disease	DC	155	10.7	99 (95-100)	100 (99-100)	153	1290	2	2
	VC	72	30.0	99 (91-100)	98 (94-100)	71	165	3	1
Myocardial infarction	DC	235	16.2	97 (94-99)	99 (98-100)	229	1204	8	6
	VC	33	13.8	97 (83-100)	100 (97-100)	32	206	1	1
Peptic ulcer	DC	127	8.8	99 (95-100)	100 (99-100)	126	1319	1	1
	VC	19	7.9	95 (72-100)	100 (98-100)	18	221	0	1
Peripheral vascular disease	DC	111	7.7	99 (94-100)	100 (99-100)	110	1331	5	1
	VC	23	9.6	100 (82-100)	100 (98-100)	23	217	0	0

CI = confidence interval; DC = derivation cohort; NA = not applicable (comorbidity was not present in the validation cohort); VC = validation cohort.

TABLE 3. Performance of the Automated Digital Algorithm and ICD-9 Code Search Compared With Reference Standard in the Validation Cohort

Charlson variable	Automated digital algorithm				ICD-9 code search			
	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%)	NPV (%)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%)	NPV (%)
AIDS		100		100		100	0	100
Cerebrovascular disease	95 (82-100)	99 (96-100)	95	99	55 (39-70)	90 (85-94)	52	91
Chronic pulmonary disease	98 (90-100)	99 (96-100)	97	99	56 (43-68)	94 (90-97)	78	86
Congestive heart failure	100 (90-100)	99 (96-100)	96	100	81 (66-90)	86 (80-90)	58	95
Connective tissue disease	93 (67-100)	100 (98-100)	100	100	67 (39-87)	98 (95-99)	71	98
Dementia	92 (62-100)	100 (97-100)	92	100	8 (0.4-38)	100 (97-100)	50	95
Diabetes mellitus	92 (81-97)	99 (96-100)	96	97	52 (40-65)	98 (95-100)	92	85
Diabetes with end-organ damage	100 (82-100)	99 (96-100)	89	100	92 (72-99)	96 (92-98)	71	99
Hemiplegia	100 (40-100)	100 (97-100)	80	100	75 (22-99)	97 (94-99)	33	100
Leukemia	100 (63-100)	100 (98-100)	100	100	100 (63-100)	100 (98-100)	100	100
Lymphoma	91 (57-100)	100 (98-100)	100	100	45 (18-75)	99 (96-100)	63	97
Malignant tumor	96 (89-99)	99 (95-100)	97	98	48 (37-60)	91 (85-95)	71	79
Metastatic solid cancer	100 (68-100)	100 (98-100)	100	100	73 (39-93)	91 (86-94)	28	99
Mild liver disease	100 (63-100)	100 (98-100)	100	100	33 (9-69)	95 (90-97)	20	97
Moderate-severe liver disease	100 (67-100)	99 (97-100)	83	100	90 (54-99)	97 (94-99)	60	100
Moderate-severe renal disease	99 (91-100)	98 (94-100)	96	99	75 (63-84)	95 (90-97)	86	90
Myocardial infarction	97 (83-100)	100 (97-100)	97	100	64 (45-79)	97 (94-99)	78	94
Peptic ulcer	95 (72-100)	100 (98-100)	100	100	68 (43-86)	97 (94-99)	68	97
Peripheral vascular disease	100 (82-100)	100 (98-100)	100	100	65 (43-83)	90 (85-94)	42	96

CI = confidence interval; ICD-9 = International Classification of Diseases, Ninth Revision; NPV = negative predictive value; PPV = positive predictive value.

cally extract CCI comorbidities.^{10,33,34} However, the ICD-9-coded administrative databases lack a clinical definition for diagnoses, causing variability in coding practices.³⁵ Our results revealed that ICD-9 codes underreport the comorbidities that substantiate the finding of the previous studies.^{19,20,36,37} The underreporting could be attributable to extra emphasis on the procedures and complications on admission, compared with the comorbidities, for monetary reasons.²⁰ Romano et al³⁸ also found that the CCI comorbidities were not accurately defined in ICD-9 codes, which produced interobserver variations in ICD-9 codes assigned to the comorbidities. Although automatic searches using ICD-9 codes to identify comorbidities has been used in many research projects, the lack of accuracy in criteria used by the staff who code medical records may differ from physicians' criteria in diagnosing a medical condition, which significantly limits the broad use of this method. The automatic note search strategies were derived from the algorithm-incorporated keyword and program for a query within the specific note section. This approach enhanced the use of the patient database query and tremendously reduced the time when compared with the manual medical

record review (mean time taken to manually review 1 patient note for the CCI comorbidities ranged from 5 to 10 minutes). The implementation of an electronic strategy to extract information is not only useful for research purposes but also may be helpful for the treatment of patients.³⁹ Because an automated digital algorithm provides accurate information about a patient's comorbidities, it will help physicians to recognize comorbidity information early and might help in better treatment. Comorbidities act as a prognosticating factor for patient survival and treatment-related outcomes. Patients with higher CCI scores are at increased risk for readmissions and hospitalizations; thus, using automated digital algorithms to identify comorbidities early will certainly be an important factor and might well be used for early palliative consultations if needed in the future. The high sensitivity and specificity of the automated digital algorithm make it an important tool for physicians and investigators in accurately estimating comorbidities and might help in making early decisions and avoiding medical errors.

Another alternative search strategy to identify comorbidities is the Systematized Nomenclature of Medicine—CLINICAL Terms (SNOMED-CT). AL-

though this method produced better performance than the ICD-9 code search strategy, there were also significant limitations for broad use in clinical research.⁴⁰ Chiang et al⁴¹ suggested that SNOMED-CT coding is imperfect and unreliable and requires physician training and repeated testing. Furthermore, SNOMED-CT does not satisfactorily distinguish the exact terms at the clinical interface level for the study template at the current stage.⁴²

Our search strategies also had certain limitations. First, performance of the automated digital algorithm and coding of the CCI is dependent on the quality of the database and consistency of the text entries, which limits the applicability of this approach to units with this database or one similar. However, the logic and the free text search concept could be generalized to other institutions; it provides potential for diffusion of the method at sites willing to replicate the programming effort because the medical documentation training is similar across the country. Because electronic clinical notes are becoming a standard feature of the modern era, our approach will become more generalizable. Second, we only focused on a pertinent section (medical and surgical history) of clinical notes to search for comorbidities, which might have caused us to miss some information provided in other note sections, although the same validation process can be extended to other sections of clinical notes because the concept remains the same. Third, the data can be missed because of errors or corruption in the data warehouse.⁴³ However, this will only account for a small proportion of the database. Fourth, some of the CCI comorbidities definitions are outdated. Since the original CCI was developed in 1987, medicine has undergone a vast amount of change. Certain diseases, such as AIDS, no longer have the same relative risk of mortality as when the CCI was developed. Similarly, criterion for untreated thoracic and abdominal aneurysm 6 cm or larger for diagnosis of peripheral vascular disease needs reassessment. The latest guidelines advocate surgery when the aneurysm is 5.5 cm or larger.⁴⁴ However, we could refine our search strategy to identify variables according to any new definitions by modifying the algorithm for new definitions. Finally, because of the retrospective nature of the study, we only included documented comorbidity in the definite diagnostic criteria.

In conclusion, CCI comorbidities can be correctly identified using the automated digital algorithm. The combination of good sensitivity, specificity, and easy calculation should encourage physicians to implement the automated digital algorithm in their clinical practice and medical research.

ACKNOWLEDGMENTS

We thank all members of the Multidisciplinary Epidemiology and Translational Research in Intensive Care group for constant and constructive feedback.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at <http://www.mayoclinicproceedings.org>.

Abbreviations and Acronyms: CCI = Charlson Comorbidity Index; CI = confidence interval; DDQB = Data Discovery and Query Builder; EMR = electronic medical record; ICD-9 = International Classification of Disease, Ninth Revision; ICU = intensive care unit; IQR = interquartile range; MCLSS = Mayo Clinic Life Sciences System NPV = negative predictive value; PPV = positive predictive value; SNOMED-CT = Systematized Nomenclature of Medicine—Clinical Terms

Grant Support: This work was supported by the National Institutes of Health grant R01 LM10468Z-01.

Correspondence: Address to Guangxi Li, MD, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (li.guangxi@mayo.edu).

REFERENCES

1. Feinstein AR. The pre-therapeutic classification of co-morbidity in chronic disease. *J Chronic Dis*. 1970;23(7):455-468.
2. Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality using administrative data. *J Clin Epidemiol*. 1997;50(8):903-908.
3. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol*. 2003;56(3):221-229.
4. Poses RM, McClish DK, Smith WR, Bekes C, Scott WE. Prediction of survival of critically ill patients by admission comorbidity. *J Clin Epidemiol*. 1996;49(7):743-747.
5. Durairaj L, Will JG, Torner JC, Doebbeling BN. Prognostic factors for mortality following interhospital transfers to the medical intensive care unit of a tertiary referral center. *Crit Care Med*. 2003;31(7):1981-1986.
6. Keenan SP, Dodek P, Chan K, et al. Intensive care unit admission has minimal impact on long-term mortality. *Crit Care Med*. 2002;30(3):501-507.
7. Charlson ME, Pompei P, Ales KA, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
8. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173(6):676-682.
9. Brooks SE, Ahn J, Mullins CD, Baquet CR. Resources and use of the intensive care unit in patients who undergo surgery for ovarian carcinoma. *Cancer*. 2002;95(7):1457-1462.
10. Needham DM, Scales DC, Laupacis A, Pronovost PJ. A systematic review of the Charlson comorbidity index using Canadian administrative databases: a perspective on risk adjustment in critical care research. *J Crit Care*. 2005;20(1):12-19.

11. Hsiao C, Hing E, Socey C, Cai B. Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-based Physician Practices: United States, 2001-2011. National Center for Health Statistics (NCHS) Data Brief; No. 79: November 2011. <http://www.cdc.gov/nchs/data/databriefs/DB79.pdf>. Accessed March 2, 2012.
12. Bays RA, Kaelin LD. Electronic medical records for the office. *J Vasc Surg*. 2010;51(5):1302-1308.
13. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Eng J Med*. 2010;363(6):501-504.
14. 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.
15. McCarthy EP, Iezzoni LI, Davis RB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care*. 2000;38(8):868-876.
16. Kern EF, Maney M, Miller DR, et al. Failure of ICD-9-CM codes to identify patients with comorbid chronic kidney disease in diabetes. *Health Serv Res*. 2006;41(2):564-580.
17. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias: does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267(16):2197-2203.
18. Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K; Ischemic Heart Disease Patient Outcomes Research Team. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. *J Clin Epidemiol*. 1994;47(3):249-260.
19. Hawker GA, Coyte PC, Wright JG, Paul JE, Bombardier C. Accuracy of administrative data for assessing outcomes after knee replacement surgery. *J Clin Epidemiol*. 1997;50(3):265-273.
20. Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease: an Australian perspective. *J Clin Epidemiol*. 2001;54(7):687-693.
21. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Med Care*. 2002;40(8):675-685.
22. Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. *JAMA*. 2011;305(2):169-172.
23. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. *Crit Care Med*. 2011;39(7):1626-1634.
24. Cartin-Ceba R, Kojicic M, Li G, et al. Epidemiology of critical care syndromes, organ failures and life support interventions in a suburban U.S. community. *Chest*. 2011;140(6):1447-1455.
25. Paturi A, Valerio-Rojas JC, Kashyap R, Sathick IJ, Subramanian A, Cartin-Ceba R. Statin therapy is associated with decreased hospital mortality in patients with severe sepsis and septic shock [abstract]. *Chest*. 2011;140(4_MeetingAbstracts):1063A.
26. Alsara A, Warner DO, Li G, Herasevich V, Gajic O, Kor DJ. Derivation and validation of automated electronic search strategies to identify pertinent risk factors for postoperative acute lung injury. *Mayo Clin Proc*. 2011;86(5):382-388.
27. Ho KM, Knuiman M, Finn J, Webb SA. Estimating long-term survival of critically ill patients: the PREDICT model. *PLoS One*. 2008;3(9):e3226.
28. Johnston JA, Wagner DP, Timmons S, Welsh D, Tsevat J, Render ML. Impact of different measures of comorbid disease on predicted mortality of intensive care unit patients. *Med Care*. 2002;40(10):929-940.
29. Pittet D, Thievent B, Wenzel RP, Li N, Gurman G, Suter PM. Importance of pre-existing co-morbidities for prognosis of septicemia in critically ill patients. *Intensive Care Med*. 1993;19(5):265-272.
30. Chu YT, Ng YY, Wu SC. Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMC Health Serv Res*. 2010;10:140.
31. Christensen S, Johansen MB, Christiansen CF, Jensen R, Leme-show S. Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care. *Clin Epidemiol*. 2011;3:203-211.
32. D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol*. 1996;49(12):1429-1433.
33. Sundararajan V, Quan H, Halfon P, et al. Cross-national comparative performance of three versions of the ICD-10 Charlson index. *Med Care*. 2007;45(12):1210-1215.
34. 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.
35. Iezzoni LI. Risk adjustment for medical effectiveness research: an overview of conceptual and methodological considerations. *J Invest Med*. 1995;43(2):136-150.
36. Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol*. 1999;52(2):137-142.
37. Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. *J Clin Epidemiol*. 1994;47(9):1027-1032.
38. 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(10):1075-1090.
39. Kutney-Lee A, Kelly D. The effect of hospital electronic health record adoption on nurse-assessed quality of care and patient safety. *J Nurs Adm*. 2011;41(11):466-472.
40. Elkin PL, Brown SH, Husser CS, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc*. 2006;81(6):741-748.
41. Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA Annu Symp Proc*. 2006:131-135.
42. Brown SH, Rosenbloom ST, Bauer BA, et al. Direct comparison of MEDCIN and SNOMED CT for representation of a general medical evaluation template. *AMIA Annu Symp Proc*. 2007:75-79.
43. Wisniewski MF, Kieszowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. *J Am Med Inform Assoc*. 2003;10(5):454-462.
44. Brewster DC, Cronenwett JL, Hallett JW, Jr., Johnston KW, Krupski WC, Matsumura JS. Guidelines for the treatment of abdominal aortic aneurysms: report of a subcommittee of the Joint Council of the American Association for Vascular Surgery and Society for Vascular Surgery. *J Vasc Surg*. 2003;37(5):1106-1117.