Observational Study

Medicine®

OPEN

The derivation and validation of a simple model for predicting in-hospital mortality of acutely admitted patients to internal medicine wards

Ali Sakhnini, MD[a], Walid Saliba, MD, MPH[b,c], Naama Schwartz, PhD[d], Naiel Bisharat, MD, PhD[a,c,*]

Abstract

Limited information is available about clinical predictors of in-hospital mortality in acute unselected medical admissions. Such information could assist medical decision-making.

To develop a clinical model for predicting in-hospital mortality in unselected acute medical admissions and to test the impact of secondary conditions on hospital mortality.

This is an analysis of the medical records of patients admitted to internal medicine wards at one university-affiliated hospital. Data obtained from the years 2013 to 2014 were used as a derivation dataset for creating a prediction model, while data from 2015 was used as a validation dataset to test the performance of the model. For each admission, a set of clinical and epidemiological variables was obtained. The main diagnosis at hospitalization was recorded, and all additional or secondary conditions that coexisted at hospital admission or that developed during hospital stay were considered secondary conditions.

The derivation and validation datasets included 7268 and 7843 patients, respectively. The in-hospital mortality rate averaged 7.2%. The following variables entered the final model; age, body mass index, mean arterial pressure on admission, prior admission within 3 months, background morbidity of heart failure and active malignancy, and chronic use of statins and antiplatelet agents. The c-statistic (ROC-AUC) of the prediction model was 80.5% without adjustment for main or secondary conditions, 84.5%, with adjustment for the main diagnosis, and 89.5% with adjustment for the main diagnosis and secondary conditions. The accuracy of the predictive model reached 81% on the validation dataset.

A prediction model based on clinical data with adjustment for secondary conditions exhibited a high degree of prediction accuracy. We provide a proof of concept that there is an added value for incorporating secondary conditions while predicting probabilities of in-hospital mortality. Further improvement of the model performance and validation in other cohorts are needed to aid hospitalists in predicting health outcomes.

Abbreviations: ACG = adjusted clinical groups, AUC = area under the curve, BI = Business Intelligence, BMI = body mass index, CCBs = calcium channel blockers, CHF = congestive heart failure, COPD = chronic obstructive pulmonary disease, CVD = cerebrovascular disease, EHRs = electronic health records, ICD-9 = International Classification of Diseases—Ninth Revision, IDI = integrated discrimination improvement, IRB = institutional review board, LASSO = least absolute shrinkage and selection operator, MAP = mean arterial pressure, NRI = net reclassification index, ROC = receiver operating characteristic, SAP = systems–applications–products, UTI = urinary tract infection.

Keywords: hospital mortality, multivariate analysis, patient admission, prediction model

1. Introduction

Predicting in-hospital mortality has been the focus of numerous studies aimed to provide clinicians with simple and reproducible risk assessment models. The Charlson comorbidity index,[1] which was originally formulated to predict 1-year mortality based on coexistent comorbidities, has been adapted to predict various outcomes such as in-hospital mortality, disability, hospital readmissions, higher hospital costs, and length of stay.[2,3] Numerous studies have consistently shown that comorbidity is one of the main factors associated with adverse outcomes among hospitalized patients.[4,5] To date, nearly all prediction model studies that investigated the association between various variables and in-hospital mortality focused on specific patient populations such as patients with pneumonia,[6] congestive heart failure (CHF),[7] acute myocardial infarction,[8] chronic obstructive pulmonary disease (COPD),[9] stroke,[10] infective endocarditis,[11] and cancer.[12] In addition, these studies mainly addressed associations between the main diagnosis at hospital discharge notes and various variables, without adjustment for secondary diagnoses or concomitant conditions that coexisted at the time of admission or that developed during hospital stay. Such conditions could have a substantial impact on the predicted probability of in-hospital mortality and are not always accounted for during data analysis.

Our objective was to develop a simple and reliable model that relies only on clinical characteristics to predict in-hospital

Editor: Ting Shuo Huang.

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

[a] *Department of Medicine D, Emek Medical Center, Clalit Health Services, Afula,* [b] *Department of Community Medicine and Epidemiology, Carmel Medical Center, Clalit Health Services,* [c] *Ruth and Bruce Rappaport Faculty of Medicine, Technion—Israel Institute of Technology, Haifa,* [d] *Research Authority, Emek Medical Center, Clalit Health Services, Afula, Israel.*

[*] *Correspondence: Naiel Bisharat, Department of Medicine D, Emek Medical Center, 21 Rabin Avenue, Afula 18341, Israel (e-mail: bisharat_na@clalit.org.il).*

Medicine (2017) 96:25(e7284)

Received: 21 November 2016 / Received in final form: 1 June 2017 / Accepted: 4 June 2017

http://dx.doi.org/10.1097/MD.0000000000007284

mortality among patients admitted to internal medicine wards and to examine the impact of secondary conditions on overall hospital mortality.

2. Materials and methods

We conducted a retrospective cohort study of patients who were admitted to one of the 5 internal medicine wards at Emek Medical Center over a 3-year period. Data obtained from the first 2 years (2013–2014) were used as a derivation dataset for creating a prediction model, while data from 2015 were used as a validation dataset to test the performance of the model.

Emek Medical Center is a university-affiliated hospital located in northeastern Israel, with an annual 130,000 visits to the emergency departments. Our hospital policy for the past 15 years dictates the review of all discharge notes of hospitalized patients by specially trained administrative staff. This is to ensure the completeness and the compatibility of the main and any secondary diagnosis with the medical notes registered during the patient's hospitalization. Data for the present study were gathered using 2 separate datasets, Chameleon Medical Record (Elad Health, Tel-Aviv, Israel), a web-based electronic medical record information system that includes inpatient administrative and clinical information and SAP (systems–applications–products) Business Objects-Business Intelligence (BI) platform (SAP, Walldorf, Germany), a suite of front-end applications that assembles clinical and administrative data from both in-hospital and outpatient sources. Admissions that lasted for at least 24 hours were eligible for inclusion in the study. Cohort assembly was carried out while ensuring that only one admission per patient was included. Thus, for patients who had multiple admissions during the study period (2013–2014) and survived, one admission was randomly selected. For patients who had multiple admissions and died we chose the admission that ended with death, we did not include any admission before that. If the patient died during the study period (2013–2014), had multiple admissions, but did not die during hospital stay, we chose only the last admission before his death. For each patient's admission the following variables were obtained: age, gender, body mass index (BMI), mean arterial pressure (MAP) on admission, date of admission, date of discharge, date of death, length of stay, month of admission, weekend or weekday admission, admission within 3 months before index admission, and time of admission according to nurses' shifts (08:00–15:59, 16:00–23:59, and 24:00–07:59).

The Chameleon Medical Record system uses the International Classification of Diseases—Ninth Revision (ICD-9) for assigning medical diagnosis and requires differentiation between the main diagnosis, secondary conditions, and background morbidity. For each patient, the main diagnosis was recorded and all other conditions that coexisted at hospital admission or that developed during hospital stay were considered to be secondary conditions. The main diagnosis and secondary conditions listed in the discharge notes were grouped into 13 categories, matching the ICD-9 classification with some modifications (see Supplementary file S1, http://links.lww.com/MD/B763). Conditions that could not be grouped into any of the 13 categories, or that had less than 5% frequency, were grouped under "other diagnosis." This grouping of diagnoses was intended to decrease misclassifications and increase the utility, as well as feasibility, of the prediction model.

From the outpatient datasets the following variables were recorded: the adjusted clinical groups (ACG) score[13] (in the year before index admission), which measures morbidity burden based on disease patterns, age and gender as a constellation of morbidities, not as individual diseases. The ACG system automatically collapses the full set of ACG categories into 6 simplified morbidity categories (nonusers, healthy-users; and low, moderate, high, and very high morbidity). For each ACG, a relative weight was determined, which is the ratio of the mean ambulatory cost for each ACG to the mean ambulatory cost for the entire population. In addition, the number and type of specific chronic comorbidities was noted; CHF, diabetes mellitus, COPD, hypertension, chronic renal failure, cerebrovascular disease (CVD), and active malignancy excluding localized basal or squamous cell skin cancer and cervical carcinoma in situ; and the number and type of chronic medications: diuretics, β blockers, calcium channel blockers (CCBs), statins, antiplatelet therapy, antithrombotic therapy, angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, and proton pump inhibitors.

Patients were excluded from the analysis according to the following criteria: short admissions (<24 hours), admissions classified under symptoms, signs, and ill-defined conditions (ICD-9 codes 780–799), and admissions classified under "observation for" (ICD-9 codes V71, V71.2, V29.0, V29.1, V29.2, V29.8, and V29.9).

2.1. Statistical analysis

We used the Chi-square test (or Fisher exact test) to investigate the association between categorical variables and death. Continuous variables were examined using the Student *t* test (or Wilcoxon 2 sample test). The effect of the independent risk factors on the odds of death was estimated by implementing multivariate stepwise logistic regression. The stepwise algorithm was used after looking for confounders and interactions in the stratified analyses. The threshold probability for entering variables into the model was $P < .10$. Removal threshold was $P > .05$. The model accuracy and goodness of fit were estimated using receiver operating characteristic (ROC) and area under the curve (AUC), and the Hosmer and Lemeshow Goodness-of-Fit Test.

To explore the impact of main and secondary conditions on the accuracy of the model, 3 multivariate models were compared: without main or secondary diagnoses (model A); with main diagnoses only (model B), with main and secondary conditions (model C). The models were compared using the AUC as well as the integrated discrimination improvement (IDI),[14] and the net reclassification index (NRI).[15] The application of NRI and IDI is intended to assess the added discrimination offered by the addition of a marker to a prediction model. The NRI uses risk categories to quantify the reclassification improvement of the new model over the other. The IDI assesses the improvement in sensitivity without sacrificing average specificity. The model was validated using the validation dataset; for each patient a predicted probability for in-hospital mortality was obtained, using the final model coefficients. To assess the accuracy of the prediction model, ROC curve was constructed and AUC was calculated. The data management and statistical analyses were performed using the SAS (version 9.4) software.

The study was approved by the hospital's institutional review board (IRB; approval number: EMC-14-0109).

3. Results

In 2013 to 2014, 12,499 patients were admitted to the 5 participating internal medicine wards, resulting in 21,794

Table 1

Clinical and epidemiological characteristics according to survival, together with crude odds ratios for mortality of each covariate (derivation dataset, n = 7268).

Variable	Survived, n = 6386	Died, n = 882	OR [95% CI]	*P*
Mean age (range)	67.8 [18–105]	77.5 [23–105]	1.05 [1.04–1.05]	<.001
Female sex (%)	2984 (46.7)	453 (51.4)	1.20 [1.04–1.39]	.010
BMI (kg/m^2) ± SD	28.3 ± 6.2	25.8 ± 6.3	0.93 [0.91–0.94]	<.001
MAP on admission ± SD	94.4 ± 17.4	82.6 ± 25.2	0.97 [0.96–0.99]	<.001
Mean length of stay (d)	5.9	7.8	1.03 [1.02–1.04]	<.001
Time of admission				
08:00–12:00	1586 (25.0)	235 (27.0)	1	
12:00–16:00	1691 (26.7)	230 (26.5)	0.92 [0.76–1.12]	.39
16:00–19:00	1043 (16.5)	116 (13.4)	0.75 [0.59–0.95]	.017
19:00–24:00	1301 (20.5)	181 (20.8)	0.94 [0.76–1.16]	.55
24:00–08:00	717 (11.3)	107 (12.3)	1.01 [0.79–1.29]	.95
Weekend admission				
No	4386 (68.7)	601 (68.1)	1	
Yes	2000 (31.3)	281 (31.9)	1.03 [0.88–1.19]	.74
Admission within prior 3 months				
No	5027 (78.7)	479 (54.3)	1	
Yes	1359 (21.3)	403 (45.7)	3.11 [2.69–3.6]	<.001
Background morbidity				
COPD	814 (12.8)	158 (17.9)	1.49 [1.24–1.8]	<.001
CRF	1083 (16.9)	246 (27.9)	1.89 [1.61–2.23]	<.001
DM	2365 (37.0)	378 (42.9)	1.28 [1.11–1.47]	.001
HTN	3640 (57.0)	601 (68.1)	1.61 [1.39–1.88]	<.001
HF	1304 (20.4)	284 (32.2)	1.85 [1.59–2.16]	<.001
CVD	679 (10.6)	107 (12.1)	1.16 [0.93–1.44]	.18
Malignancy	816 (12.8)	256 (29.0)	2.79 [2.37–3.29]	<.001
Chronic medication				
Diuretics	956 (14.9)	204 (23.2)	1.7 (1.44–2.02)	<.001
Beta blockers	1783 (27.9)	214 (24.3)	0.7 [0.53–0.95]	.019
Antithrombotic therapy	522 (8.2)	52 (5.9)	0.7 [0.53–0.95]	.019
PPIs	1442 (22.6)	238 (26.9)	1.27 [1.08–1.49]	.004
Antiplatelet agents	2513 (39.4)	270 (30.6)	0.68 [0.58–0.79]	<.001
Statins	2159 (33.8)	152 (17.2)	0.41 [0.34–0.49]	<.001
ACEI and ARBs	2239 (35.06)	212 (24.04)	0.59 [0.50–0.69]	<.001
CCBs	1248 (19.5)	130 (14.7)	0.71 [0.59–0.87]	.001
ACG category				
Nonusers	30 (0.61)	2 (0.27)	1	
Healthy-users	385 (7.79)	63 (8.54)	2.45 [0.57–10.52]	.23
Low morbidity	872 (17.65)	118 (15.99)	2.03 [0.48–8.6]	.34
Moderate morbidity	1313 (26.58)	170 (23.04)	1.94 [0.46–8.2]	.37
High morbidity	1510 (30.57)	231 (31.3)	2.29 [0.55–9.66]	.26
Very high morbidity	830 (16.8)	154 (20.87)	2.78 [0.66–11.76]	.16
Mean ACG weight	3.3 ± 2.8	3.6 ± 2.9	1.04 [1.01–1.07]	.004

ACEI = angiotensin converting enzyme inhibitors, ACG = adjusted clinical groups, ARBs = angiotensin receptor blockers, BMI = body mass index, CCBs = calcium channel blockers, CI = confidence interval, COPD = chronic obstructive pulmonary disease, CRF = chronic renal failure, CVD = cerebrovascular disease, DM = diabetes mellitus, HF = heart failure, HTN = hypertension, MAP = mean arterial pressure, OR = odds ratio, PPIs = proton pump inhibitors, SD = standard deviation.

admissions. Readmission episodes ranged from 2 to 33. The in-hospital mortality rate was 7.1% (n = 882). After excluding patients who did not meet study inclusion criteria (n = 4723), and patients with missing discharge diagnoses (n = 508), the 7268 eligible patients who remained comprised the derivation dataset (Table 1). Altogether, nearly 80% of the main diagnoses fit the 13 categories, and more than 93% of the diagnoses of patients who died fit these categories. The most frequent main diagnoses were pneumonia, heart failure, and CVD (Table 2). The most common main diagnoses among the deceased in decreasing frequency were pneumonia, sepsis and septicemia, malignant neoplasms, and heart failure (Table 2).

Univariate analysis based on the derivation dataset resulted in 22 variables that were considered candidates for the logistic regression (Table 1), with 8 remaining in the final model (Table 3). These included age, BMI, MAP on admission, previous admission within 3 months before the index admission, background diagnoses of heart failure and active malignancy, and chronic use of statins and antiplatelet agents. In addition, 6 secondary conditions entered the final model; these included heart failure, pneumonia, sepsis and septicemia, renal failure, malignant neoplasm, and acute coronary syndrome (Table 3, and Supplementary file-S2, http://links.lww.com/MD/B764). The analysis was executed again without patients with missing BMI values (n = 939), no significant differences were found from the original cohort (data not shown).

A prediction model was created based on estimates from the final model. The c-statistic (ROC-AUC) of the prediction model was 80.5% without adjustment for main or secondary

Table 2

Main diagnoses from hospital discharge notes according to survival, classified into 13 categories, with crude odds ratio for mortality.

Group no.	Main diagnosis	Survived, n = 6386	Died, n = 882	OR [95% CI]	*P*
1	Deficiency and other anemia	324 (5.1)	11 (1.3)	0.8 [0.4–1.48]	.42
2	Heart failure	772 (12.1)	85 (9.6)	2.5 [1.76–3.51]	<.001
3	Chronic obstructive pulmonary disease and allied conditions	313 (4.9)	22 (2.5)	1.6 [0.96–2.63]	.07
4	Cerebrovascular disease	730 (11.4)	66 (7.5)	2 [1.42–2.94]	.001
5	Malignant neoplasms	204 (3.2)	123 (14.1)	13.6 [9.63–19.22]	<.001
6	Acute coronary syndrome	555 (8.7)	29 (3.3)	1.2 [0.75–1.86]	.48
7	Cardiac dysrhythmias	411 (6.4)	10 (1.1)	0.6 [0.28–1.08]	.08
8	Pneumonia	722 (11.3)	206 (23.4)	6.4 [4.75–8.74]	<.001
9	Acute and chronic renal failure	318 (5.0)	59 (6.7)	4.2 [2.86–6.14]	<.001
10	Sepsis and septicemia	168 (2.6)	128 (14.5)	17.2 [12.13–24.40]	<.001
11	Septic shock	6 (0.1)	56 (6.4)	210.6 [87.2–508.8]	<.001
12	Venous thrombosis and embolism	154 (2.4)	8 (0.9)	1.2 [0.55–2.50]	.68
13	Urinary tract infections	400 (6.3)	21 (2.4)	1.2 [0.71–1.98]	.52
14	Other	1309 (20.5)	58 (6.6)		

See Supplementary file S1, http://links.lww.com/MD/B763 for diagnosis/medical conditions included in each group.
CI = confidence interval, OR = odds ratio.

conditions, 84.5%, with adjustment for the main diagnosis, and 89.5% with adjustment for the main diagnosis and secondary conditions (Fig. 1). Similarly, both IDI and NRI measures showed statistically significant discrimination ability for the addition of secondary conditions (Table 4).

The validation dataset consisted of 7843 patients with 11,508 admissions; 6323 patients were admitted only once. Readmission episodes ranged from 2 to 13. In-hospital mortality rate was 7.4% (n = 582). For each patient, the predictive probability for in-hospital death was calculated according to the established

Table 3

Multivariate stepwise logistic regression to assess the contribution of variables to mortality.

Variable	βi	SE (βi)	Adjusted OR (95% CI)	*P*
Intercept	−5.24	0.579		<.001
Age	0.045	0.005	1.05 (1.04–1.06)	<.001
BMI	−0.023	0.010	0.98 (0.96–1)	.025
MAP on admission	−0.01	0.003	0.99 (0.98–0.99)	.002
Admission within 3 prior months	0.862	0.116	2.37 (1.89–2.97)	<.001
Heart failure background	0.397	0.135	1.49 (1.14–1.94)	.003
Underlying malignancy	0.304	0.147	1.36 (1.02–1.81)	.039
Prior use of antiplatelet agents	−0.314	0.137	0.73 (0.56–0.96)	.022
Prior use of statins	−0.569	0.152	0.57 (0.42–0.76)	<.001
Main diagnosis				
Deficiency and other anemia	−0.602	0.441	0.55 (0.2–1.3)	.18
Heart failure	−0.753	0.309	0.47 (0.2–0.86)	.015
Chronic obstructive pulmonary disease and allied conditions	−0.023	0.362	0.98 (0.5–1.99)	.90
Cerebrovascular disease	1.288	0.255	3.63 (2.2–5.9)	<.001
Malignant neoplasms	1.657	0.333	5.24 (2.7–10.1)	<.001
Acute coronary syndrome	−0.348	0.424	0.71 (0.3–1.6)	.42
Cardiac dysrhythmias	−0.556	0.412	0.57 (0.2–1.3)	.17
Pneumonia	−0.259	0.28	0.77 (0.45–1.3)	.36
Acute and chronic renal failure	0.061	0.308	1.06 (0.58–1.95)	.84
Sepsis and septicemia	−0.121	0.318	0.89 (0.48–1.65)	.70
Septic shock	1.99	0.587	7.31 (2.3–23.1)	.001
Venous thrombosis and embolism	0.359	0.557	1.43 (0.48–4.27)	.52
Urinary tract infections	0.296	0.432	1.34 (0.58–3.14)	.49
Other			1	
Secondary conditions				
Heart failure	1.132	0.213	3.1 (2.0–4.7)	<.001
Sepsis and septicemia	2.651	0.233	14.17 (8.9–22.4)	<.001
Pneumonia	1.757	0.209	5.79 (3.8–8.7)	<.001
Acute coronary syndrome	1.026	0.319	2.79 (1.5–5.2)	.001
Renal failure	1.306	0.17	3.69 (2.6–5.2)	<.001
Malignant neoplasm	1.215	0.284	3.37 (1.9–5.9)	<.0001

BMI = body mass index, CI = confidence interval, MAP = mean arterial pressure, OR = odds ratio, SE = standard error.

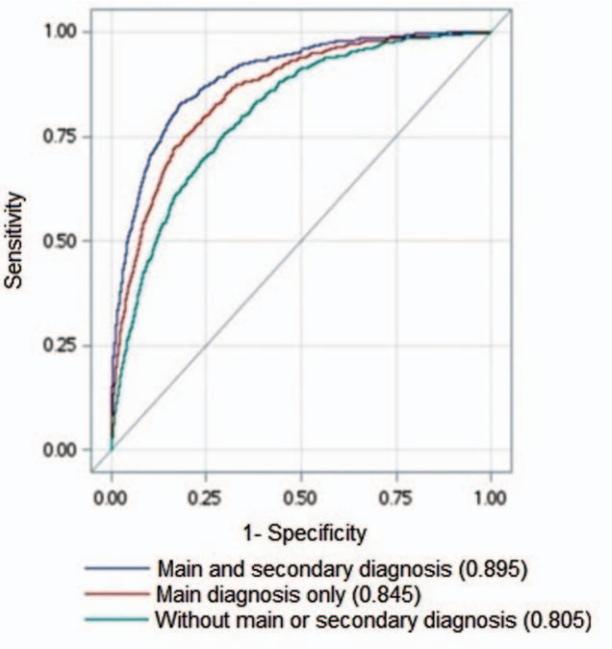

Figure 1. Receiver operating characteristic (ROC) curves and relevant area under curve (AUC) applied for 3 possible models using standardized weighting coefficients.

model. The accuracy of the developed predictive model reached 81%.

4. Discussion

The present study showed that a simple clinically based model can reasonably predict the risk of in-hospital mortality of acutely admitted medical patients. Our objective was to integrate secondary conditions coexistent at hospital admission or developing during hospital stay and to examine their impact on the prediction accuracy of in-hospital mortality. The addition of secondary conditions increased the c-statistic of the model form 84.5% to 89.5%. The derived model includes 8 variables, 13 possible main diagnoses, and 6 secondary conditions. Testing the accuracy of the prediction model on a separate cohort revealed a c-statistic of 81%.

Our prediction model would be easy for clinicians to use as it relies on basic variables that include age, BMI, MAP on admission, history of prior admission (3 months), chronic morbidity (heart failure and malignancy), and the use of certain chronic medications (statins and antiplatelet agents). Previous publications have described models for predicting in-hospital mortality with clinical and laboratory variables using large datasets in the United States.[16–19] However, these studies did not evaluate the role of additional clinical data, such as vital signs, in predicting mortality.[16,17] Some targeted specific patient populations (male patients admitted to intensive care units),[17] or limited their analysis to 6 common clinical conditions.[18] A recent study described an automated disease-specific risk adjustment system using clinical data,[19] while incorporating a wide spectrum of clinical conditions, 2 dozen numerical laboratory tests, and administrative data. The average c-statistic for the automated clinical models was 0.83. Nevertheless, none of these studies adjusted their analyses for secondary conditions. The c-statistic of our model was 89.5%, indicating an excellent correlation with in-hospital mortality, and an excellent predictive accuracy of 81%. Compared with other predictive models for in-hospital mortality, our simple model provides better, or at least comparable, predictive accuracy.[16,18–20]

Table 4

Summary statistics comparing the discrimination ability of the different risk prediction models.

	Difference in AUC, %	IDI [95% CI]	NRI [95% CI]
Model C vs model A	9	19.5% [18.5–20.5]	33% [30.5–35.6]
Model C vs model B	5	10.1% [9.3–10.9]	19.1% [16.8–21.4]
Model B vs model A	4	9.4% [8.7–10.1]	16.6% [14.4–18.7]

Model A: without main diagnosis or secondary conditions.
Model B: with main diagnosis only.
Model C: with main diagnosis and secondary conditions.
The application of NRI and IDI is intended to assess the incremental value of one model versus another.
AUC = area under curve, CI = confidence interval, IDI = integrated discrimination improvement, NRI = net reclassification index.

A study published 3 decades ago estimated that nearly 17% of patients suffer from new complications during hospitalization.[21] A more recent study reported that in-hospital medical complications developed among 25% of patients admitted with stroke.[22] A limitation of the present study is that we could not determine whether the secondary conditions listed in the computerized discharge notes were complications that developed during hospital stay or were present at hospital admission. Several recent studies evaluated rates of specific medical complications among patients admitted with some common medical conditions. For example, rates of cardiac events among patients admitted with community acquired pneumonia ranged between 8% and 19%.[23,24] Likewise, rates of cardiac dysrhythmias among patients with sepsis have been estimated to range between 8% and 46%.[25] Urinary tract infection (UTI) and pneumonia developed in 15% and 9%, respectively, of patients with stroke, according to a Danish study[22]; and cardiac arrest developed in 3.9% of patients with acute ischemic stroke in a recently published study.[26]

Heart failure was the most common diagnoses among the survivors of the present study, in both the derivation and validation cohorts, but less so among the deceased. This was also the most common diagnosis among readmitted patients. Heart failure diagnosis entered the final model in 2 different categories, as a background chronic morbid condition, and as a secondary condition. However, heart failure as the main diagnosis was not associated with increased probability of in-hospital mortality. This confusing observation could imply that admission criteria for patients with heart failure in our institution may not have been properly implemented and that some of these patients could have been treated in an outpatient setting.

Data from the derivation and validation datasets imply that prior use of statins, and antiplatelet agents are associated with lower rates of in-hospital mortality. These observations may reflect healthier user effect, yet are to be interpreted cautiously. Further investigations are needed to extend these observations to additional patients. Nevertheless, recent studies suggested that prior use of statins are associated with lower in-hospital mortality among patients admitted with acute ischemic stroke,[27] intracerebral hemorrhage,[28] and sepsis.[29] In addition, the use of statins has been suggested to be associated with declining in-hospital mortality from acute myocardial infarction[30] and heart failure,[28] and a modest reduction in pneumonia mortality

in wards outside of intensive care.[29] Finally, a recent systematic review and meta-analysis of observational studies that evaluated the effect of statins use on mortality in cancer patients concluded that statins use may be beneficial for overall survival and cancer-specific survival.[31] Similarly, prior use of antiplatelet agents has been shown to be associated with lower in-hospital mortality in critically ill[32] and septic patients.[33]

Risk prediction using data from electronic health records (EHRs) has become popular in the past 20 years with the increased availability of EHRs in hospitals and other healthcare providers.[34,35] Clinical research using EHRs is typically carried out using either association analysis[36] or prediction analysis,[37] while combining both, as implemented in the present study, is uncommon. Our design was impacted by our goal to provide a proof of concept that there is an added value of incorporating secondary conditions while predicting probabilities of in-hospital mortality. The selection of covariates in present study was carried out using multivariate stepwise logistic regression, this method, though criticized in some publications[38–41] and expert opinion[42] due to possible bias, model over-fitting, and lack of generalizability, was the most popular method of selecting covariates in epidemiological studies published in 2008.[42] Our decision to use multivariate stepwise logistic regression was driven by its simplicity, ease of use, reproducibility, and to make our model accessible and understood to the readers. In this regard, logistic regression with regularization is considered the preferred method of statistical analysis in epidemiological studies.[42–45] Regularized logistic regression has several advantages over standard logistic regression. First, it helps prevent the model from over-fitting the data, second, it makes the first step of analysis (association tests) unnecessary since it allows automatic selection of the most informative covariates, and last, it has better model generalizability. One technique of regularized logistic regression is the least absolute shrinkage and selection operator (LASSO),[46] also named "shrinkage with selection." This technique corrects the extremes in the distribution of all variables and thus shrinks very unstable estimates toward zero. This effectively excludes some variables without the need for formal statistical testing. Despite its powerful and important features, LASSO and similar methods, were not applied in any study of 171 selected articles from 4 leading epidemiological journals in 2008.[42] Perhaps there are several reasons for that. One may be that implementing stepwise methods is much simpler than the modern techniques (LASSO and other shrinkage models). Another is the lack of familiarity of medical researchers with these methods, and third, it is possible that statisticians may have not adequately promoted and addressed the method's feasibility.[42]

Our study has a number of limitations; first, the retrospective study design confers limitations, including the potential for misclassification and incomplete data. Second, the lack of accuracy and uniformity in hospital discharge notes may have impacted data analysis as main diagnoses may have been mistakenly classified as secondary and vice versa. However, this misclassification is nondifferential and is expected to bias our results toward the null. The high proportion of diagnoses that fits the 13 established categories, and the particularly high proportion that fits the categories of those who died, indicate good, though not complete, categorization of the diagnoses recorded on discharge notes. Third, it could be argued that the utility of a prediction model based on discharge diagnoses may not be ideal for patients at the start of their admission, due to the discrepancy between admission and discharge diagnoses or due to the occurrence of discharge diagnoses toward the end of a patient's admission. We believe that the impact of such bias is limited. In support of our view, a study of adults admitted to general internal medicine wards of a large medical center showed that the 10 most common admitting diagnoses that did not match the principal discharge diagnosis were classified as ill-defined conditions (ICD-9 codes 780–799)[47]; all these conditions were excluded from the current analysis. Fourth, our cohort lacked data concerning smoking and alcohol drinking habits which could influence the predictive probability of in-hospital death. Fifth, the generalizability of the proposed model could be limited for other populations as the stepwise algorithm may produce irreproducible estimates.[42] And last, our study is from a single-institution which could limit generalizability of our findings.

Despite these limitations, our study has several strengths. First, this is a first-ever study to provide a clinical model for predicting in-hospital mortality of unselected acute medical admissions while incorporating secondary conditions. Second, the model requires simple and readily available measures. Third, it performs equally well to other more complicated models. Further improvement of the model's performance and validation in other larger cohorts are needed to aid hospitalists in predicting health outcomes.

5. Conclusion

A prediction model based on clinical and epidemiological data with adjustment for secondary conditions exhibited 81% prediction accuracy of in-hospital mortality among unselected acute medical admissions. We provide a proof of concept that there is an added value for incorporating secondary conditions while predicting probabilities of in-hospital mortality. Further improvement of the model performance and validation in other cohorts are needed to aid hospitalists in predicting health outcomes. The calculator for predicting in-hospital mortality is available in the Supplementary file (S2), http://links.lww.com/MD/B764.

References

[1] Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–83.
[2] de Groot V, Beckerman H, Lankhorst GJ, et al. How to measure comorbidity. a critical review of available methods. J Clin Epidemiol 2003;56:221–9.
[3] Kieszak SM, Flanders WD, Kosinski AS, et al. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. J Clin Epidemiol 1999;52:137–42.
[4] Frey C, Zhou H, Harvey D, et al. Co-morbidity is a strong predictor of early death and multi-organ system failure among patients with acute pancreatitis. J Gastrointest Surg 2007;11:733–42.
[5] Mnatzaganian G, Ryan P, Norman PE, et al. Accuracy of hospital morbidity data and the performance of comorbidity scores as predictors of mortality. J Clin Epidemiol 2012;65:107–15.
[6] Marrie TJ, Wu L. Factors influencing in-hospital mortality in community-acquired pneumonia: a prospective study of patients not initially admitted to the ICU. Chest 2005;127:1260–70.
[7] Pocock SJ, Wang D, Pfeffer MA, et al. Predictors of mortality and morbidity in patients with chronic heart failure. Eur Heart J 2006;27:65–75.
[8] Falcao FJ, Alves CM, Barbosa AH, et al. Predictors of in-hospital mortality in patients with ST-segment elevation myocardial infarction undergoing pharmacoinvasive treatment. Clinics (Sao Paulo) 2014;68:1516–20.
[9] Ho TW, Tsai YJ, Ruan SY, et al. In-hospital and one-year mortality and their predictors in patients hospitalized for first-ever chronic obstructive pulmonary disease exacerbations: a nationwide population-based study. PLoS ONE 2014;9:e114866.

[10] Linfante I, Walker GR, Castonguay AC, et al. Predictors of mortality in acute ischemic stroke intervention: analysis of the North American Solitaire Acute Stroke Registry. Stroke 2015;46:2305–8.
[11] Park LP, Chu VH, Peterson G, et al. Validated risk score for predicting 6-month mortality in infective endocarditis. J Am Heart Assoc 2016;5: e003016.
[12] Ganti S, Hadad S. Comparison of 30-day, 90-day and in-hospital postoperative mortality for eight different cancer types (Br J Surg 2012; 99: 1149–1154). Br J Surg 2012;99:1602; author reply 1602–3.
[13] Starfield B, Weiner J, Mumford L, et al. Ambulatory care groups: a categorization of diagnoses for research and management. Health Serv Res 1991;26:53–74.
[14] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72; discussion 112–207.
[15] Kerr KF, Wang Z, Janes H, et al. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology 2013;25:114–21.
[16] Escobar GJ, Greene JD, Scheirer P, et al. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Med Care 2008;46:232–9.
[17] Render ML, Deddens J, Freyberg R, et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. Crit Care Med 2008;36:1031–42.
[18] Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. Med Care 2007;45: 789–805.
[19] Tabak YP, Sun X, Derby KG, et al. Development and validation of a disease-specific risk adjustment system using automated clinical data. Health Serv Res 2010;45(6 Pt 1):1815–35.
[20] Asadollahi K, Hastings IM, Gill GV, et al. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. Emerg Med Australas 2011;23:354–63.
[21] Charlson ME, Sax FL, MacKenzie CR, et al. Morbidity during hospitalization: can we predict it? J Chronic Dis 1987;40:705–12.
[22] Ingeman A, Andersen G, Hundborg HH, et al. In-hospital medical complications, length of stay, and mortality among stroke unit patients. Stroke 2011;42:3214–8.
[23] Musher DM, Rueda AM, Kaka AS, et al. The association between pneumococcal pneumonia and acute cardiac events. Clin Infect Dis 2007;45:158–65.
[24] Viasus D, Garcia-Vidal C, Manresa F, et al. Risk stratification and prognosis of acute cardiac events in hospitalized adults with community-acquired pneumonia. J Infect 2013;66:27–33.
[25] Kuipers S, Klein Klouwenberg PM, Cremer OL. Incidence, risk factors and outcomes of new-onset atrial fibrillation in patients with sepsis: a systematic review. Crit Care 2014;18:688.
[26] Joundi RA, Rabinstein AA, Nikneshan D, et al. Cardiac arrest in acute ischemic stroke: incidence, predisposing factors, and clinical outcomes. J Stroke Cerebrovasc Dis 2016;25:1644–52.
[27] Hong KS, Lee JS. Statins in acute ischemic stroke: a systematic review. J Stroke 2015;17:282–301.
[28] Wang JQ, Wu GR, Wang Z, et al. Long-term clinical outcomes of statin use for chronic heart failure: a meta-analysis of 15 prospective studies. Heart Lung Circ 2014;23:105–13.
[29] Jia M, Huang W, Li L, et al. Statins reduce mortality after non-severe but not after severe pneumonia: a systematic review and meta-analysis. J Pharm Pharm Sci 2015;18:286–302.
[30] Ezekowitz JA, Kaul P, Bakal JA, et al. Declining in-hospital mortality and increasing heart failure incidence in elderly patients with first myocardial infarction. J Am Coll Cardiol 2009;53:13–20.
[31] Zhong S, Zhang X, Chen L, et al. Statin use and mortality in cancer patients: systematic review and meta-analysis of observational studies. Cancer Treat Rev 2015;41:554–67.
[32] Wang L, Li H, Gu X, et al. Effect of antiplatelet therapy on acute respiratory distress syndrome and mortality in critically ill patients: a meta-analysis. PLoS ONE 2016;11:e0154754.
[33] Tsai MJ, Ou SM, Shih CJ, et al. Association of prior antiplatelet agents with mortality in sepsis patients: a nationwide population-based cohort study. Intensive Care Med 2015;41:806–13.
[34] Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. J Intern Med 2013;274:547–60.
[35] Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc 2013;20:e206–11.
[36] Chen Y, Xie W, Gunter CA, et al. Inferring clinical workflow efficiency via electronic medical record utilization. AMIA Annu Symp Proc 2015;2015:416–25.
[37] Zheng T, Xie W, Xu L, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform 2016;97:120–7.
[38] Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. J Clin Epidemiol 2004;57:1138–46.
[39] Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. Am J Epidemiol 2008;167:523–9; discussion 521–30.
[40] Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002;155:176–84.
[41] Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol 1989;129:125–37.
[42] Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. Eur J Epidemiol 2009;24:733–6.
[43] Li W, Liu H, Yang P, et al. Supporting regularized logistic regression privately and efficiently. PLoS ONE 2016;11:e0156479.
[44] Stiglic G, Povalej Brzan P, Fijacko N, et al. Comprehensible predictive modeling using regularized logistic regression and comorbidity based features. PLoS ONE 2015;10:e0144439.
[45] Wu MY, Zhang XF, Dai DQ, et al. Regularized logistic regression with network-based pairwise interaction for biomarker identification in breast cancer. BMC Bioinformatics 2016;17:108.
[46] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.
[47] Johnson T, McNutt R, Odwazny R, et al. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. J Hosp Med 2009;4:234–9.