

A new prediction model for assessing the clinical outcomes of ICU patients with community-acquired pneumonia: a decision tree analysis

Shufang Zhang^a, Kai Zhang^b, Yang Yu^b, Baoping Tian^b, Wei Cui^b and Gensheng Zhang^b

^aDepartment of Cardiology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, PR China;

^bDepartment of Critical Care Medicine, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, PR China

ABSTRACT

Purpose: We aimed to develop a new scoring index based on decision-tree analysis to predict clinical outcomes of patients with community-acquired pneumonia (CAP) admitted to the intensive care unit (ICU).

Methods: Data of 3519 ICU patients with CAP were obtained from the Medical Information Mart for Intensive Care III (MIMIC III) 2001–2012 database and analysed between 30-d survivors and non-survivors. Accuracy, sensitivity, and specificity of the new decision tree model were compared with those of CURB-65 and SOAR.

Results: The newly developed classification and regression tree (CART) model identified coexisting illnesses as the most important single discriminating factor between survivors and non-survivors. The CART model area under the curve (AUC) 0.661 was superior to that of CURB-65 (0.609) and SOAR (0.589). CART sensitivity was 73.4%, and specificity 49.0%. CURB-65 and SOAR sensitivity for predicting 30-d mortality were 74.5 and 80.7%, and specificity was 42.3 and 33.9%, respectively. After smoothing, the CART model had higher sensitivity and specificity than both CURB-65 and SOAR.

Conclusions: The new CART prediction model has higher specificity and better receiver operating characteristics (ROC) curves than CURB-65 and SOAR score indices although its accuracy and sensitivity are only moderately better than the other systems.

KEY MESSAGES

- The new CART prediction model has higher specificity and better ROC curves than CURB-65 and SOAR score indices.
- However, accuracy and sensitivity of the new CART prediction model are only moderately better than the other systems in predicting 30-day mortality in CAP patients.

ARTICLE HISTORY

Received 22 January 2018

Revised 4 July 2018

Accepted 6 July 2018

KEYWORDS

Intensive care;
mortality; community-
acquired pneumonia

Background

Community-acquired pneumonia (CAP) is a frequent cause of intensive care unit (ICU) admissions internationally and is the leading cause of deaths due to infectious disease in Western countries [1]. Incidence of CAP is about 20–30% in developing countries, which is significantly higher than the 3–4% reported for developed countries and varies markedly with age, occurring most commonly in the very young and very old [2]. The disease course and outcomes of CAP vary widely, presenting with high risk for respiratory failure or sepsis-associated organ dysfunction in hospitalized patients [3,4]. The overall mortality rate among hospitalized patients with CAP is 13% but this rises to more than 35% in severe CAP (SCAP) patients [5]. Thirty-day

mortality is extremely high, especially in patients with comorbidities such as cancer or renal disease [6].

Investigators have aimed to identify critical factors for predicting ICU admission and prognosis in patients with CAP [1–3]. Scoring systems have been developed that combine multiple serum biomarkers and clinical parameters by which to assess CAP and predict outcomes [2,4,7,8]. For example, the Infectious Diseases Society of America/American Thoracic Society (IDSA/ATS) system, developed in 2001 (2) and revised in 2007(8), identifies CAP by the presence of 1 of 2 major and at least 3 of 9 minor criteria. Other existing scoring systems, include pneumonia severity scoring indices (PSI); score based on confusion, urea, respiratory rate, blood pressure, age 65 (CURB 65); severe

community acquired pneumonia (SCAP) score; score based on systolic blood pressure, oxygenation, age, respiratory rate (SOAR); and score based on systolic blood pressure, multilobar involvement, albumin, respiratory rate, tachycardia, confusion, oxygenation, pH (SMART-COP), which are widely used today for predicting clinical outcomes of CAP [3,6–9]. However, these systems have certain shortcomings, including performing poorly in predicting patients at higher risk, and showing low positive rates at recommended cut-off points for predicting 30-d mortality [3]. PSI, for example, was found to have only modest utility for discriminating between patients with fatal and nonfatal pneumonia [6]. Also, the generalizability of the existing scoring systems is not consistent when applied in different clinical situations. For example, Gonzalez et al. [6] concluded that the performance of CURB-65 and PSI systems are inadequate for predicting pneumonia-related mortality in immunocompromised cancer patients. Fang et al. [10] compared scoring indices in patients with healthcare-associated pneumonia, finding that PSI, CURB-65, IDSA/ATS, SCAP, SOAR and SMART-COP were not ideal for determining the need for intensive care in this patient population. However, those authors did state that the SCAP score was as accurate as or better than other scoring systems (e.g. CURB-65 and PSI) in predicting adverse outcomes in CAP patients.

Given the inadequate and inconsistent performance of existing scoring systems for predicting clinical outcomes of CAP patients, including 30-d mortality, it seems clear that an effective prediction model is still needed. We hypothesized that a prediction model based on decision tree analysis would be better able to predict the probability of mortality of CAP patients. We planned to analyse a wider range of variables that may impact 30-d mortality in CAP patients. Therefore, the present study aimed to develop a new decision tree-based scoring model to predict clinical outcomes, especially 30-d mortality, of CAP patients admitted to ICU, and to evaluate performance of the new scoring index compared to other scoring systems in current use.

Methods

Data source

All data for the present study were obtained from the Medical Information Mart for Intensive Care III (MIMIC-III) 2001–2012 [11]. MIMIC-III is a large, freely accessible database comprising de-identified health-related data collected from over forty thousand patients

admitted to critical care units of the Beth Israel Deaconess Medical Center (BIDMC), Boston, MA, between 2001 and May 2012. The MIMIC research database is a joint venture managed by researchers from the Laboratory for Computational Physiology at Massachusetts Institute of Technology (MIT), Cambridge, MA. The Department of Medicine at BIDMC is supported by grants from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under award numbers R01-EB001659 (2003–2013) and R01-EB017205 (2014–2018). Use of the MIMIC database for research purposes has been approved by the Institutional Review Boards of BIDMC and MIT.

Study population

Patients admitted to ICU who were diagnosed with CAP were eligible for inclusion. After screening the MIMIC-III 2000–2012 database, the data of 3519 patients with CAP were included for analysis and were divided into two groups, including 952 30-d survivors and 2567 non-survivors. Pneumonia patients with other types of pneumonia were designated as CAP to differentiate cases from institution-acquired pneumonia or healthcare-associated pneumonia (acquired prior to hospital admission through association with healthcare but not otherwise community-acquired) [12].

Subjects' demographic and clinical characteristics

Demographics include subjects' age (grouped as <40, 40–64, 65–79 and 80+) and gender. Clinical characteristics include coexisting illness (neoplastic disease, liver disease, congestive heart failure, cerebrovascular disease and renal disease), invasive mechanical ventilation, septic shock needing vasopressor support, respiratory rate, confusion, leukopenia, thrombocytopenia, hypothermia, hyperthermia, hypotension, pulse rate, arterial pH for arterial blood gas (ABG) test, and laboratory values (blood urea nitrogen (BUN), sodium, glucose and haematocrit), oxygen, and pleural effusion. Clinical characteristics correspond to variables criteria for model development (shown below with definitions).

Primary study outcomes

The endpoint of this study is 30-d all-cause mortality. Patients' 30-d mortality was defined as all-cause death within 30 d after admission to ICU, calculated using

the date of death and admission date. Deaths may have occurred during or after hospitalization.

Study variables

The new scoring index was developed using the classification and regression tree (CART) approach. The following criteria shown with defined values were used for model construction: invasive mechanical ventilation, septic shock needing vasopressor support, respiratory rate \geq at breaths/min, confusion, BUN level \geq ev mg/dl, leukopenia (white blood cell [WBC] count $<$ 4000 cells/mm), thrombocytopenia (platelet count $<$ 100,000 cells/mm), hypothermia (core temperature $<$ 36 °C or 96.8 °F), hyperthermia (core temperature \geq em °C or 104.0 °F), hypotension (mean arterial pressure of 40–50 mmHg or a systolic blood pressure $<$ 90 mmHg or diastolic BP \leq 60 mmHg), coexisting illness (neoplastic disease, liver disease, congestive heart failure, cerebrovascular disease, and renal disease), pulse rate \geq 125 beats/min, arterial pH for ABG test (ABG) $<$ 7.35, sodium $<$ 130 mmol/l, glucose \geq 250 mg/dl (14 mmol/l), haematocrit $<$ 30%, PaO₂ $<$ 60 mmHg or oxygen saturation $<$ 90% or PaO₂/FiO₂ ratio $<$ 250, pleural effusion, age (80+, 65–79, 40–64 and $<$ 40) and gender.

Comparison of scoring indices

Accuracy, sensitivity and specificity of the newly constructed decision tree model were compared with those of two existing scoring systems, CURB-65 score and SOAR score, used currently to predict mortality outcomes in hospitalized CAP patients.

CURB-65 index: CURB-65 is an expanded version of CRB-65 that offers a simplified scoring system with eight variables to assess severity in patients with CAP [7]. The CURB-65 index identifies high-risk patients using the following criteria (definitions of variables for data extraction are also listed below): confusion, BUN \geq 20mg/dl, respiratory rate \geq 30 breaths/min, systolic blood pressure $<$ 90 mmHg or diastolic blood pressure \leq 60 mmHg, and age \geq 65. Severity is divided into three classes for scores of 0–1, 2 and 3–5.

SOAR score: The SOAR score comprises severity assessment criteria recommended by the British Thoracic Society, namely systolic blood pressure, oxygenation, age and respiratory rate [13]. The SOAR score identifies severe CAP using the following criteria (definitions of variables for data extraction are also listed below): systolic blood pressure $<$ 90 mmHg, PaO₂/FiO₂ ratio $<$ 250, age \geq 65 and respiratory rate

\geq 30 breaths/min. Severity is divided into two classes, for scores of 0–1 and 2+.

Statistical analysis

Differences in categorical variables between the two groups (survivors and non-survivors) were determined using the Chi-square test and data are expressed as number and percentage. All potential risk factors were screened by univariate regression analyses to evaluate associations with 30-d all-cause mortality. Significant risk factors from univariate analysis were included in CART analysis [14]. Nodes in CART were controlled to have a minimum size of 100 records in parent nodes and 50 records in final child nodes. A 10-fold cross-validation was used to assess the predictive ability of the regression tree model. Overall model discrimination for the new CART model, CURB-65 and SOAR scores was assessed by sensitivity, specificity and area under the receiver operating characteristic (ROC) curve (AUC), with and without smoothing of binormal distribution. Sensitivity and specificity cut-off was determined using Youden's J statistic [15]. AUC pairwise comparisons between all three scores were done using the DeLong test [16]. Smoothed ROC curve AUC comparisons were done using the bootstrap method. AUC confidence intervals were calculated using the DeLong test [16] for non-smoothed ROC curves and bootstrap for smoothed curves. Finally, logistic regression models were used to analyse differences in probability of 30-d mortality between severe and non-severe CAP patients. All statistical assessments were two sided and evaluated at the 0.05 level of significance. Statistical analyses were performed using the statistical software package SPSS complex sample module version 22.0 (IBM Corp, Armonk, NY) and R for statistical computing version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) [17,18].

Results

The initial search of the MIMIC-III database version 1.4 identified 58,976 ICU admissions between 2001 and 2012. Among these, 3519 patients had CAP, including 952 survivors and 2567 non-survivors. Table 1 shows a comparison of clinical characteristics between 30-d survivors and non-survivors. Results of univariate analysis indicated that 12 risk factors (age, invasive mechanical ventilation use, septic shock with the need for vasopressor portal, respiratory rate, BUN level, thrombocytopenia, hypothermia, hypotension,

Table 1. Comparison of demographic and clinical characteristics between 30-d survivors and non-survivors.

Variables	Total (%)	Survivors (n = 2567)	Non-survivors (n = 952)	χ^2
Age, n (%)				79.81*
<65 years	1398 (39.7)	1135 (44.2)	263 (27.6)	
≥65 years	2121 (60.3)	1432 (55.8)	689 (72.4)	
Gender, n (%)				0.336
Male	1931 (54.9)	1401 (54.6)	530 (55.7)	
Female	1588 (45.1)	1166 (45.4)	422 (44.3)	
Invasive mechanical ventilation use, n (%)				16.16*
No	954 (27.1)	743 (28.9)	211 (22.2)	
Yes	2565 (72.9)	1824 (71.1)	741 (77.8)	
Septic shock with the need for vasopressors, n (%)				66.86*
No	1934 (55.0)	1518 (59.1)	416 (43.7)	
Yes	1585 (45.0)	1049 (40.9)	536 (56.3)	
Respiratory rate, n (%)				4.31*
<30 breaths/min,	3034 (87.6)	2226 (88.3)	808 (85.7)	
≥30 breaths/min,	430 (12.4)	295 (11.7)	135 (14.3)	
Confusion, n (%)				1.96
No	908 (46.0)	681 (47.0)	227 (43.4)	
Yes	1065 (54.0)	769 (53.0)	296 (56.6)	
BUN level, n (%)				61.94*
<20 mg/d	1246 (35.6)	1007 (39.4)	239 (25.1)	
≥20 mg/d	2258 (64.4)	1546 (60.6)	712 (74.9)	
Leukopenia, n (%)				2.64
No	3500 (99.5)	2550 (99.3)	950 (99.8)	
Yes	19 (0.5)	17 (0.7)	2 (0.2)	
Thrombocytopenia, n (%)				41.44*
No	3253 (92.6)	2418 (94.3)	835 (87.9)	
Yes	261 (7.4)	146 (5.7)	115 (12.1)	
Hypothermia, n (%)				19.24*
No	2751 (80.0)	2046 (81.9)	705 (75.2)	
Yes	686 (20.0)	453 (18.1)	233 (24.8)	
Hyperthermia, n (%)				0.269
No	3427 (99.7)	2491 (99.7)	936 (99.8)	
Yes	10 (0.3)	8 (0.3)	2 (0.2)	
Hypotension, n (%)				11.75*
No	3180 (91.5)	2340 (92.5)	840 (88.9)	
Yes	294 (8.5%)	189 (7.5)	105 (11.1)	
Coexisting illnesses				80.79*
No	728 (20.7)	627 (24.4)	101 (10.6)	
Yes	2791 (79.3)	1940 (75.6)	851 (89.4)	
Pulse rate, n (%)				0.207
<125 beats/min	1335 (90.7)	960 (90.9)	375 (90.1)	
≥125 beats/min	137 (9.3)	96 (9.1)	41 (9.9)	
Arterial pH for arterial blood gas test, n (%)				10.23*
≥7.35	1003 (58.2)	758 (60.5)	245 (52.0)	
<7.35	720 (41.8)	494 (39.5)	226 (48.0)	
Sodium, n (%)				2.91
≥130 mmol/l	3277 (93.5)	2398 (93.9)	879 (92.3)	
<130 mmol/l	228 (6.5)	155 (6.1)	73 (7.7)	
Glucose, n (%)				0.04
<250 mg/dl	3181 (90.8)	2319 (90.9)	862 (90.6)	
≥250 mg/dl	322 (9.2)	233 (9.1)	89 (9.4)	
Haematocrit, n (%)				17.70*
≥30%	2658 (75.9)	1984 (77.7)	674 (70.9)	
<30%	846 (24.1)	569 (22.3)	277 (29.1)	
PaO ₂ <60 mmHg or oxygen saturation <90% or PaO ₂ /FiO ₂ ratio <250				0.02
No	1559 (90.4)	1133 (90.5)	426 (90.3)	
Yes	165 (9.6)	119 (9.5)	46 (9.7)	
<i>Streptococcus pneumoniae</i> , n (%)				9.62*
No	3402 (96.7)	2467 (96.1)	935 (98.2)	
Yes	117 (3.3)	100 (3.9)	17 (1.8)	
<i>Klebsiella pneumoniae</i> , n (%)				0.01
No	3311 (94.1)	2416 (94.1)	895 (94.0)	
Yes	208 (5.9)	151 (5.9)	57 (6.0)	
<i>Legionella pneumophila</i> , n (%)				0.33
No	3513 (99.8)	2562 (99.8)	951 (99.9)	
Yes	6 (0.2)	5 (0.2)	1 (0.1)	

*Significant difference between survivors and non-survivors, $p < .05$.

coexisting illnesses, ABG test, haematocrit and positive *Streptococcus pneumoniae*) were independently associated with 30-d mortality (Table 1).

Figure 1 depicts the CART analysis for predicting 30-d mortality. Among the 12 variables evaluated, the CART method identified coexisting illnesses, septic

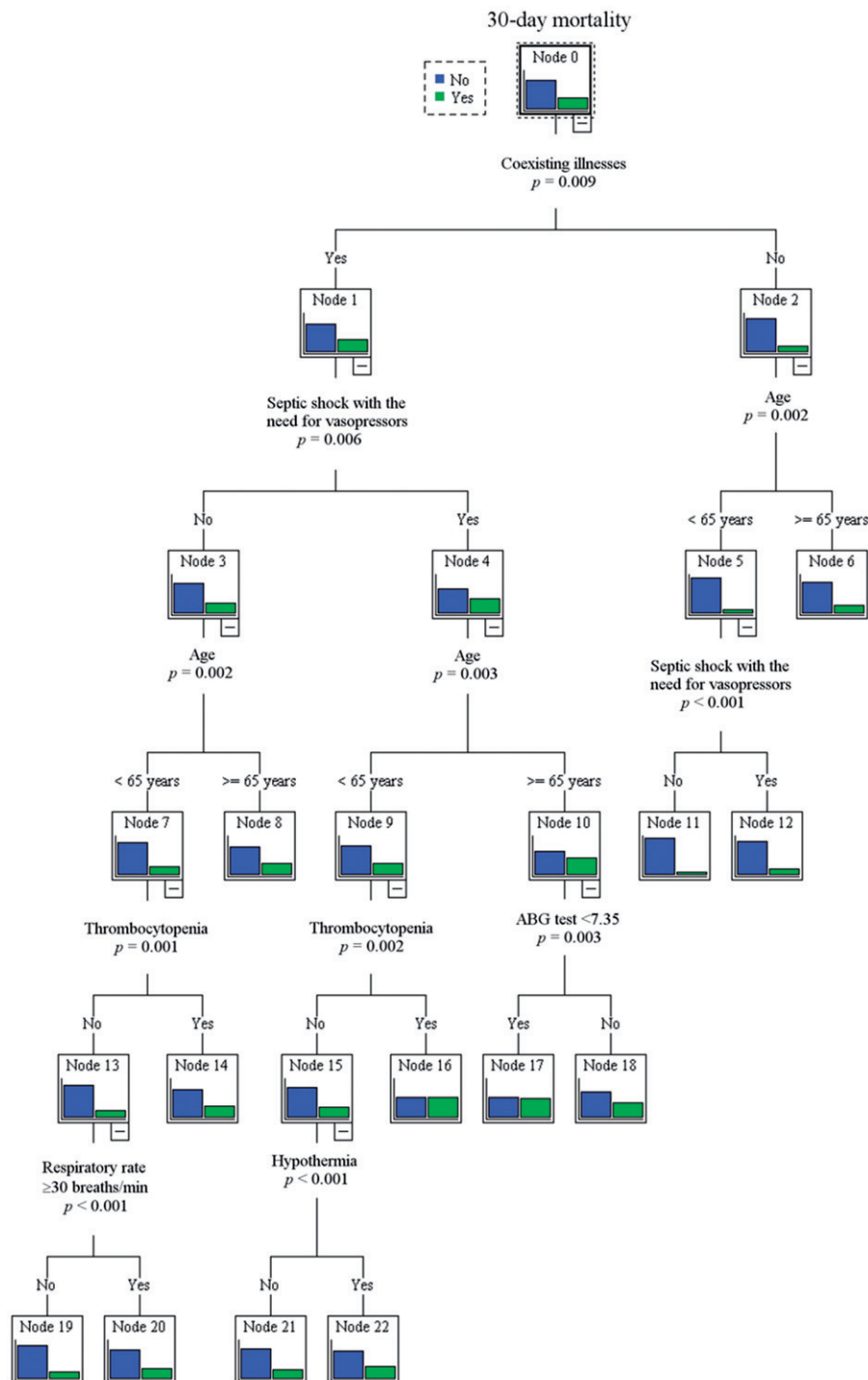


Figure 1. Decision tree model for predicting 30-d mortality.

shock with the need for vasopressors, age, ABG test, hypothermia, thrombocytopenia and respiratory rate as significant predictors for 30-d mortality. Coexisting illnesses was the most important single discriminating factor between survivors and non-survivors. The second most important predictors of 30-d mortality in patients with coexisting illnesses was age and septic shock with the need for vasopressor portal.

Analysis of the discrimination power of the CART model, CURB-65 scoring system and SOAR scoring system for predicting 30-d mortality using ROC curves are shown in Figure 2 and Table 2. AUC of the CART model for predicting 30-d mortality was 0.661. CART analysis for predicting mortality, at the cut-off determined by Youden's J statistic, had a sensitivity of 73.4% and a specificity of 49.0%. AUC of the CART

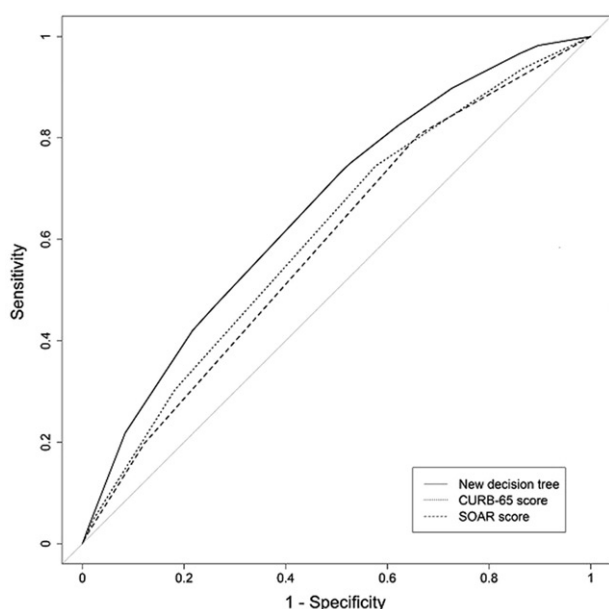


Figure 2. ROC curves for new decision tree and two scoring systems in the main cohort.

Table 2. Accuracy, sensitivity and specificity of the CART risk model, CURB-65 score and SOAR score.

	New decision tree	CURB-65 score	SOAR score
ROC curve			
AUC	0.661	0.608	0.589
Sensitivity	73.4%	74.5%	80.7%
Specificity	49.0%	42.3%	33.9%
ROC curve with smoothing			
AUC	0.667	0.612	0.611
Sensitivity	67.7%	65.0%	64.4%
Specificity	56.6%	51.2%	51.7%

Sensitivity and specificity cut-off point was calculated using Youden's J statistic.

model was superior to that of the other scoring systems (CURB-65 0.609; SOAR 0.589). This was true even after smoothing of binormal distribution (CART 0.667; CURB-65 0.612; SOAR 0.611). Sensitivity of the CURB-65 and SOAR scoring systems for predicting 30-d mortality with Youden's J statistic cut-off was 74.5 and 80.7%, respectively, and specificity was 42.3 and 33.9%, respectively. The new CART model had higher specificity compared to both CURB-65 and SOAR scores. After smoothing, the CART model had higher sensitivity (67.7%) and specificity (56.6%) than both CURB-65 and SOAR scores (Table 2).

Figure 3 shows the AUCs, AUC 95% confidence intervals, and results of pairwise DeLong tests between ROC curves of all three scores, non-smoothed and smoothed. Results indicate that the new CART model has significantly higher AUC than both CURB-65 and SOAR scores.

Logistic regression models were used to compare the probability for 30-d mortality between severe and

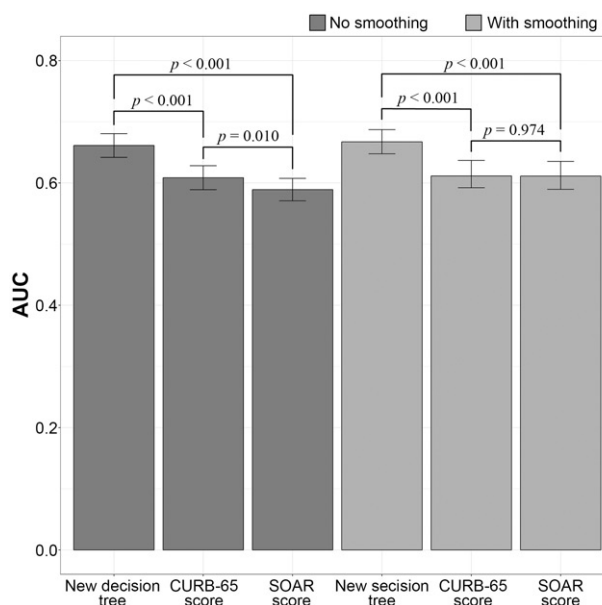


Figure 3. ROC curve AUC pairwise comparisons.

non-severe CAP patients as identified by each of the three scores (Table 3). For comparing, all three scores were dichotomized into severe and non-severe categories. Category cut-offs for CURB-65 and SOAR were derived from Subramanian et al. [19]. The third quartile ($Q_3 = 3$) was used as cut-off for the new CART model. Results of regression analysis indicate that in all three scoring systems, severe CAP patients have significantly higher probabilities for 30-d mortality compared to non-severe patients. However, among severe patients, the OR of the new CART model was higher than the OR for CURB-65 and SOAR scores. Pseudo r-squared measures were also higher for the CART logistic regression model compared to that for CURB-65 and SOAR score regression models, indicating that the new CART model is better at predicting 30-d mortality between severe and non-severe CAP patients.

Discussion

This study evaluated the effectiveness of a newly developed prediction model to assess 30-d mortality of CAP patients in the ICU specifically, and the results were compared to those of two existing scoring systems, CURB-65 and SOAR score. Since CAP is an exceptionally severe disease that often requires ICU admission and has an extremely high 30-d mortality, we assumed that CAP severity in patients already admitted to ICU would likely correlate with 30-day mortality. Most of the existing scoring systems were designed originally to predict severity but not necessarily admission to ICU [7]. Of the two comparator

Table 3. Probability of 30-d mortality for severe vs. non-severe CAP patients for CART risk model, CURB-65 score and SOAR score.

	OR	95% CI	p Value
New decision tree			
Non-severe (<4)	Ref.		
Severe (≥ 4)	2.802	2.346–3.347	<.001
CURB-65 score			
Non-severe (<3)	Ref.		
Severe (≥ 3)	1.960	1.653–2.325	<.001
SOAR score			
Non-severe (<2)	Ref.		
Severe (≥ 2)	1.779	1.458–2.169	<.001
	–2LL	Cox and Snell's r^2	Nagelkerke's r^2
New decision tree	3982.9	0.035	.051
CURB-65 score	4050.7	0.016	.024
SOAR score	4077.6	0.009	.013

CURB-65 and SOAR score severity cut-off determined according to Subramanian et al. [19]. The third quartile ($Q_3=3$) was used as cut-off for the new decision tree CART risk model.

95% CI: 95% confidence interval.

–2LL: –2 log likelihood.

models in this study, CURB-65 is used primarily to determine whether CAP patients require hospitalization and SOAR is used to assess risk in hospitalized patients whether in ICU or not. Results of this study show that the newly designed CART model has higher specificity and a better ROC curve compared to those of the existing CURB-65 and SOAR score indices. Logistic regression analysis revealed that in all three models severe CAP patients have significantly higher probabilities for 30-d mortality compared to non-severe patients, but the new CART model is better at predicting 30-d mortality between severe and non-severe CAP patients. However, the sensitivity and accuracy of the new decision-tree model were only moderately better than those of the two existing models. Nevertheless, after smoothing, both the sensitivity and specificity of the new CART model were higher than those of CURB-65 and SOAR score indices.

The new prediction model was built on a foundation of decision-tree analysis, following the CART analysis described previously by Brims et al. [20] The CART approach has been used previously for simple prognostic models in acetaminophen-induced acute liver failure and is reported to offer improved sensitivity and model performance even though accuracy and specificity were shown to be equal or “negligibly worse” than other prognostic models for this clinical condition [21]; while the CART model offered only slightly better predictive accuracy at admission of patients with acute liver failure, higher predictive accuracy was found later during post-admission. The CART prediction model examines the interaction of multiple variables with a given outcome. In developing the new model for CAP, we took more risk factors into consideration compared to

existing systems. CURB-65, for example, uses eight risk factors in assessing severity, and SOAR uses the four factors on which it is based – systolic blood pressure, oxygenation, age and respiratory rate. In developing the new index in this study, univariate analysis identified 12 risk factors associated with 30-d mortality, including age, invasive mechanical ventilation use, septic shock needing vasopressor support, respiratory rate, BUN level, thrombocytopenia, hypothermia, hypotension, coexisting illnesses, ABG test, haematocrit and positive *Streptococcus pneumoniae*. In general, expanding the risk factors is intended to improve the ability of the index to identify patients at high risk for 30-d mortality. However, the number of included indicators may not make a significant difference in determining prognosis for all clinical conditions; use of a decision tree model for acute pancreatitis had only three clinical indicators and, though deficiencies of the indicators were noted, sensitivity and specificity were 88.6 and 90.0%, respectively, and the authors did not suggest expanding the indicators in future system development [22]. Nevertheless, since the risk factors we identified were also associated with 30-d mortality, we expected the expansion to help promote good performance of the new predictive model in terms of accuracy, sensitivity and specificity, which would indicate an enhanced ability to identify patients at highest risk. In this study, the purpose was to determine 30-d mortality in ICU patients, but not to determine which patients should be admitted to ICU. Another study of mortality prediction among older adults hospitalized for CAP expanded the CURB score inventory of risk factors with age and comorbidities, including confusion, urea, respiratory rate and blood pressure, reporting finally that predictive accuracy for 30-d mortality was comparable to that of PSI [23]. It must be noted, however, that the new scoring model was evaluated in severely ill patients who were all hospitalized in ICU. Therefore, it can be expected that the new decision tree model with an expanded risk factor inventory will perform somewhat better than the two scores with fewer indicators of severity, CURB-65 and SOAR, which were used for comparison.

The ROC curve is a prime outcome for assessing predictability. Area under the ROC curve, or AUC, indicates the discrimination capability in a logistic regression model. That is, knowing the value of the predicted probability of a specific outcome such as 30-d mortality occurring allows us to establish a threshold. In this study, the ROC of the newly developed CART system was better than the ROC of either CURB-65 or SOAR, indicating that the new system is

better able to predict 30-d mortality in CAP patients. Sensitivity and specificity, however, must be determined to test the predictive ability indicated by ROC values. Sensitivity identifies true positives (true positives + false negatives) [24]. If the test methodology, in this case the CART evaluation, is highly sensitive and the test result is negative, it is nearly certain that the outcome (30-d mortality) has not been predicted. Specificity identifies true negatives (true negative + false positives) [19]. If the test methodology is highly specific and the test result is positive, it is nearly certain that the outcome (30-d mortality) has been predicted. In this case, accuracy, sensitivity and specificity refer to the ability of the scoring system to identify patients with CAP who are at high risk of mortality within 30 d. Results of this study showed that sensitivity of the CART index was only somewhat, or moderately, better than sensitivity of the CURB-65 and SOAR systems. In another study, pooled sensitivity for CURB-65 was only 49% and pooled specificity for PSI was only 48%, indicating that neither scoring system was sufficiently accurate to predict 30-d mortality [25]. On the other hand, specificity of the newly designed CART system was superior to that of CURB-65 and SOAR scores. CURB-65 was reported by Liu et al. [7] to have a deficiency in predictive specificity. In predicting CAP 30-d mortality, such a deficiency in specificity can result in classifying patients incorrectly as low risk. Although it may be reasonable to suggest that predicting 30-d mortality should be sufficiently accurate to determine if intensive management strategies are needed, all patients in this study were already in ICU and the goal was to identify 30-d mortality, not whether intensive management was needed.

Kolditz et al. [3] have suggested that accurate mortality prediction may not always identify patients who are likely to develop severe CAP and who should be admitted to ICU for intensified management. Indeed, this study aimed to predict clinical outcomes, specifically 30-d mortality, of patients with CAP admitted to ICU rather than determining whether ICU admission was necessary. Although admission to ICU is consistent with high-risk prediction, all scoring systems are not able to predict high risk in all patients. This is further complicated by the policies of the ICU in different medical centres and the decisions of individual physicians. One study, for example, showed that 35% of patients had contraindications for ICU admission according to the criteria applied by the hospital [1]. In a review study that evaluated several severity scales for CAP, the authors noted that PSI, CRB-65 and CURB-65 had different strengths and weaknesses, with the greatest

common strength being good negative predictive values for mortality in populations that have a relatively low prevalence of death [26]. In a clinical situation like CAP, however, which has a high mortality rate, more accurate risk classification is needed to support appropriate management decisions whether patients are in ICU or not. Acknowledging that disease severity and chronic pulmonary disease are strong predictors of ICU admission, Vohra et al. [27] found that every 10-point increase in the PSI index was predictive of 20.7% increased odds of ICU admission. Consequently, those authors suggested that greater use of severity indices such as PSI may optimize management decisions and ultimately minimize mortality in the ICU setting where length of stay and mortality are already significantly higher than in the general medicine setting.

This study has associated strengths and limitations. We used the high-quality MIMIC-III database that encompasses a diverse and exceptionally large population of ICU patients. It provided high temporal resolution data, including lab results, electronic documentation, and bedside monitoring trends and waveforms, which all helped to give credence to our analysis and results. Nevertheless, the present study has several limitations. First, data were obtained from only a single medical centre in Boston, MA, which means that our results may not be generalizable to other populations in other locations. Second, the MIMIC database does not include lifestyle and dietary information, environmental exposure, or family medical history, all factors that may have influenced our results if they had been included in analysis. Although CAP was defined for this study based on Kaplan et al. [12], the MIMIC III database generates diagnosis codes at the end of the hospital stay so this study could not include the duration of all cases of CAP in its analyses and the exclusion criteria could not identify healthcare-associated pneumonia. Also, no information was available on medical utilization after discharge, only in-hospital data during patients' ICU stays were included. All data for the present study came from ICU patients, which is a well-defined portion of CAP patients, but in clinical practice, ICU patients do not represent all CAP patients and other populations must be studied. Further prospective, multicentre study is still needed to further validate the newly developed CART model, and independent risk factors must be expanded to include those lacking as variables in this study.

Conclusions

The new CART prediction model has higher specificity and better ROC curves than CURB-65 and SOAR score

indices, although its accuracy and sensitivity are only moderately better than the other systems in predicting 30-d mortality in CAP patients admitted to ICU. CART is a simple, effective system for assessing the severity of CAP patients admitted to ICU, however, ICU patients do not represent all CAP patients and the CART model must be studied in other populations. Further prospective study is needed with a large sample from multiple settings to corroborate results of this initial development study.

Ethical approval

Use of the MIMIC-III database for research purposes has been approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center and Massachusetts Institute of Technology. Since patient data in MIMIC-III database are deidentified, signed informed consent of subjects is not required for this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by grants from the National Natural Science Foundation of China (No. 81570017, GS Zhang) and the Medical and Health Research Program of Zhejiang Province (Core Talents Plan) of Zhejiang Province [No. 2016RCA014, GS Zhang].

References

- [1] Chalmers JD, Taylor JK, Mandal P, et al. Validation of the infectious diseases society of America/American thoracic society minor criteria for intensive care unit admission in community-acquired pneumonia patients without major criteria or contraindications to intensive care unit care. *Clin Infect Dis*. 2011;53: 503–511.
- [2] Shah BA, Ahmed W, Dhobi GN, et al. Validity of pneumonia severity index and CURB-65 severity scoring systems in community acquired pneumonia in an Indian setting. *Indian J Chest Dis Allied Sci*. 2010;52: 9–17.
- [3] Kolditz M, Ewig S, Höffken G. Management-based risk prediction in community-acquired pneumonia by scores and biomarkers. *Eur Respir J*. 2013;41:974–984.
- [4] Bauer TT, Ewig S, Marre R, CAPNETZ Study Group, et al. CRB-65 predicts death from community-acquired pneumonia. *J Intern Med*. 2006;260:93–101.
- [5] Ewig S, Birkner N, Strauss R, et al. New perspectives on community-acquired pneumonia in 388,406 patients. Results from a nationwide mandatory performance measurement programme in healthcare quality. *Thorax*. 2009;64:1062–1069.
- [6] Gonzalez C, Johnson T, Rolston K, et al. Predicting pneumonia mortality using CURB-65, PSI, and patient characteristics in patients presenting to the emergency department of a comprehensive cancer centre. *Cancer Med*. 2014;3:962–970.
- [7] Liu JL, Xu F, Zhou H, et al. Expanded CURB-65: a new score system predicts severity of community-acquired pneumonia with superior efficiency. *Sci Rep*. 2016;6: 22911.
- [8] Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336:243–250.
- [9] Charles PG, Wolfe R, Whitby M, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis*. 2008;47:375–384.
- [10] Fang WF, Yang KY, Wu CL, et al. Application and comparison of scoring indices to predict outcomes in patients with healthcare-associated pneumonia. *Crit Care*. 2011;15:R32.
- [11] Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016; 3:160035.
- [12] Kaplan V, Angus DC, Griffin MF, et al. Hospitalized community-acquired pneumonia in the elderly: age- and sex-related patterns of care and outcomes in the United States. *Am J Respir Crit Care Med*. 2002;165: 766–772.
- [13] Myint PK, Kamath AV, Vowler SL, et al. Severity assessment criteria recommended by the British Thoracic Society for community-acquired pneumonia (CAP) and older patients. Should SOAR (systolic blood pressure, oxygenation, age and respiratory rate criteria be used in older people? A compilation study of two prospective cohorts. *Age Ageing*. 2006;35:286–291.
- [14] Fonarow GC, Adams KF, Abraham WT, et al. ADHERE Scientific Advisory Committee, Study Group, and Investigators FT. Risk stratification for in-hospital mortality in acutely decompensated heart failure classification and regression tree analysis. *JAMA*. 2005;293: 572–580.
- [15] Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–35.
- [16] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
- [17] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. [cited 2018 May 20]. Available from: <https://www.R-project.org/>
- [18] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze ROC curves. *BMC Bioinformatics*. 2011;12:77. doi: 10.1186/1471-2105-12-77.
- [19] Subramanian DN, Musonda P, Sankaran P, et al. Performance of SOAR (systolic blood pressure, oxygenation, age, respiratory rate) scoring criteria in community-acquired pneumonia: a prospective multi-centre study. *Age Ageing*. 2013;42:94–97.

- [20] Brims FJ, Meniawy TM, Duffus I, et al. A novel clinical prediction model for prognosis in malignant pleural mesothelioma using decision tree analysis. *J Thorac Oncol.* 2016;11:573–582.
- [21] Speiser JL, Lee WM, Karvellas CJ, et al. Predicting outcome on admission and post-admission for acetaminophen-induced acute liver failure using classification and regression tree models. *PLoS One.* 2015;10:e0122929.
- [22] Yang Z, Dong L, Zhang Y, et al. Prediction of severe acute pancreatitis using a decision tree model based on the revised Atlanta Classification of Acute Pancreatitis. *PLoS One.* 2015;10:e0143486.
- [23] Abisheganaden J, Ding YY, Chong WF, et al. Predicting mortality among older adults hospitalized for community-acquired pneumonia: an enhanced confusion, urea, respiratory rate and blood pressure score compared with pneumonia severity index. *Respirology.* 2012;17:969–975.
- [24] Emory University. Sensitivity and specificity. [Cited 2017 Oct 1]. Available from: <https://www.med.emory.edu/EMAC/curriculum/diagnosis/sensand.htm>
- [25] Chalmers JD, Mandal P, Singanayagam A, et al. Severity assessment tools to guide ICU admission in community-acquired pneumonia: systematic review and meta-analysis. *Intensive Care Med.* 2011;37:1409–1420.
- [26] Loke YK, Kwok CS, Niruban A, et al. Value of severity scales in predicting mortality from community-acquire pneumonia: a systematic review and meta-analysis. *Thorax.* 2010;65:884–890.
- [27] Vohra AS, Tak HJ, Shah MB, et al. Intensive care unit admission with community-acquired pneumonia. *Am J Med Sci.* 2015;350:380–386.