BMJ Open

Evaluation of prognostic utility of serum CRP levels in combination with CURB-65 and validation of the CURB-65 for prediction of mortality in patients with clinically suspected sepsis using decision curve analysis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007049
Article Type:	Research
Date Submitted by the Author:	29-Oct-2014
Complete List of Authors:	Yamamoto, Shungo; School of Public Health in the Graduate School of Medicine, Kyoto University, Department of Healthcare Epidemiology; Kobe University Hospital, Division of Infectious Diseases Yamazaki, Shin; Kyoto University, Department of Healthcare Epidemiology Shimizu, Tsunehiro; Kyoto City Hospital, Department of Infectious Diseases Takeshima, Taro; Jichi Medical University, Division of Community and Family Medicine, Center for Community Medicine; Kyoto University, Department of Healthcare Epidemiology Fukuma, Shingo; Kyoto University, Department of Healthcare Epidemiology Yamamoto, Yosuke; Kyoto University, Department of Healthcare Epidemiology Tochitani, Kentaro; Kyoto City Hospital, Department of Infectious Diseases Tsuchido, Yasuhiro; Kyoto City Hospital, Department of Infectious Diseases Shinohara, Koh; National Center for Global Health and Medicine, Disease Control and Prevention Center Fukuhara, Shunichi; Kyoto University, Department of Healthcare Epidemiology
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Diagnostics, Emergency medicine, Medical management
Keywords:	PRIMARY CARE, STATISTICS & RESEARCH METHODS, INFECTIOUS DISEASES

Evaluation of prognostic utility of serum CRP levels in combination with CURB-65 and validation of the CURB-65 for prediction of mortality in patients with clinically suspected sepsis using decision curve analysis

Shungo Yamamoto MD, DTM&H[1], Shin Yamazaki PhD[1], Tsunehiro Shimizu MD, PhD[2], Taro Takeshima MD[1,3], Shingo Fukuma MD, PhD[1,5], Yosuke Yamamoto MD, PhD[1], Kentaro Tochitani MD[2], Yasuhiro Tsuchido MD[2], Koh Shinohara MD[4], Shunichi Fukuhara MD, DMSc[1,5]

[1]Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Infectious Diseases, Kyoto City Hospital, Kyoto, Japan

[3]Division of Community and Family Medicine, Center for Community Medicine, Jichi Medical University, Tochigi, Japan

[4]Disease Control and Prevention Center, National Center for Global Health and Medicine, Tokyo, Japan

[5]Center for Innovative Research for Communities and Clinical Excellence (CIRC2LE), Fukushima Medical University, Fukushima, Japan

Corresponding author:

Shunichi Fukuhara

Department of Healthcare Epidemiology

School of Public Health in the Graduate School of Medicine

Kyoto University

Yoshida Konoemachi

Sakyo-ku, Kyoto, 606 8501, Japan

E-mail: fukuhara.shunichi.6m@kyoto-u.ac.jp

Phone: +81-75-753-4646, Fax: +81-75-753-4644

Conflict of interests

All authors certify that they have no conflict of interests.

Keywords: CURB-65; CRP; sepsis; decision curve analysis; net benefit

Abstract

Objectives: The prognostic utility of serum C reactive protein (CRP) alone in sepsis is controversial. Here, we evaluated the prognostic utility of serum CRP levels in combination with the CURB-65 score.

Design: Retrospective cohort study.

Setting: Emergency department (ED) of an urban teaching hospital in Japan.

Participants: Consecutive ED patients over 15 years old who were admitted to the hospital after having a blood culture taken in the ED between 1th January 2010 and 31th December 2012.

Main outcome measures: 30-day in-hospital mortality.

Results: 1262 patients ultimately analyzed for score evaluation. The 30-day in-hospital mortality was 8%. Multivariable analysis showed that serum CRP ≥150 mg/L was an independent predictor of death (adjusted odds ratio: 1.8; 95% confidence interval [CI]: 1.1 to 2.8). We assessed the predictive performance of the CURB-65 as well as a modified CURB-65 with CRP (≥15mg/dL) added. The area under the receiver operating characteristics curve (AUC) of the CURB-65 and modified CURB-65 were 0.76 (95% CI: 0.72 to 0.80) and 0.77 (95% CI: 0.72 to 0.81), respectively. Both models presented good calibration for mortality and were useful among threshold probabilities in the range of 0% to 30%. However, while incorporating CRP into CURB-65 yielded a significant category-free net reclassification improvement (NRI) of 0.387 (95% CI: 0.193 to 0.582) and integrated discrimination improvement (IDI) of 0.015 (95% CI: 0.004 to 0.027), decision curve analyses showed the CURB-65 and the modified CURB-65 scores had comparable net benefits for

prediction of mortality.

Conclusions: Measurement of serum CRP added limited clinical usefulness to the CURB-65 in predicting mortality in patients with clinically suspected sepsis, regardless of the source.

Article summary:

Strengths and limitations of this study

- A major strength of this study is using decision curve analysis for evaluation of clinical usefulness. That revealed the additive clinical usefulness of CRP to the CURB-65 was limited, although reclassification metrics such as net reclassification improvement and integrated discrimination improvement were improved.
- The CURB-65 was validated in Asian population with clinically suspected infection, regardless of the source, with adequate sample size as an external validation study. The AUC was 0.76 (95%CI: 0.72 to 0.80) and it was useful among threshold probabilities in the range of 0% to 30%. Our data enhanced the generalizability of the CURB-65 score.
- The limitations of this study are the possibility of selection bias of the eligible patients and the retrospective nature of the study in a single hospital.

Introduction

Sepsis is a major cause of morbidity and mortality worldwide, with in-hospital mortality still at 18% or more, according to recent surveys in resource-rich countries.[1-3] Early identification of high-risk patients and timely intervention for sepsis are therefore crucial to improving outcomes.

Severity assessment is important in the management of patients, including decision-making regarding choice of treatment and patient disposition. To encourage implementation, a clinical prediction rule must be user-friendly.[4] The CURB-65 is a simple prediction rule originally developed as a prognostic scoring system for community-acquired pneumonia (CAP).[5] The rule has been well validated in patients with CAP, and CURB-65-guided antibiotic therapy has safely reduced broad-spectrum antibiotic use in this population.[6 7] In addition to its utility among CAP patients, CURB-65 has also been correlated with mortality in patients with suspected sepsis, regardless of the source, and in patients admitted for non-surgical illness in the United States, Spain, and the United Kingdom;[8-10] however, the system has not yet been validated in such populations in Asian countries.

Serum C reactive protein (CRP) is an acute phase protein often evaluated as a marker of systemic inflammation.[11] In Japan, serum CRP levels have been used as a diagnostic and prognostic marker of infection in daily clinical practice and clinical trials of new drugs.[12] However, evidence demonstrating its value is insufficient at present for routine application of serum CRP levels to assess severity of infection. As a prognostic marker, some have reported that serum CRP on admission is associated with mortality.[13 14] However, a systematic review reported conflicting findings, noting that serum CRP levels were not significantly different

between survivor and nonsurvivor, suggesting that these levels may have limited value in reflecting the severity of sepsis.[11] As a diagnostic marker, the sensitivity and specificity of serum CRP for discriminating bacterial from non-infectious inflammation was only 75% and 67%, respectively, according to a meta-analysis.[15] However, while the diagnostic performance of serum CRP alone is limited, serum CRP has been reported to contribute some additional information to a prediction rule involving a patient's symptoms and physical examination in diagnosis of pneumonia.[16] In this respect, the additive prognostic value of serum CRP to an existing severity score is unknown.

Here, we evaluated the prognostic performance of serum CRP in combination with the CURB-65 score. In addition, we also validated the CURB-65 score in patients with suspected sepsis, regardless of the source of infection, as a geographical and domain validation.[5]

Materials and Methods

Study Design, Setting and Patients

We performed a retrospective cohort study at Kyoto City Hospital, an urban teaching hospital with 548 beds in Japan. Consecutive emergency department (ED) patients over 15 years old admitted to the hospital after having a blood culture taken in the ED between 1 January 2010 and 31 December 2012 were included. Doctors' decision to order a blood culture was used as a surrogate marker for suspected sepsis as in previous studies.[10 17] To facilitate data independence, only the index admission was included for patients with multiple admissions during the study period. Patients transferred from another hospital or who had cardiopulmonary arrest on arrival at the hospital were excluded.

Data collection

The following data were extracted from electronic medical records for each patient: age, gender, underlying disease, vital signs (heart rate, blood pressure, respiratory rate, mental confusion and body temperature), laboratory findings (white blood cell [WBC] count, platelet count, and blood urea nitrogen [BUN] and serum CRP levels), and outcome. For vital signs and laboratory data, initial values at the hospital visit were recorded. The items of the CURB-65 score were as follows: mental confusion (present/absent), BUN > 20 mg/dL (7 mmol/L), respiratory rate (RR) ≥ 30/min, either or both systolic blood pressure (SBP) < 90 mmHg or diastolic blood pressure (DBP) ≤ 60 mmHg, and age ≥ 65 years.[5] Mental confusion was defined as disorientation in person, place or time or being in a stupor or coma.[5] We defined other covariates as follows: abnormal body temperature (BT) as > 38 or < 36 ℃, tachycardia as heart rate (HR) > 90/min, leukocytosis as WBC > 15,000/μL, and low platelet count as platelet count < 150,000/μL, based on a previous study.[17]

The main outcome measure was 30-day in-hospital mortality. Patients who were discharged or transferred from the hospital within 30 days of admission or who remained in the hospital for more than 30 days were considered alive in this analysis.[18]

Statistical analysis

First, we explored the cut-off point of serum CRP level for prediction of death in patients with suspected sepsis. We graphically checked whether or not the relationship between serum CRP level and mortality was linear in the logit with a smoothing curve using a locally weighted

least squares (Lowess) regression.[19] Because the optimal cut-off point was unknown, serum CRP results were first divided into quartiles and rounded to the nearest integer. Each patient was then assigned to one of four categories corresponding to the CRP quartiles. We conducted multivariable logistic regression analysis to adjust predictors of death by sequentially introducing pre-specified groups of variables, such as demographics, vital signs, laboratory findings and items of the CURB-65. We assessed multicollinearity using the variance inflation factor (VIF). VIFs greater than 2.5 may be problematic.[20] We also computed the unadjusted odds ratio of covariates using univariable logistic regression to show the influence of adjustment for predictors.

Second, we validated the CURB-65 model. We graphically assessed the calibration of the CURB-65 model with calibration plot and tested it with the Hosmer-Lemeshow test. A p value < 0.05 indicates a lack of good fit for the model. Regarding the model discrimination, we also computed the area under the receiver operating characteristic curve (AUC) with a 95% confidence interval (CI) using 500 bootstrap resampling.[21] The predicted mortalities with 95% CI were calculated by introducing the CURB-65 score as a continuous variable into univariable logistic regression.

Third, we assessed the model performance of the modified CURB-65 score, which was made by incorporating CRP information into the CURB-65 model, using a calibration plot and Hosmer-Lemeshow test for calibration and AUC for discrimination. Additive information of CRP was evaluated by category-free net reclassification improvement (NRI) and integrated discrimination improvement (IDI).[22] With regard to clinical usefulness, we examined the net benefit using decision curve analysis,[23] a novel method of evaluating diagnostic tests or

prediction models established by Vickers and Elkin. Briefly, the method is based on the principle that the relative harms of false positives and false negatives can be expressed in terms of a probability threshold.[24] The net benefit is obtained by subtracting the proportion of patients who are false positive from the proportion who are true positive, weighting by the relative harm of a false-positive and a false-negative result. The net benefit of making a decision based on the model can be calculated with the following formula:

$$\text{Net benefit} = \frac{True\ Positives}{n} - \left(\frac{Pt}{1-Pt}\right)\frac{False\ positives}{n}$$

where n is the total number of patients in the study and P_t is a given threshold probability.[23-25]

With regard to sample size estimation, at least 8 to 10 events per variable are required for reliable multiple logistic regression analysis,[26 27] and 100 events and 100 nonevents are required for an external validation study.[28] We assumed 30 to 40 deaths per year among eligible patients, collecting 3 years' worth of data (90 to 120 estimated deaths) to appropriately conduct multiple logistic regressions with 11 variables and ensure adequate statistical power.

In terms of handling missing values, we planned to perform a complete case analysis if missing values were below 5%, as such an analysis might be feasible then. [29] If missing values were above 5%, we planned to apply an appropriate imputation method.

Data were analyzed with R software 3.0.1 (www.r-project.org) and Stata software, version 13 (StataCorp., College Station, TX, USA) including programs of Decision Curve Analysis provided by Vickers.[30]

Ethical approval

This protocol was approved by the Ethics Committee of Kyoto University Graduate School and Faculty of Medicine and Kyoto City Hospital.

Results

Patient characteristics

Among 1310 eligible patients over 3 years of study, 108 deaths (8.2% mortality) were recorded. Demographics, underlying diseases, vital signs, laboratory findings and chief diagnosis for admission were presented in Table 1. Diagnosis was unclear in 96 patients (7.3%). Respiratory rates data were missing for 21 patients, and CRP data were missing for 28 (Table 2). Overall cases with any missing predictor were 48 (3.7%), so we conducted a complete case analysis, leaving 1262 patients (106 deaths, 8.4% mortality) for model evaluation analyses.

Evaluation of CRP as a predictor of mortality

The relationship between serum CRP level and mortality was almost linear in the logit. Serum CRP results were divided into quartiles and rounded to the nearest integer, and we set interim cut-off points as 20, 70, and 150 mg/L. Unadjusted odds ratios for mortality of each CRP group and other covariates are shown in Table 2. We set 150 mg/L as the cut-off point to dichotomize serum CRP levels and then sequentially introduced groups of variables into the multivariable logistic regression model.

Adjusted odds ratios for mortality are shown in Table 3. We found no evidence of multicollinearity because the variance inflation factors for predictors in all models in Table 3 were less than 1.3. We identified serum CRP level ≥ 150 mg/L as an independent predictor of death in patients with clinically suspected sepsis.

Validation of the CURB-65

Observed 30-day in-hospital mortalities and predicted mortalities computed by the CURB-65 score are shown in Table 4. The CURB-65 showed good calibration for mortality, with a Hosmer-Lemeshow test of 4.08 (df=5, P=0.538), indicating good fit. The calibration plot is shown in Supplementary Figure 1a. The AUC for the CURB-65 was 0.76 (95% CI: 0.72 to 0.80; Supplementary Figure 2).

Additive information of CRP to CURB-65

Because the adjusted odds ratio (i.e. regression coefficient) of CRP was comparable to each item in the CURB-65, we made a modified CURB-65 score by adding one point to the CURB-65 score when serum CRP level was ≥ 150 mg/L. Table 4 shows observed 30-day in-hospital mortalities and predicted mortalities stratified by the modified CURB-65 score. The modified CURB-65 also showed good calibration for mortality, with a Hosmer-Lemeshow test of 4.52 (df=6, P=0.607). The calibration plot is shown in Supplementary Figure 1b. The AUC for the modified CURB-65 was 0.77 (95% CI: 0.72 to 0.81; Supplementary Figure 2). By incorporating CRP into CURB-65, event NRI was −0.151 and nonevent NRI was 0.538, giving an overall category-free NRI of 0.387 (95% CI: 0.193 to

0.582, p<0.0001). Further, IDI for events was 0.014 and IDI for nonevents was 0.001, giving an overall IDI of 0.015 (95% CI: 0.004 to 0.027, p<0.01). These findings were statistically significant.

Decision curve analysis

Figure 1 demonstrates the decision curve for the CURB-65 and modified CURB-65 to predict 30-day in-hospital mortality in patients with clinically suspected sepsis. Both the CURB-65 and the modified CURB-65 are useful between threshold probabilities of 0% to 30%. However, both curves depict little difference in net benefit. The net benefits at each point of the CURB-65 score and the modified CURB-65 score are shown in Table 4. The summary of findings of this study is presented in Table 5.

Discussion

We determined that having high CRP levels was independently associated with high mortality in our population. We also conducted geographical and domain validation of the CURB-65 in patients with clinically suspected sepsis, which comprised an external validation in a different geographical area and in a population including different category of patients from CAP.[4] Although reclassification metrics such as NRI and IDI were improved on incorporation of CRP level, the additive clinical usefulness of CRP to the CURB-65 was admittedly limited (Table 5).

The utility of serum CRP as a prognostic marker has been found to vary.[11] In Japan,

universal health coverage allows people to consult a doctor soon after they recognize any symptoms, with no particular limitations.[31] Given that secretion of CRP peaks at 36-50 h after inflammatory stimulus,[11] serum CRP level might be useful as a surrogate marker of duration from disease onset to consulting a doctor as well as a marker reflecting intensity of inflammation. We believe that the association between serum CRP level and mortality will be more easily identified in countries such as Japan where the population has easy access to hospitals, due to the wide distribution in duration from disease onset to visiting a hospital.

Strengths and limitations of the study

A major strength of this study was our evaluation using decision curve analysis. Performance of a prediction model has traditionally been evaluated by discrimination and calibration. Discrimination is "a measure of how well a prognostic model discriminates individuals with and without the outcome of interest" and calibration is "the ability to correctly estimate the likelihood of a future event across the whole range of prognostic estimates."[32] There is a tradeoff between discrimination and calibration, and a model typically cannot be perfect with respect to both.[33 34] Calibration is more relevant than discrimination when estimating absolute risk of mortality in decision-making. However, having good calibration alone is not sufficient to show that a model would improve decision-making.[35] As metrics of reclassification, NRI and IDI have enjoyed increasing usage in evaluating improvement in prediction models. However, these improvements in NRI and IDI are also not sufficient for evaluating clinical usefulness.[36-38] Decision curve analysis can take into account risk threshold, weighting benefits and harms, and is useful in evaluating clinical utility of a

prediction model.[23 38 39] For instance, the net benefit of 0.061 at a threshold probability of 3% can be interpreted as meaning that making a decision based on the prediction model, compared to assuming that all patients would be alive, leads to the equivalent of a net 6.1 true-positive results per 100 patients with no corresponding increase in the number of false-positive results (Table 4).[23] Addition of CRP to the CURB-65 improved the prognostic classification in terms of NRI and IDI; however, its clinical usefulness was limited when considering net benefit examined by decision curve analysis.

Several limitations to the present study warrant mention. We cannot rule out the possibility of selection bias, as only inpatients who had a blood culture taken were included. We may therefore have missed patients with infection who did not undergo blood culture in the ED, and contrarily, some patients without infection were included in the study. However, clinicians must routinely make decisions in the ED despite being unsure as to whether or not a patient is actually infected; we therefore considered it important to evaluate a clinical prediction rule accounting for such clinical uncertainty. Inclusion of patients without infectious diseases, we feel, reflects a real-world scenario. Another limitation is the retrospective nature of the study and the fact that it was conducted in a single hospital. Given the study's retrospective design, patients with high CRP might have received more intensive therapy than those with relatively low CRP. Such bias might have lowered the predictive ability of CRP. However, as an external validation study, we believe our data enhanced the generalizability of the CURB-65 score with adequate statistical power.[28]

Conclusions

The CURB-65 correlated well with 30-day in-hospital mortality in patients with clinically suspected infection, and it was useful among threshold probabilities in the range of 0% to 30%. While serum CRP level ≥ 150 mg/L was found to be associated with high mortality, its clinical usefulness in combination with the CURB-65 was limited. CURB-65 score may prove helpful in making decisions regarding the management of clinically suspected sepsis, regardless of the source.

Contributors

S. Yamamoto had full access to all of the data in the study and takes responsibility for integrity of the data and accuracy of the data analysis and wrote the first draft.

T. Shimizu, K. Tochitani, Y. Tsuchido and K. Shinohara collected and interpreted the data, and drafted the paper.

S. Yamazaki, T. Takeshima, S. Fukuma, Y. Yamamoto and S. Fukuhara supervised the research, interpreted the data and contributed to the writing of the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests

None.

Ethics approval

The Ethics Committee of Kyoto University Graduate School and Faculty of Medicine and Kyoto City Hospital.

Data sharing statement

No additional data are available.

References

1. Kumar G, Kumar N, Taneja A, et al. Nationwide trends of severe sepsis in the 21st century (2000-2007). Chest 2011;**140**:1223-31.

2. Kaukonen K-M, Bailey M, Suzuki S, et al. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. JAMA 2014;**311**:1308-16.

3. Fujishima S, Gando S, Saitoh D, et al. A multicenter, prospective evaluation of quality of care and mortality in Japan based on the Surviving Sepsis Campaign guidelines. Journal of infection and chemotherapy 2014;**20**:115-20.

4. Toll DB, Janssen KJM, Vergouwe Y, et al. Validation, updating and impact of clinical

prediction rules: a review. Journal of clinical epidemiology 2008;**61**:1085-94.

5. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;**58**:377-82.

6. Chalmers JD, Singanayagam A, Akram AR, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. Thorax 2010;**65**:878-83.

7. Chalmers JD, Singanayagam A, Akram AR, et al. Safety and efficacy of CURB65-guided antibiotic therapy in community-acquired pneumonia. Journal of Antimicrobial Chemotherapy 2011;**66**:416-23.

8. Howell MD, Donnino MW, Talmor D, et al. Performance of severity of illness scoring systems in emergency department patients with infection. Academic emergency medicine 2007;**14**:709-14.

9. Armiñanzas C, Velasco L, Calvo N, et al. CURB-65 as an initial prognostic score in Internal Medicine patients. European journal of internal medicine 2013;**24**:416-19.

10. Marwick CA, Guthrie B, Pringle JE, et al. Identifying which septic patients have increased mortality risk using severity scores: a cohort study. BMC anesthesiology 2014;**14**:1.

11. Mitaka C. Clinical laboratory differentiation of infectious versus non-infectious systemic inflammatory response syndrome. Clinica chimica acta 2005;**351**:17-29.

12. Saito A, Miki F, Oizumi K, et al. Clinical evaluation methods for new antimicrobial agents to treat respiratory infections: Report of the Committee for the Respiratory

System, Japan Society of Chemotherapy. Journal of infection and chemotherapy 1999;**5**:110-23.

13. Chalmers JD, Singanayagam A, Hill AT. C-reactive protein is an independent predictor of severity in community-acquired pneumonia. The American journal of medicine 2008;**121**:219-25.

14. Lobo SMA, Lobo FRM, Bota DP, et al. C-reactive protein levels correlate with mortality and organ failure in critically ill patients. Chest 2003;**123**:2043-49.

15. Simon L, Gauvin F, Amre DK, et al. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clinical infectious diseases 2004;**39**:206-17.

16. van Vugt SF, Broekhuizen BDL, Lammens C, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. BMJ 2013;**346**:f2450-f50.

17. Shapiro NI, Wolfe RE, Moore RB, et al. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. Critical Care Medicine 2003;**31**:670-75.

18. Graham PL, Cook DA. Prediction of risk of death using 30-day outcome: a practical end point for quality auditing in intensive care. Chest 2004;**125**:1458-66.

19. Hosmer DW. *Applied Logistic Regression*. 3 ed. Hoboken, New Jersey: Wiley, 2013.

20. Katz MH. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*. 3rd ed. New York: Cambridge University Press, 2011.

21. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology 2003;**56**:441-47.

22. Cook NR, Paynter NP. Performance of reclassification statistics in comparing risk prediction models. Biometrical journal 2011;**53**:237-58.

23. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical decision making 2006;**26**:565-74.

24. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC medical informatics and decision making 2008;**8**:53.

25. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. The American statistician 2008;**62**:314-20.

26. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology 1996;**49**:1373-79.

27. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. American journal of epidemiology 2003;**158**:280-87.

28. Vergouwe Y, Steyerberg EW, Eijkemans MJC, et al. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of clinical epidemiology 2005;**58**:475-83.

29. Royston P, Moons KGM, Altman DG, et al. Prognosis and prognostic research:

Developing a prognostic model. BMJ 2009;**338**:b604.

30. Vickers AJ. Decision Curve Analysis. 2013. http://www.mskcc.org/research/epidemiology-biostatistics/health-outcomes/decision-curve-analysis-0. Accessed 19 April, 2014.

31. Ikegami N, Yoo B-K, Hashimoto H, et al. Japanese universal health coverage: evolution, achievements, and challenges. Lancet 2011;**378**:1106-15.

32. Tripepi G, Jager KJ, Dekker FW, et al. Statistical methods for the assessment of prognostic biomarkers(part II): calibration and re-classification. Nephrology Dialysis Transplantation 2010;**25**:1402-05.

33. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation 2007;**115**:928-35.

34. Jiang X, Menon A, Wang S, et al. Doubly Optimized Calibrated Support Vector Machine (DOC-SVM): an algorithm for joint optimization of discrimination and calibration. PLoS ONE 2012;**7**:e48823.

35. Holmberg L, Vickers A. Evaluation of prediction models for decision-making: beyond calibration and discrimination. PLoS Medicine 2013;**10**:e1001491.

36. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology (Cambridge, Mass) 2010;**21**:128-38.

37. Leening MJG, Vedder MM, Witteman JCM, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Annals of internal medicine 2014;**160**:122-31.

38. Vickers AJ, Pepe M. Does the net reclassification improvement help us evaluate models and markers? Annals of internal medicine 2014;**160**:136-37.

39. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Annals of internal medicine 2012;**157**:294-95.

Table 1. Patient characteristics, chief diagnosis for admission, and outcome

Characteristics	n=1310*
Demographics	
Age (years), median (range)	76 (15-100)
Female, n (%)	588 (44.9)
Nursing home resident, n (%)	37 (2.8)
Underlying diseases, n (%)	
Cerebrovascular disease	164 (12.5)
Congestive heart failure	106 (8.1)
Chronic respiratory disease	159 (12.1)
Chronic kidney disease	107 (8.2)
Chronic liver disease	80 (6.1)
Diabetes mellitus	252 (19.2)
Malignancy	230 (17.6)
Dementia	125 (9.5)
Autoimmune disorder	60 (5.0)
Human immunodeficiency virus positive	3 (0.2)
Vital signs	
Heart rate (beats/min), median (SD)	98 (20.4)
Systolic blood pressure (mmHg), median (SD)	132 (28.3)
Respiratory rate (breaths/min), median (SD)	20 (6.7)
Body temperature (°C), median (SD)	38.1 (1.4)
Mental confusion, n (%)	223 (17.0)
Laboratory data	

White blood cell count (/μL), median (SD)	10400 (6503)
Platelet count (×10⁴/μL), median (SD)	19.6 (9.9)
C reactive protein (mg/L), median (SD)	72.3 (90)
Blood urea nitrogen (mg/dL), median (SD)	19.3 (19.5)
Chief diagnosis for admission	
Pneumonia	405 (30.9)
Urinary tract infection	196 (15.0)
Skin and Soft tissue infection	69 (5.3)
Acute cholangitis	48 (3.7)
Acute cholecystitis	34 (2.6)
Bowel perforation	21 (1.6)
Other bacterial infection	154 (11.8)
Non-bacterial infection	94 (7.2)
Non-infection	193 (14.7)
Unclear	96 (7.3)
Bacteremia, n (%)	217 (16.6)
30-day in-hospital mortality, n (%)	108 (8.2)

Abbreviation: SD, standard deviation

*48 were excluded due to missing predictors in analyses for score evaluation.

Table 2. Unadjusted odds ratios for mortality calculated by univariable logistic regression

Variables	Unadjusted Odds Ratio (95% CI)	Missing n, (%)
CRP		
0.1-19.9 mg/L	1 [reference]	
20-69.9 mg/L	1.7 (0.8-3.5)	28 (2.1)
70-149.9 mg/L	2.3 (1.2-4.5)	
≥ 150 mg/L	4.1 (2.1-7.7)	
Age ≥ 65 years old	3.5 (1.9-6.6)	0 (0)
Female	0.6 (0.4-0.9)	0 (0)
Mental confusion	3.0 (2.0-4.6)	0 (0)
Hypotension (SBP < 90 or DBP ≤ 60 mmHg)	3.0 (2.0-4.5)	0 (0)
Abnormal body temperature (BT >38 or < 36 °C)	0.9 (0.6-1.3)	0 (0)
Tachycardia (HR > 90 /min)	1.5 (0.97-2.3)	0 (0)
Tachypnea (RR ≥ 30 /min)	3.2 (2.1-5.0)	21 (1.6)
Platelet count < 15 × 10^4/μL	2.1 (1.4-3.1)	0 (0)
BUN > 20 mg/dL	4.8 (3.0-7.8)	0 (0)
WBC > 15000/μL	1.7 (1.1-2.6)	0 (0)
Overall	-	48 (3.7)

Abbreviation: CRP, C reactive protein; SBP, systolic blood pressure; DBP, diastolic blood pressure; BT, body temperature; HR, heart rate; RR, respiratory rate; BUN, blood urea nitrogen; WBC, white blood cell count

Table 3. Adjusted odds ratios for mortality in multivariable logistic regression analyses

Variables	Adjusted Odds Ratios (95% CI)				
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
CRP ≥ 150 mg/L	2.5 (1.6-3.7)	2.3 (1.5-3.5)	2.1 (1.4-3.3)	1.8 (1.1-2.8)	2.0 (1.3-3.1)
Age ≥ 65 years	-	3.7 (1.9-7.2)	3.0 (1.5-5.9)	2.5 (1.3-5.1)	2.3 (1.1-4.6)
Female	-	0.6 (0.4-0.9)	0.6 (0.4-0.9)	0.6 (0.4-1.02)	-
Mental confusion	-	-	2.4 (1.5-3.8)	2.1 (1.3-3.3)	2.1 (1.3-3.4)
Hypotension (SBP < 90 or DBP ≤ 60 mmHg)	-	-	2.7 (1.7-4.1)	2.3 (1.5-3.5)	2.1 (1.3-3.2)
Abnormal body temperature (BT >38 or < 36 °C)	-	-	1.0 (0.6-1.5)	1.0 (0.7-1.6)	-
Tachycardia (HR > 90/min)	-	-	1.4 (0.9-2.3)	1.5 (0.9-2.5)	-
Tachypnea (RR ≥ 30/min)	-	-	2.4 (1.5-3.9)	2.3 (1.4-3.7)	2.4 (1.5-3.9)
Platelet count < 15 × 10^4/μL	-	-	-	1.8 (1.1-2.8)	-
BUN > 20 mg/dL	-	-	-	2.4 (1.4-4.0)	2.7 (1.6-4.5)
WBC > 15000/μL	-	-	-	1.5 (0.9-2.4)	-

Abbreviation: CRP, C reactive protein; SBP, systolic blood pressure; DBP, diastolic blood pressure; BT, body temperature; HR, heart rate; RR, respiratory rate; BUN, blood urea nitrogen; WBC, white blood cell count
Model 1: Unadjusted
Model 2: Adjusted for age and gender
Model 3: Adjusted for age, gender, and vital signs
Model 4: Adjusted for age, gender, vital signs, and lab data
Model 5: Adjusted for items of the CURB-65

Table 4. Observed mortality, predicted mortality, and net benefit in the CURB-65 and the modified CURB-65*

CURB-65 score	Observed 30-day mortality, % (death/total)	Predicted mortality† (95% CI) %	Net benefit‡	Modified CURB-65 score	Observed 30-day mortality, % (death/total)	Predicted mortality† (95% CI) %	Net benefit‡
0	0 (0/190)	1 (0.7-2.1)	0.075	0	0 (0/152)	1 (0.6-2.0)	0.075
1	3 (9/334)	3 (2.0-4.2)	0.061	1	2 (7/287)	2 (1.6-3.6)	0.068
2	8 (33/409)	7 (5.2-8.3)	0.034	2	6 (23/381)	5 (4.0-6.9)	0.048
3	13 (34/254)	14 (11.9-17.2)	0.015	3	12 (32/265)	11 (9.2-13.3)	0.024
4	30 (25/84)	28 (22.5-35.2)	0.004	4	17 (22/129)	22 (17.7-26.5)	0.012
5	39 (7/18)	48 (37.0-60.1)	0	5	44 (18/41)	38 (29.9-47.8)	0.005
				6	57 (4/7)	58 (45.2-70.4)	0

* The modified CURB-65 score was made by addition of 1 point if CRP ≥ 150 mg/L.

† Predicted mortality was calculated by introducing the CURB-65 and modified CURB-65 score as a continuous variable into univariable logistic regression.

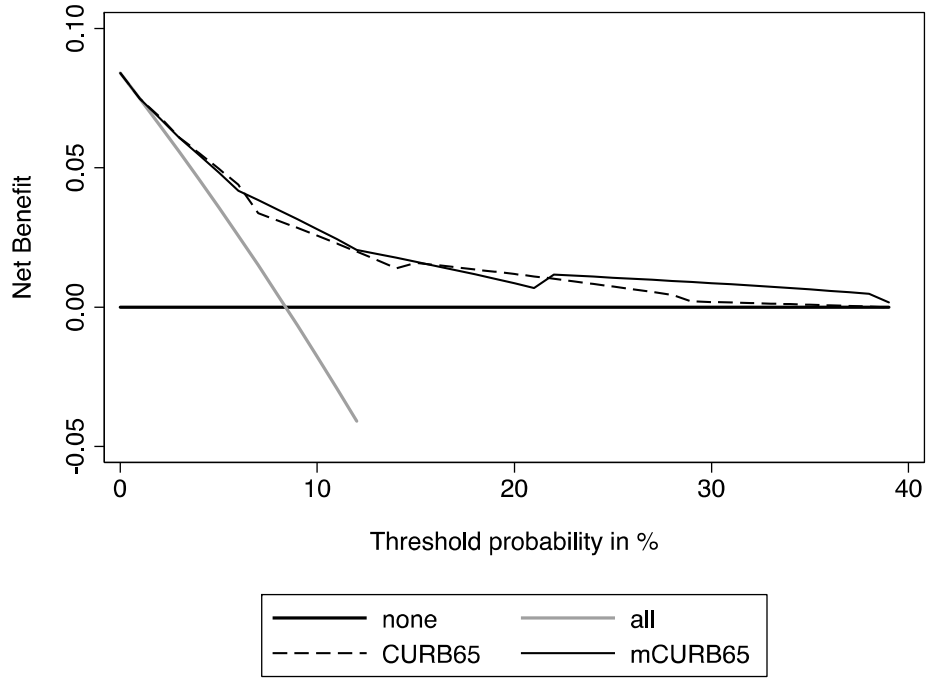
‡ Net benefits were calculated for each predicted mortality as a threshold probability.

Table 5. Summary of findings of this study

	CURB-65	Modified CURB-65
Discrimination		
AUC	0.76 (95% CI: 0.72 to 0.80)	0.77 (95% CI: 0.72 to 0.81)
Calibration		
Hosmer-Lemeshow test and calibration plot	good	good
Reclassification		
Overall category-free NRI	0.387 (95% CI: 0.193 to 0.582, p<0.0001)	
Overall IDI	0.015 (95% CI: 0.004 to 0.027, p<0.01)	
Clinical usefulness		
NB examined by DCA	comparable	

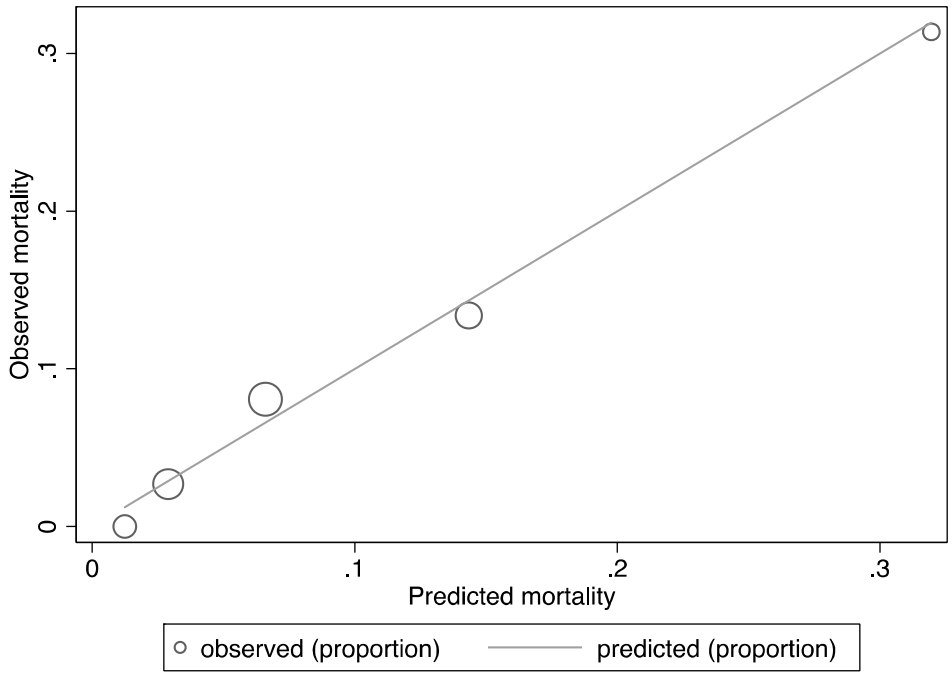
Abbreviation: AUC, area under the receiver operating characteristic curve; NRI, net reclassification improvement; IDI, integrated discrimination improvement; NB, net benefit; DCA, decision curve analysis

Figure 1. Decision curve for the CURB-65 and modified CURB-65 (mCURB-65) to predict 30-day mortality in patients with suspected sepsis

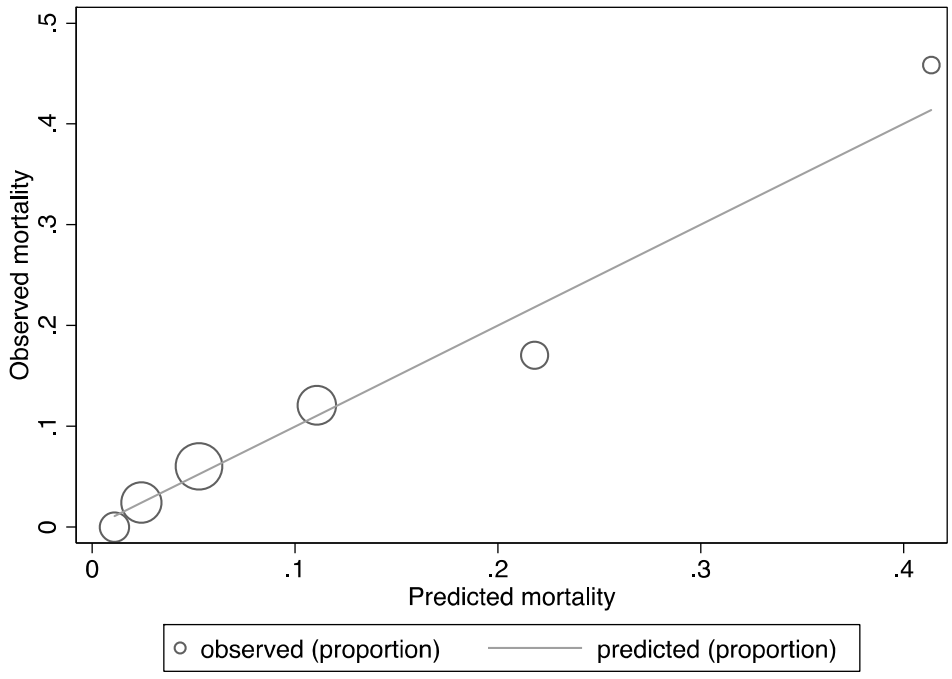


The thick black line is the net benefit of treating no patients, assuming that all would be alive; the thin grey line is the net benefit of treating all patients similarly regardless of their severity, assuming that all would die; the long dashed line is the net benefit of treating patients according to the CURB-65 score; and the thin black line is the net benefit of treating patients based on the modified CURB-65 score.

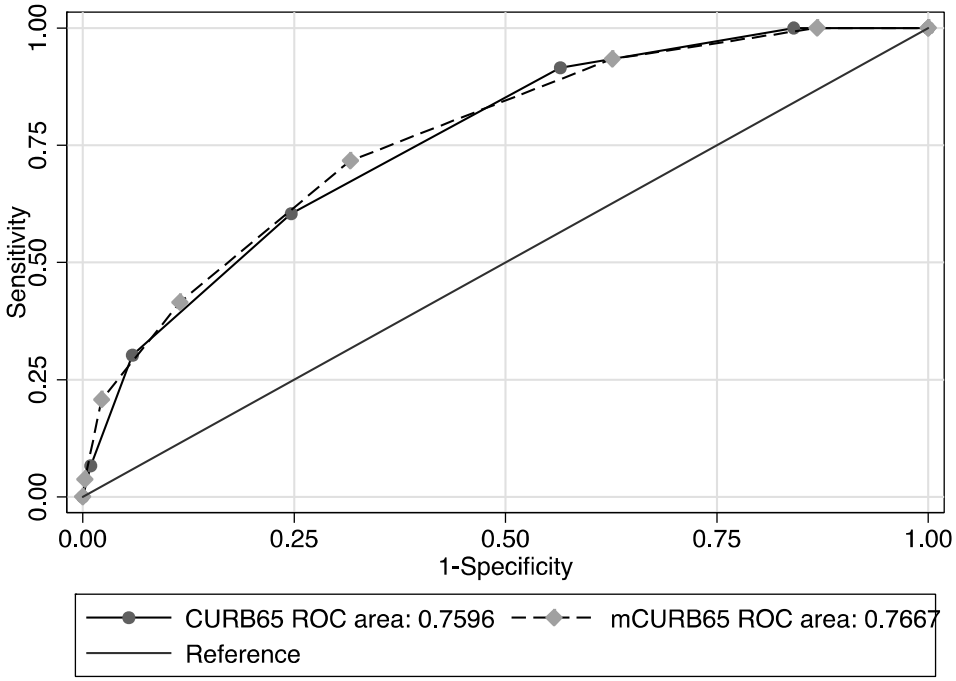
Supplementary Figure 1a. Calibration plot of the CURB-65



Supplementary Figure 1b. Calibration plot of the modified CURB-65



Supplementary Figure 2. The receiver operating characteristics curves of the CURB-65 and the modified CURB-65



STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[Within the method section of the abstract page 3]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See results section of abstract page 3-4]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[page 5-6]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[page 6]**
Methods		
Study design	4	Present key elements of study design early in the paper **[Methods page 6-7]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods page 6-7]**
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods page 6-7]** *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls **[N/A]** *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants **[N/A]**
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed **[N/A]** *Case-control study*—For matched studies, give matching criteria and the number of controls per case **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[page 7]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[page 7]**
Bias	9	Describe any efforts to address potential sources of bias **[page 6-7]**
Study size	10	Explain how the study size was arrived at **[page 9]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[page 7]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[page 7-9]**
		(*b*) Describe any methods used to examine subgroups and interactions **[N/A]**
		(*c*) Explain how missing data were addressed **[page 9]**
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed **[page 7]** *Case-control study*—If applicable, explain how matching of cases and controls was addressed **[N/A]** *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy **[N/A]**
		(*e*) Describe any sensitivity analyses **[N/A]**

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[page 10 and table 1]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[N/A information in page 10 and table 1]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[page 10 and table 1]**
		(b) Indicate number of participants with missing data for each variable of interest **[table 2]**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) **[page 10]**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time **[page 10 and table 1]**
		Case-control study—Report numbers in each exposure category, or summary measures of exposure **[N/A]**
		Cross-sectional study—Report numbers of outcome events or summary measures **[N/A]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[page 10-11; table 2 and 3]**
		(*b*) Report category boundaries when continuous variables were categorized **[page 10-11; table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[N/A]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[decision curve analysis in page 11-12; table 4; figure 1]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[page 12]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[page 14]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[page 14-15]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[page 14]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[N/A]**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Prognostic utility of serum CRP levels in combination with CURB-65 in patients with clinically suspected sepsis; decision curve analysis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007049.R1
Article Type:	Research
Date Submitted by the Author:	24-Mar-2015
Complete List of Authors:	Yamamoto, Shungo; School of Public Health in the Graduate School of Medicine, Kyoto University, Department of Healthcare Epidemiology; Kobe University Hospital, Division of Infectious Diseases Yamazaki, Shin; Kyoto University, Department of Healthcare Epidemiology Shimizu, Tsunehiro; Kyoto City Hospital, Department of Infectious Diseases Takeshima, Taro; Kyoto University, Department of Healthcare Epidemiology; Jichi Medical University, Division of Community and Family Medicine, Center for Community Medicine Fukuma, Shingo; Kyoto University, Department of Healthcare Epidemiology Yamamoto, Yosuke; Kyoto University, Department of Healthcare Epidemiology Tochitani, Kentaro; Kyoto City Hospital, Department of Infectious Diseases Tsuchido, Yasuhiro; Kyoto City Hospital, Department of Infectious Diseases Shinohara, Koh; National Center for Global Health and Medicine, Disease Control and Prevention Center Fukuhara, Shunichi; Kyoto University, Department of Healthcare Epidemiology
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Diagnostics, Emergency medicine, Medical management
Keywords:	PRIMARY CARE, STATISTICS & RESEARCH METHODS, INFECTIOUS DISEASES

Prognostic utility of serum CRP levels in combination with CURB-65 in patients with clinically suspected sepsis; decision curve analysis

Shungo Yamamoto MD, DTM&H[1], Shin Yamazaki PhD[1], Tsunehiro Shimizu MD, PhD[2], Taro Takeshima MD[1,3], Shingo Fukuma MD, PhD[1,5], Yosuke Yamamoto MD, PhD[1], Kentaro Tochitani MD[2], Yasuhiro Tsuchido MD[2], Koh Shinohara MD[4], Shunichi Fukuhara MD, DMSc[1,5]

[1]Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Infectious Diseases, Kyoto City Hospital, Kyoto, Japan

[3]Division of Community and Family Medicine, Center for Community Medicine, Jichi Medical University, Tochigi, Japan

[4]Disease Control and Prevention Center, National Center for Global Health and Medicine, Tokyo, Japan

[5]Center for Innovative Research for Communities and Clinical Excellence (CIRC2LE),

Fukushima Medical University, Fukushima, Japan

Corresponding author:

Shunichi Fukuhara

Department of Healthcare Epidemiology

School of Public Health in the Graduate School of Medicine

Kyoto University

Yoshida Konoemachi

Sakyo-ku, Kyoto, 606 8501, Japan

E-mail: fukuhara.shunichi.6m@kyoto-u.ac.jp

Phone: +81-75-753-4646, Fax: +81-75-753-4644

Conflict of interests

All authors certify that they have no conflict of interests.

Keywords: CURB-65; CRP; sepsis; decision curve analysis; net benefit

Abstract

Objectives: The prognostic utility of serum C reactive protein (CRP) alone in sepsis is controversial. We used decision curve analysis (DCA) to evaluate the clinical usefulness of combining serum CRP levels with the CUBR-65 score in patients with suspected sepsis.

Design: Retrospective cohort study.

Setting: Emergency department (ED) of an urban teaching hospital in Japan.

Participants: Consecutive ED patients over 15 years old who were admitted to the hospital after having a blood culture taken in the ED between 1 January 2010 and 31 December 2012.

Main outcome measures: 30-day in-hospital mortality.

Results: Data from 1262 patients were analyzed for score evaluation. The 30-day in-hospital mortality was 8.4%. Multivariable analysis showed that serum CRP ≥150 mg/L was an independent predictor of death (adjusted odds ratio: 2.0; 95% confidence interval [CI]: 1.3 to 3.1). We compared the predictive performance of the CURB-65 with the performance of a modified CURB-65 with that included CRP (≥150 mg/L) to

quantify the clinical usefulness of combining serum CRP to the CURB-65. The areas under the receiver operating characteristics curves (AUC) of the CURB-65 and modified CURB-65 were 0.76 (95% CI: 0.72 to 0.80) and 0.77 (95% CI: 0.72 to 0.81), respectively. Both models had good calibration for mortality and were useful among threshold probabilities from 0% to 30%. However, while incorporating CRP into CURB-65 yielded a significant category-free net reclassification improvement of 0.387 (95% CI: 0.193 to 0.582) and integrated discrimination improvement of 0.015 (95% CI: 0.004 to 0.027), DCA showed that the CURB-65 and the modified CURB-65 score had comparable net benefits for prediction of mortality.

Conclusions: Measurement of serum CRP added limited clinical usefulness to the CURB-65 in predicting mortality in patients with clinically suspected sepsis, regardless of the source.

Article summary:

Strengths and limitations of this study

- Combining serum CRP with the CURB-65 gave statistically significant values of

net reclassification improvement and integrated discrimination improvement. In contrast, decision curve analysis showed that combining serum CRP with the CURB-65 was of only limited clinical usefulness.

- The limitations of this study are the possibility of selection bias of the eligible patients and the retrospective nature of the study in a single hospital.

Introduction

Sepsis is a major cause of morbidity and mortality worldwide, with in-hospital mortality still at 18% or more, according to recent surveys in resource-rich countries.[1-3] Early identification of high-risk patients and timely intervention for sepsis are therefore crucial to improving outcomes.

Severity assessment is important in the management of patients, including decision-making regarding choice of treatment and patient disposition. To encourage implementation, a clinical prediction rule must be user-friendly.[4] While there are a lot of well-known scoring systems for severity of illness such as the Acute Physiology and Chronic Health Evaluation (APACHE), Sepsis-related Organ Failure Assessment (SOFA), and Multiple Organ Dysfunction Score (MODS), they have too many items to use conveniently in the emergency department (ED).[5] In addition, these scores have not been well validated in settings other than the intensive care unit (ICU). The CURB-65 is a simple prediction rule originally developed as a prognostic scoring system for community-acquired pneumonia (CAP).[6] The rule has been well validated in patients with CAP, and CURB-65-guided antibiotic therapy has safely reduced broad-spectrum

antibiotic use in this population.[7 8] In addition to its utility among CAP patients, CURB-65 has also been correlated with mortality in patients with suspected sepsis, regardless of the source, and in patients admitted for non-surgical illness.[9-11]

Serum C reactive protein (CRP) is an acute phase protein often evaluated as a marker of systemic inflammation.[12] In Japan, serum CRP levels have been used as a diagnostic and prognostic marker of infection in daily clinical practice and clinical trials of new drugs.[13] However, evidence demonstrating its value is insufficient at present for routine application of serum CRP levels to assess severity of infection. As a prognostic marker, some have reported that serum CRP on admission is associated with mortality.[14 15] However, a systematic review reported conflicting findings, noting that serum CRP levels were not significantly different between survivor and nonsurvivor, suggesting that these levels may have limited value in reflecting the severity of sepsis.[12] As a diagnostic marker, the sensitivity and specificity of serum CRP for discriminating bacterial from non-infectious inflammation was only 75% and 67%, respectively, according to a meta-analysis.[16] However, while the diagnostic performance of serum CRP alone is limited, serum CRP has been reported to contribute some additional information to a

prediction rule involving a patient's symptoms and physical examination in diagnosis of pneumonia.[17] In this respect, the additive prognostic value of serum CRP to an existing severity score is unknown.

Performance of a prediction model has traditionally been evaluated by discrimination and calibration. However, having good discrimination and calibration alone is not sufficient to show that a model would improve decision-making.[18] As metrics of reclassification, net reclassification improvement (NRI) and integrated discrimination improvement (IDI) have enjoyed increasing usage in evaluating improvement in prediction models. However, these improvements quantified as NRI and IDI are also not sufficient for evaluating clinical usefulness.[19-21] Decision curve analysis (DCA), which was first described by Vickers and Elkin, can be used to incorporate the clinical consequences of a decision into evaluations of diagnostic tests or prediction models.[22] To our knowledge, there have been no studies in which DCA is employed to evaluate the clinical usefulness of serum CRP levels and CURB-65 score in patients with suspected sepsis.

Here, our objective was to use DCA to evaluate clinical usefulness of combining

serum CRP levels to the CURB-65 score in patients with suspected sepsis, regardless of the source of infection.

Materials and Methods

Study Design, Setting and Patients

We performed a retrospective cohort study at Kyoto City Hospital, an urban teaching hospital with 548 beds in Japan. Consecutive emergency department (ED) patients over 15 years old admitted to the hospital after having a blood culture taken in the ED between 1 January 2010 and 31 December 2012 were included. Doctors' decision to order a blood culture was used as a surrogate marker for suspected sepsis as in previous studies.[11 23] To facilitate data independence, only the index admission was included for patients with multiple admissions during the study period. Patients transferred from another hospital or who had cardiopulmonary arrest on arrival at the hospital were excluded.

Data collection

The following data were extracted from electronic medical records for each patient: age,

gender, underlying disease, vital signs (heart rate, blood pressure, respiratory rate, and mental confusion and body temperature), laboratory findings (white blood cell [WBC] count, platelet count, and blood urea nitrogen [BUN] and serum CRP levels), and outcome. For vital signs and laboratory data, initial values at the hospital visit were recorded. Blood pressure was measured with noninvasive cuff. Serum CRP was measured with latex turbidimetric immunoassay. The items of the CURB-65 score were as follows: mental confusion (present/absent), BUN > 7 mmol/L (20 mg/dL), respiratory rate (RR) ≥ 30/min, either or both systolic blood pressure (SBP) < 90 mmHg or diastolic blood pressure (DBP) ≤ 60 mmHg, and age ≥ 65 years.[6] Mental confusion was defined as disorientation in person, place or time or being in a stupor or coma as with a previous study.[6]

The main outcome measure was 30-day in-hospital mortality. Patients who were discharged or transferred from the hospital within 30 days of admission or who remained in the hospital for more than 30 days were considered alive in this analysis.[24]

Statistical analysis

First, we validated the CURB-65 model. We graphically assessed the calibration of the CURB-65 model with calibration plot and tested it with the Hosmer-Lemeshow test. A p value < 0.05 indicates a lack of good fit for the model. Regarding the model discrimination, we also computed the area under the receiver operating characteristic curve (AUC) with a 95% confidence interval (CI) using 500 bootstrap resampling.[25] The predicted mortalities with 95% CI were calculated by introducing the CURB-65 score as a continuous variable into univariable logistic regression.

Second, we examined additional value of serum CRP. We graphically checked whether or not the relationship between serum CRP level and mortality was linear in the logit with a smoothing curve using a locally weighted least squares (Lowess) regression.[26] We conducted logistic regression analysis after adding CRP as a continuous variable to the CURB-65 system. User-friendliness is important for clinical prediction rules and dichotomized test results (normal *vs* abnormal) are easy to use and interpret. We explored the cut-off point of serum CRP level for prediction of death in patients with suspected sepsis because the optimal cut-off point was unknown. Serum CRP results were first divided into quartiles and rounded to the nearest 10. Each patient

was then assigned to one of four categories corresponding to the CRP quartiles. We assessed the most optimal cut-off point from the AUC. We conducted multivariable logistic regression analysis to adjust predictors of death by introducing pre-specified variables: items of the CURB-65. We assessed multicollinearity using the variance inflation factor (VIF). VIFs greater than 2.5 may be problematic.[27] We also computed the unadjusted odds ratio of covariates using univariable logistic regression to show the influence of adjustment for predictors.

Third, we assessed the model performance of the modified CURB-65 score, which was made by incorporating CRP information into the CURB-65 model, using a calibration plot and Hosmer-Lemeshow test for calibration and AUC for discrimination. Additive information of CRP was evaluated by category-free NRI and IDI.[28] With regard to clinical usefulness, we examined the net benefit using DCA.[22] Briefly, the method is based on the principle that the relative harms of false positives and false negatives can be expressed in terms of a probability threshold.[29] The net benefit is obtained by subtracting the proportion of patients who are false positive from the proportion who are true positive, weighting by the relative harm of a false-positive and

a false-negative result. The net benefit of making a decision based on the model can be calculated with the following formula:

$$\text{Net benefit} = \frac{True\ Positives}{n} - \left(\frac{Pt}{1-Pt}\right)\frac{False\ positives}{n}$$

where n is the total number of patients in the study and P_t is a given threshold probability.[22 29 30]

With regard to sample size estimation, at least 8 to 10 events per variable are required for reliable multiple logistic regression analysis,[31 32] and 100 events and 100 nonevents are required for an external validation study.[33] We assumed 30 to 40 deaths per year among eligible patients, collecting 3 years' worth of data (90 to 120 estimated deaths) to appropriately conduct multiple logistic regressions with 11 variables and ensure adequate statistical power.

In terms of handling missing values, we planned to perform a complete case analysis if missing values were below 5%, as such an analysis might be feasible then. [34] If missing values were above 5%, we planned to apply an appropriate imputation

method.

Data were analyzed with R software 3.0.1 (www.r-project.org) and Stata software, version 13 (StataCorp., College Station, TX, USA) including programs of Decision Curve Analysis provided by Vickers.[35]

Ethical approval

This protocol was approved by the Ethics Committee of Kyoto University Graduate School and Faculty of Medicine and Kyoto City Hospital.

Results

Patient characteristics

Among 1310 eligible patients over 3 years of study, 108 deaths (8.2% mortality) were recorded. Demographics, underlying diseases, vital signs, laboratory findings and chief diagnosis for admission were presented in Table 1. Diagnosis was unclear in 92 patients (7.3%). Respiratory rates data were missing for 21 patients, and CRP data were missing for 28 (Table 3). Overall cases with any missing predictor were 48 (3.7%), so we conducted a complete case analysis, leaving 1262 patients (106 deaths, 8.4% mortality)

for model evaluation analyses.

Validation of the CURB-65 in our population

Observed 30-day in-hospital mortalities and predicted mortalities computed by the CURB-65 score are shown in Table 4. The CURB-65 showed good calibration for mortality, with a Hosmer-Lemeshow test of 4.08 (df=5, P=0.538), indicating good fit. The calibration plot is shown in Supplementary Figure 1a. The AUC for the CURB-65 was 0.76 (95% CI: 0.72 to 0.80; Supplementary Figure 2).

Evaluation of CRP as a predictor of mortality

The relationship between serum CRP level and mortality was almost linear in the logit (Supplementary Figure 3). An unadjusted odds ratio for mortality was 1.05 (95% CI: 1.03 to 1.07) per 10 mg/L rise in serum CRP level. Addition of continuous serum CRP level to the CURB-65 system revealed an adjusted odds ratio for mortality was 1.04 (95% CI: 1.01 to 1.06) per 10 mg/L increase in concentration. Because the optimal cut-off point was unknown, serum CRP results were divided into quartiles: the quartile

points were 18.2, 72.3 and 149.2 mg/L. Then, they were rounded to the nearest 10, and we set interim cut-off points as 20, 70, and 150 mg/L. Observed mortality and unadjusted odds ratios for mortality of each CRP group are shown in Table 2. We repeated regression analyses adding serum CRP as a dichotomized variable with each interim cut-off point. We found 150 mg/L as the most optimal threshold to dichotomize serum CRP levels.

Adjusted odds ratios for mortality are shown in Table 3. We found no evidence of multicollinearity because the variance inflation factors for predictors in the model in Table 3 were less than 1.2. We identified serum CRP level ≥ 150 mg/L as an independent predictor of death in patients with clinically suspected sepsis.

Additive information of CRP to CURB-65

Because the adjusted odds ratio (i.e. regression coefficient) of CRP was comparable to each item in the CURB-65, we made a modified CURB-65 score by adding one point to the CURB-65 score when serum CRP level was ≥ 150 mg/L. Table 4 shows observed

30-day in-hospital mortalities and predicted mortalities stratified by the modified CURB-65 score. The modified CURB-65 also showed good calibration for mortality, with a Hosmer-Lemeshow test of 4.52 (df=6, P=0.607). The calibration plot is shown in Supplementary Figure 1b. The AUC for the modified CURB-65 was 0.77 (95% CI: 0.72 to 0.81; Supplementary Figure 2). By incorporating CRP into CURB-65, event NRI was −0.151 and nonevent NRI was 0.538, giving an overall category-free NRI of 0.387 (95% CI: 0.193 to 0.582, p<0.0001). Further, IDI for events was 0.014 and IDI for nonevents was 0.001, giving an overall IDI of 0.015 (95% CI: 0.004 to 0.027, p<0.01). These findings were statistically significant.

Decision curve analysis

Figure 1 demonstrates the decision curves for the CURB-65 and modified CURB-65 to predict 30-day in-hospital mortality in patients with clinically suspected sepsis. Both the CURB-65 and the modified CURB-65 are useful between threshold probabilities of 0% to 30%. However, both curves cross and depict little difference in net benefit. The net benefits at each point of the CURB-65 score and the modified CURB-65 score are

shown in Table 4. The comparison of discrimination, calibration, reclassification metrics and clinical usefulness between CURB-65 and modified CURB-65 are summarized in Table 5.

To assess the robustness of our findings, we repeated DCA with changing the cut-off level of serum CRP as 20 mg/L and 70 mg/L, respectively, in sensitivity analyses. Similarly, we found the additive clinical usefulness of serum CRP was unremarkable.

Discussion

We determined that having high CRP levels was independently associated with high mortality in our population. We also confirmed geographical and domain validation of the CURB-65 in our patients, which comprised an external validation in a different geographical area and in a population including different category of patients from CAP.[4]

The utility of serum CRP as a prognostic marker has been found to vary.[12] In Japan, universal health coverage allows people to consult a doctor soon after they recognize any symptoms, with no particular limitations.[36] Given that secretion of CRP

peaks at 36-50 h after inflammatory stimulus,[12] serum CRP level might be useful as a surrogate marker of duration from disease onset to consulting a doctor as well as a marker reflecting intensity of inflammation. We believe that the association between serum CRP level and mortality will be more easily identified in countries such as Japan where the population has easy access to hospitals, due to the wide distribution in duration from disease onset to visiting a hospital. Although reclassification metrics such as NRI and IDI were statistically improved on incorporation of CRP level, the additive clinical usefulness of CRP to the CURB-65 was admittedly limited (Table 5).

Strengths and limitations of the study

A major strength of this study is our evaluation using DCA. To the best of our knowledge, this is the first study, which examined clinical usefulness of serum CRP and the CURB-65 score in septic patients using DCA. DCA can take into account risk threshold, weighting benefits and harms, and is useful in evaluating clinical utility of a prediction model.[21 22 37] For instance, the net benefit of 0.061 at a threshold probability of 3% in the CURB-65 score can be interpreted as meaning that making a decision

based on the prediction model, compared to assuming that all patients would be alive, leads to the equivalent of a net 6.1 true-positive results per 100 patients with no corresponding increase in the number of false-positive results (Table 4).[22] In our population, 8.4% overall 30-day in-hospital mortality means a maximum net benefit of 0.084, which is calculated if we use a threshold probability of 0%. There are no universally accepted criteria on patient's risk threshold for suspected sepsis to make a decision about patient disposition or therapeutic indication. If we extrapolate the data on community-acquired pneumonia, low, intermediate and high risk of mortality are considered to be about 1-2%, 8-9% and 20-30%, respectively.[6] The ability to make better decisions with serum CRP than without was considered to be unremarkable in this range of risk threshold.

Our study is a type of external validation study with model updating to assess whether serum CRP level has additive value to the CURB-65 or not.[4] Another strengths of this study might be our sample size with adequate statistical power for an external validation study. [33]

Several limitations to the present study warrant mention. We cannot rule out the

possibility of selection bias, as only inpatients who had a blood culture taken were included. We may therefore have missed patients with infection who did not undergo blood culture in the ED, and contrarily, some patients without infection were included in the study. However, clinicians must routinely make decisions in the ED despite being unsure as to whether or not a patient is actually infected; we therefore considered it important to evaluate a clinical prediction rule accounting for such clinical uncertainty. Inclusion of patients without infectious diseases, we feel, reflects a real-world scenario. Another limitation is the retrospective nature of the study and the fact that it was conducted in a single hospital. Given the study's retrospective design, patients with high CRP might have received more intensive therapy than those with relatively low CRP. Such bias might have lowered the predictive ability of CRP.

Conclusions

While serum CRP level ≥ 150 mg/L was found to be associated with high mortality, its additive clinical usefulness to the CURB-65 was limited based on DCA. The CURB-65

correlated well with 30-day in-hospital mortality in patients with clinically suspected infection, and it was useful among threshold probabilities in the range of 0% to 30%. Measurement of serum CRP may contribute little to making decisions regarding the management of clinically suspected sepsis patients.

Contributors

S. Yamamoto had full access to all of the data in the study and takes responsibility for integrity of the data and accuracy of the data analysis and wrote the first draft.

T. Shimizu, K. Tochitani, Y. Tsuchido and K. Shinohara collected and interpreted the data, and drafted the paper.

S. Yamazaki, T. Takeshima, S. Fukuma, Y. Yamamoto and S. Fukuhara supervised the research, interpreted the data and contributed to the writing of the manuscript.

Funding

This research received no specific grant from any funding agency in the public,

commercial or not-for-profit sectors.

Competing interests

None.

Ethics approval

The Ethics Committee of Kyoto University Graduate School and Faculty of Medicine and Kyoto City Hospital.

Data sharing statement

No additional data are available.

References

1. Kumar G, Kumar N, Taneja A, et al. Nationwide trends of severe sepsis in the 21st

century (2000-2007). Chest 2011;**140**:1223-31.

2. Kaukonen K-M, Bailey M, Suzuki S, et al. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. JAMA 2014;**311**:1308-16.

3. Fujishima S, Gando S, Saitoh D, et al. A multicenter, prospective evaluation of quality of care and mortality in Japan based on the Surviving Sepsis Campaign guidelines. Journal of infection and chemotherapy 2014;**20**:115-20.

4. Toll DB, Janssen KJM, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. Journal of clinical epidemiology 2008;**61**:1085-94.

5. Calle P, Cerro L, Valencia J, et al. Usefulness of severity scores in patients with suspected infection in the emergency department: a systematic review. The Journal of emergency medicine 2012;**42**:379-91.

6. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;**58**:377-82.

7. Chalmers JD, Singanayagam A, Akram AR, et al. Severity assessment tools for

predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. Thorax 2010;**65**:878-83.

8. Chalmers JD, Singanayagam A, Akram AR, et al. Safety and efficacy of CURB65-guided antibiotic therapy in community-acquired pneumonia. Journal of Antimicrobial Chemotherapy 2011;**66**:416-23.

9. Howell MD, Donnino MW, Talmor D, et al. Performance of severity of illness scoring systems in emergency department patients with infection. Academic emergency medicine 2007;**14**:709-14.

10. Armiñanzas C, Velasco L, Calvo N, et al. CURB-65 as an initial prognostic score in Internal Medicine patients. European journal of internal medicine 2013;**24**:416-19.

11. Marwick CA, Guthrie B, Pringle JE, et al. Identifying which septic patients have increased mortality risk using severity scores: a cohort study. BMC anesthesiology 2014;**14**:1.

12. Mitaka C. Clinical laboratory differentiation of infectious versus non-infectious systemic inflammatory response syndrome. Clinica chimica acta

2005;**351**:17-29.

13. Saito A, Miki F, Oizumi K, et al. Clinical evaluation methods for new antimicrobial agents to treat respiratory infections: Report of the Committee for the Respiratory System, Japan Society of Chemotherapy. Journal of infection and chemotherapy 1999;**5**:110-23.

14. Chalmers JD, Singanayagam A, Hill AT. C-reactive protein is an independent predictor of severity in community-acquired pneumonia. The American journal of medicine 2008;**121**:219-25.

15. Lobo SMA, Lobo FRM, Bota DP, et al. C-reactive protein levels correlate with mortality and organ failure in critically ill patients. Chest 2003;**123**:2043-49.

16. Simon L, Gauvin F, Amre DK, et al. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clinical infectious diseases 2004;**39**:206-17.

17. van Vugt SF, Broekhuizen BDL, Lammens C, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic

study. BMJ 2013;**346**:f2450-f50.

18. Holmberg L, Vickers A. Evaluation of prediction models for decision-making: beyond calibration and discrimination. PLoS Medicine 2013;**10**:e1001491.

19. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology (Cambridge, Mass) 2010;**21**:128-38.

20. Leening MJG, Vedder MM, Witteman JCM, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Annals of internal medicine 2014;**160**:122-31.

21. Vickers AJ, Pepe M. Does the net reclassification improvement help us evaluate models and markers? Annals of internal medicine 2014;**160**:136-37.

22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical decision making 2006;**26**:565-74.

23. Shapiro NI, Wolfe RE, Moore RB, et al. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. Critical Care Medicine 2003;**31**:670-75.

24. Graham PL, Cook DA. Prediction of risk of death using 30-day outcome: a practical end point for quality auditing in intensive care. Chest 2004;**125**:1458-66.

25. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology 2003;**56**:441-47.

26. Hosmer DW. 4 Independent variables in multivariable analysis. Applied Logistic Regression. 3 ed. Hoboken, New Jersey: Wiley, 2013.

27. Katz MH. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*. 3rd ed. New York: Cambridge University Press, 2011.

28. Cook NR, Paynter NP. Performance of reclassification statistics in comparing risk prediction models. Biometrical journal 2011;**53**:237-58.

29. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC medical informatics and decision making 2008;**8**:53.

30. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. The American statistician 2008;**62**:314-20.

31. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology 1996;**49**:1373-79.

32. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. American journal of epidemiology 2003;**158**:280-87.

33. Vergouwe Y, Steyerberg EW, Eijkemans MJC, et al. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of clinical epidemiology 2005;**58**:475-83.

34. Royston P, Moons KGM, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. BMJ 2009;**338**:b604.

35. Vickers AJ. Decision Curve Analysis. Secondary Decision Curve Analysis. 2013. http://www.mskcc.org/research/epidemiology-biostatistics/health-outcomes/decision-curve-analysis-0. Accessed 19 April, 2014.

36. Ikegami N, Yoo B-K, Hashimoto H, et al. Japanese universal health coverage: evolution, achievements, and challenges. Lancet 2011;**378**:1106-15.

37. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Annals of internal medicine 2012;**157**:294-95.

Table 1. Patient characteristics, chief diagnosis for admission, and outcome

Characteristics	n=1262
Demographics	
Age (years), median (IQR)	76 (64-83)
Female, n (%)	560 (44.4)
Nursing home resident, n (%)	37 (2.9)
Underlying diseases, n (%)	
Cerebrovascular disease	156 (12.4)
Congestive heart failure	101 (8.0)
Chronic respiratory disease	155 (12.3)
Chronic kidney disease	100 (7.9)
Chronic liver disease	77 (6.1)
Diabetes mellitus	243 (19.3)
Malignancy	222 (17.6)
Dementia	121 (9.6)
Autoimmune disorder	63 (5.0)
Human immunodeficiency virus positive	3 (0.2)
Vital signs	
Heart rate (beats/min), median (IQR)	98 (85-156)
Systolic blood pressure (mmHg), median (IQR)	131 (113-150)
Respiratory rate (breaths/min), median (IQR)	20 (18-24)
Body temperature (°C), median (IQR)	38.1 (37.1-39)
Mental confusion, n (%)	215 (17.0)
Laboratory data	
White blood cell count (10^9/L), median (IQR)	10.5 (7.6-14.6)
Platelet count (×10^9/L), median (IQR)	196 (150-256)
C reactive protein (mg/L), median (IQR)	72.3 (18.2-149.2)
Blood urea nitrogen (mmol/L), median (IQR)	6.9 (5.0-10.2)
Chief diagnosis for admission	
Pneumonia	393 (33.6)
Urinary tract infection	188 (16.1)
Skin and Soft tissue infection	62 (5.3)

Acute cholangitis	47 (4.0)
Acute cholecystitis	33 (2.8)
Bowel perforation	21 (1.8)
Other bacterial infection	150 (12.8)
Non-bacterial infection	103 (8.8)
Non-infection	174 (14.9)
Unclear	92 (7.3)
Bacteremia, n (%)	210 (16.6)
30-day in-hospital mortality, n (%)	106 (8.4)

Abbreviation: IQR, interquartile range

Table 2. Observed mortality and unadjusted odds ratios for mortality stratified by serum CRP categories

Variables	Observed 30-day mortality, % (death/total)	Unadjusted Odds Ratio* (95% CI)
CRP		
0.1-19.9 mg/L	4.0 (13/326)	1 [reference]
20-69.9 mg/L	6.6 (19/289)	1.7 (0.8-3.5)
70-149.9 mg/L	8.7 (29/335)	2.3 (1.2-4.5)
≥ ***150 mg/L***	14.4 (45/312)	4.1 (2.1-7.7)

Abbreviation: CRP, C reactive protein

*Calculated by univariable logistic regression

Table 3. Adjusted odds ratios for mortality in multivariable logistic regression analyses

Variables	Unadjusted Odds Ratios (95% CI)	Adjusted Odds Ratios* (95% CI)	Missing n, (%)
CRP ≥ 150 mg/L	2.5 (1.6-3.7)	2.0 (1.3-3.1)	28 (2.1)
Age ≥ 65 years	3.7 (1.9-7.3)	2.3 (1.1-4.6)	0 (0)
Mental confusion	2.9 (1.9-4.5)	2.1 (1.3-3.4)	0 (0)
Hypotension (SBP < 90 or DBP ≤ 60 mmHg)	3.0 (2.0-4.5)	2.1 (1.3-3.2)	0 (0)
Tachypnea (RR ≥ 30/min)	3.1 (2.0-4.8)	2.4 (1.5-3.9)	21 (1.6)
BUN > 7 mmol/L	4.7 (2.9-7.6)	2.7 (1.6-4.5)	0 (0)
Overall	-	-	48 (3.7)

Abbreviation: CRP, C reactive protein; SBP, systolic blood pressure; DBP, diastolic blood pressure; RR, respiratory rate; BUN, blood urea nitrogen

*Adjusted for items of the CURB-65

Table 4. Observed mortality, predicted mortality, and net benefit in the CURB-65 and the modified CURB-65*

CURB-65 score	Observed 30-day mortality, % (death/total)	Predicted mortality† (95% CI) %	Net benefit‡	Modified CURB-65 score	Observed 30-day mortality, % (death/total)	Predicted mortality† (95% CI) %	Net benefit ‡
0	0 (0/190)	1 (0.7-2.1)	0.075	**0**	0 (0/152)	1 (0.6-2.0)	0.075
1	3 (9/334)	3 (2.0-4.2)	0.061	**1**	2 (7/287)	2 (1.6-3.6)	0.068
2	8 (33/409)	7 (5.2-8.3)	0.034	**2**	6 (23/381)	5 (4.0-6.9)	0.048
3	13 (34/254)	14 (11.9-17.2)	0.015	**3**	12 (32/265)	11 (9.2-13.3)	0.024
4	30 (25/84)	28 (22.5-35.2)	0.004	**4**	17 (22/129)	22 (17.7-26.5)	0.012
5	39 (7/18)	48 (37.0-60.1)	0	**5**	44 (18/41)	38 (29.9-47.8)	0.005
				6	57 (4/7)	58 (45.2-70.4)	0

* The modified CURB-65 score was made by addition of 1 point if CRP ≥ 150 mg/L.

† Predicted mortality was calculated by introducing the CURB-65 and modified CURB-65 score as a continuous variable into univariable logistic regression.

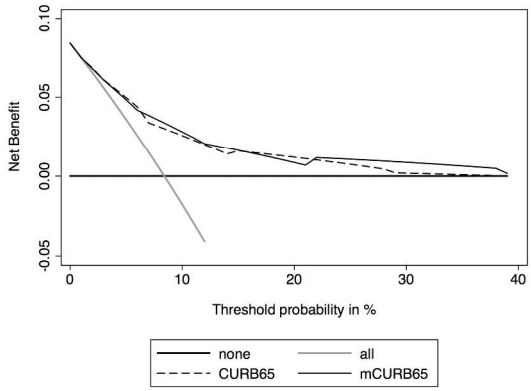
‡ Net benefits were calculated for each predicted mortality as a threshold probability.

Table 5. Comparison of discrimination, calibration, reclassification metrics and clinical usefulness between CURB-65 and modified CURB-65

	CURB-65	Modified CURB-65
Discrimination		
AUC	0.76 (95% CI: 0.72 to 0.80)	0.77 (95% CI: 0.72 to 0.81)
Calibration		
Hosmer-Lemeshow test and calibration plot	good	good
Reclassification		
Overall category-free NRI	0.387 (95% CI: 0.193 to 0.582, p<0.0001)	
Overall IDI	0.015 (95% CI: 0.004 to 0.027, p<0.01)	
Clinical usefulness		
NB examined by DCA	comparable	

Abbreviation: AUC, area under the receiver operating characteristic curve; NRI, net reclassification improvement; IDI, integrated discrimination improvement; NB, net benefit; DCA, decision curve analysis

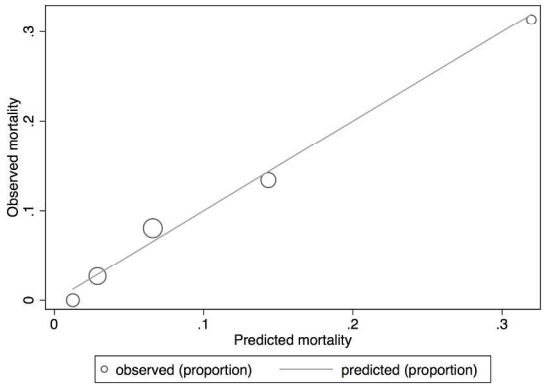
Figure 1. Decision curves for the CURB-65 and modified CURB-65 (mCURB-65) to predict 30-day mortality in patients with suspected sepsis



The thick black line is the net benefit of treating no patients, assuming that all would be alive; the thin grey line is the net benefit of treating all patients similarly regardless of their severity, assuming that all would die; the long dashed line is the net benefit of treating patients according to the CURB-65 score; and the thin black line is the net benefit of treating patients based on the modified CURB-65 score.

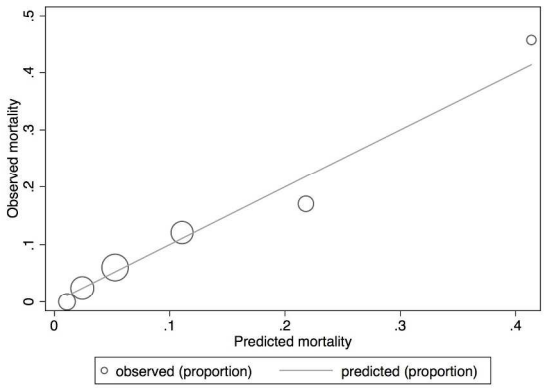
209x297mm (300 x 300 DPI)

Supplementary Figure 1a. Calibration plot of the CURB-65



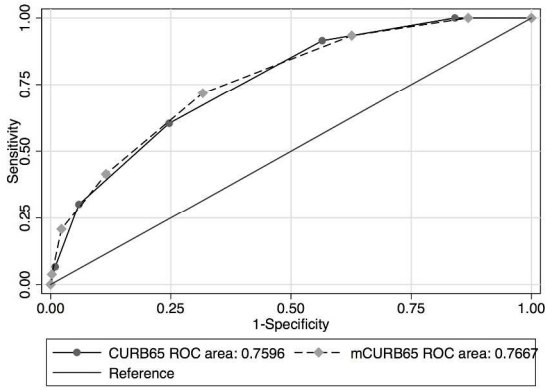
209x297mm (300 x 300 DPI)

Supplementary Figure 1b. Calibration plot of the modified CURB-65



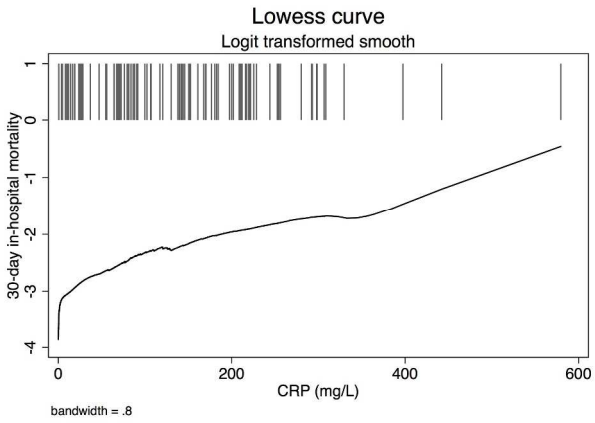
209x297mm (300 x 300 DPI)

Supplementary Figure 2. The receiver operating characteristics curves of the CURB-65 and the modified CURB-65



209x297mm (300 x 300 DPI)

Supplementary Figure 3. The relationship between serum CRP level and mortality; locally weighted scatterplot smoothing (Lowess) curve



209x297mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[Within the method section of the abstract page 3]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See results section of abstract page 3-4]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[page 5-7]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[page 7]**
Methods		
Study design	4	Present key elements of study design early in the paper **[Methods page 6]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods page 6-7]**
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods page 6-7]** *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls **[N/A]** *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants **[N/A]**
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed **[N/A]** *Case-control study*—For matched studies, give matching criteria and the number of controls per case **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[page 7]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[page 7]**
Bias	9	Describe any efforts to address potential sources of bias **[page 6-7]**
Study size	10	Explain how the study size was arrived at **[page 10]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[page 7-8]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[page 7-9]**
		(*b*) Describe any methods used to examine subgroups and interactions **[N/A]**
		(*c*) Explain how missing data were addressed **[page 11]**
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed **[page 7-8]** *Case-control study*—If applicable, explain how matching of cases and controls was addressed **[N/A]** *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy **[N/A]**
		(*e*) Describe any sensitivity analyses **[N/A]**

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[page 11 and table 1]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[N/A information in page 11 and table 1]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[page 11 and table 1]**
		(b) Indicate number of participants with missing data for each variable of interest **[table 3]**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) **[page 11]**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time **[page 11 and table 1]**
		Case-control study—Report numbers in each exposure category, or summary measures of exposure **[N/A]**
		Cross-sectional study—Report numbers of outcome events or summary measures **[N/A]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[page 11-13; table 2 and 3]**
		(*b*) Report category boundaries when continuous variables were categorized **[page 11-12; table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[N/A]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[decision curve analysis in page 13; table 4; figure 1]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[page 13-14]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[page 15-16]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[page 14-16]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[page 15]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[N/A]**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.