AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

New Sepsis Definition (Sepsis-3) and Community-Acquired Pneumonia Mortality: a validation and clinical decision-making study

Running title: *Sepsis-3 in community-acquired pneumonia*

Otavio T. Ranzani MD, MSc[1,2], Elena Prina MD[1], Rosario Menéndez MD, PhD, FERS[3], Adrian Ceccato MD[1,4], Catia Cilloniz PhD[1], Raul Méndez MD[3], Albert Gabarrus MSc[1], Enric Barbeta MD[1], Gianluigi Li Bassi MD PhD[1], Miquel Ferrer MD PhD FERS[1], Antoni Torres MD PhD FERS[1]

1- Department of Pulmonology, Hospital Clinic of Barcelona, University of Barcelona, Institut D'investigacions August Pi I Sunyer (IDIBAPS); Centro de Investigación Biomedica En Red-Enfermedades Respiratorias (CibeRes, CB06/06/0028), Barcelona, Spain;
2- Respiratory Intensive Care Unit, Pulmonary Division, Heart Institute, Hospital das Clínicas, University of Sao Paulo, Sao Paulo, Brazil;
3- Pneumology Department, ISS/Hospital Universitario y Politecnico La Fe, CIBER Enfermedades Respiratorias (CIBERES), Valencia, Spain;
4-Seccion Neumologia, Hospital Nacional Prof. Alejandro Posadas, Palomar, Argentina.

Corresponding author:
Dr. Antoni Torres
UVIR, *Servei de Pneumologia, Hospital Clínic, Villarroel* 170
08036 Barcelona, Spain.
Phone/Fax: +34 93 227 55 49
E-mail: ATORRES@clinic.ub.es.
Website: www.idibapsrespiratoryresearch.org

Funding/Support: Centro de Investigación Biomedica En Red-Enfermedades Respiratorias (CibeRes); OTR and AC are supported by the European Respiratory Society Research Fellowships (ERS-LTRF).

Author Contributions: Study concept and design: OTR, EP and AT. Acquisition, analysis, or interpretation of data: OTR, EP, RoM, AC, CC, RM, AG, EB, GLB, MF and AT. Drafting of the manuscript: OTR, EP. Critical revision of the manuscript for important intellectual content: RoM, AC, CC, RM, AG, EB, GLB, MF and AT. Statistical analysis: OTR. Administrative, technical, or material support: RoM, AC, CC, RM, AG, EB, GLB, MF and AT. Study supervision: AT.

Conflict of Interest Disclosures: All authors declare that they have no conflicts of interest.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Descriptor: 10.12 Pneumonia: Bacterial Infections

"At a Glance Commentary"

Scientific Knowledge on the Subject:

In 2016, the Sepsis-3 Task Force updated the clinical criteria for sepsis, excluding the need for SIRS and introducing a flowchart that comprises the qSOFA and SOFA scores. However, the clinical decision-making process cannot rely on risk stratification scores, because a decision-aid tool must account for the benefits and harms of clinicians incorporating that tool into clinical practice. A clinical decision-making analysis of Sepsis-3 is not yet available.

What This Study Adds to the Field:

We demonstrated that qSOFA outperformed SIRS and presented better clinical usefulness in patients with community-acquired pneumonia. Among the tools for initial assessment, SIRS presented the worst net benefit versus qSOFA and CRB, significantly increasing the risk of over-treatment and being comparable to the "treat-all" strategy. Among the tools for a comprehensive assessment, PSI had better predictive performance and net benefit for mortality than mSOFA and CURB-65; while mSOFA was more useful when considering mortality/ICU admission. Finally, following the Sepsis-3 flowchart resulted in better identification of patients at high risk of worse outcomes.

Word count: 3,523

"This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org"

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

ABSTRACT

Rationale: Sepsis-3 Task Force updated the clinical criteria for sepsis, excluding the need for systemic inflammatory response syndrome (SIRS) criteria. The clinical implications of the proposed flowchart including the quick Sequential (Sepsis-related) Organ Failure Assessment (qSOFA) and SOFA scores are unknown.

Objective: To perform a clinical decision-making analysis of Sepsis-3 in patients with community-acquired pneumonia (CAP).

Methods: Cohort study including adult patients with CAP from two Spanish university-hospitals. SIRS, qSOFA, CRB (Confusion, Respiratory rate and Blood pressure), mSOFA, CURB-65 and Pneumonia Severity Index (PSI) were calculated with data from the emergency department. We used decision-curve analysis to evaluate the clinical usefulness of each score and the primary outcome was in-hospital mortality.

Measurements and Main Results: Of 6,874 patients, 442 (6.4%) died in hospital. SIRS presented the worst discrimination, followed by qSOFA, CRB, mSOFA, CURB-65 and PSI. Overall, overestimation of in-hospital mortality and mis-calibration was more evident for qSOFA and mSOFA. SIRS had lower net benefit than qSOFA and CRB, significantly increasing the risk of over-treatment and being comparable to the "treat-all" strategy. PSI had higher net benefit than mSOFA and CURB-65 for mortality; while mSOFA seemed more applicable when considering mortality/ICU admission. Sepsis-3 flowchart resulted in better identification of patients at high risk of mortality.

Conclusions: qSOFA and CRB outperformed SIRS and presented better clinical usefulness as prompt tools for CAP patients in the emergency department. Among the tools for a comprehensive patient assessment, PSI had the best decision-aid tool profile.

Word count: 241

Keywords: qSOFA; SIRS; validation, pneumonia, sepsis

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

INTRODUCTION

Community-acquired pneumonia (CAP) represents a significant infection burden worldwide, and it is often complicated by sepsis (1-4). Early recognition of sepsis is fundamental to guide treatment, improve outcomes and decrease costs (5-7). In contrast, in patients with uncomplicated infection, over-treatment should be avoided to prevent unnecessary harm.

Sepsis is a syndrome characterized by a dysregulated host response to infection leading to life-threatening organ dysfunction (5). In 2016, the Sepsis-3 Task Force updated previous recommendations primarily aiming to accurately differentiate between sepsis and uncomplicated infection (5). By applying a data-driven approach to identify patients at risk of worse outcomes, the Task Force proposed a new clinical definition, removing the need for systemic inflammatory response syndrome (SIRS) criteria. Thus, in infected patients, sepsis was clinically defined by an increase in Sequential (Sepsis-related) Organ Failure Assessment (SOFA) score of 2 points or more. Additionally, a bedside score for risk stratification, namely the quick SOFA (qSOFA), has been proposed, which incorporates hypotension, altered mental status and tachypnea (5, 8).

In patients with CAP, several scores have been developed to identify high-risk patients and support therapeutic decisions (4, 9). Two of these scores, CURB-65 (Confusion, Urea, Respiratory rate, Blood pressure and Age) and PSI (Pneumonia Severity Index) are well-validated scores to support CAP management and prognosis (9, 10). Simplifications of CURB-65 (i.e., CRB-65 and CRB) (11) have been developed and validated to facilitate the risk stratification process; these simplified scores do not require blood tests (12), as in the qSOFA. Yet the definitions for hypotension and tachypnea parameters on the CRB tool differ from those of the qSOFA.

Sepsis-3 will change clinical practice and influence medical decisions. However, clinical decision-making cannot rely only on predictive performance measures, such as discrimination and calibration (13, 14). Indeed, decision-aid tools must also account for the benefits and harms resulting from clinicians' choice (13, 14). To date, no clinical decision-making analysis of Sepsis-3 is available, including the proposed bedside tool (qSOFA) and the Sepsis-3 Flowchart, which includes qSOFA and SOFA scores. Therefore, the aim of our study was to evaluate three tools for initial assessment (SIRS, qSOFA and CRB) and three tools for a comprehensive assessment (SOFA, CURB-65 and PSI) as decision-aid prognostic tools in CAP using decision-curves methodology. Additionally, the Sepsis-3 flowchart was also applied in this population.

Some of the results of this study have been previously reported in the form of an abstract (15).

METHODS

Study design and patients

We retrospectively analyzed patients from two cohorts, which prospectively included patients aged ≥16 years with a clinical diagnosis of CAP from two Spanish University-Hospitals (Hospital Clinic, Barcelona from 1996 to 2015; Hospital Universitario y Politecnico La Fe, Valencia from 2012 to 2015). These cohorts had comparable inclusion and exclusion criteria and definitions of the variables. Local Research Ethics Committees approved both cohorts (Barcelona, Spain – Register: 2009/5451; Valencia, Spain – Register: 2011/0219).

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

CAP was defined as a new pulmonary infiltrate on chest radiograph upon hospital admission and acute symptoms of lower respiratory tract infection (e.g., fever, cough, sputum production, pleuritic chest pain). Immunosuppression (i.e., patients taking more than 10 mg of prednisone-equivalent per day for at least 2 weeks, on cytotoxic therapy or with acquired immunodeficiency syndrome) and active tuberculosis were exclusion criteria. We included patients from nursing home. Demographic variables, comorbidities, and physiologic parameters were collected in the emergency department (ED). All patients had a complete microbiologic evaluation and microbiologic confirmation of CAP was defined according to current guidelines (16, 17). In each institution, a dedicated clinical researcher prospectively included patients, under the supervision of an experienced pulmonary physician. Patients were followed up until hospital discharge, and all survivors were re-examined or contacted by phone 30 days after hospital discharge. Further details are reported in previous publications (16, 17).

Outcomes

Our primary outcome was all-cause in-hospital mortality (5, 8). We also explored two secondary outcomes: 1) in-hospital mortality and/or need for critical care support ≥3 days ("composite outcome") (5, 8); 2) 30-day mortality. We defined need for critical care support as admission to an intensive care unit (ICU) or high-dependency unit (HDU).

Scores definition

We clustered the six scores in those that might facilitate: a) the clinician's initial decision (SIRS, qSOFA and CRB); and b) clinician's decision after initial management and additional exams (SOFA, CURB-65 and PSI). We adapted the Sepsis-3 flowchart illustrating

this approach and the timeline of the clinical decision-making processes involved in the ED (**Figure 1**).

We defined SIRS, qSOFA, CRB, CURB-65 and PSI as originally described (**Table E1**) (5, 8, 9, 12). For SOFA score, we calculated the respiratory, hematological, hepatic, and renal systems as originally described. However, we adapted the SOFA calculation for neurologic and cardiovascular parameters, using a conservative approach similar to Sepsis-3 (modified SOFA–mSOFA, **Table E1**). We used the first clinical signs/symptoms documented in the ED for all scores. For mSOFA, we used the first reported data, comprising the early resuscitation phase, as previously validated (18). For missing mSOFA values, we attributed a normal value (i.e., zero points), reflecting clinical practice and as widely reported (5, 8). In a sensitivity analysis, we used multiple imputation (5, 8). We also compared qSOFA and CRB with their corresponding qSOFA-65 and CRB-65, by adding the age component.

Statistical Analysis

We assessed the predictive performance of SIRS, qSOFA, CRB, mSOFA, CURB-65 and PSI for the primary and secondary outcomes (19). We evaluated calibration with calibration plots and two complementary goodness-of-fit statistics (*Hosmer-Lemeshow* and the *le Cessie-van Houwelingen-Copas-Hosmer* tests*)* (20). Calibration curves were built with a smoothed non-parametric method (20, 21). We used the area under the receiver operating characteristic curve (AUROC) to assess discrimination. The 95% confidence interval (CI) estimation for the AUROCs and their comparisons were performed using bootstrapping methods in 10,000 samples (21, 22). Overall fit was assessed using scaled Brier score and Nagelkerke R-square (19, 21). To incorporate important information that clinicians might have at the bedside (8), we evaluated the additional predictive contribution of SIRS, qSOFA, CRB and mSOFA to a baseline risk for in-hospital mortality estimated by a multivariate

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

logistic regression model. The baseline risk model included age, gender, chronic respiratory disease, chronic neurologic disease, liver disease, heart failure, diabetes mellitus, neoplasia, chronic renal disease, and microbiologic confirmation. The baseline and additional risk models were fitted after multiple imputation.

For a score to be clinically useful, it must have good discrimination and be well-calibrated but those alone are not enough (14, 23, 24). Indeed, discrimination and calibration may not reflect clinical utility (25). The main barrier to translating discrimination and calibration to clinical practice is that sensitivity, specificity and prediction errors are weighted equally (e.g., true-positive and false-positive rates), while clinicians usually apply different weights during the decision-making process (23). Decision-curve analysis is a method that depicts the predicted net benefit ("NB = benefit x true-positive classifications minus harm/cost x false-positive classifications") of a prediction tool over a range of threshold probabilities. Threshold probabilities quantify how over-treatment is considered against treatment benefits (19, 23, 25-28). For instance, if a clinician weights the harm/cost of overtreatment versus the benefit of appropriated treatment at 1:19, we have a threshold probability of 5% and a number willing to treat (NWT) of 20 (26, 29). Decision curves have the advantage of being able to plot a plausible range of threshold probabilities. We defined 100 to 5 NWTs as a plausible range (i.e., threshold probabilities from 0 to 20%), because it is unlikely that clinicians will use a score to make decisions about treatment of infected patients for higher threshold probabilities. At any given NWT, the score with the higher net benefit is the preferred one. The NB of each score was estimated for the primary and secondary outcomes and compared with the "treat-none" and "treat-all" strategies. The "treat-all" strategy assumes everyone will develop the event and receive the intervention independent of any score. The associated intervention comprises the initial treatment of septic patients in the ED, such as additional blood sampling, aggressive resuscitation, intensive monitoring,

invasive procedures, and, place of treatment. We hypothesized harm, at patient and hospital levels, associated with over-treatment and overuse of hospital resources, such as adverse events of broad-spectrum antibiotics and aggressive resuscitation/invasive procedures, ICU admission for patients unlikely to benefit and hospital costs (**Figure 1**) (4, 5, 30). Finally, we described the distribution and outcomes of patients based on combinations between SIRS (resembling Sepsis-2 definition), qSOFA (Sepsis-3 flowchart) and CRB with mSOFA.

Sensitivity, specificity, positive and negative predictive values were calculated as shown elsewhere. As we expected few missing values for SIRS, qSOFA and CRB, our main analysis was conducted on the complete-case data; for sensitivity analysis, we conducted multiple imputation. We pre-specified two subgroups, defined by age (<65, ≥65yo), and chronic comorbidities (without chronic comorbidities, ≥1 chronic comorbidity). All statistical analyses were performed using R software, version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria) (31). We followed the TRIPOD guidelines (32) and further information about methods and statistical analysis are in the appendix.

RESULTS

Patient Characteristics

The "Barcelona cohort" included 6,304 patients and the "Valencia cohort" 570 patients, totaling 6,874 patients. The patients had a mean age of 66 (19) years, 62.2% were males and 65.5% presented ≥1 chronic comorbidity and 2,860 (41.8%) had microbiologic confirmation (**Table 1**). Seven hundred seventy-eight (11.3%) patients were discharged after a short-stay period in the ED, while 5,146 (74.9%) and 950 (13.8%) were admitted to the ward and ICU/HDU, respectively. Overall, in-hospital mortality occurred in 442 (6.4%)

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

patients, in-hospital mortality or ≥3 ICU/HDU days in 716 (10.4%) and, 30-day mortality in 477/6,377 (7.5%) (**Table 1**).

Scores distribution

Our complete-case analysis comprised 6,024 patients (87.6%) (**Table E2**, **Table E3, Figure E1**). There was a clear association between qSOFA, CRB, mSOFA, CURB-65 and PSI with in-hospital mortality. Nevertheless, higher SIRS points poorly predicted in-hospital mortality (**Figure 2**). Similar results were found in the imputed data (**Figure E2**) and for secondary outcomes (**Figure E3**, **Figure E4**).

Very few patients who were discharged after a short ED stay had qSOFA and CRB ≥2 points (4 and 2%, respectively), whereas 61% had SIRS≥2 points (**Figure 3**). These patients had very low 30-day mortality (3/744, 0.4%). In contrast, patients admitted to the ICU/HDU had in-hospital mortality of 15.7%, and higher scores. Sepsis (Infection+mSOFA≥2 points) was present in 17% of patients discharged after a short ED stay, 64% of those admitted to the ward and 89% of patients treated in the ICU/HDU (**Figure 3**).

Scores performance

SIRS presented the lowest discrimination value (AUROC 0.579, 95% CI 0.551-0.605), followed by qSOFA (AUROC 0.697, 95% CI 0.671-0.722), CRB (AUROC 0.716, 95% CI 0.690-0.741), CURB-65 (AUROC 0.746, 95% CI 0.722-0.769), mSOFA (AUROC 0.748, 95% CI 0.721-0.774) and PSI (AUROC 0.780, 95% CI 0.760-0.799)(**Table 2**, **Figure E5**). All scores presented worse discrimination for in-hospital mortality in patients ≥65 years old. In those patients without chronic comorbidities, the discrimination of all scores improved (**Table E4**). Regarding calibration, in general scores overestimated in-hospital mortality, and mis-calibration was more evident for qSOFA, mSOFA and CURB-65 (**Table 2**, **Figure E5**).

The overall performance measured by the scaled Brier score and R-square increased from SIRS to qSOFA, CRB, CURB-65, mSOFA and PSI (**Table 2**). We observed similar results when analyzing Barcelona and Valencia cohorts separately, but mSOFA and CURB-65 had better discrimination in the Valencia cohort (**Table E5**). We found similar results when analyzing the imputed data (**Table E6**, **Figure E6**) and, for secondary outcomes (**Table E7**, **Figure E7**, **Table E8**, **Figure E8**). Nevertheless, for the composite outcome, CRB had better discrimination than qSOFA and mSOFA had the highest discrimination and best calibration. CRB-65 outperformed qSOFA, CRB and qSOFA-65 for in-hospital mortality (**Table E9**, **Figure E9**).

Scores additional contribution to the baseline risk model

When analyzing the predictive performance for in-hospital mortality in the multiple imputed data, SIRS contributed very little to the baseline model discrimination ($AUROC_{Baseline}$ 0.745, 95% CI 0.722-0.766 vs. $AUROC_{Baseline+SIRS}$ 0.752, 95% CI, 0.731-0.774). In contrast, qSOFA ($AUROC_{Baseline+qSOFA}$ 0.780, 95% CI, 0.760-0.800), CRB ($AUROC_{Baseline+CRB}$ 0.794, 95% CI, 0.775-0.813), and mSOFA ($AUROC_{Baseline+mSOFA}$ 0.836, 95% CI, 0.818-0.854) notably improved the model discrimination. These improvements in discrimination were also observed in the IDI measures (**Table E10, Figure E10**).

Clinical usefulness and decision-curve analysis

Among the tools for the initial assessment, SIRS≥2 presented high sensitivity and low specificity, while qSOFA≥2 and CRB≥2 presented moderate sensitivity and high specificity for in-hospital mortality (**Table 2**). Among the follow-up tools, mSOFA≥2 presented high sensitivity and low specificity and CURB-65≥2 and PSI≥4 presented a good compromise between sensitivity (78 and 92%, respectively) and specificity (60 and 47%, respectively). CRB had the highest positive likelihood ratio (3.05, 95% CI, 2.65-3.51) and PSI the lowest

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

negative likelihood ratio (0.16, 95% CI, 0.12-0.23) (**Table 2**). We observed the same pattern in the imputed data (**Table E6**), and for secondary outcomes (**Table E7**, **Table E8**). In the sub-group analysis for in-hospital mortality, we observed similar findings except that mSOFA≥2 had higher specificity in the subgroup of patients aged<65 years (sensitivity 94, specificity 51%) and without chronic comorbidities (sensitivity 88, specificity 51%) (**Table E11**).

The net benefit of qSOFA and CRB outperformed SIRS for in-hospital mortality, and SIRS showed a NB close to the "treat-all" strategy for the majority of the NWT values (**Figure 4-A**). For NWT between 15-30 and lower than 8, CRB had higher NB than qSOFA. PSI had the highest NB over the whole NWT range, except for values lower than 8, when mSOFA outperformed PSI for in-hospital mortality. When translating these findings to the number of avoided interventions in a hypothetical population of 100 patients with pneumonia, assuming a physician weights the harm/cost of overtreatment versus the benefit of appropriated treatment at 1:19 (NWT=20), the number of interventions could have been decreased by 7% without missing any death using SIRS; 16% using qSOFA; 27% using CRB or mSOFA; 30% using CURB-65 and 35% using PSI (**Figure 4-B**). We observed similar findings on NB for secondary outcomes, except that mSOFA outperformed other scores for a wide range of NWT for the composite outcome (**Figure 4 C-D**, **Figure 4 E-F**). The NB of the full models showed that "Baseline model+SIRS" had virtually no advantage compared with the "Baseline model" alone. The models "Baseline+qSOFA" and "Baseline+CRB" had higher NB than previous models for NWTs between 25 and 7. In contrast, the model "Baseline+mSOFA" presented the highest NB over the whole NWT range (**Figure E11**).

Patients positive for SIRS/mSOFA (n=3,274, 54%) had 9.0% (95% CI, 8.5-9.5) in-hospital mortality, while those positive for qSOFA/mSOFA (n=1,090, 18%) and CRB/mSOFA (n=788, 13%) had 16.6% (95% CI, 15.5-17.7) and 18.0% (95% CI, 16.7-19.4)

in-hospital mortality, respectively (**Figure 5**). Similar findings were observed in the imputed data (**Figure E12**), and among patients with or without chronic comorbidities (**Figure E13**).

DISCUSSION

In a population of patients with community-acquired pneumonia, qSOFA outperformed SIRS for in-hospital mortality risk stratification and presented better clinical usefulness virtually in all evaluations. CRB had slightly better predictive performance than qSOFA for discrimination and calibration measures, but presented similar clinical usefulness for the majority of scenarios. For a comprehensive assessment of CAP, mSOFA and PSI had the best predictive performance and highest net benefit. The combination of qSOFA or CRB with mSOFA better selected high-risk patients, while potentially decreasing the burden of intensive monitoring and overtreatment.

The Sepsis-2 definitions, published in 2001, raised awareness of sepsis syndrome and have been associated with better care and outcomes (6, 7). However, SIRS criteria weakly predicted patient outcomes (3, 33), which associated with its high sensitivity and low specificity, likely classify SIRS as an unreliable tool for bedside clinical decision-making, and research (5, 8, 34, 35). Our current analysis in CAP patients confirmed these limitations (3) and highlighted risks of overtreatment, demonstrating that the NB of SIRS is comparable to the "treat-all" strategy. Indeed, the decision-curve analysis showed that when different weights for true-positive and false-positive classifications were applied, SIRS did not provide any additional benefit for decision-making. In contrast, we found a positive NB if clinicians incorporated qSOFA or CRB for the initial assessment, decreasing the number of unnecessary interventions while not missing any death. qSOFA and CRB were better than

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

SIRS or a "treat-all" strategy for NWT values below 40, which seems reasonable for use in the ED (5, 30), where qSOFA and CRB can be easily assessed. Given that CRB and CRB-65 were specifically developed for CAP patients, they had better calibration and discrimination than qSOFA, as well as higher specificity. Thus, rather than qSOFA, physicians could consider CRB or CRB-65 for the initial risk stratification of CAP patients.

For a comprehensive assessment of CAP, PSI had the best mortality prediction and highest NB from high NWT values down to an NWT of 8, reinforcing its pivotal role on CAP management. mSOFA seemed to be more applicable for NWT values below 12, mainly when considering ICU admission. This might be because PSI comprises 20 variables and has age as a main determinant for risk classification; while mSOFA measures acute organ dysfunctions in 6 domains. Further studies should investigate whether both scores are complementary in CAP management. Of note, the Baseline+mSOFA model, which could be analogous to PSI+mSOFA, had higher discrimination and NB than PSI alone.

Our results are in line with those of the pivotal Sepsis-3 clinical criteria study (5, 8), which showed better discrimination for qSOFA and mSOFA compared with SIRS. In contrast, mSOFA clearly outperformed qSOFA in our population. The discrimination of qSOFA in our study was lower than that reported originally in Sepsis-3 (5), which might be because of the differences in the populations included and because we measured qSOFA and mSOFA using ED data. Sepsis-3 aimed to identify infected patients with ≥10% of mortality (5, 8, 36). In our study that goal was achieved: 18% of the patients presented positive qSOFA/mSOFA and in-hospital mortality in these patients was 16.6%. Interestingly, when describing the prevalence of each score categorized by place of treatment, it seems that clinicians relied on the parameters hypotension, altered mental status and tachypnea for decision-making. Indeed, only 2% and 4% of patients who were not hospitalized had qSOFA≥2 and CRB≥2, respectively. However, SIRS≥2 was present in the vast majority of

the promptly discharged patients (61%). Interestingly, 46% of patients had qSOFA<2/mSOFA≥2; in-hospital mortality in these patients was low (5.4%); this might indicate that patients with qSOFA<2 presented some points on mSOFA, but ultimately not associated with death. Among the scores we evaluated, qSOFA was recently developed by a data-driven process from large databases. As with CRB, it attributes one point to each clinical parameter, is promptly available at bedside and is easily repeated without invasive measures. Yet it is important to emphasize that the suggested cut-off of 2 points for qSOFA had low sensitivity, being inappropriate if applied as a single screening tool, resulting in delayed recognition of sepsis (37).

Our study has some strengths that must be highlighted. First, we described challenges in decision-making that could be faced by clinicians on a daily basis, not only during evaluation of hospitalized patients, but also in those rapidly discharged following ED evaluation. Additionally, it is known that predictive performance measures have disadvantages (19-21, 32, 38) and are difficult to translate into clinical-practice (14). Thus, we used clinical decision-making analyses (13) to complement predictive performance evaluations, which are fundamental to better support clinicians' decision (23, 24, 39).

This study has also some limitations. First, we analyzed one type of infection, from only two Spanish institutions, potentially limiting generalizability of our results. However the data came from two prospective CAP cohorts, increasing our ability to capture data granularity. Second, although our data were prospective collected from consecutive patients and had few missing values, misclassification and selection bias could have occurred. We expect both to be low, due to the standard procedures for prospective data collection and researchers' extensive expertise in this field. Moreover, our outcomes were objective (mortality/ICU admission) and we had few losses to follow-up, decreasing the possibility of outcome bias. Third, we could not fully calculate the SOFA score for the cardiovascular and

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

neurological parameters; thus, by adopting a conservative approach we may have hampered the SOFA performance. However, the mSOFA score maintained its high predictive power, confirming feasibility of SOFA score calculation outside the ICU (18). Fourth, we could not differentiate between acute and chronic organ dysfunction; however, our analysis excluding patients with chronic comorbidities showed similar findings. Fifth, we observed score mis-calibration, which can influence clinical decision based on NB (40). Finally, we did not incorporate clinical judgment into the models, which could ultimately improve the performance of the Sepsis-3 flowchart.

CONCLUSIONS

We demonstrated that for initial assessment, qSOFA outperformed SIRS and presented better clinical usefulness in CAP patients in the ED. Moreover, CRB and CRB-65 had better predictive performance than qSOFA for initial stratification of CAP patients in some scenarios, including higher net benefit for some values of NWT. For the comprehensive assessment of CAP, PSI had the best predictive performance and net benefit for mortality, while mSOFA seemed more suitable when considering ICU admission. Finally, the Sepsis-3 flowchart provided an improved, feasible approach for identifying patients with CAP at higher risk of death. Further studies, including other CAP cohorts and other sources of infection, should be conducted to corroborate our findings.

ACKNOWLEDGMENTS We would like to thank clinicians and healthcare professionals who assiduously work in the collaborating institutions, and who helped in the development of both cohorts.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

References:

1. Alberti C, Brun>Buisson C, Chevret S, Antonelli M, Goodman SV, Martin C, Moreno R, Ochagavia AR, Palazzo M, Werdan K, Le Gall JR. Systemic inflammatory response and progression to severe sepsis in critically ill infected patients. *Am J Respir Crit Care Med* 2005; 171: 461>468.
2. Restrepo MI, Jorgensen JH, Mortensen EM, Anzueto A. Severe community>acquired pneumonia: current outcomes, epidemiology, etiology, and therapy. *CurrOpinInfectDis* 2001; 14: 703> 709.
3. Dremsizov T, Clermont G, Kellum JA, Kalassian KG, Fine MJ, Angus DC. Severe sepsis in community> acquired pneumonia: when does it happen, and do systemic inflammatory response syndrome criteria help predict course? *Chest* 2006; 129: 968>978.
4. Prina E, Ranzani OT, Torres A. Community>acquired pneumonia. *Lancet* 2015; 386: 1097>1108.
5. Singer M, Deutschman CS, Seymour CW, Shankar>Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis>3). *JAMA* 2016; 315: 801>810.
6. Rhodes A, Phillips G, Beale R, Cecconi M, Chiche JD, De Backer D, Divatia J, Du B, Evans L, Ferrer R, Girardis M, Koulenti D, Machado F, Simpson SQ, Tan CC, Wittebole X, Levy M. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPreSS study). *Intensive Care Med* 2015; 41: 1620>1628.
7. Noritomi DT, Ranzani OT, Monteiro MB, Ferreira EM, Santos SR, Leibel F, Machado FR. Implementation of a multifaceted sepsis education program in an emerging country setting: clinical outcomes and cost>effectiveness in a long>term follow>up study. *Intensive Care Med* 2014; 40: 182>191.
8. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar>Hari M, Singer M, Deutschman CS, Escobar GJ, Angus DC. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis>3). *JAMA* 2016; 315: 762>774.
9. Singanayagam A, Chalmers JD. Severity assessment scores to guide empirical use of antibiotics in community acquired pneumonia. *Lancet Respir Med* 2013; 1: 653>662.
10. Capelastegui A, Espana PP, Quintana JM, Areitio I, Gorordo I, Egurrola M, Bilbao A. Validation of a predictive rule for the management of community>acquired pneumonia. *EurRespir J* 2006; 27: 151>157.
11. Bauer TT, Ewig S, Marre R, Suttorp N, Welte T, Group CS. CRB>65 predicts death from community>acquired pneumonia. *J Intern Med* 2006; 260: 93>101.
12. Lim WS, Baudouin SV, George RC, Hill AT, Jamieson C, Le JI, Macfarlane JT, Read RC, Roberts HJ, Levy ML, Wani M, Woodhead MA. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009; 64 Suppl 3: iii1>55.
13. Moskowitz A, Andersen LW, Cocchi M, Donnino MW. The Misapplication of Severity>of>Illness Scores toward Clinical Decision Making. *Am J Respir Crit Care Med* 2016; 194: 256>258.
14. Holmberg L, Vickers A. Evaluation of prediction models for decision>making: beyond calibration and discrimination. *PLoS Med* 2013; 10: e1001491.
15. Bos L, Schouten L, van Vught L, Wiewel M, Ong D, Cremer O, Artigas A, Martin>Loeches I, Hoogendijk A, van der Poll T, Horn J, Juffermans N, Schultz M, de Prost N, Pham T, Carteaux G, Dessap AM, Brun>Buisson C, Fan E, Bellani G, Laffey J, Mercat A, Brochard L, Maitre B, Howells PA, Thickett DR, Knox C, Park DP, Gao F, Tucker O, Whitehouse T, McAuley DF, Perkins GD, Pham T, Laffey J, Bellani G, Fan E, Pisani L, Roozeman JP, Simonis FD, Giangregorio A, Schouten LR, Van der Hoeven SM, Horn J, Neto AS, Festic E, Dondorp AM, Grasso S, Bos LD, Schultz MJ, Koster>Brouwer M, Verboom D, Scicluna B, van de Groep K, Frencken J, Schultz M, van der Poll T, Bonten M, Cremer O, Ko JI, Kim KS, Suh GJ, Kwon WY,

Kim K, Shin JH, Ranzani OT, Prina E, Menendez R, Ceccato A, Mendez R, Cilloniz C, Gabarrus A, Ferrer M, Torres A, Urbano A, Zhang LA, Swigon D, Pike F, Parker RS, Clermont G, Scheer C, Kuhn SO, Modler A, Vollmer M, Fuchs C, Hahnenkamp K, Rehberg S, Gründling M, Taggu A, Darang N, Öveges N, László I, Tánczos K, Németh M, Lebák G, Tudor B, Érces D, Kaszaki J, Huber W, Trásy D, Molnár Z, Ferrara G, Edul VSK, Canales HS, Martins E, Canullán C, Murias G, Pozo MO, Eguillor JFC, Buscetti MG, Ince C, Dubin A, Aya HD, Rhodes A, Fletcher N, Grounds RM, Cecconi M, Jacquet>Lagrèze M, Riche M, Schweizer R, Portran P, Fornier W, Lilot M, Neidecker J, Fellahi JL, Escoresca>Ortega A, Gutiérrez>Pizarraya A, Charris>Castro L, Corcia>Palomo Y, Fernandez>Delgado E, Garnacho>Montero J, Roger C, Muller L, Elotmani L, Lipman J, Lefrant JY, Roberts JA, Muñoz>Bermúdez R, Samper M, Climent C, Vasco F, Sara V, Luque S, Campillo N, Cerrato SG, Masclans JR, Alvarez>Lerma F, Brugger SC, Jimenez GJ, Torner MM, Cabello JT, Garrido BB, Casals XN, Gaite FB, Vidal MV, Martínez MP, Gusarov V, Shilkin D, Dementienko M, Nesterova E, Lashenkova N, Kuzovlev A, Zamyatin M, Demoule A, Carreira S, Lavault S, Palancca O, Morawiec E, Mayaux J, Arnulf I, Similowski T, Rasmussen BS, Maltesen RG, Hanifa M, Pedersen S, Kristensen SR, Wimmer R, Panigada M, Bassi GL, Ranzani OT, Kolobow T, Zanella A, Cressoni M, Berra L, Parrini V, Kandil H, Salati G, Livigni S, Amatu A, Andreotti A, Tagliaferri F, Moise G, Mercurio G, Costa A, Vezzani A, Lindau S, Babel J, Cavana M, Consonni D, Pesenti A, Gattinoni L, Torres A, Mansouri P, Zand F, Zahed L, Dehghanrad F, Bahrani M, Ghorbani M, Cambiaghi B, Moerer O, Mauri T, Kunze>Szikszay N, Ritter C, Pesenti A, Quintel M, Vilander LM, Kaunisto MA, Vaara ST, Pettilä V, Mulier JLGH, Rozemeijer S, Spoelstra>de Man AME, Elbers PE, Tuinman PR, de Waard MC, Oudemans>van Straaten HM, Liberatore AMA, Souza RB, Martins AMCRPF, Vieira JCF, Koh IHJ, Martínez MG, Sánchez RJ, Gascón LM, Mulero MDR, Freire AO, Muñoz AO, Acebes SR, Martínez ÁF, Aliaga SM, Para LH, Payá JM, Mulero FR, Guerci P, Ince Y, Heeman P, Ergin B, Ince C, Uz Z, Massey M, Ince Y, Papatella R, Bulent E, Guerci P, Toraman F, Ince C, Longbottom ER, Torrance HD, Owen HC, Hinds CJ, Pearse RM, O'Dywer MJ, Trogrlic Z, van der Jagt M, Lingsma H, Ponssen HH, Schoonderbeek JF, Schreiner F, Verbrugge SJ, Duran S, van Achterberg T, Bakker J, Gommers DAMPJ, Ista E, Krajčová A, Waldauf P, Duška F, Shah A, Roy N, McKechnie S, Doree C, Fisher S, Stanworth SJ, Jensen JF, Overgaard D, Bestle MH, Christensen DF, Egerod I, Pivkina A, Gusarov V, Zhivotneva I, Pasko N, Zamyatin M, Jensen JF, Egerod I, Bestle MH, Christensen DF, Alklit A, Hansen RL, Knudsen H, etal. ESICM LIVES 2016: part one. A8 Validation of quick SOFA in a large population of patients with community>acquired pneumonia. *Intensive Care Medicine Experimental* 2016; 4: 27.

16. Prina E, Ranzani OT, Polverino E, Cillóniz C, Ferrer M, Fernandez L, Puig de la Bellacasa J, Menéndez R, Mensa J, Torres A. Risk factors associated with potentially antibiotic>resistant pathogens in community>acquired pneumonia. *Ann Am Thorac Soc* 2015; 12: 153>160.
17. Cilloniz C, Ewig S, Polverino E, Marcos MA, Esquinas C, Gabarrus A, Mensa J, Torres A. Microbial aetiology of community>acquired pneumonia and its relation to severity. *Thorax* 2011; 66: 340>346.
18. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 2009; 37: 1649>1654.
19. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
20. Harrell Jr F. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. USA: Springer; 2001.
21. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21: 128>138.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

22. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54: 774>781.
23. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6.
24. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. *Ann Intern Med* 2012; 157: 294>295.
25. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565>574.
26. Dorresteijn JA, Visseren FL, Ridker PM, Wassink AM, Paynter NP, Steyerberg EW, van der Graaf Y, Cook NR. Estimating treatment effects for individual patients based on the results of randomised clinical trials. *BMJ* 2011; 343: d5888.
27. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, Field JK. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case>control and cohort validation study. *Ann Intern Med* 2012; 157: 242>250.
28. Siddiqui MM, Rais>Bahrami S, Turkbey B, George AK, Rothwax J, Shakir N, Okoro C, Raskolnikov D, Parnes HL, Linehan WM, Merino MJ, Simon RM, Choyke PL, Wood BJ, Pinto PA. Comparison of MR/ultrasound fusion>guided biopsy with ultrasound>guided biopsy for the diagnosis of prostate cancer. *JAMA* 2015; 313: 390>397.
29. van der Leeuw J, Ridker PM, van der Graaf Y, Visseren FL. Personalized cardiovascular disease prevention by applying individualized prediction of treatment effects. *Eur Heart J* 2014; 35: 837>843.
30. Welker JA, Huston M, McCue JD. Antibiotic timing and errors in diagnosing pneumonia. *Arch Intern Med* 2008; 168: 351>356.
31. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2015.
32. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1>73.
33. Kaukonen KM, Bailey M, Pilcher D, Cooper DJ, Bellomo R. Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med* 2015; 372: 1629>1638.
34. Churpek MM, Zadravecz FJ, Winslow C, Howell MD, Edelson DP. Incidence and Prognostic Value of the Systemic Inflammatory Response Syndrome and Organ Dysfunctions in Ward Patients. *Am J Respir Crit Care Med* 2015; 192: 958>964.
35. Vincent JL. Dear SIRS, I'm sorry to say that I don't like you. *Crit Care Med* 1997; 25: 372>374.
36. Besen BAMP, Romano TG, Nassar AP, Taniguchi LU, Azevedo LCP, Mendes PV, Zampieri FG, Park M. Sepsis>3 definitions predict ICU mortality in a low–middle>income country. *Annals of Intensive Care* 2016; 6: 107.
37. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, Edelson DP. Quick Sepsis>related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and Early Warning Scores for Detecting Clinical Deterioration in Infected Patients outside the Intensive Care Unit. *Am J Respir Crit Care Med* 2017; 195: 906>911.
38. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest* 2012; 42: 216>228.
39. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA* 2015; 313: 409>410.
40. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision>analytic performance. *Med Decis Making* 2015; 35: 162>169.

FIGURE LEGENDS:

Figure 1. **Flowchart about the decision-making process for community-acquired pneumonia management at the emergency department**

*First clinical decision encompasses the decision to assess organ dysfunction and pneumonia severity with additional laboratory and/or invasive procedures

†Second clinical decision encompasses the decision, after full assessment, to admit the patient to the ward/ICU, consider additional treatment not yet started or change treatments already started

The flowchart does not regulate timing for institute life-saving treatments or, for instance, prompt starting of empiric antibiotic treatment.

Figure 2. **In-hospital mortality stratified by each score in patients with community-acquired pneumonia**

Panel A: systemic inflammatory response syndrome (SIRS) criteria; Panel B: quick Sequential (Sepsis-related) Organ Failure; Panel C: Confusion, Respiratory rate and Blood pressure (CRB) points; Panel D: modified Sequential (Sepsis-related) Organ Failure Assessment (SOFA) points; Panel E: Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) points and Panel F: Pneumonia Severity Index (PSI) risk class.
Error bars denote 95% confidence intervals. The x-axis for mSOFA score was winsorized for values higher than 7 points for illustration.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure 3. **Distribution of SIRS, qSOFA, CRB, mSOFA, CURB-65 and PSI scores, place of treatment and mortality for community-acquired pneumonia patients.**

CAP – community-acquired pneumonia; CRB - Confusion, Respiratory rate and Blood pressure; CURB-65 – Confusion, Urea, Respiratory rate, Blood pressure and Age; ED – emergency department; HDU – high-dependency unit; ICU – intensive care unit; NWT – number willing to treat; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS – systemic inflammatory response syndrome; mSOFA – modified Sequential (Sepsis-related) Organ Failure Assessment; PSI – Pneumonia Severity Index.

* Patients were followed up until hospital discharge, and all survivors were re-examined or contacted by phone 30 days after hospital discharge.

Figure 4. **Decision curves showing the net benefit of SIRS criteria, qSOFA, CRB, mSOFA, CURB-65 and PSI scores in the treatment of community-acquired pneumonia patients at risk of in-hospital mortality, composite outcome and 30-day mortality. Panels A, C and E** show the net benefit (NB = benefit x true-positive classifications minus harm/cost x false-positive classifications) of each score and the strategy to treat-none and treat-all patients over the plausible range of number willing to treat (NWT) (i.e., threshold probabilities). **Panels B, D and F** show the net reduction in interventions in a theoretical population of 100 patients by using the scores to make clinical decisions.

Illustrative example: if a clinician weights the harm/cost of overtreatment versus the benefit of appropriated treatment at 1:19 for in-hospital mortality, we have a threshold probability of 5% and a NWT of 20. This choice specifies that death of a community-acquired pneumonia

patient who remained untreated is 19 times worse than the consequences of overtreatment of an unnecessarily-treated patient. At a NWT of 19, the net benefit of the SOFA, qSOFA and CRB scores outperform SIRS and, treat-all strategies. At the same time, at a NWT of 20, we could reduce the number of interventions without missing any in-hospital death by 7% using SIRS criteria, 16% using qSOFA, 27% using CRB or mSOFA scores, 30% for CURB-65 and 35% for PSI.

CRB – Confusion, Respiratory rate and Blood pressure; CURB-65 – Confusion, Urea, Respiratory rate, Blood pressure and Age; NWT – number willing to treat; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS – systemic inflammatory response syndrome; mSOFA – modified Sequential (Sepsis-related) Organ Failure Assessment; PSI – Pneumonia Severity Index.

Figure 5. **In-hospital mortality of 6,024 patients with community-acquired pneumonia, stratified by SIRS, qSOFA or CRB combined with mSOFA**. **Panel A** shows the in-hospital mortality stratified by SIRS criteria and **mSOFA** scores combinations. **Panel B** shows the in-hospital mortality stratified by qSOFA and **mSOFA** scores combinations. **Panel C** shows the in-hospital mortality stratified by CRB and **mSOFA** scores combinations. Error bars denote 95% confidence interval. The numbers and percentages between parentheses represent the frequency distribution of each combination in the whole cohort.

CRB - Confusion, Respiratory rate and Blood pressure; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS – systemic inflammatory response syndrome; mSOFA – modified Sequential (Sepsis-related) Organ Failure Assessment.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table 1 – Characteristics of community-acquired patients from two cohorts in Spain

	Overall (n=6,874)
Demographic variables, n(%) or mean (SD)	
Age, years	66.1 (19)
Age ≥65 years	4,170 (60.9%)
Gender Male	4,259 (62.2%)
Co-morbidities, n (%)	
Chronic respiratory disease	2,045 (30.3%)
Chronic heart failure	1,127 (16.5%)
Diabetes mellitus	1,299 (19.1%)
Chronic liver disease	371 (5.6%)
Chronic renal disease	499 (7.3%)
Neurologic disease	1,135 (17.0%)
Neoplasia	489 (7.3%)
Nursing home	361 (5.3%)
Vital signs upon presentation, mean (SD)	
Respiratory rate, per min	26 (8)
Heart rate, beats per min	98 (19)
Systolic blood pressure, mmHg	131 (27)
Diastolic blood pressure, mmHg	73 (14)
Mean arterial pressure, mmHg	92 (16)
Temperature °C	37.5 (1)
Laboratory findings upon presentation, median [IQR] or n (%)	
Leukocyte, cells/mm^3	12,400 [8,570-17,280]
C-reactive protein, mg/dl	16.7 [7.8-26.3]
Creatinine, mg/dl	1.0 [0.8-1.4]
Total bilirubin, mg/dl	0.50 [0.40-0.80]
Platelets, cells/mm^3	234 [187-302]
PaO_2/FiO_2 ratio, mmHg	281 [238-328]
Microbiology confirmed	2,860 (41.8%)
Scores upon diagnosis, mean (SD) / median [IQR]	
SIRS, points	2.3 (1) / 2 [2-3]
SIRS ≥ 2	4,908 (78.8%)
qSOFA, points	1.0 (0.7) / 1 [0-1]
qSOFA ≥ 2	1,260 (20.5%)
CRB, points	0.7 (0.8) / 1 [0-1]
CRB ≥ 2	902 (14.6%)
mSOFA, points	2.2 (2) / 2 [1-3]
mSOFA ≥ 2	4,288 (62.4%)
CURB-65, points	1.4 (1) / 1 [1-2]
CURB ≥ 2	2,592 (42.0%)
PSI, points	3.4 (1) / 4 [2-4]
PSI ≥ 4	3,379 (55.4%)
Outcomes, n (%) or median [IQR]	
In-hospital mortality	442 (6.4%)
In-hospital mortality or 3 days of ICU stay	716 (10.4%)
30-days mortality	477/6,377 (7.5%)
Hospital stay, days	7 [4-10]

Abbreviations: CRB: Confusion, Respiratory rate and Blood pressure; CURB-65: Confusion, Urea, Respiratory rate , Blood pressure and Age; ICU: intensive care unit; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; PSI: Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.

Table 2. Scores performance and clinical utility for in-hospital mortality in 6,024 patients with community-acquired pneumonia (complete-case analysis)

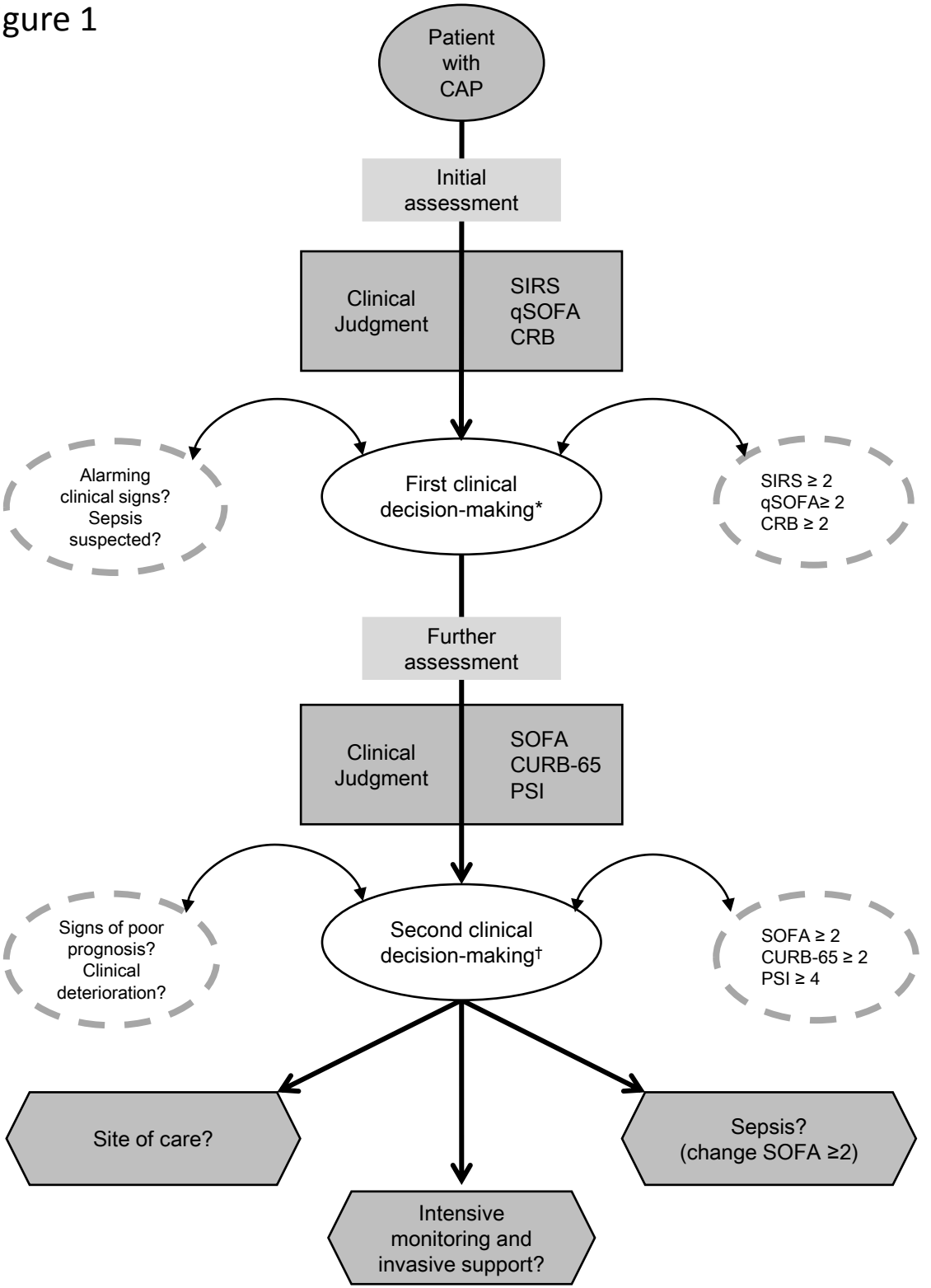
	Discrimination	Calibration		Overall performance		Clinical utility						
Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
SIRS	0.579 (0.551*0.605)	0.776	0.061	0.9%	1.3%	**SIRS**≥2	89 (85-92)	22 (20-23)	7 (6-8)	97 (96-98)	1.13 (1.09-1.18)	0.52 (0.39-0.69)
qSOFA	0.697 (0.671-0.722)	0.494	0.036	4.3%	8.8%	**qSOFA**≥2	50 (45-55)	81 (80-82)	15 (13-17)	96 (96-97)	2.70 (2.41-3.03)	0.61 (0.55-0.68)
CRB	0.716 (0.690-0.741)	0.997	0.224	4.3%	10.7%	**CRB**≥2	40 (35-45)	87 (86-88)	17 (14-20)	96 (95-96)	3.05 (2.65-3.51)	0.69 (0.64-0.75)
mSOFA	0.748 (0.721-0.774)	0.999	0.040	7.7%	14.3%	**mSOFA**≥2	88 (85-91)	37 (36-38)	9 (8-9)	98 (97-99)	1.41 (1.35-1.47)	0.31 (0.24-0.42)
CURB-65	0.746 (0.722-0.769)	0.801	0.008	5.9%	12.9%	**CURB-65**≥2	78 (74-82)	60 (59-61)	12 (10-13)	98 (97-98)	1.96 (1.84-2.09)	0.36 (0.30-0.44)
PSI	0.780 (0.760-0.799)	0.998	0.221	7.8%	17.4%	**PSI** ≥4	92 (89-95)	47 (46-48)	10 (09-11)	99 (98-99)	1.74 (1.68-1.81)	0.16 (0.12-0.23)

AUROC curve comparisons:
SIRS vs: qSOFA (p<0.001), CRB (p<0.001), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001)
qSOFA vs: CRB (p=0.055), mSOFA (p=0.001), CURB-65 (p<0.001), PSI (p<0.001)
CRB vs: mSOFA (p=0.035), CURB-65 (p<0.001), PSI (p<0.001)
mSOFA vs: CURB-65 (p=0.924), PSI (p=0.031)
CURB-65 vs: PSI (p=0.002)

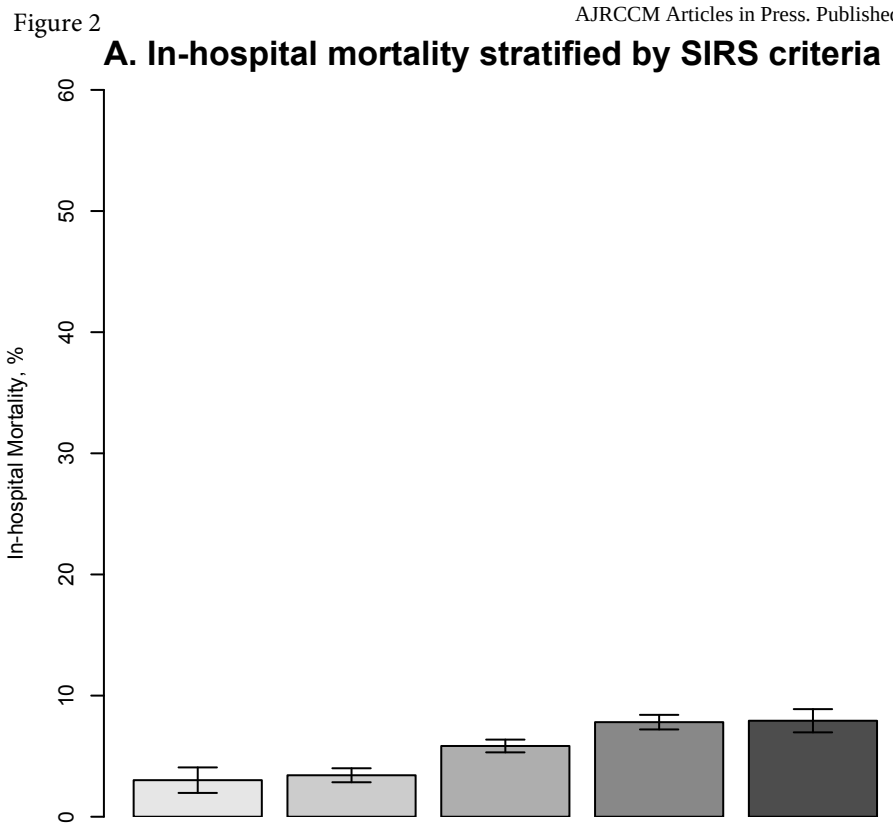
AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65: Confusion, Urea, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; NPV: negative predictive value; PSI: Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; PPV: positive predictive value; SIRS: systemic inflammatory response syndrome;. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

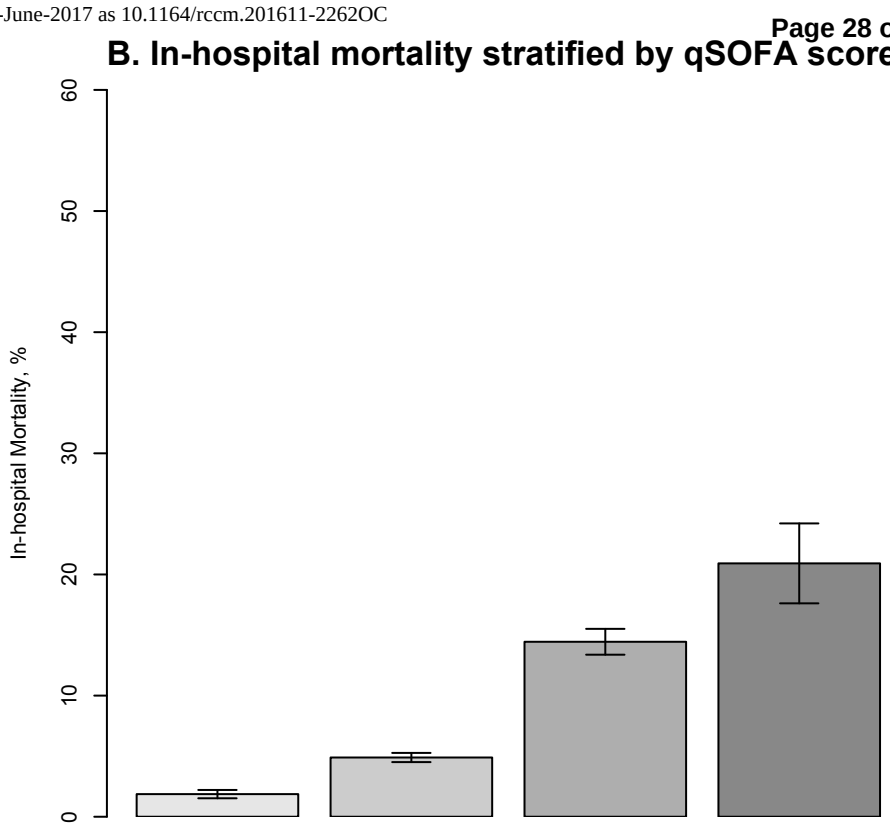
AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

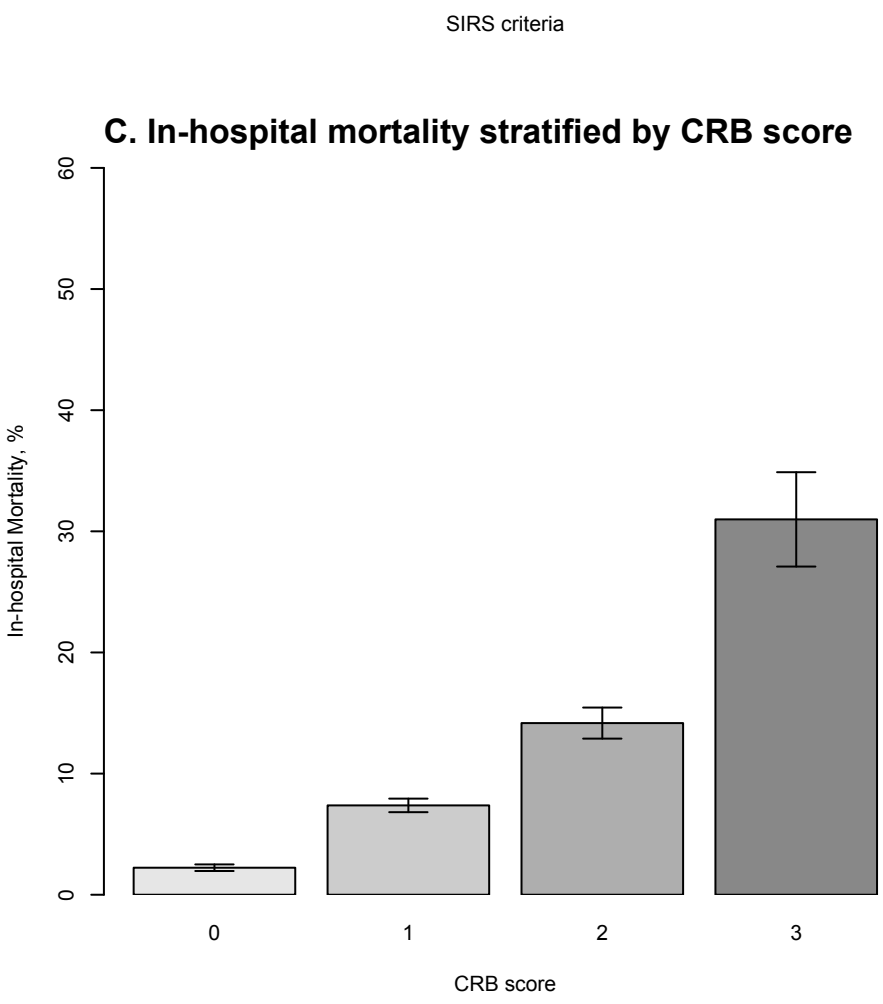
Figure 1

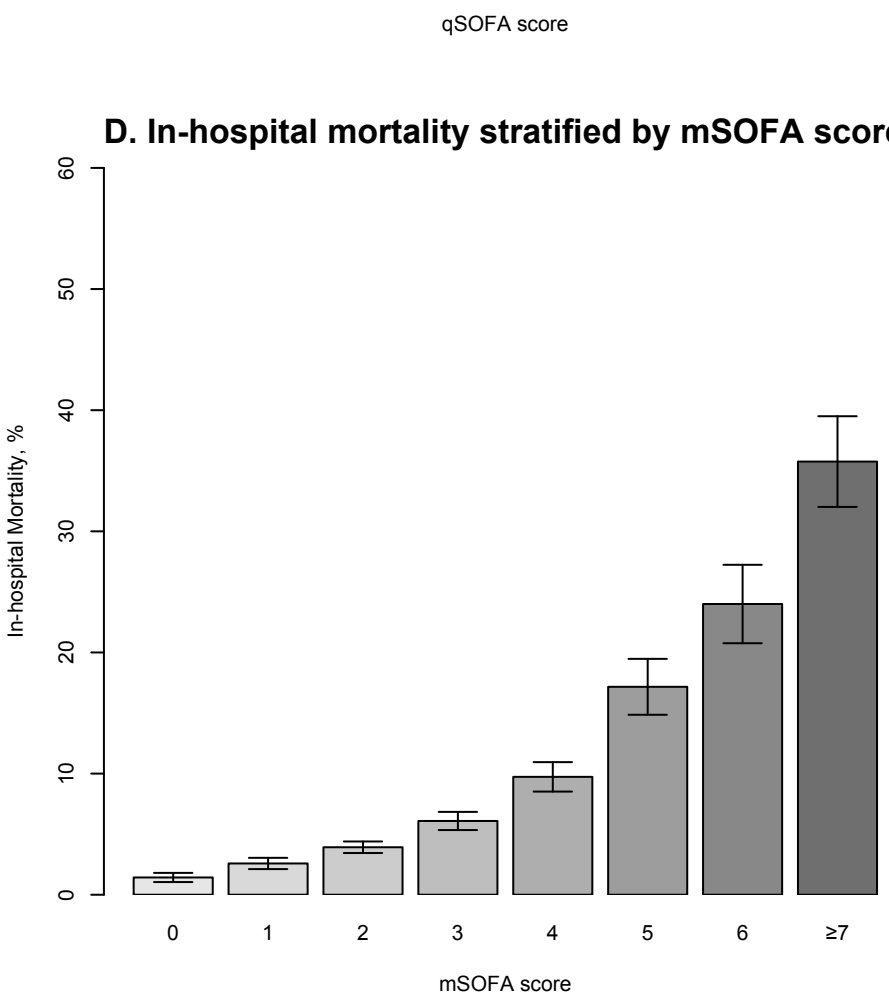


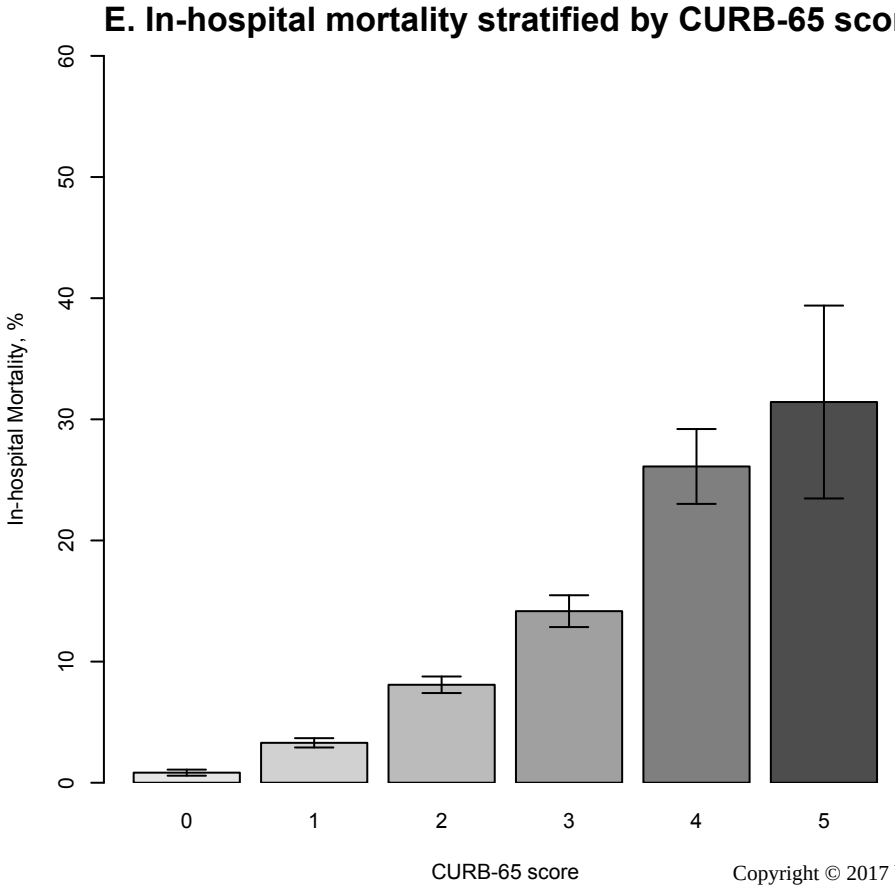
AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

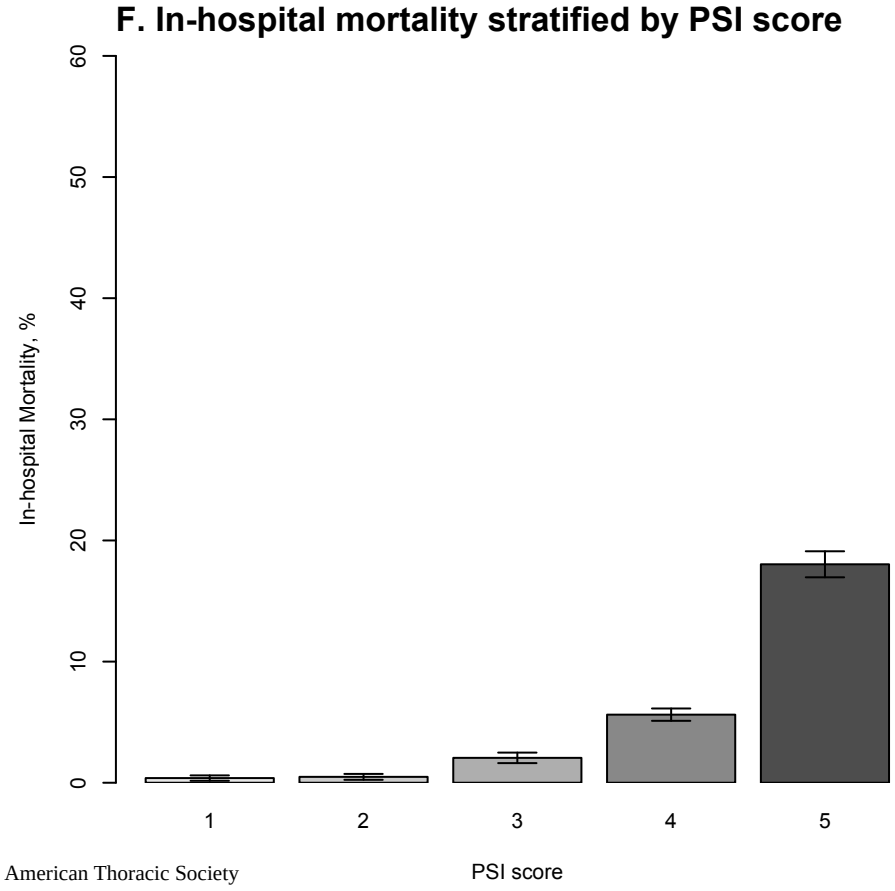












AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure 3

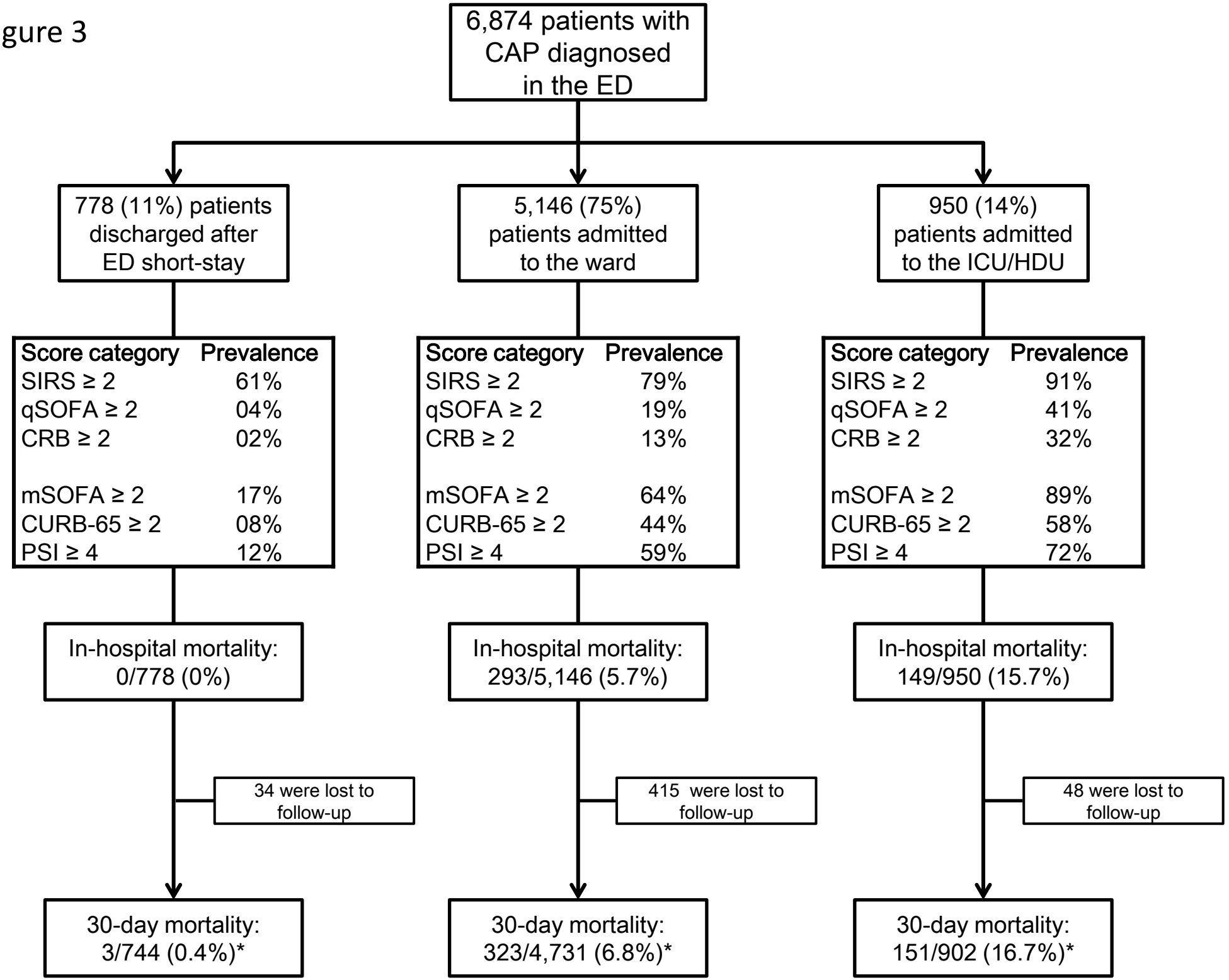
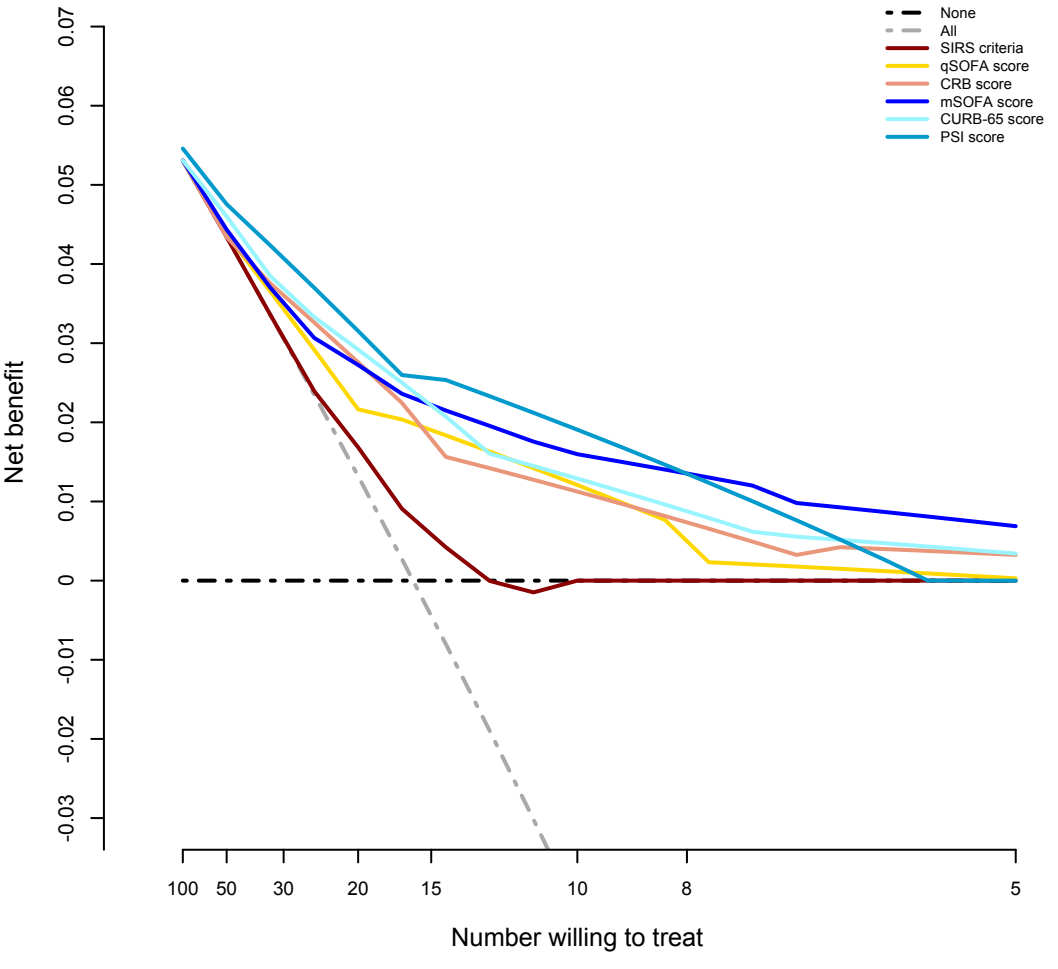
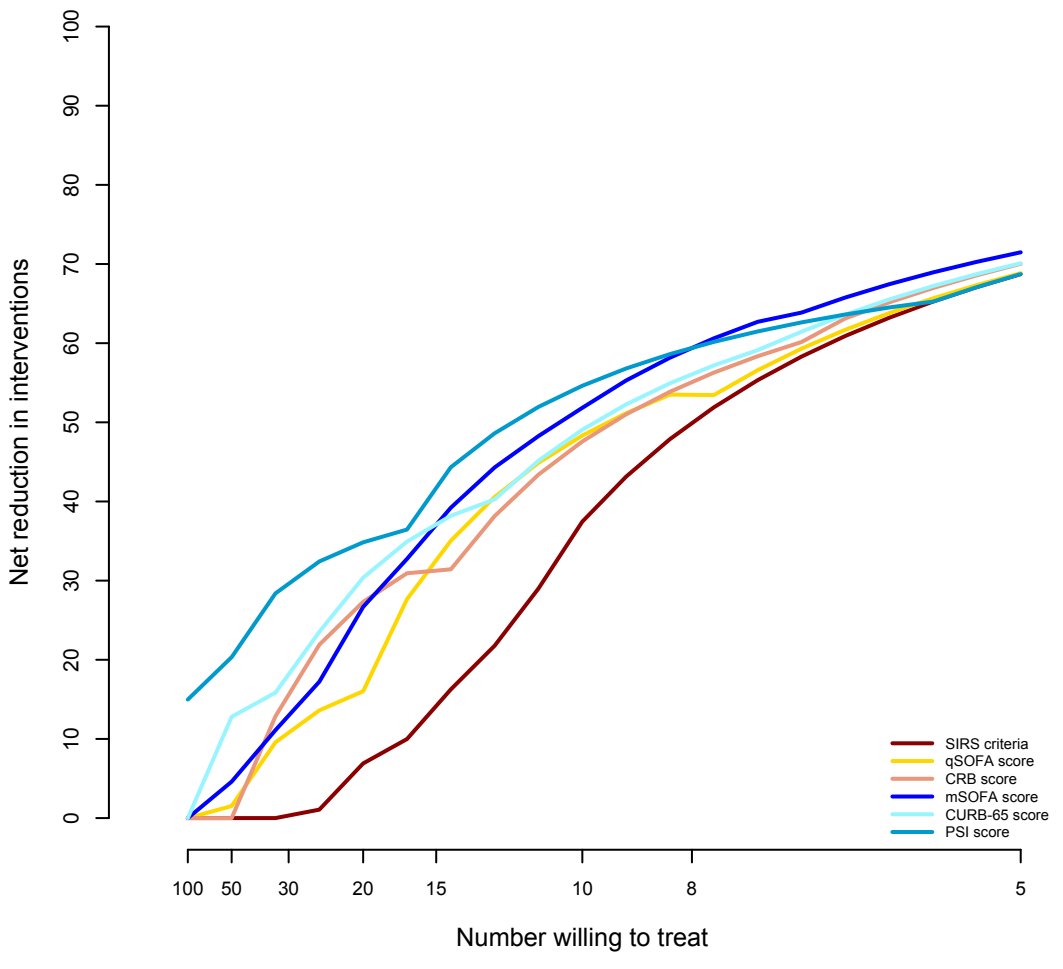


Figure 4

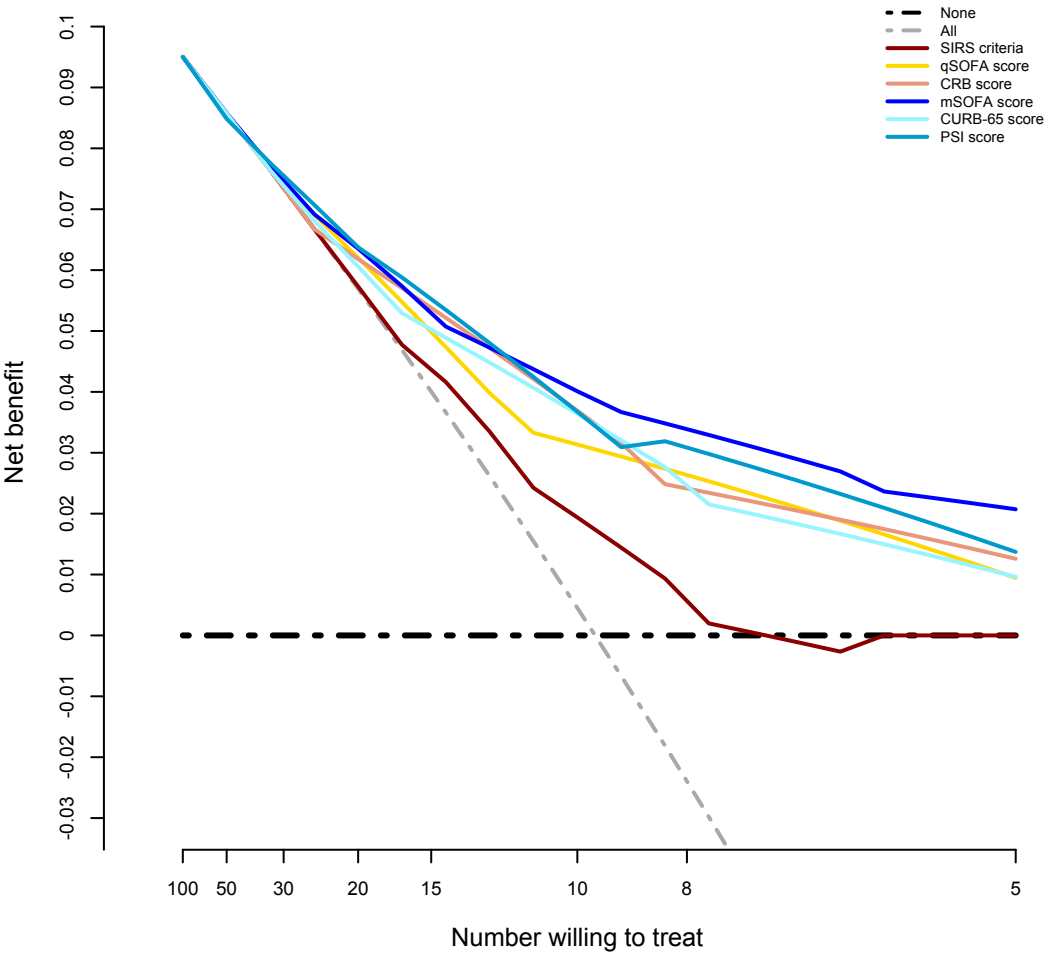
A. Decision-curve analysis for in-hospital mortality



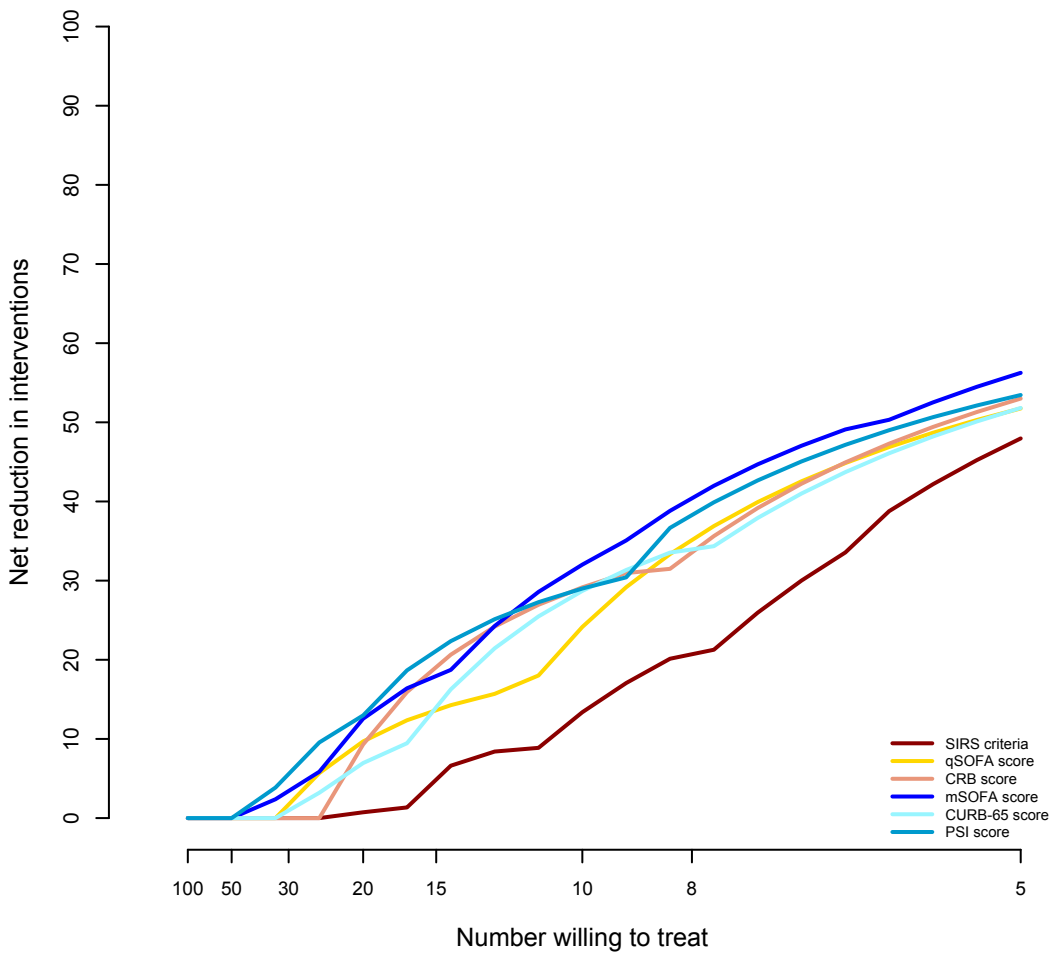
B. Interventions avoided for in-hospital mortality



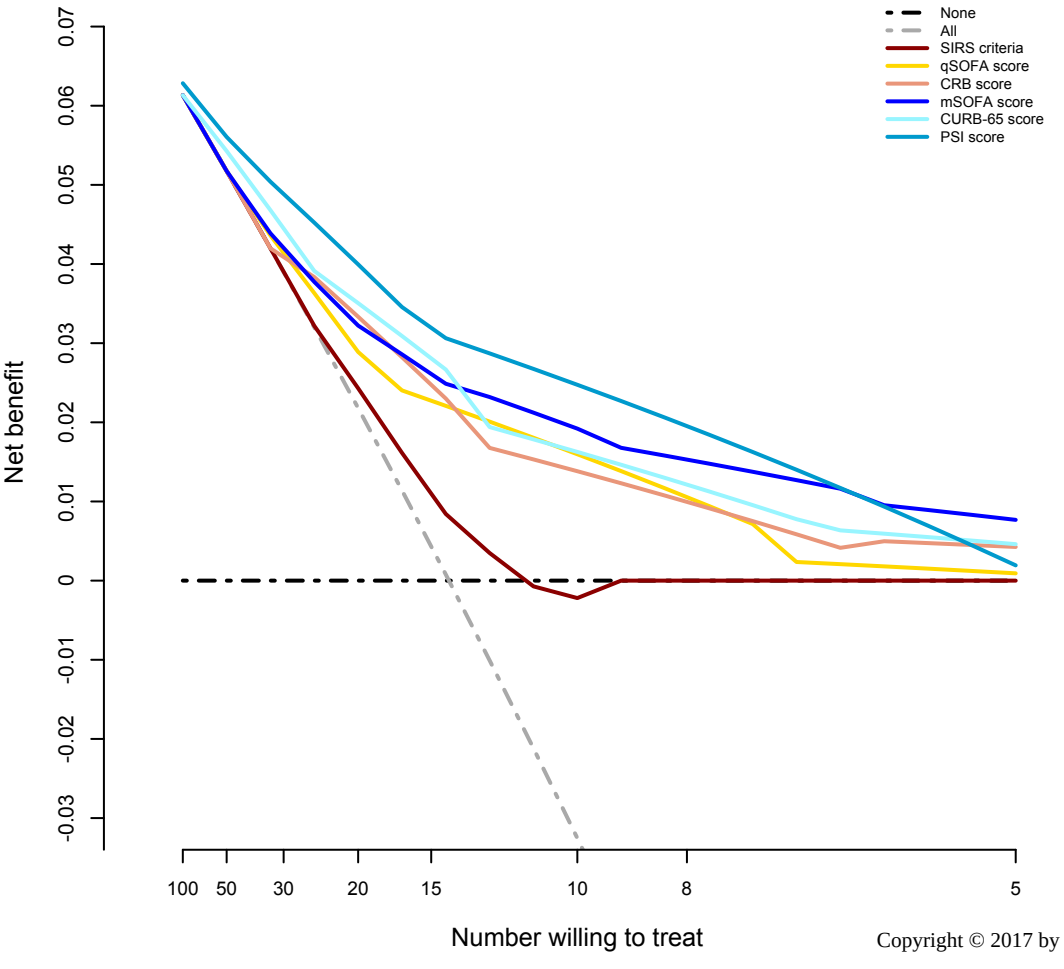
C. Decision-curve analysis for in-hospital mortality/3 days of ICU stay



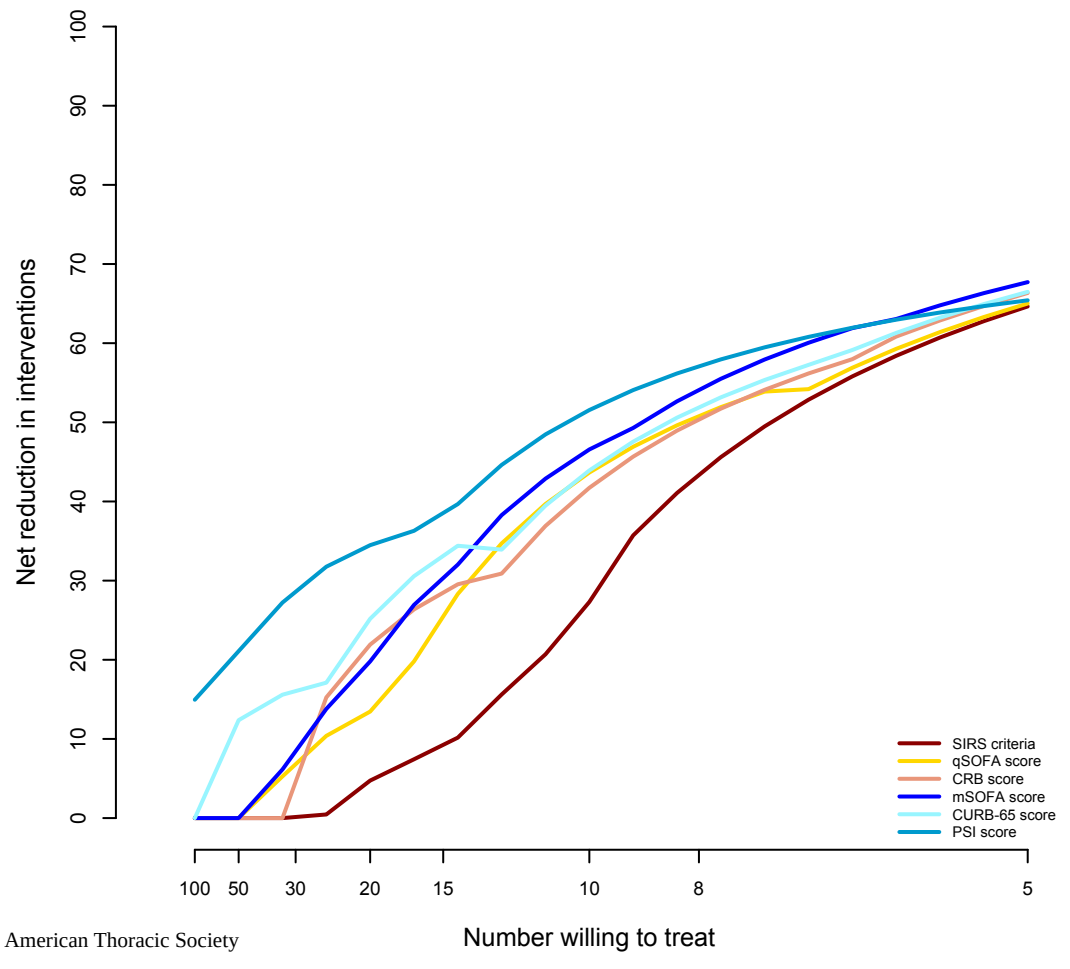
D. Interventions avoided for in-hospital mortality/3 days of ICU stay



E. Decision-curve analysis for 30-day mortality

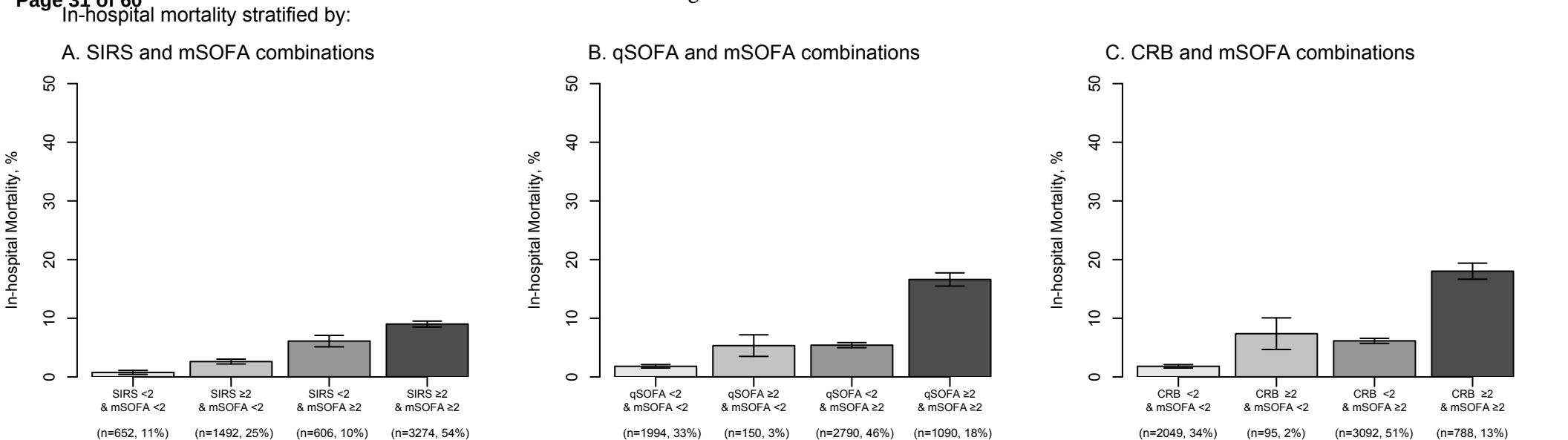


F. Interventions avoided for 30-day mortality



AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure 5



Online Supplement Material

Ranzani OT, Prina E, Menéndez R, et al. New Sepsis Definition (Sepsis-3) and Community-Acquired Pneumonia Mortality: a validation and clinical decision-making study

Table of Contents

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table E1. SIRS criteria, qSOFA, qSOFA-65, CRB, CRB-65, mSOFA, CURB-65 and PSI definitions

Table E1 - A: SIRS criteria definition

	SIRS(1)	Points
Temperature	>38°C or <36°C	1 point
Heart rate	>90/min	1 point
Respiratory rate or $PaCO_2$	>20/min or $PaCO_2$ <32 mm Hg	1 point
White blood cell count	>12.000/mm^3 or <4.000/mm^3 or 10% immature bands	1 point

Table E1 - B: qSOFA, qSOFA-65, CRB and CRB-65 scores definition

	qSOFA(2)	qSOFA-65(3)	CRB(4, 5)	CRB-65(5, 6)	Points
Respiratory rate	≥22/min	≥22/min	≥30/min	≥30/min	1 point
Altered mentation	Presence	Presence	Presence	Presence	1 point
Hemodynamic status	Systolic blood pressure ≤100 mm Hg	Systolic blood pressure ≤100 mm Hg	Systolic blood pressure <90 mm Hg or Diastolic blood pressure ≤ 60 mm Hg	Systolic blood pressure <90 mm Hg or Diastolic blood pressure ≤ 60 mm Hg	1 point
Age	-	≥ 65	-	≥ 65	1 point

Table E1 - C: modified SOFA score**(2, 7-9)**

	Original Definition	Barcelona cohort definition	Valencia cohort definition
Respiratory			
Points	**PaO_2/FiO_2 (mmHg)**	**PaO_2/FiO_2 (mmHg)**	**PaO_2/FiO_2 (mmHg)**
0	≥400	≥400	≥400
1	300-399	300-399	300-399
2	200-299	200-299	200-299
3	100-199 and Respiratory support	100-199 and Respiratory support	100-199 and Respiratory support
4	<100 and Respiratory Support	<100 and Respiratory Support	<100 and Respiratory Support
Hematologic			
Points	**Platelets count ($x10^3/mm^3$)**	**Platelets count ($x10^3/mm^3$)**	**Platelets count ($x10^3/mm^3$)**
0	≥150	≥150	≥150
1	100-149	100-149	100-149
2	50-99	50-99	50-99
3	20-49	20-49	20-49
4	<20	<20	<20
Hepatic			
Points	**Total serum Bilirubin mg/dl**	**Total serum Bilirubin mg/dl**	**Total serum Bilirubin mg/dl**
0	<1.2	<1.2	<1.2
1	1.2-1.9	1.2-1.9	1.2-1.9
2	2.0-5.9	2.0-5.9	2.0-5.9
3	6.0-11.9	6.0-11.9	6.0-11.9
4	>12.0	>12.0	>12.0
Cardiovascular			
Points	**MAP or vasoactive drugs**	**MAP or shock status**	**MAP or shock status**
0	≥ 70 mmHg	≥ 70 mmHg	≥ 70 mmHg
1	< 70 mmHg	< 70 mmHg	< 70 mmHg
2	Dopamine <5 or Dobutamine (any dose)	Refractory hypotension attributed to infection after adequate fluid resuscitation and use of any vasopressor	Refractory hypotension attributed to infection after adequate fluid resuscitation and use of any vasopressor
3	Dopamine 5.1 -15 or Epinephrine ≤0.1 or Norepinephrine ≤0.1	-	-
4	Dopamine >15 or Epinephrine >0.1 or Norepinephrine >0.1	-	-
Neurologic			
Points	**Glasgow Coma Scale**	**Mental status**	**Mental status**
0	15	Normal	Normal
1	13-14	Acute altered mental status (confusion, decreased consciousness)	Acute altered mental status (confusion, decreased consciousness)
2	10-12	-	-
3	6-9	-	-
4	<6	-	-
Renal			
Points	**Creatinine (mg/dL) or Urinary Output (mL)**	**Creatinine (mg/dL)**	**Creatinine (mg/dL) or Urinary Output (mL)**
0	sCr <1.2	sCr <1.2	sCr <1.2
1	sCr 1.2-1.9	sCr 1.2-1.9	sCr 1.2-1.9
2	sCr 2.0-3.4	sCr 2.0-3.4	sCr 2.0-3.4
3	sCr 3.5-4.9 or UO<500 ml/24h	sCr 3.5-4.9	sCr 3.5-4.9 or UO <0.5 ml/Kg/h for 2h
4	sCr ≥5.0 or UO≤ 200ml/24h	sCr ≥5.0	sCr ≥5.0

sCR – serum creatinine; UO – urinary output

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table E1 - D: CURB-65 score definition

	CURB-65(4)	**Points**
Respiratory rate	≥30/min	1 point
Altered mentation	Presence	1 point
Hemodynamic status	Systolic blood pressure <90 mm Hg or Diastolic blood pressure ≤ 60 mm Hg	1 point
Age	≥65 years	1 point
Urea	>7 mmol/l	1 point

Table E1 - E: PSI score definition(10)

		Points
Age > 50 years	No	No for all: PSI risk class I
Neoplastic disease	No	
Liver disease	No	
Congestive heart failure	No	
Cerebrovascular disease	No	
Renal disease	No	
Altered mental status	No	
Respiratory rate ≥ 30 breaths/minute	No	
Systolic blood pressure < 90 mm Hg	No	
Temperature <30º or ≥ 40ºC	No	
Pulse ≥ 125 beats/minute	No	
If yes for some condition above:		
Age	Male	Age (yr)
	Female	Age - 10
Nursing home resident	Yes	+10
Neoplastic disease	Yes	+30
Liver disease	Yes	+20
Congestive heart failure	Yes	+10
Cerebrovascular disease	Yes	+10
Renal disease	Yes	+10
Altered mental status	Yes	+20
Respiratory rate ≥ 30 breaths/minute	Yes	+20
Systolic blood pressure < 90 mm Hg	Yes	+20
Temperature <30º or ≥ 40ºC	Yes	+15
Pulse ≥ 125 beats/minute	Yes	+10
Arterial pH < 7.35	Yes	+30
BUN ≥ 11 mmol/L	Yes	+20
Sodium <130 mmol/L	Yes	+20
Glucose ≥ 250 mg/dL (14 mmol/L)	Yes	+10
Hematocrit < 30%	Yes	+10
PaO2 < 60 mm Hg or oxygen saturation <90% on pulse oximetry	Yes	+10
Pleural effusion	Yes	+10

Risk class II: ≤70 points,
Risk class III: 71-90 points;
Risk class IV: 91-130 points;
Risk class V: >130 points

Additional description of statistical methods

We performed all analysis in a merged dataset as a whole cohort, because the cohorts were from two research groups with intense collaboration, having similar standardized definitions and methods of data collection. Before merging, we re-checked definitions, implausible values and standardized each variable.

Scores performance: Predictive performance measures had advantages and disadvantages(11-15); to overcome this problem we explored different and complementary methods. Calibration was mainly evaluated by Calibration plots. Calibration curves were built with a smoothed nonparametric method(13, 15). We also evaluated two complementary goodness-of-fit statistics to assess calibration (*Hosmer-Lemeshow* and the *le Cessie-van Houwelingen-Copas-Hosmer* tests*)*(15). We used two goodness-of-fit statistics because the Hosmer-Lemeshow test divides the population in 10 groups, which is an arbitrary cut-off, as well it is sensible to the sample size (14, 16). We used the area under the receiver-operating characteristic curve (AUROC) to assess discrimination. The 95% confidence interval (CI) estimation for the AUROCs and theirs comparisons were performed using bootstrapping method in 10,000 samples (13, 17). We assessed overall fit using two measures: 1) scaled Brier score (mean squared error between the outcome and prediction; Brier score was scaled on its maximum value for an interpretation similar to the r-square), and 2) Nagelkerke R-square (13, 14). For better comparison of the additional prediction contribution of each score to the baseline risk model, we used the integrated discrimination improvement (IDI) index (13, 14). Net reclassification measures were not used because there is not an established cut-off risk for mortality in septic patients.

Baseline risk model: after multiple imputation of 10 datasets, we fit a multivariate logistic model. In-hospital mortality was the dependent variable and age, gender, chronic respiratory disease, chronic neurologic disease, liver disease, heart failure, diabetes mellitus, neoplasia, chronic renal disease, and etiologic diagnosis. Age was modeled through a restricted cubic spline.

Multiple imputation: As we expected few missing values for SIRS, qSOFA and CRB, our main analysis was conducted in the complete-case data (i.e., patients for whom we had all data for the SIRS, qSOFA and, CRB). We investigated the missing data patterns for covariates and scores, and we assumed missing at random (MAR) conditioned in the covariates (18). We used multiple imputation (aregImpute routine)(14) to generate 10 datasets to evaluate the prediction performance for the primary outcome (in-hospital mortality). The model for multiple imputation included all covariates of the baseline risk model, and the SIRS, qSOFA, CRB, SOFA, CURB-65 and PSI variables. As recommended, the model also included the outcome (in-hospital mortality), and the cohort indicator ("Barcelona"/ "Valencia") as an auxiliary variable. For simplicity, we filled in missing values with the first set of imputed values from the multiple imputation for the predictive performance evaluation (14, 19).

The original scores and those constructed after multiple imputation followed a normal distribution for SIRS and non-normal for qSOFA, CRB, CURB-65 and PSI.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Decision-curve analysis

A well-calibrated score is necessary but not enough for clinical usefulness, as well good discrimination (20, 21). To support clinical decision making, the above statistical measures may not reflect clinical utility (Can I make better decisions by using a score than without it?).(22) The main difficult to translate discrimination and calibration to clinical practice is because they equally weight sensitivity, specificity and prediction errors (e.g., true-positive and false-positive rates).(20) Nevertheless, in clinical practice, clinicians usually apply different weights during the decision-making process; furthermore, the weights depend on the patient characteristics and resources availability (20). Decision-curve analysis was recently introduced in clinical research and depicts the predicted net benefit ("NB = benefit x true-positive classifications minus harm/cost x false-positive classifications") of a prediction tool over a range of threshold probabilities. Threshold probabilities quantify how overtreatment is considered against treatment benefits (14, 20, 22-25). For instance, whether a clinician weights the harm/cost of overtreatment versus the benefit of appropriated treatment (e.g., early antibiotic, aggressive resuscitation and intensive monitoring) at 1:19, we have a threshold probability of 5% and a number willing to treat (NWT) of 20.(23, 26) Thus, the 5% threshold probability specifies that death of a septic patient who remained untreated is 19 times worse than the consequences of overtreatment of an unnecessarily treated patient.(26) There is not an established cut-off risk for treating sepsis; and decision-curves have the advantage to plot a plausible range of threshold probabilities. We defined 100 to 5 NWTs as plausible range (i.e., threshold probabilities from 0 to 20%), because it is unlikely clinicians will use a score to make decisions about treatment of septic patients for higher threshold probabilities.

It is worth noting that decision-curve analysis cannot be used to choose a threshold probability (20, 27). Second, prediction tools with higher NB perform better than other at that risk threshold, independently by how much it is higher (20, 27). Third, decision-curve interpretation also varies according to the outcome prevalence (27). As expected, for common outcomes, the NB of treat-all strategy is higher than when the outcome occurrence is low. The NB is higher as the risk model identifies high-risk patients without classifying patients without the event as high-risk patients (27).

All statistical analyses were performed using R software, version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria)(28), with the adds-on packages: *rms*, *Hmisc*, *epiR*, *foreign*, *pROC*, *sciplot* and the function *dca* (available at https://www.mskcc.org/departments/epidemiology-biostatistics/health-outcomes/decision-curve-analysis-01).

Table E2. SIRS, qSOFA, CRB, mSOFA, CURB-65 and PSI distribution among cohorts

	Overall	Barcelona	Missing	Valencia	Missing values
SIRS			626 (9.9%)		18 (3.2%)
0	284 (4.6%)	235 (4.1%)		49 (8.9%)	
1	1038 (16.7%)	903 (15.9%)		135 (24.5%)	
2	2058 (33.0%)	1859 (32.7%)		199 (36.1%)	
3	2035 (32.7%)	1903 (33.5%)		132 (29.3%)	
4	815 (13.1%)	778 (13.7%)		37 (6.7%)	
qSOFA			696 (11.0%)		17 (3.0%)
0	1590 (25.8%)	1251 (22.3%)		339 (61.3%)	
1	3311 (53.7%)	3129 (55.8%)		182 (32.9%)	
2	1102 (17.9%)	1076 (19.2%)		26 (4.7%)	
3	158 (2.6%)	152 (2.7%)		6 (1.1%)	
CRB			696 (11.0%)		17 (3.0%)
0	3029 (49.2%)	2631 (46.9%)		398 (72.0%)	
1	2230 (36.2%)	2105 (37.5%)		125 (22.6%)	
2	756 (12.2%)	727 (13.0%)		29 (5.2%)	
3	146 (2.4%)	145 (2.6%)		1 (0.2%)	
mSOFA*			-		-
0	1269 (18.5%)	1142 (18.1%)		127 (22.3%)	
1	1317 (19.2%)	1220 (19.4%)		97 (17.0%)	
2	1847 (26.9%)	1691 (26.8%)		156 (27.4%)	
3	1116 (16.2%)	1028 (16.3%)		88 (15.4%)	
4	657 (9.5%)	622 (9.9%)		35 (6.1%)	
5	293 (4.3%)	272 (4.3%)		21 (3.7%)	
6	190 (2.8%)	165 (2.6%)		25 (4.4%)	
>=7	185 (2.7%)	164 (2.6%)		21 (3.7%)	
CURB-65			707 (11.2%)		17 (3.0%)
0	1,351 (22.0%)	1,217 (21.7%)		134 (24.2%)	
1	2,211 (36.0%)	1,973 (35.3%)		238 (43.0%)	
2	1,625 (26.4%)	1,500 (26.8%)		125 (22.6%)	
3	720 (11.7%)	677 (12.1%)		43 (7.8%)	
4	208 (3.4%)	195 (3.5%)		13 (2.4%)	
5	35 (0.6%)	35 (0.6%)		-	
PSI			756 (12.0%)		17 (3.0%)
Class Risk I	807 (13.2%)	733 (13.2%)		74 (13.4%)	
Class Risk II	831 (13.6%)	724 (13.1%)		107 (19.4%)	
Class Risk III	1,084 (17.8%)	934 (16.8%)		150 (27.1%)	
Class Risk IV	2,065 (33.9%)	1,893 (34.1%)		172 (31.1%)	
Class Risk V	1,314 (21.5%)	1,264 (22.8%)		50 (9.0%)	

* Normal range (0 points) was imputed for missing values in mSOFA

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table E3. mSOFA score missing pattern and mSOFA distribution

Table E3 - A: mSOFA score missing pattern

Missing systems for mSOFA	General
No missing	1978 (28.8%)
1 system missing	2829 (41.2%)
2 system missing	1708 (24.9%)
3 system missing	260 (3.8%)
4 system missing	54 (0.8%)
5 system missing	20 (0.3%)
6 system missing	25 (0.4%)

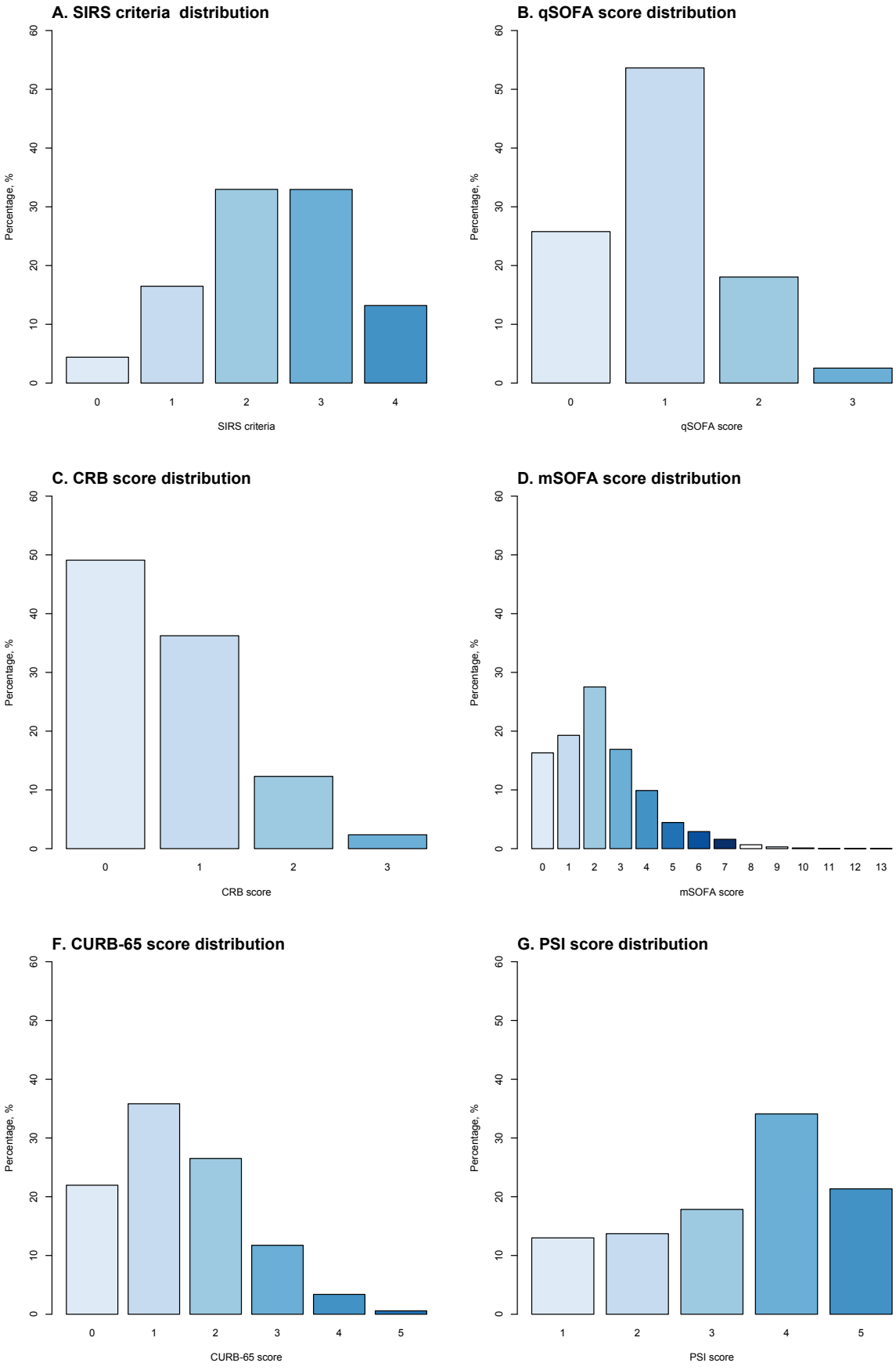
We adapted the components for Cardiovascular, Neurological and Renal components.

Table E3 - B: mSOFA distribution originally and after imputation

mSOFA score	Overall	Barcelona	Valencia
6 systems available (n=1,978)			
Mean (SD)	2.8 (2)	2.8 (2)	3.2 (2)
Median [IQR]	2 [2-4]	2 [2-4]	2 [2-4]
Simple imputation*			
Mean (SD)	2.2 (2)	2.2 (2)	2.2 (2)
Median [IQR]	2 [1-3]	2 [1-3]	2 [1-3]
Multiple imputation			
Mean (SD)	2.7 (2)	2.7	2.8
Median [IQR]	2 [2-4]	2 [2-4]	2 [2-3]

* Missing values imputed as zero(7)

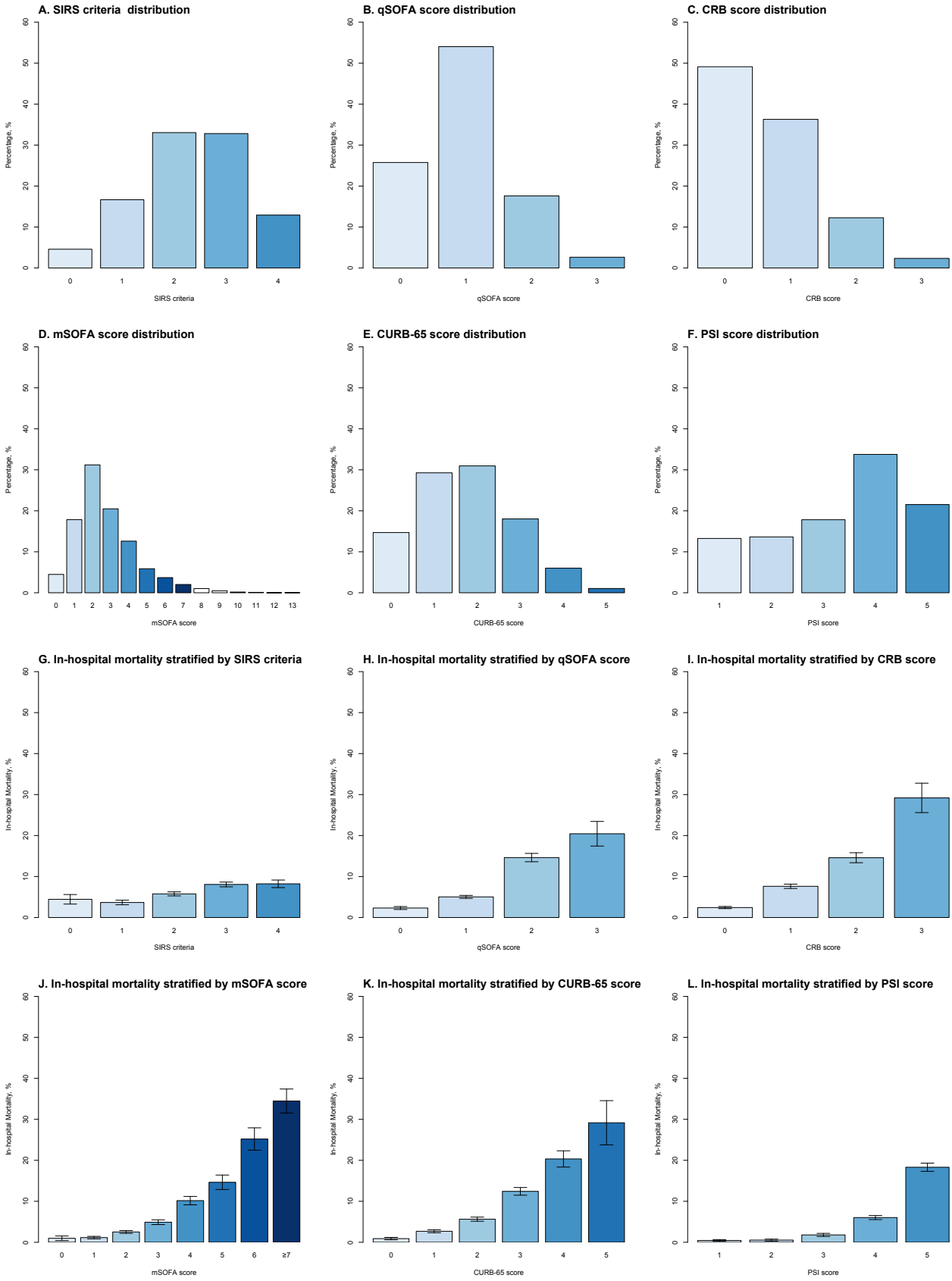
Figure E1. Distribution of scores in 6,024 patients with community-acquired pneumonia.



Panel A: systemic inflammatory response syndrome (SIRS) criteria; Panel B: quick Sequential (Sepsis-related); Panel C: Confusion, Respiratory rate and Blood pressure (CRB) points; Panel D: modified Sequential (Sepsis-related) Organ Failure Assessment (mSOFA) points; Panel E: Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) points and Panel F: Pneumonia Severity Index (PSI) risk class.
Error bars denote 95% confidence intervals. The x-axis for mSOFA score was winsorized for values higher than 7 points for illustration.

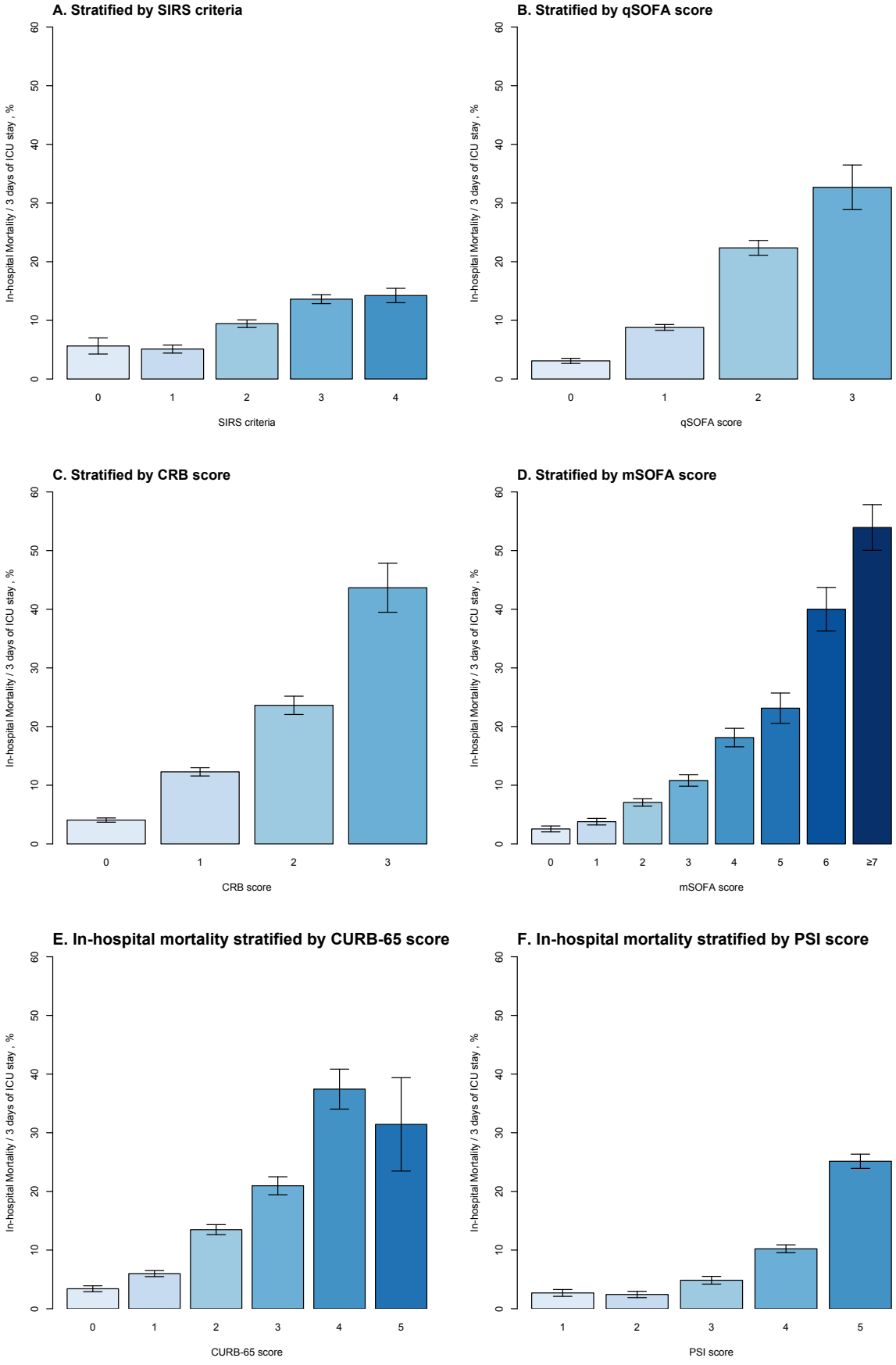
AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure E2. Distribution of scores and in-hospital mortality stratified by score points in 6,874 patients with community-acquired pneumonia (imputed data)



Panels A, B, C, D, E and F: scores distribution for systemic inflammatory response syndrome (SIRS) criteria; quick Sequential (Sepsis-related); Respiratory rate and Blood pressure (CRB), modified Sequential (Sepsis-related) Organ Failure Assessment (mSOFA); Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) and Pneumonia Severity Index (PSI). Panels G, H, I, J, K and L: in-hospital mortality stratified by systemic inflammatory response syndrome (SIRS) criteria; quick Sequential (Sepsis-related); Respiratory rate and Blood pressure (CRB), modified Sequential (Sepsis-related) Organ Failure Assessment (mSOFA); Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) and Pneumonia Severity Index (PSI) Error bars denote 95% confidence intervals. The x-axis for mSOFA score was winsorized for values higher than 7 points for illustration.

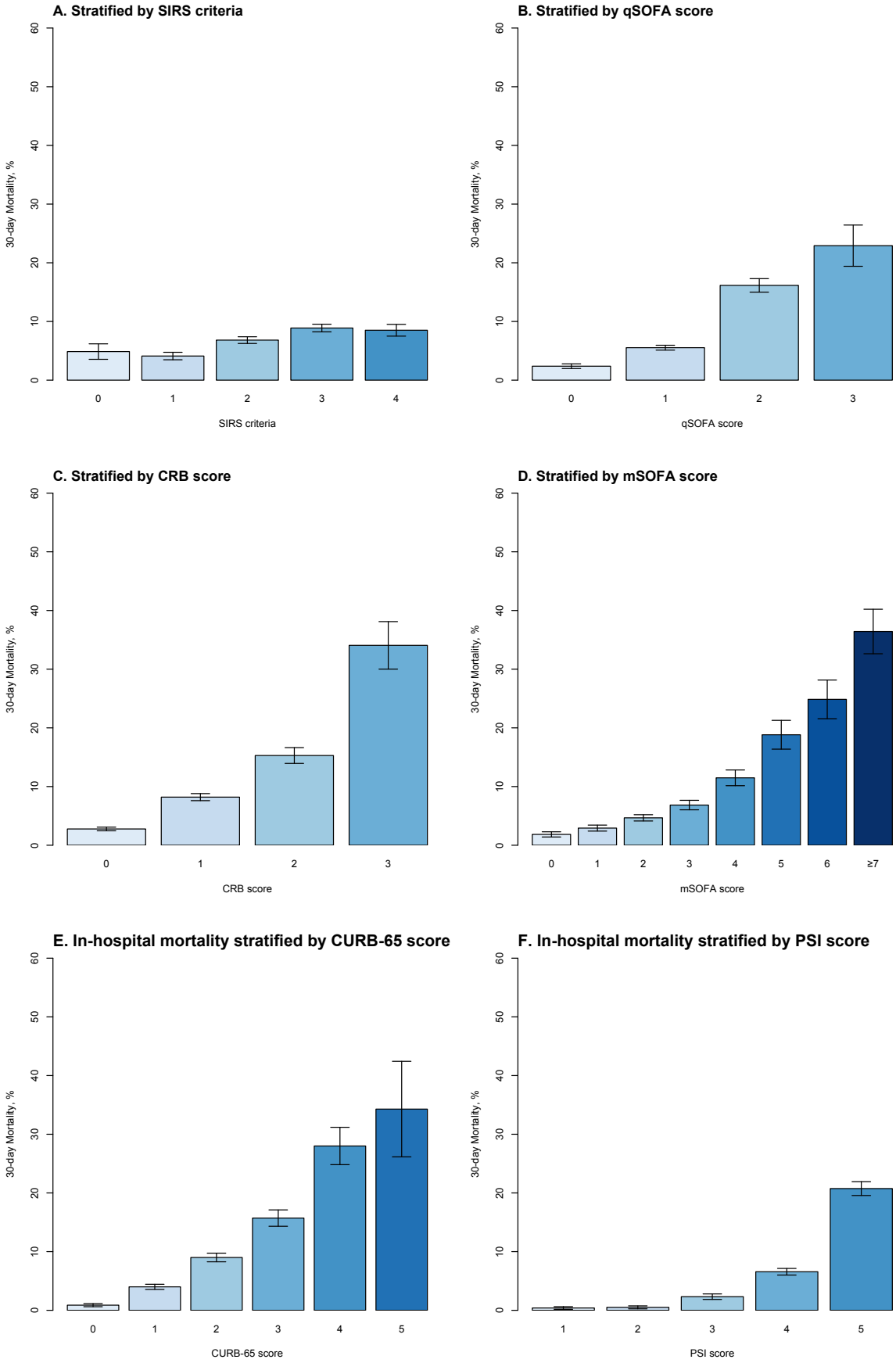
Figure E3. Scores and in-hospital mortality / 3 days of ICU stay (complete-case analysis)



Panel A: systemic inflammatory response syndrome (SIRS) criteria; Panel B: quick Sequential (Sepsis-related); Panel C: Confusion, Respiratory rate and Blood pressure (CRB) points; Panel D: modified Sequential (Sepsis-related) Organ Failure Assessment (mSOFA) points; Panel E: Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) points and Panel F: Pneumonia Severity Index (PSI) risk class.
Error bars denote 95% confidence intervals. The x-axis for mSOFA score was winsorized for values higher than 7 points for illustration.

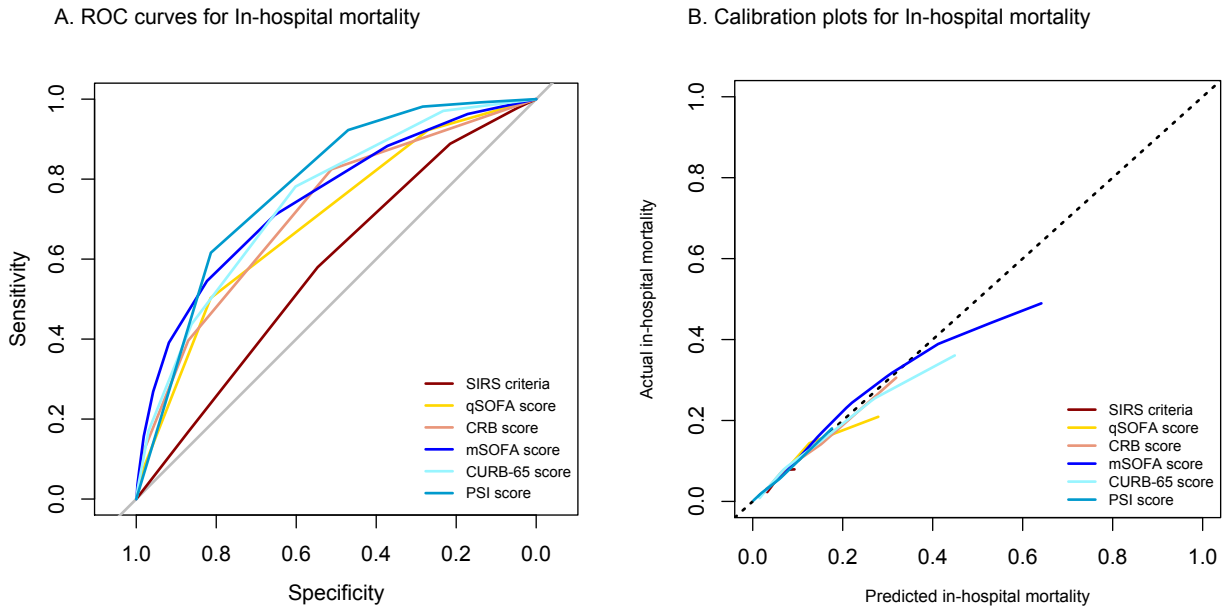
AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure E4. Scores and 30-day mortality (complete-case analysis)



Panel A: systemic inflammatory response syndrome (SIRS) criteria; Panel B: quick Sequential (Sepsis-related); Panels C: Confusion, Respiratory rate and Blood pressure (CRB) points, Panel D: modified Sequential (Sepsis-related) Organ Failure Assessment (mSOFA) points; Panel E: Confusion, Urea, Respiratory rate, Blood pressure and Age (CURB-65) points and Panel F: Pneumonia Severity Index (PSI). Error bars denote 95% confidence intervals. The x-axis for mSOFA score was winsorized for values higher than 7 points for illustration.

Figure E5. Discrimination and Calibration of SIRS, qSOFA, CRB, mSOFA, CURB-65 and PSI scores and their additional contribution to a baseline risk model for in-hospital mortality



Panel A shows the scores discrimination evaluated through the area under the receiver operating characteristic curve (AUROC) for in-hospital mortality in 6,024 patients with community-acquired pneumonia. **Panel B** shows the scores calibration for in-hospital mortality in 6,024 patients with community-acquired pneumonia. Calibration curves assessed by nonparametric smoothed methods.

CRB - Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; mSOFA - Sequential (Sepsis-related) Organ Failure Assessment; PSI – Pneumonia Severity Index; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table E4. Subgroups analysis regarding discrimination for in-hospital mortality

	Age		Chronic comorbidity	
	Age <65	Age ≥65	Without chronic comorbidities	At least 1 chronic c comorbidity
	AUROC (95% CI)	**AUROC (95% CI)**	**AUROC (95% CI)**	**AUROC (95% CI)**
SIRS	0.591 (0.524-0.656)	0.573 (0.543-0.603)	0.594 (0.517-0.668)	0.578 (0.547-0.608)
qSOFA	0.733 (0.674-0.791)	0.669 (0.640-0.696)	0.708 (0.642-0.772)	0.676 (0.647-0.706)
CRB	0.767 (0.709-0.821)	0.685 (0.655-0.713)	0.731 (0.662-0.793)	0.699 (0.669-0.729)
mSOFA	0.874 (0.829-0.913)	0.687 (0.653-0.720)	0.786 (0.723-0.843)	0.733 (0.701-0.763)
CURB-65	0.772 (0.714-0.826)	0.687 (0.658-0.717)	0.782 (0.727-0.835)	0.713 (0.683-0.741)
PSI	0.848 (0.801-0.894)	0.715 (0.690-0.740)	0.839 (0.784-0.886)	0.730 (0.704-0.754)
ROC curve comparisons:	SIRS vs. qSOFA (p<0.001), SIRS vs. CRB (p<0.001), SIRS vs. mSOFA (p<0.001), SIRS vs. CURB-65 (p<0.001), SIRS vs. PSI (p<0.001), qSOFA vs. CRB (p=0.108), qSOFA vs. mSOFA (p<0.001), qSOFA vs. CURB-65 (p=0.057), qSOFA vs. PSI (p<0.001), CRB vs. mSOFA (p=0.001), CRB vs. CURB-65 (p=0.545), CRB vs. PSI (p=0.005), mSOFA vs. CURB-65 (p=0.001), mSOFA vs. PSI (p=0.323), CURB-65 vs. PSI (p=0.007)	SIRS vs. qSOFA (p<0.001), SIRS vs. CRB (p<0.001), SIRS vs. mSOFA (p<0.001), SIRS vs. CURB-65 (p<0.001), SIRS vs. PSI (p<0.001), qSOFA vs. CRB (p=0.169), qSOFA vs. mSOFA (p=0.320), qSOFA vs. CURB-65 (p=0.142), qSOFA vs. PSI (p=0.003), CRB vs. mSOFA (p=0.916) CRB vs. CURB-65 (p=0.659), CRB vs. PSI (p=0.039), mSOFA vs. CURB-65 (p=0.967) mSOFA vs. PSI (p=0.116) CURB-65 vs. PSI (p=0.053)	SIRS vs. qSOFA (p=0.003), SIRS vs. CRB (p=0.002), SIRS vs. mSOFA (p<0.001), SIRS vs. CURB-65 (p<0.001), SIRS vs. PSI (p<0.001), qSOFA vs. CRB (p=0.436), qSOFA vs. mSOFA (p=0.035), qSOFA vs. CURB-65 (p=0.010), qSOFA vs. PSI (p<0.001), CRB vs. mSOFA (p=0.180) CRB vs. CURB-65 (p=0.019), CRB vs. PSI (p=0.004), mSOFA vs. CURB-65 (p=0.923) mSOFA vs. PSI (p=0.195) CURB-65 vs. PSI (p=0.037)	SIRS vs. qSOFA (p<0.001), SIRS vs. CRB (p<0.001), SIRS vs. mSOFA (p<0.001), SIRS vs. CURB-65 (p<0.001), SIRS vs. PSI (p<0.001), qSOFA vs. CRB (p=0.032), qSOFA vs. mSOFA (p=0.001), qSOFA vs. CURB-65 (p=0.005), qSOFA vs. PSI (p=0.001), CRB vs. mSOFA (p=0.062) CRB vs. CURB-65 (p=0.145), CRB vs. PSI (p=0.043), mSOFA vs. CURB-65 (p=0.280) mSOFA vs. PSI (p=0.911) CURB-65 vs. PSI (p=0.183)

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; PSI – Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.

Table E5. Scores performance and clinical utility for in-hospital mortality by each cohort.

		Discrimination	Calibration		Overall performance		Clinical utility						
	Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
Barcelona	**SIRS**	0.575 (0.547-0.602)	0.778	0.028	0.1%	1.2%	**SIRS**≥2	89 (86-92)	20 (19-21)	7 (7-8)	96 (95-98)	1.12 (1.08-1.17)	0.52 (0.39-0.71)
	qSOFA	0.691 (0.665-0.718)	0.302	0.054	3.1%	8.5%	**qSOFA**≥2	52 (47-57)	80 (79-81)	15 (13-17)	96 (95-97)	2.59 (2.31-2.91)	0.60 (0.54-0.67)
	CRB	0.712 (0.686-0.738)	0.996	0.312	4.8%	10.5%	**CRB**≥2	40 (35-46)	86 (85-87)	17 (14-20)	95 (95-96)	2.91 (2.52-3.36)	0.69 (0.64-0.75)
	mSOFA	0.746 (0.718-0.773)	0.990	0.512	8.1%	17.8%	**mSOFA**≥2	88 (84-91)	37 (36-38)	9 (8-10)	98 (97-98)	1.39 (1.34-1.46)	0.33 (0.25-0.43)
	CURB-65	0.741 (0.717-0.764)	0.883	0.010	6.3%	12.5%	**CURB-65**≥2	78 (73-82)	59 (58-61)	12 (10-13)	97 (97-98)	1.91 (1.79-2.03)	0.37 (0.31-0.46)
	PSI	0.777 (0.757-0.796)	0.999	0.562	6.5%	17.3%	**PSI** ≥4	93 (90-96)	46 (44-47)	11 (10-12)	99 (98-99)	1.71 (1.65-1.78)	0.15 (0.10-0.22)
	AUROC curve comparisons:	SIRS vs: qSOFA (p<0.001), CRB (p<0.001), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001) qSOFA vs: CRB (p=0.049), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001) CRB vs: mSOFA (p=0.032), CURB-65 (p<0.001), PSI (p<0.001) mSOFA vs: CURB-65 (p= 0.799), PSI (p=0.047) CURB-65 vs: PSI (p=0.001)											
Valencia	**SIRS**	0.547 (0.413-0.675)	0.999	0.361	1.1%	0.1%	**SIRS**≥2	79 (54-94)	33 (29-37)	4 (2-7)	98 (94-99)	1.18 (0.93-1.50)	0.64 (0.27-1.54)
	qSOFA	0.702 (0.587-0.807)	0.999	0.393	1.9%	7.7%	**qSOFA**≥2	21 (6-46)	95 (92-96)	12 (4-29)	97 (95-98)	3.95 (1.54-10.15)	0.83 (0.66-1.05)
	CRB	0.705 (0.585-0.819)	0.998	0.545	1.9%	9.5%	**CRB**≥2	26 (9-51)	95 (93-97)	17 (6-36)	97 (95-99)	5.77 (2.47-13.47)	0.77 (0.59-1.01)
	mSOFA	0.791 (0.694-0.877)	0.994	0.037	2.0%	12.7%	**mSOFA**≥2	95 (74-100)	40 (35-44)	5 (3-8)	100 (97-100)	1.57 (1.38-1.78)	0.13 (0.02-0.90)
	CURB-65	0.792 (0.665-0.897)	0.999	0.668	7.9%	16.9%	**CURB-65**≥2	84 (60-97)	70 (66-74)	9 (5-14)	99 (98-100)	2.79 (2.20-3.52)	0.23 (0.08-0.64)
	PSI	0.773 (0.653-0.877)	0.774	0.022	7.9%	14.6%	**PSI** ≥4	74 (49-91)	61 (57-66)	6 (4-11)	98 (96-100)	1.91 (1.43-2.55)	0.43 (0.20-0.91)
	AUROC curve comparisons:	SIRS vs: qSOFA (p=0.023), CRB (p=0.037), mSOFA (p=0.004), CURB-65 (p=0.001), PSI (p=0.001) qSOFA vs: CRB (p=0.950), mSOFA (p=0.141), CURB-65 (p=0.029), PSI (p=0.250) CRB vs: mSOFA (p=0.168), CURB-65 (p=0.065), PSI (p=0.315) mSOFA vs: CURB-65 (p= 0.980), PSI (p=0.547) CURB-65 vs: PSI (p=0.684)											

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; NPV: negative predictive value; PPV: positive predictive value; PSI – Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

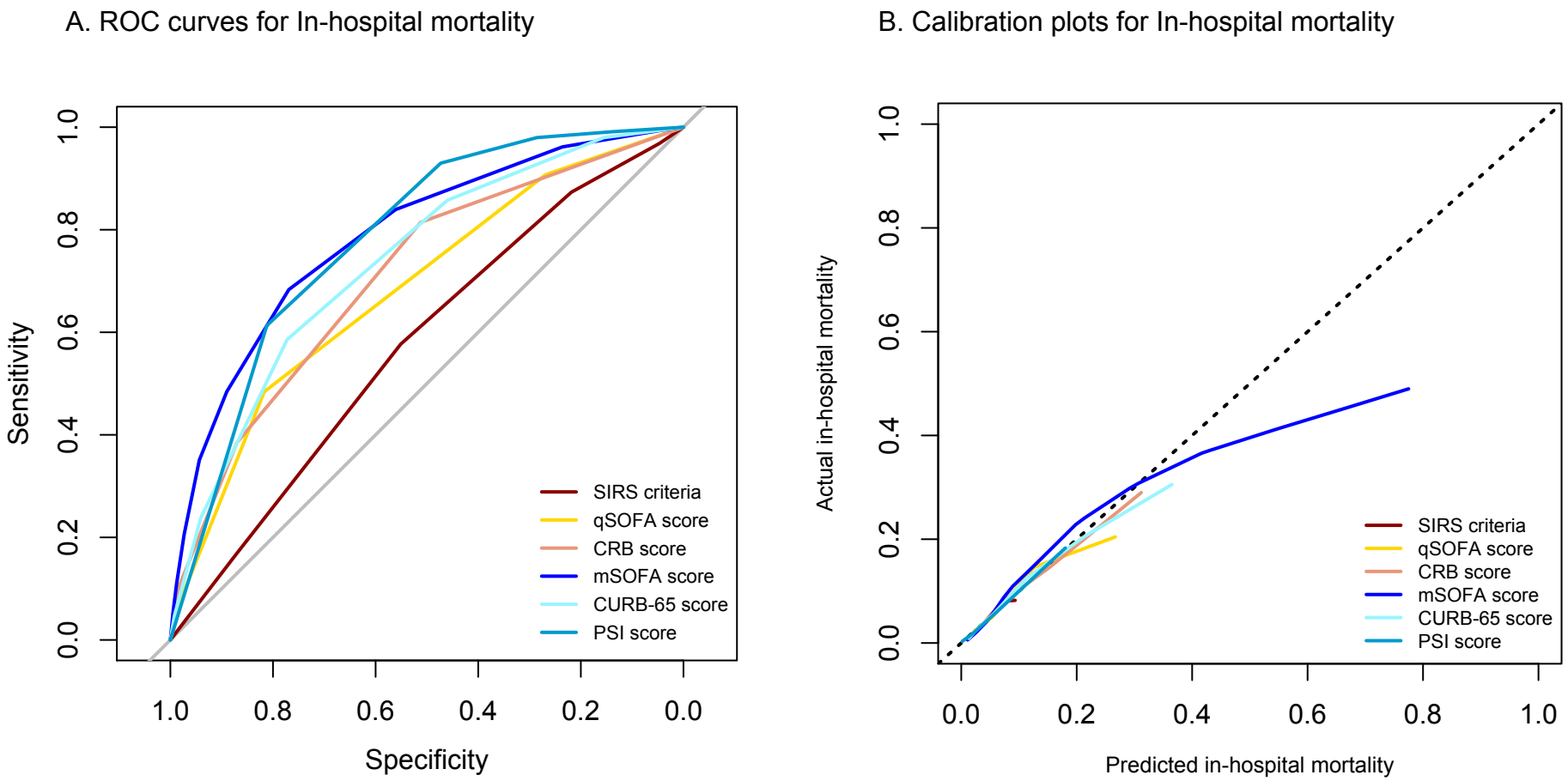
Table E6. Scores performance and clinical utility for in-hospital mortality in 6,874 patients with community-acquired pneumonia (imputed analysis)

	Discrimination	Calibration		Overall performance		Clinical utility						
Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
SIRS	0.575 (0.550-0.601)	0.880	0.317	0.3%	1.1%	**SIRS**≥2	87 (84-90)	22 (21-23)	7 (6-8)	96 (95-97)	1.12 (1.08-1.16)	0.58 (0.45-0.74)
qSOFA	0.684 (0.660-0.708)	0.366	0.091	3.2%	7.9%	**qSOFA**≥2	48 (44-53)	82 (81-83)	15 (14-17)	96 (95-96)	2.65 (2.37-2.95)	0.63 (0.58-0.69)
CRB	0.708 (0.684-0.731)	0.707	0.109	4.8%	10.2%	**CRB**≥2	38 (34-43)	87 (86-88)	17 (15-19)	95 (95-96)	2.97 (2.59-3.39)	0.71 (0.66-0.76)
mSOFA	0.787 (0.765-0.809)	0.162	<0.001	8.2%	17.8%	**mSOFA**≥2	96 (94-98)	24 (23-25)	8 (7-9)	99 (98-99)	1.26 (1.23-1.29)	0.16 (0.10-0.26)
CURB-65	0.736 (0.714-0.758)	0.860	0.040	6.9%	11.9%	**CURB-65**≥2	86 (82-89)	46 (45-47)	10 (9-11)	98 (97-98)	1.59 (1.52-1.66)	0.31 (0.25-0.39)
PSI	0.781 (0.763-0.799)	0.999	0.301	7.0%	17.7%	**PSI**≥4	93 (90-95)	47 (46-49)	11 (10-12)	99 (99-99)	1.76 (1.70-1.83)	0.15 (0.11-0.21)

AUROC curve comparisons:
SIRS vs: qSOFA (p<0.001), CRB (p<0.001), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001)
qSOFA vs: CRB (p=0.014), mSOFA (p<0.001), CURB-65 (p=0.002), PSI (p<0.001)
CRB vs: mSOFA (p<0.001), CURB-65 (p=0.090), PSI (p<0.001)
mSOFA vs: CURB-65 (p=0.002), PSI (p=0.684)
CURB-65 vs: PSI (p<0.001)

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; NPV: negative predictive value; PPV: positive predictive value; PSI – Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

Figure E6. Scores discrimination and calibration for in-hospital mortality in imputed data (n=6,874)



Panel A shows the scores discrimination evaluated through the area under the receiver operating characteristic curve (AUROC) for in-hospital mortality in 6,874 patients with community-acquired pneumonia. **Panel B** shows the scores calibration for in-hospital mortality in 6,874 patients with community-acquired pneumonia. Calibration curves assessed by nonparametric smoothed methods. Calibration curves assessed by nonparametric smoothed methods.

CRB-Confusion, Respiratory rate and Blood pressure; CURB-65-Confusion, Urea, Respiratory rate, Blood pressure and Age; mSOFA–modified Sequential (Sepsis-related) Organ Failure Assessment; PSI–Pneumonia Severity Index; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

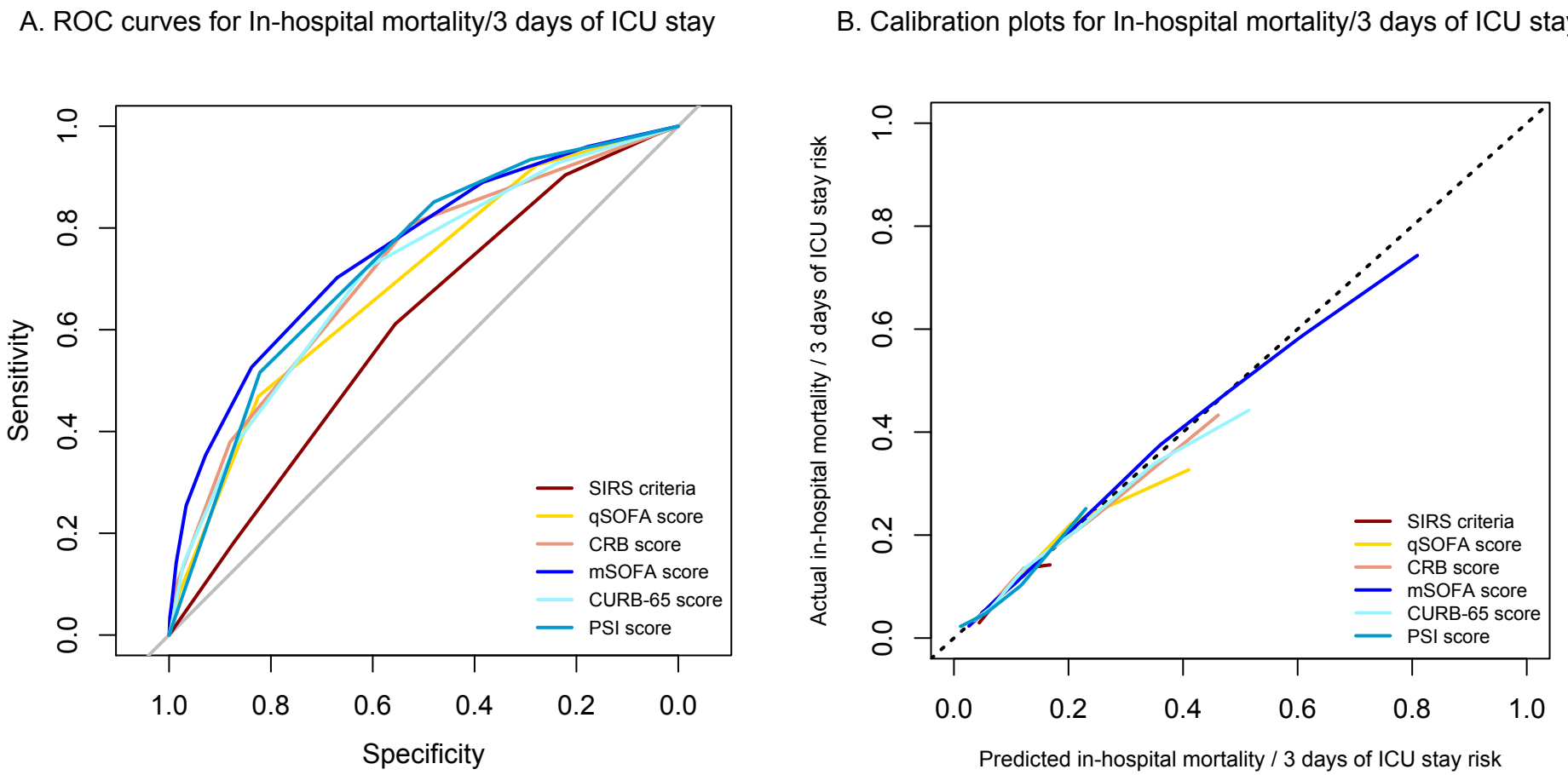
Table E7. Scores performance and clinical utility for in-hospital mortality / 3 ICU/HDU days in 6,024 patients with community-acquired pneumonia (complete-case analysis)

	Discrimination	Calibration		Overall performance		Clinical utility						
Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
SIRS	0.602 (0.581-0.623)	0.120	0.007	1.1%	2.6%	**SIRS**≥2	90 (88-93)	22 (21-23)	12 (11-13)	95 (94-96)	1.16 (1.13-1.20)	0.43 (0.34-0.55)
qSOFA	0.692 (0.672-0.711)	0.579	0.020	5.4%	10.2%	**qSOFA**≥2	47 (43-51)	82 (81-83)	24 (21-26)	93 (92-94)	2.67 (2.41-2.96)	0.64 (0.60-0.69)
CRB	0.712 (0.691-0.731)	0.997	0.070	7.5%	12.4%	**CRB**≥2	38 (34-42)	88 (87-89)	27 (24-30)	92 (92-93)	3.17 (2.80-3.59)	0.71 (0.66-0.75)
mSOFA	0.752 (0.731-0.773)	0.972	0.323	11.8%	17.7%	**mSOFA**≥2	89 (86-91)	38 (37-40)	14 (13-15)	97 (96-97)	1.44 (1.40-1.50)	0.29 (0.23-0.36)
CURB-65	0.708 (0.686-0.728)	0.883	0.081	6.4%	11.0%	**CURB-65**≥2	72 (68-76)	61 (60-63)	18 (16-19)	95 (94-96)	1.86 (1.75-1.97)	0.46 (0.40-0.52)
PSI	0.729 (0.710-0.749)	0.115	0.001	6.6%	13.0%	**PSI**≥4	85 (82-88)	48 (47-49)	16 (15-17)	97 (96-97)	1.64 (1.57-1.71)	0.31 (0.26-0.38)

AUROC curve comparisons:
SIRS vs: qSOFA (p<0.001), CRB (p<0.001), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001)
qSOFA vs: CRB (p=0.015), mSOFA (p<0.001), CURB-65 (p=0.106), PSI (p=0.001)
CRB vs: mSOFA (p=0.001), CURB-65 (p=0.421), PSI (p=0.093)
mSOFA vs.: CURB-65 (p<0.001), PSI (p=0.062)
CURB-65 vs.: PSI (p=0.009)

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; NPV: negative predictive value; PPV: positive predictive value; PSI – Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

Figure E7. Scores discrimination and calibration for in-hospital mortality / 3 days of ICU stay in complete-case analysis (n=6,024)



Panel A shows the scores discrimination evaluated through the area under the receiver operating characteristic curve (AUROC) for in-hospital mortality / 3 days of ICU stay in 6,024 patients with community-acquired pneumonia. **Panel B** shows the scores calibration for in-hospital mortality / 3 days of ICU stay in 6,024 patients with community-acquired pneumonia. Calibration curves assessed by nonparametric smoothed methods.

CRB - Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; mSOFA-modified Sequential (Sepsis-related) Organ Failure Assessment; PSI–Pneumonia Severity Index; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

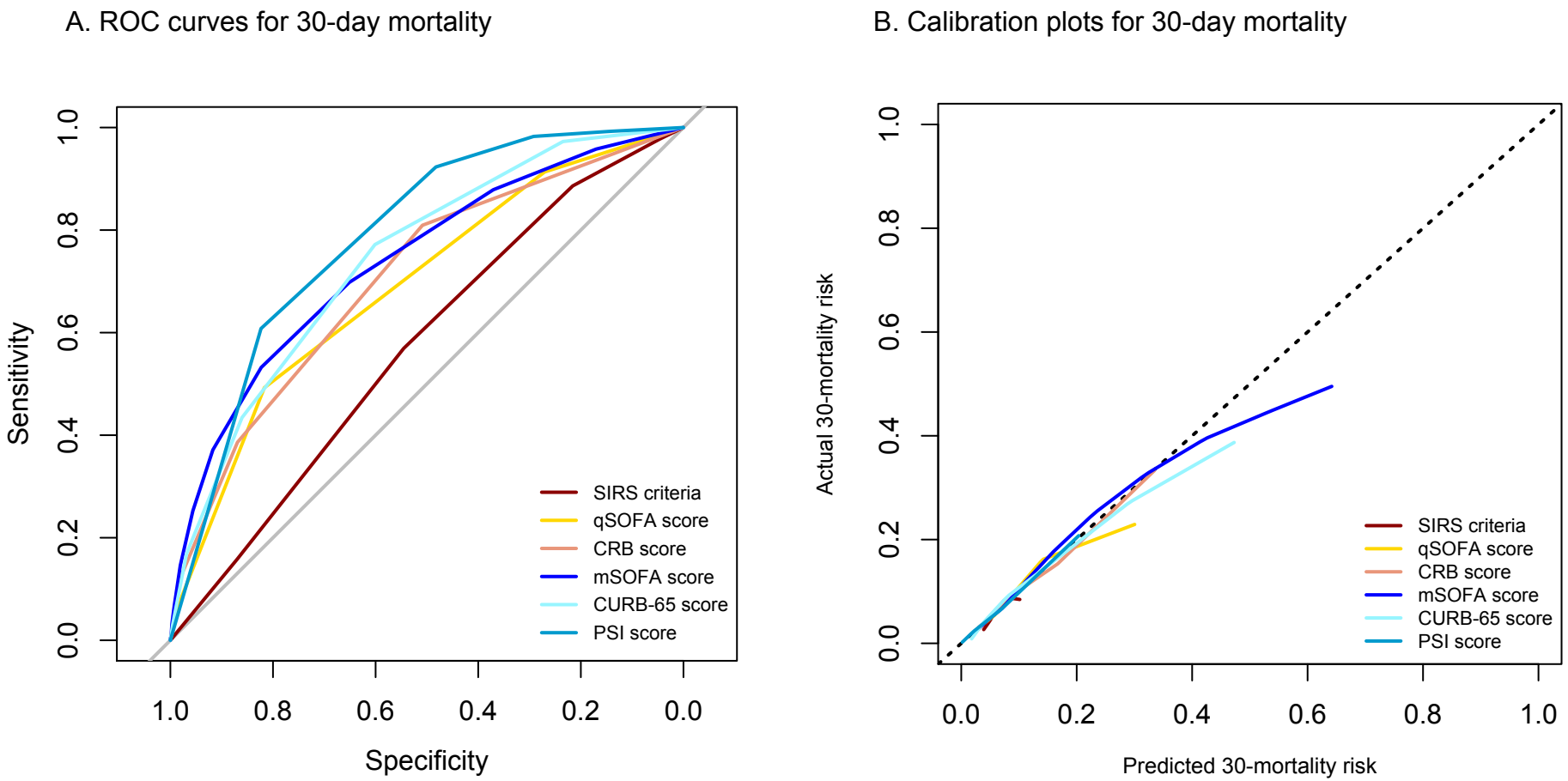
Table E8. Scores performance and clinical utility for 30-day mortality in 6,024 patients with community-acquired pneumonia (complete case analysis)

	Discrimination	Calibration		Overall performance		Clinical utility						
Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
SIRS	0.572 (0.546-0.598)	0.578	0.028	1.0%	1.1%	**SIRS**≥2	89 (85-92)	22 (20-23)	8 (7-9)	96 (95-97)	1.13 (1.09-1.17)	0.53 (0.40-0.70)
qSOFA	0.690 (0.665-0.715)	0.473	0.086	4.1%	8.7%	**qSOFA**≥2	49 (44-54)	82 (81-83)	17 (15-19)	95 (95-96)	2.69 (2.40-3.01)	0.62 (0.56-0.68)
CRB	0.705 (0.680-0.730)	0.995	0.492	5.6%	10.4%	**CRB**≥2	39 (34-44)	87 (86-88)	18 (16-21)	95 (94-96)	2.95 (2.56-3.40)	0.71 (0.65-0.76)
mSOFA	0.739 (0.712-0.764)	0.975	0.036	7.1%	13.5%	**mSOFA**≥2	88 (84-91)	37 (36-38)	10 (9-11)	98 (97-98)	1.40 (1.34-1.46)	0.33 (0.25-0.43)
CURB-65	0.743 (0.720-0.766)	0.877	0.006	5.5%	13.1%	**CURB-65**≥2	77 (73-81)	60 (59-61)	13 (12-14)	97 (97-98)	1.94 (1.82-2.06)	0.38 (0.32-0.45)
PSI	0.786 (0.767-0.804)	0.999	0.304	8.7%	19.0%	**PSI**≥4	92 (89-95)	48 (47-50)	12 (11-13)	99 (98-99)	1.78 (1.72-1.85)	0.16 (0.11-0.22)

AUROC curve comparisons:
SIRS vs: qSOFA (p<0.001), CRB (p<0.001), mSOFA (p<0.001), CURB-65 (p<0.001), PSI (p<0.001)
qSOFA vs: CRB (p=0.127), mSOFA (p=0.001), CURB-65 (p<0.001), PSI (p<0.001)
CRB vs: mSOFA (p=0.027), CURB-65 (p<0.001), PSI (p<0.001)
mSOFA vs.: CURB-65 (p=0.746), PSI (p=0.001)
CURB-65 vs.: PSI (p<0.001)

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; NPV: negative predictive value; PPV: positive predictive value; PSI – Pneumonia Severity Index; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome.. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

Figure E8. Scores discrimination and calibration for 30-day mortality in complete-case analysis (n=6,024)



Panel A shows the scores discrimination evaluated through the area under the receiver operating characteristic curve (AUROC) for 30-day mortality in 6,024 patients with community-acquired pneumonia. **Panel B** shows the scores calibration for 30-day mortality in 6,024 patients with community-acquired pneumonia. Calibration curves assessed by nonparametric smoothed methods.

CRB - Confusion, Respiratory rate and Blood pressure; CURB-65 - Confusion, Urea, Respiratory rate, Blood pressure and Age; mSOFA - modified Sequential (Sepsis-related) Organ Failure Assessment; PSI – Pneumonia Severity Index; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

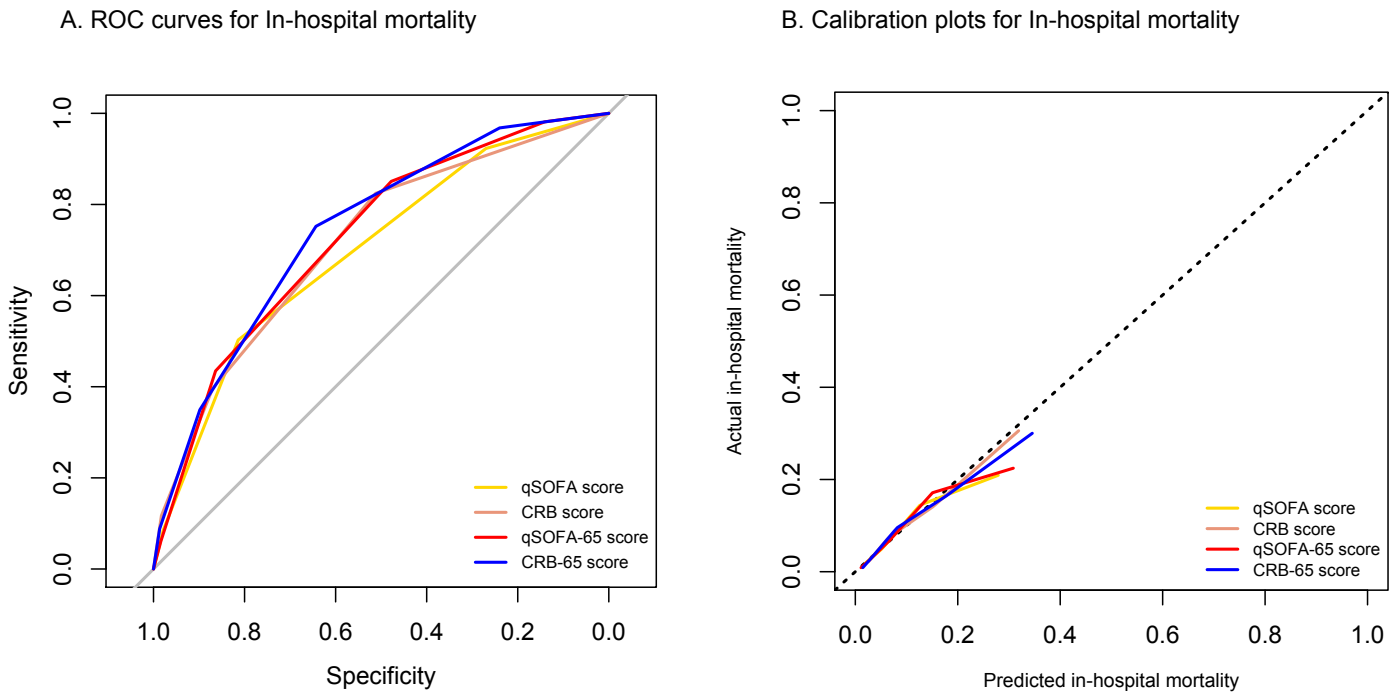
AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Table E9. qSOFA, CRB, qSOFA-65, CRB-65 performance and clinical utility for in-hospital mortality in 6,024 patients with community-acquired pneumonia

	Discrimination	Calibration		Overall performance		Clinical utility						
Score	AUROC (95% CI)	Hosmer-Lemeshow test	Global calibration test*	Scaled Brier score	Nagelkerke R-square	**Score category**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**LR+** (95% CI)	**LR-** (95% CI)
qSOFA	0.697 (0.671-0.722)	0.494	0.036	4.3%	8.8%	**qSOFA**≥2	50 (45-55)	81 (80-82)	15 (13-17)	96 (96-97)	2.70 (2.41-3.03)	0.61 (0.55-0.68)
CRB	0.716 (0.690-0.741)	0.997	0.224	4.3%	10.7%	**CRB**≥2	40 (35-45)	87 (86-88)	17 (14-20)	96 (95-96)	3.05 (2.65-3.51)	0.69 (0.64-0.75)
qSOFA-65	0.726 (0.702-0.750)	0.198	0.009	4.2%	10.9%	**qSOFA-65**≥2	85 (81-89)	48 (46-49)	10 (9-11)	98 (97-98)	1.63 (1.55-1.71)	0.31 (0.25-0.40)
CRB-65	0.743 (0.720-0.766)	0.592	0.015	5.9%	12.7%	**CRB-65**≥2	75 (71-79)	64 (63-66)	12 (11-14)	98 (97-98)	2.11 (1.97-2.26)	0.39 (0.32-0.46)

AUROC curve comparisons:
qSOFA vs: CRB (p=0.055), qSOFA-65 (p<0.001), CRB-65 (p<0.001)
CRB vs: qSOFA-65 (p=0.378), CRB-65 (p<0.001)
qSOFA-65 vs: CRB-65 (p=0.037)

Figure E9. qSOFA, CRB, qSOFA-65, CRB-65 discrimination and calibration for in-hospital mortality in 6,024 patients with community-acquired pneumonia.



AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; CRB-65 - Confusion, Respiratory rate, Blood pressure and Age; LR+: positive likelihood ratio; LR-: negative likelihood ratio; NPV: negative predictive value; PPV: positive predictive value; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; qSOFA-65: quick Sequential (Sepsis-related) Organ Failure Assessment plus Age . * *le Cessie-van Houwelingen-Copas-Hosmer* test.

Table E10. Additional predictive performance contribution of scores to a baseline risk model with demographic and chronic comorbidities

	Discrimination	Discrimination improvement	Calibration		Overall performance	
	AUROC (95% CI)	IDI Baseline + Score vs. Baseline (95% CI)	Hosmer Lemeshow test	Global calibration Test*	Scaled Brier score	Nagelkerke R square
Baseline model	0.745 (0.723-0.766)	-	0.031	<0.001	4.8%	11.8%
Baseline + SIRS	0.752 (0.731-0.774)	0.007 (0.004-0.010)	0.077	0.022	4.8%	13.0%
Baseline + qSOFA	0.780 (0.760-0.800)	0.022 (0.016-0.028)	0.004	<0.001	6.5%	15.7%
Baseline + CRB	0.794 (0.775-0.813)	0.033 (0.025-0.040)	0.010	<0.001	6.5%	17.4%
Baseline + mSOFA	0.836 (0.818-0.854)	0.095 (0.081-0.110)	0.768	0.035	13.2%	25.4%

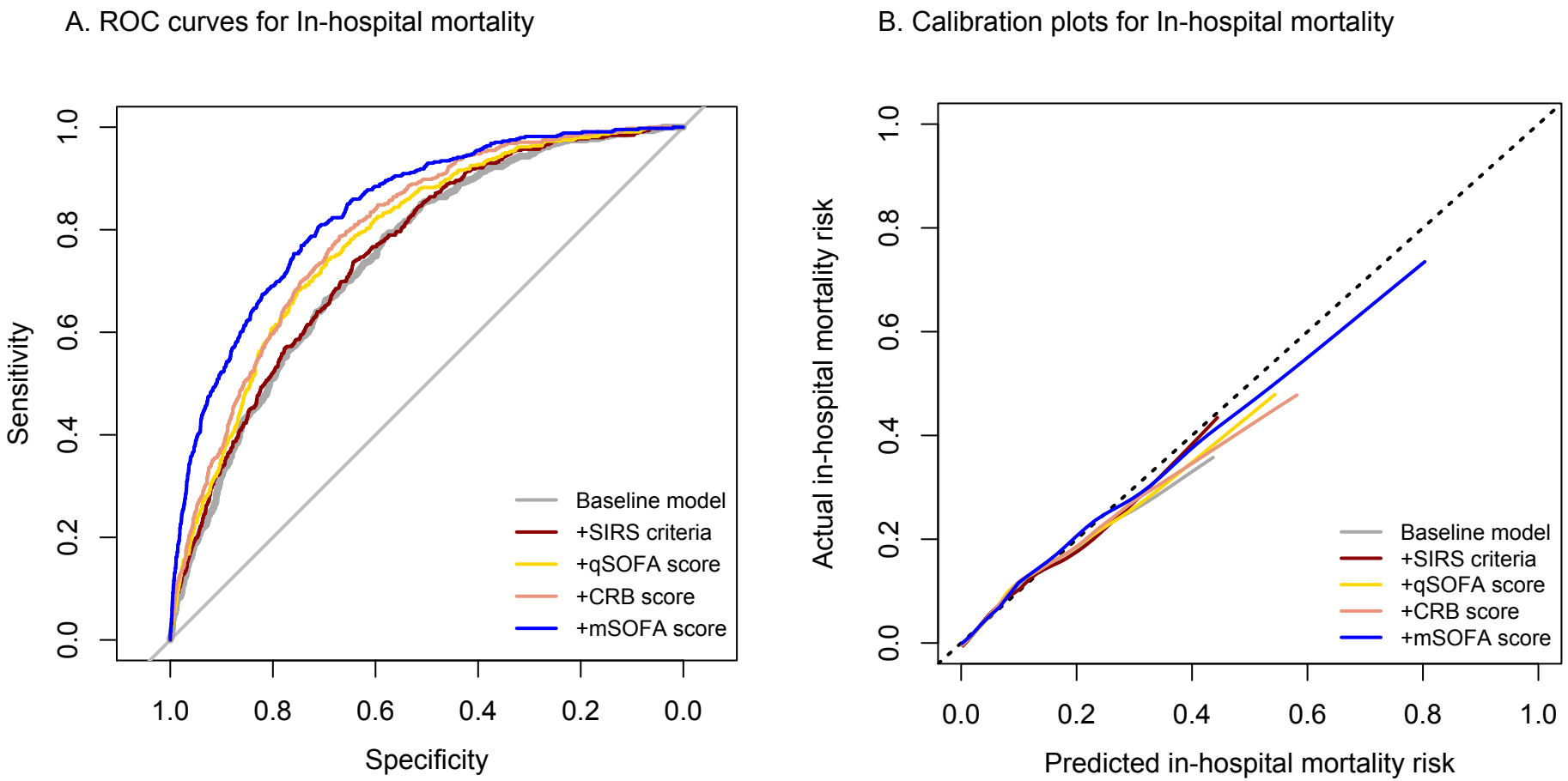
AUROC comparisons:
Baseline vs. +SIRS (p=0.065),
Baseline vs. +qSOFA (p<0.001),
Baseline vs. +CRB (p<0.001),
Baseline vs. +mSOFA (<0.001),
Baseline + SIRS vs. +qSOFA (p<0.001),
Baseline + SIRS vs. +CRB (p<0.001),
Baseline + SIRS vs. +mSOFA (p<0.001),
Baseline + qSOFA. vs. +CRB (0.007),
Baseline + qSOFA. vs. +mSOFA (p<0.001),
Baseline + CRB vs. +mSOFA (p<0.001).

IDI comparison among models
+ SIRS vs. +qSOFA: 0.015 (0.008-0.021),
+ SIRS vs. +CRB: 0.025 (0.017-0.034),
+ SIRS vs. +mSOFA: 0.088 (0.073-0.103),
+ qSOFA vs. +CRB: 0.011 (0.005-0.016),
+qSOFA vs. +mSOFA: 0.073 (0.060-0.086),
+ CRB vs. +mSOFA: 0.063 (0.050-0.076).

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; CRB: Confusion, Respiratory rate and Blood pressure; IDI: Integrated Discrimination Index; mSOFA: modified Sequential (Sepsis-related) Organ Failure Assessment; qSOFA: quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS: systemic inflammatory response syndrome. * *le Cessie-van Houwelingen-Copas-Hosmer* test.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure E10. Discrimination and Calibration of SIRS, qSOFA, CRB and mSOFA scores and their additional contribution to a baseline risk model for in-hospital mortality



Panel A shows the discrimination of baseline risk model and additional models with each score, estimated through a logistic regression model, for in-hospital mortality in the multiple imputed data (n=6,874 patients). **Panel B** shows the calibration of the baseline risk model and additional models with each score for in-hospital mortality in the multiple imputed data (n=6,874 patients). Calibration curves assessed by nonparametric smoothed methods. The baseline risk model included demographic (age, gender) and, comorbidity variables (chronic respiratory disease, chronic neurologic disease, liver disease, heart failure, diabetes mellitus, neoplasia, chronic renal disease, and etiologic diagnosis).

CRB - Confusion, Respiratory rate and Blood pressure; mSOFA - modified Sequential (Sepsis-related) Organ Failure Assessment; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

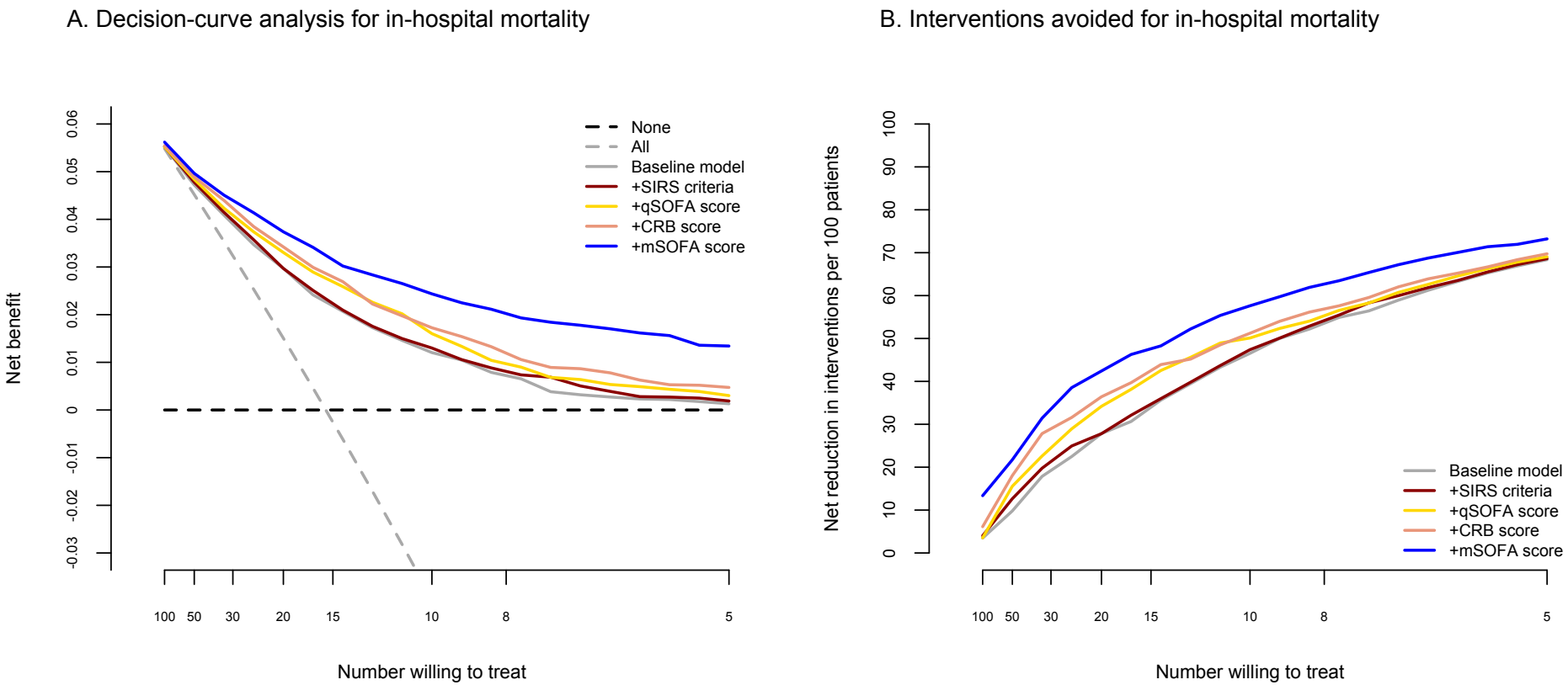
Table E11. Clinical utility of SIRS, qSOFA, CRB, mSOFA, CURB-65 and PSI among subgroups for in-hospital mortality (complete-case analysis, n=6,024)*

	<65 years					
	Sensitivity	**Specificity**	**PPV**	**NPV**	**LR+**	**LR-**
SIRS≥2	86 (75-93)	24 (22-25)	3 (3-4)	98 (97-99)	1.12 (1.01-1.24)	0.62 (0.35-1.10)
qSOFA≥2	51 (38-63)	86 (85-88)	10 (7-14)	98 (98-99)	3.71 (2.88-4.79)	0.57 (0.45-0.73)
CRB≥2	42 (30-55)	92 (91-93)	14 (9-19)	98 (97-99)	5.26 (3.86-7.17)	0.63 (0.52-0.77)
mSOFA≥2	94 (86-98)	51 (49-53)	6 (4-7)	100 (99-100)	1.94 (1.80-2.08)	0.11 (0.04-0.29)
CURB-65≥2	46 (34-59)	90 (88-91)	12 (8-17)	98 (98-99)	4.51 (3.41-5.98)	0.60 (0.48-0.74)
PSI≥4	81 (70-90)	79 (77-81)	11 (8-14)	99 (99-100)	3.89 (3.38-4.47)	0.24 (0.15-0.39)
	>=65 years					
	Sensitivity	**Specificity**	**PPV**	**NPV**	**LR+**	**LR-**
SIRS≥2	90 (86-93)	20 (19-22)	9 (8-10)	95 (94-97)	1.12 (1.08-1.17)	0.52 (0.37-0.72)
qSOFA≥2	50 (45-56)	78 (77-79)	17 (15-20)	95 (94-95)	2.29 (2.01-2.60)	0.64 (0.57-0.71)
CRB≥2	39 (34-45)	84 (82-85)	18 (15-21)	94 (93-95)	2.39 (2.04-2.80)	0.73 (0.66-0.80)
mSOFA≥2	87 (83-90)	28 (26-29)	10 (9-11)	96 (94-97)	1.20 (1.14-1.26)	0.48 (0.35-0.64)
CURB-65≥2	85 (81-89)	40 (39-42)	11 (10-13)	97 (96-98)	1.43 (1.35-1.50)	0.37 (0.28-0.48)
PSI≥4	95 (92-97)	25 (24-27)	10 (9-12)	98 (97-99)	1.27 (1.23-1.31)	0.21 (0.13-0.34)
	Without chronic comorbidities					
	Sensitivity	**Specificity**	**PPV**	**NPV**	**LR+**	**LR-**
SIRS≥2	84 (71-93)	25 (23-27)	3 (2-4)	98 (97-99)	1.11 (0.98-1.26)	0.65 (0.34-1.24)
qSOFA≥2	42 (28-57)	87 (85-89)	8 (5-12)	98 (98-99)	3.24 (2.30-4.58)	0.67 (0.53-0.84)
CRB≥2	32 (20-47)	91 (90-92)	9 (5-14)	98 (97-99)	3.58 (2.33-5.50)	0.75 (0.62-0.90)
mSOFA≥2	88 (76-95)	51 (48-53)	4 (3-6)	99 (99-100)	1.79 (1.60-2.00)	0.24 (0.11-0.50)
CURB-65≥2	68 (53-80)	76 (74-78)	7 (5-10)	99 (98-99)	2.85 (2.32-3.51)	0.42 (0.28-0.63)
PSI≥4	84 (71-93)	74 (71-76)	8 (6-10)	99 (99-100)	3.17 (2.75-3.66)	0.22 (0.12-0.41)
	At least 1 chronic comorbidity					
	Sensitivity	**Specificity**	**PPV**	**NPV**	**LR+**	**LR-**
SIRS≥2	90 (86-93)	20 (19-22)	9 (8-10)	96 (94-97)	1.13 (1.08-1.17)	0.50 (0.35-0.70)
qSOFA≥2	51 (46-57)	79 (77-80)	17 (15-20)	95 (94-96)	2.39 (2.10-2.71)	0.62 (0.55-0.70)
CRB≥2	41 (36-47)	85 (84-86)	19 (16-23)	94 (93-95)	2.75 (2.35-3.22)	0.69 (0.63-0.76)
mSOFA≥2	90 (86-93)	29 (28-31)	10 (9-11)	97 (96-98)	1.27 (1.21-1.32)	0.35 (0.25-0.50)
CURB-65≥2	80 (75-85)	51 (49-53)	13 (11-14)	97 (96-98)	1.64 (1.53-1.75)	0.39 (0.31-0.49)
PSI≥4	93 (89-96)	34 (32-35)	11 (10-12)	98 (97-99)	1.40 (1.35-1.46)	0.21 (0.14-0.32)

*95% confidence interval between brackets

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

Figure E11. Decision-curve analysis for in-hospital mortality using Baseline Risk Models Scores in imputed data



Panel A shows the net benefit (NB = benefit x true-positive classifications minus harm/cost x false-positive classifications) of each model and the strategy to treat-none and treat-all patients over the plausible range of number willing to treat (NWT) (i.e., threshold probabilities). Panel B shows the net reduction in interventions in a theoretical population of 100 patients by using the models to make clinical decisions.

CRB - Confusion, Respiratory rate and Blood pressure; mSOFA - modified Sequential (Sepsis-related) Organ Failure Assessment; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; ROC - receiver operating characteristic curve; SIRS – systemic inflammatory response syndrome.

Figure E12. In-hospital mortality stratified by SIRS, qSOFA CRB and mSOFA categories in the imputed data

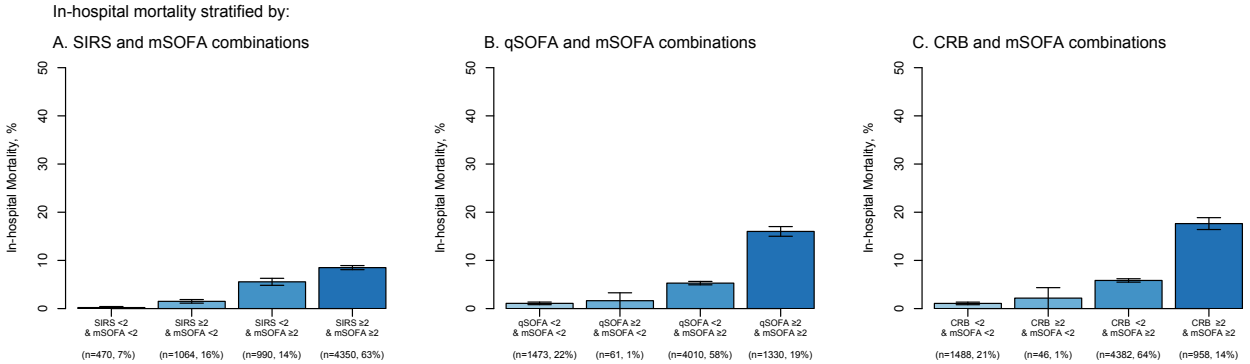
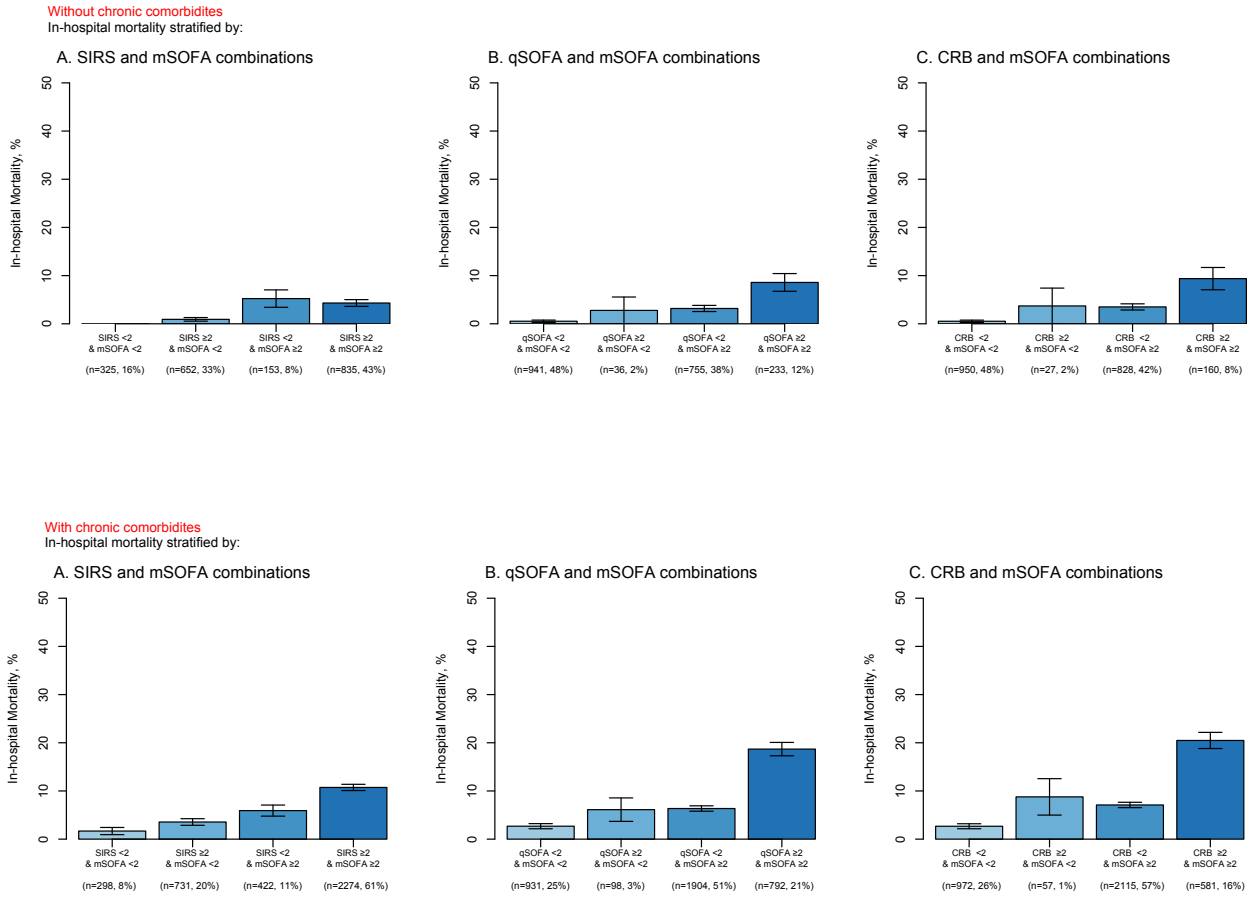


Figure E13. In-hospital mortality stratified by SIRS, qSOFA, CRB and mSOFA categories in patients without and with chronic comorbidities (complete-case analysis)



CRB - Confusion, Respiratory rate and Blood pressure; qSOFA – quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS – systemic inflammatory response syndrome; mSOFA – modified Sequential (Sepsis-related) Organ Failure Assessment.

AJRCCM Articles in Press. Published on 14-June-2017 as 10.1164/rccm.201611-2262OC

References:

1. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, Schein RMH, Sibbald WJ. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101: 1644-1655.
2. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315: 801-810.
3. Kolditz M, Scherag A, Rohde G, Ewig S, Welte T, Pletz M, Group CS. Comparison of the qSOFA and CRB-65 for risk prediction in patients with community-acquired pneumonia. *Intensive Care Med* 2016.
4. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, Lewis SA, Macfarlane JT. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003; 58: 377-382.
5. Bauer TT, Ewig S, Marre R, Suttorp N, Welte T, Group CS. CRB-65 predicts death from community-acquired pneumonia. *J Intern Med* 2006; 260: 93-101.
6. Capelastegui A, Espana PP, Quintana JM, Areitio I, Gorordo I, Egurrola M, Bilbao A. Validation of a predictive rule for the management of community-acquired pneumonia. *EurRespir J* 2006; 27: 151-157.
7. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M, Deutschman CS, Escobar GJ, Angus DC. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315: 762-774.
8. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 2009; 37: 1649-1654.
9. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, Pilcher DV, Australian, New Zealand Intensive Care Society Centre for O, Resource E. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. *JAMA* 2017; 317: 290-300.
10. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997; 336: 243-250.
11. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1-73.
12. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest* 2012; 42: 216-228.
13. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21: 128-138.
14. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
15. Harrell Jr F. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. USA: Springer; 2001.
16. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007; 35: 2052-2056.

17. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54: 774-781.
18. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393.
19. Masconi KL, Matsha TE, Erasmus RT, Kengne AP. Effects of Different Missing Data Imputation Techniques on the Performance of Undiagnosed Diabetes Risk Prediction Models in a Mixed-Ancestry Population of South Africa. *PLoS One* 2015; 10: e0139210.
20. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6.
21. Holmberg L, Vickers A. Evaluation of prediction models for decision-making: beyond calibration and discrimination. *PLoS Med* 2013; 10: e1001491.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565-574.
23. Dorresteijn JA, Visseren FL, Ridker PM, Wassink AM, Paynter NP, Steyerberg EW, van der Graaf Y, Cook NR. Estimating treatment effects for individual patients based on the results of randomised clinical trials. *BMJ* 2011; 343: d5888.
24. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, Field JK. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. *Ann Intern Med* 2012; 157: 242-250.
25. Siddiqui MM, Rais-Bahrami S, Turkbey B, George AK, Rothwax J, Shakir N, Okoro C, Raskolnikov D, Parnes HL, Linehan WM, Merino MJ, Simon RM, Choyke PL, Wood BJ, Pinto PA. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *JAMA* 2015; 313: 390-397.
26. van der Leeuw J, Ridker PM, van der Graaf Y, Visseren FL. Personalized cardiovascular disease prevention by applying individualized prediction of treatment effects. *Eur Heart J* 2014; 35: 837-843.
27. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *J Clin Oncol* 2016.
28. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2015.