In summary, much has been learned about the role of extracellular vesicles in mediating the therapeutic effects of MSCs in clinically relevant experimental studies of acute organ injury, with transfer of mitochondria, mRNA, and soluble factors that can improve the function of injured endothelial and epithelial cells and modulate macrophage function to advance repair. ∎

Author disclosures are available with the text of this article at www.atsjournals.org.

Michael A. Matthay, M.D.
Cardiovascular Research Institute
University of California, San Francisco
San Francisco, California

References

1. Matthay MA, Pati S, Lee JW. Concise review: mesenchymal stem (stromal) cells: biology and preclinical evidence for therapeutic potential for organ dysfunction following trauma or sepsis. *Stem Cells* 2017;35:316–324.
2. Morrison TJ, Jackson MV, Cunningham EK, Kissenpfennig A, McAuley DF, O'Kane CM, Krasnodembskaya AD. Mesenchymal stromal cells modulate macrophages in clinically relevant lung injury models by extracellular vesicle mitochondrial transfer. *Am J Respir Crit Care Med* 2017;196:1275–1286.
3. Islam MN, Das SR, Emin MT, Wei M, Sun L, Westphalen K, Rowlands DJ, Quadri SK, Bhattacharya S, Bhattacharya J. Mitochondrial transfer from bone-marrow-derived stromal cells to pulmonary alveoli protects against acute lung injury. *Nat Med* 2012;18:759–765.
4. Phinney DG, Di Giuseppe M, Njah J, Sala E, Shiva S, St Croix CM, Stolz DB, Watkins SC, Di YP, Leikauf GD, et al. Mesenchymal stem cells use extracellular vesicles to outsource mitophagy and shuttle microRNAs. *Nat Commun* 2015;6:8472.
5. Jackson MV, Morrison TJ, Doherty DF, McAuley DF, Matthay MA, Kissenpfennig A, O'Kane CM, Krasnodembskaya AD. Mitochondrial transfer via tunneling nanotubes is an important mechanism by which mesenchymal stem cells enhance macrophage phagocytosis in the in vitro and in vivo models of ARDS. *Stem Cells* 2016;34:2210–2223.
6. Gennai S, Monsel A, Hao Q, Park J, Matthay MA, Lee JW. Microvesicles derived from human mesenchymal stem cells restore alveolar fluid clearance in human lungs rejected for transplantation. *Am J Transplant* 2015;15:2404–2412.
7. Zhu YG, Feng XM, Abbott J, Fang XH, Hao Q, Monsel A, Qu JM, Matthay MA, Lee JW. Human mesenchymal stem cell microvesicles for treatment of *Escherichia coli* endotoxin-induced acute lung injury in mice. *Stem Cells* 2014;32:116–125.
8. Monsel A, Zhu YG, Gennai S, Hao Q, Hu S, Rouby JJ, Rosenzwajg M, Matthay MA, Lee JW. Therapeutic effects of human mesenchymal stem cell-derived microvesicles in severe pneumonia in mice. *Am J Respir Crit Care Med* 2015;192:324–336.
9. Kourembanas S. Exosomes: vehicles of intercellular signaling, biomarkers, and vectors of cell therapy. *Annu Rev Physiol* 2015;77:13–27.
10. Willis GR, Fernandez-Gonzalez A, Vitali SH, Liu X, Mitsialis SA, Kourembanas S. Mesenchymal stem cell exosome treatment restores lung architecture and ameliorates pulmonary hypertension associated with bronchopulmonary dysplasia [abstract]. *Am J Respir Crit Care Med* 2017;195:A4934.
11. Zhou Y, Xu H, Xu W, Wang B, Wu H, Tao Y, Zhang B, Wang M, Mao F, Yan Y, et al. Exosomes released by human umbilical cord mesenchymal stem cells protect against cisplatin-induced renal oxidative stress and apoptosis in vivo and in vitro. *Stem Cell Res Ther* 2013;4:34.
12. Camussi G, Deregibus MC, Bruno S, Cantaluppi V, Biancone L. Exosomes/microvesicles as a mechanism of cell-to-cell communication. *Kidney Int* 2010;78:838–848.
13. Kim DK, Nishida H, An SY, Shetty AK, Bartosh TJ, Prockop DJ. Chromatographically isolated CD63+CD81+ extracellular vesicles from mesenchymal stromal cells rescue cognitive impairments after TBI. *Proc Natl Acad Sci USA* 2016;113:170–175.
14. Laffey JG, Matthay MA. Fifty years of research in ARDS: cell-based therapy for ARDS: biology and potential therapeutic value. *Am J Respir Crit Care Med* [online ahead of print] 17 Mar 2017; DOI: 10.1164/rccm.201701-0107CP.
15. Németh K, Leelahavanichkul A, Yuen PS, Mayer B, Parmelee A, Doi K, Robey PG, Leelahavanichkul K, Koller BH, Brown JM, et al. Bone marrow stromal cells attenuate sepsis via prostaglandin E(2)-dependent reprogramming of host macrophages to increase their interleukin-10 production. *Nat Med* 2009;15:42–49.
16. Hayes M, Curley GF, Masterson C, Devaney J, O'Toole D, Laffey JG. Mesenchymal stromal cells are more effective than the MSC secretome in diminishing injury and enhancing recovery following ventilator-induced lung injury. *Intensive Care Med Exp* 2015;3:29.
17. O'Reilly M, Thébaud B. Cell-based therapies for neonatal lung disease. *Cell Tissue Res* 2017;367:737–745.

Severity Scores and Community-acquired Pneumonia
Time to Move Forward

Ever since the success of the pneumonia severity index (PSI) (1), a favorite pastime of community-acquired pneumonia (CAP) researchers has been to sort through their databases and try to prove that one score is better than another at predicting an important outcome, typically mortality. Over the past 20 years since the publication of the PSI, more than a dozen scores have been promulgated, some specific to pneumonia and others more generic across all patients with sepsis. Although the addition of a severity score

to clinical assessment has been shown to be associated with better patient outcomes, a clear consensus from the dozens of comparative analyses of different scores or even meta-analyses (2, 3) is hard to find.

In this issue of the *Journal*, Ranzani and colleagues (pp. 1287–1297) compare the performance of the criteria for systemic inflammatory response syndrome (SIRS); quick sepsis organ failure assessment (qSOFA); confusion, respiratory rate, and blood pressure (CRB); modified sepsis organ failure assessment (mSOFA); confusion, urea, respiratory rate, blood pressure, age > 65 years (CURB-65); and PSI in a large retrospective cohort of 6,874 patients with CAP from Spain (4). Not surprisingly, the best predictor of mortality

was the index with the most variables, the PSI. Even the PSI did not perform as well in older patients with comorbidities, probably because of the limitations of using dichotomous variables rather than continuous ones (e.g., there is a big difference between a left ventricular ejection fraction of 10% and one of 40%, but both are treated equally as 10 points in the PSI). The improved predictive ability that can be obtained by using continuous variables has been well documented with CURB-65 (5).

A more fundamental problem with all of these scores is that predicting who will die from pneumonia is not the same as identifying patients who need more than standard care. Most patients who die from pneumonia are elderly, with multiple comorbidities and significant limitations in care put in place upon or during admission. Patients who present needing respiratory or inotropic support are easily recognized and triaged into rapid resuscitative care. Much more problematic, and hence where there is a need for accurate tools to assist clinical judgment, is the identification of patients who will subsequently deteriorate. Identifying a patient at risk of deteriorating may not immediately indicate a need for different therapies, but certainly indicates the need for more intensive monitoring to ensure that deterioration will be recognized so that the appropriate support can be initiated as quickly as possible.

The continued use of all-cause mortality as the end point for determining performance is why most of the "this score is better than that score" studies are not particularly helpful. SMART-COP (6) is one notable exception, although it did not exclude patients with an obvious need for intensive-care support upon admission, and confined its interventions to inotropic support and mechanical ventilation. The National Early Warning Score took a similar approach, but was not particularly successful (7). The American Thoracic Society minor criteria (8) also were developed to predict patients at high risk of adverse outcomes from sepsis, and have reasonable sensitivity but a poor positive predictive value (9).

Although Ranzani and colleagues (4) also did not analyze their data specifically by looking at the subset of patients without significant care limitations who deteriorated after initial presentation (i.e., removing the "obvious" intensive-care cases), there are some key findings in their data. The fact that SIRS performed worse than qSOFA and CURB-65 demonstrates that an altered mental state is a very important clinical sign to identify in patients with CAP. Delirium can be subtle and is often underrecognized in patients in emergency departments (10, 11), but it is well established as having significant adverse prognostic features in a variety of acute illnesses, probably because it is a reflection of both host vulnerability and the degree of metabolic insult. Much more attention needs to be paid to screening for delirium in patients with CAP. That qSOFA did not perform quite as well as CRB suggests that the higher cutoff of 30 breaths per minute (rather than 22) may be more appropriate as a red flag in patients with CAP. Ranzani and colleagues (4) also used a composite end point of mortality and/or 3 or more days in critical care, but unfortunately did not present a separate analysis of just critical-care admission, which may have been more enlightening.

What, then, should clinicians do? The first thing is to ensure that all patients receive the minimum standard of care associated with best outcomes, including (1) antibiotics therapy compliant with antibiotic guidelines, (2) antibiotics delivered ideally within 3 hours or within 1 hour when shock is present, (3) adequate fluid resuscitation, (4) assessment of oxygenation, (5) rapid attention to metabolic abnormalities such as hyperglycemia and electrolyte abnormalities, and (6) attention to other comorbidities such as cardiac failure, airflow limitation, and arrhythmia. A clinical scoring tool should be used as an adjunct to (but not a replacement for) clinical judgment, predominantly to flag patients who are at risk of deterioration. The PSI remains the best predictor of mortality; however, it is not clear that this is the best predictor of the need for more intensive nursing or medical support. At a minimum, there needs to be a prompt assessment of mental state, respiratory rate, blood pressure, and oxygenation using the standard alarm values in the CURB-65 and PSI. This would be consistent with the Sepsis-3 recommendation to use an initial screening tool to flag patients at high risk (4). Once acute management is initiated, the clinician should consider the use of venous thrombosis prophylaxis, early ambulation (12, 13), and probably cardiovascular protection given the high rate of acute myocardial ischemia events in patients with CAP (14–16), although the need for the latter remains to be established.

Finally, I hope CAP researchers will stop trying to invent new scores to predict mortality or compare with the ones we already have. These are not useful data. We need tools that tell us what is not obvious on simple clinical assessment. We need tools that tell us to use an intervention we would not routinely use, or withhold one we would. We have been performing the same logistic regression and receiver operating characteristic curve analyses for two decades; it is time to move on. ∎

Grant Waterer, M.B. B.S., Ph.D.
University of Western Australia
Perth, Australia
and
Northwestern University
Chicago, Illinois

References

1. 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.
2. Chalmers JD, Singanayagam A, Akram AR, Mandal P, Short PM, Choudhury G, Wood V, Hill AT. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia: systematic review and meta-analysis. *Thorax* 2010;65: 878–883.
3. Marti C, Garin N, Grosgurin O, Poncet A, Combescure C, Carballo S, Perrier A. Prediction of severe community-acquired pneumonia: a systematic review and meta-analysis. *Crit Care* 2012;16:R141.
4. Ranzani OT, Prina E, Menéndez R, Ceccato A, Cilloniz C, Méndez R, Gabarrus A, Barbeta E, Li Bassi G, Ferrer M, *et al*. New sepsis definition (Sepsis-3) and community-acquired pneumonia mortality: a validation and clinical decision-making study. *Am J Respir Crit Care Med* 2017;196:1287–1297.
5. Jones BE, Jones J, Bewick T, Lim WS, Aronsky D, Brown SM, Boersma WG, van der Eerden MM, Dean NC. CURB-65 pneumonia severity assessment adapted for electronic decision support. *Chest* 2011;140: 156–163.
6. Charles PG, Wolfe R, Whitby M, Fine MJ, Fuller AJ, Stirling R, Wright AA, Ramirez JA, Christiansen KJ, Waterer GW, *et al*.; Australian

Community-Acquired Pneumonia Study Collaboration. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis* 2008;47: 375–384.

7. Sbiti-Rohr D, Kutz A, Christ-Crain M, Thomann R, Zimmerli W, Hoess C, Henzen C, Mueller B, Schuetz P; ProHOSP Study Group. The National Early Warning Score (NEWS) for outcome prediction in emergency department patients with community-acquired pneumonia: results from a 6-year prospective cohort study. *BMJ Open* 2016;6:e011021.

8. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, Dowell SF, File TM Jr, Musher DM, Niederman MS, et al.; Infectious Diseases Society of America; American Thoracic Society. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 2007;44:S27–S72.

9. Kolditz M, Ewig S, Klapdor B, Schütte H, Winning J, Rupp J, Suttorp N, Welte T, Rohde G; CAPNETZ Study Group. Community-acquired pneumonia as medical emergency: predictors of early deterioration. *Thorax* 2015;70:551–558.

10. Hustey FM, Meldon S, Palmer R. Prevalence and documentation of impaired mental status in elderly emergency department patients. *Acad Emerg Med* 2000;7:1166.

11. Tamune H, Yasugi D. How can we identify patients with delirium in the emergency department?: a review of available screening and diagnostic tools. *Am J Emerg Med* 2017;pii:S0735-6757(17)30407-2.

12. Carratalà J, Garcia-Vidal C, Ortega L, Fernández-Sabé N, Clemente M, Albero G, López M, Castellsagué X, Dorca J, Verdaguer R, et al. Effect of a 3-step critical pathway to reduce duration of intravenous antibiotic therapy and length of stay in community-acquired pneumonia: a randomized controlled trial. *Arch Intern Med* 2012;172: 922–928.

13. Mundy LM, Leet TL, Darst K, Schnitzler MA, Dunagan WC. Early mobilization of patients hospitalized with community-acquired pneumonia. *Chest* 2003;124:883–889.

14. Ramirez J, Aliberti S, Mirsaeidi M, Peyrani P, Filardo G, Amir A, Moffett B, Gordon J, Blasi F, Bordon J. Acute myocardial infarction in hospitalized patients with community-acquired pneumonia. *Clin Infect Dis* 2008;47:182–187.

15. Corrales-Medina VF, Serpa J, Rueda AM, Giordano TP, Bozkurt B, Madjid M, Tweardy D, Musher DM. Acute bacterial pneumonia is associated with the occurrence of acute coronary syndromes. *Medicine (Baltimore)* 2009;88:154–159.

16. Cangemi R, Calvieri C, Falcone M, Bucci T, Bertazzoni G, Scarpellini MG, Barillà F, Taliani G, Violi F; SIXTUS Study Group. Relation of cardiac complications in the early phase of community-acquired pneumonia to long-term mortality and cardiovascular events. *Am J Cardiol* 2015; 116:647–651.

Delving Deep into the Proteome of Lung Fibrosis Brings Plasma Cells to the Surface

Global research efforts continue to elucidate molecular mechanisms that underlie the development of interstitial lung diseases (ILDs) such as idiopathic pulmonary fibrosis (IPF). To a large extent, the omics revolution has helped drive forward the identification of key biological pathways that may be causal, diagnostic, or prognostic of fibrotic processes in the lung. Targeted genetic analyses and genome-wide association studies have identified mutations or polymorphisms associated with the risk of developing IPF, such as *SFTPC*, *TERT*, *MUC5B*, and *TOLLIP* (*see* Reference 1 for a recent review), while epigenetic and transcriptomic approaches have flagged gene expression signatures indicative of pathogenesis, and have given us valuable biomarkers such as MMP7, osteopontin, and the like (2, 3). Public repositories (e.g., the Gene Expression Omnibus) therefore contain an abundance of molecular information derived from patient lung tissue or cells isolated therefrom (e.g., fibroblasts, epithelial cells, and macrophages), or peripheral blood leukocytes. Genomic approaches further married with sophisticated integrative analyses to incorporate clinical phenotypes (such as that described by Kim and colleagues [4]) are therefore highly valuable for both classifying diseases and generating hypotheses for future research.

Despite these endeavors, we remain in a situation where treatment options for ILDs, including IPF, are scant. Thanks to the availability of pirfenidone and nintedanib (5, 6), some patients (but not all) will benefit from a slowing of disease progression. A number of clinical trials are currently ongoing that may expand our armament for fighting lung fibrosis, but it is crucial to broaden our understanding of the underlying causes. A key approach in this regard may be offered by cutting-edge proteomic analyses, such as that described by Schiller and colleagues (pp. 1298–1310) in this issue of the *Journal* (7), particularly as it is acknowledged that considerable disconnects can exist between gene expression signatures at the mRNA level and the protein level, since both post-transcriptional and post-translational controls will significantly affect the production of a functional protein (and hence the drug target for the majority of therapeutic approaches). This is notable, for example, in IPF fibroblasts (key effector cells in the fibrotic process), where a derangement of translational control is observed over and above transcript abundance (8).

However, proteomic profiles of lung fibrosis are still relatively few in number, not least due to the technical and financial limitations of performing large-scale analyses. With the advent of more accessible two-dimensional polyacrylamide gel electrophoresis approaches, chip-based assays, and mass spectrometry (MS), recent studies have elucidated the proteomic profiles of IPF bronchoalveolar lavage fluid (9, 10), IPF and nonspecific interstitial pneumonia lung tissue (11), and IPF plasma (12). MS has also been used to study metabolomics (13). These studies have provided useful insights into local or systemic changes in protein signatures associated with IPF (and also further validated certain genes already identified at the transcriptional level, such as MMP7 and osteopontin). Schiller and colleagues have taken this a significant step further by utilizing a novel tissue-fractionation approach termed quantitative detergent solubility profiling, which they previously applied to elegantly characterize proteomic changes in a murine model of lung fibrosis induced by bleomycin (14). Using quantitative detergent solubility profiling coupled with

ORIGINAL ARTICLE

New Sepsis Definition (Sepsis-3) and Community-acquired Pneumonia Mortality
A Validation and Clinical Decision-Making Study

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

[1]Department of Pulmonology, Hospital Clinic of Barcelona, University of Barcelona, Institut D'investigacions August Pi i Sunyer, Centro de Investigación Biomedica En Red-Enfermedades Respiratorias, Barcelona, Spain; [2]Respiratory Intensive Care Unit, Pulmonary Division, Heart Institute, Hospital das Clínicas, University of Sao Paulo, Sao Paulo, Brazil; [3]Pneumology Department, Instituto de Investigación Sanitaria/Hospital Universitario y Politecnico La Fe, Centro de Investigación Biomedica En Red-Enfermedades Respiratorias, Valencia, Spain; and [4]Seccion Neumologia, Hospital Nacional Prof. Alejandro Posadas, Palomar, Argentina

ORCID ID: 0000-0002-4677-6862 (O.T.R.).

Abstract

Rationale: The 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.

Objectives: To perform a clinical decision-making analysis of Sepsis-3 in patients with community-acquired pneumonia.

Methods: This was a cohort study including adult patients with community-acquired pneumonia from two Spanish university hospitals. SIRS, qSOFA, the Confusion, Respiratory Rate and Blood Pressure (CRB) score, modified SOFA (mSOFA), the Confusion, Urea, Respiratory Rate, Blood Pressure and Age (CURB-65) score, 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 miscalibration 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 with the "treat-all" strategy. PSI had higher net benefit than mSOFA and CURB-65 for mortality, whereas mSOFA seemed more applicable when considering mortality/intensive care unit 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 patients with community-acquired pneumonia in the emergency department. Among the tools for a comprehensive patient assessment, PSI had the best decision-aid tool profile.

Keywords: quick Sequential (Sepsis-related) Organ Failure Assessment; systemic inflammatory response syndrome; validation; pneumonia; sepsis

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 systemic inflammatory response syndrome and introducing a flowchart that comprises the quick Sequential (Sepsis-related) Organ Failure Assessment (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 systemic inflammatory response syndrome (SIRS) criteria 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 Confusion, Respiratory Rate and Blood Pressure (CRB), significantly increasing the risk of over-treatment and being comparable with the "treat-all" strategy. Among the tools for a comprehensive assessment, Pneumonia Severity Index (PSI) had better predictive performance and net benefit for mortality than modified SOFA and the Confusion, Urea, Respiratory Rate, Blood pressure and Age score (CURB-65), whereas modified SOFA was more useful when considering mortality/intensive care unit admission. Finally, following the Sepsis-3 flowchart resulted in better identification of patients at high risk of worse outcomes.

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 two 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, Confusion, Urea, Respiratory Rate, Blood Pressure and Age (CURB-65) and Pneumonia Severity Index (PSI), are well-validated scores to support CAP management and prognosis (9, 10). Simplifications of CURB-65 (i.e., Confusion, Respiratory Rate and Blood Pressure [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 greater than or equal to 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, Register 2009/5451; Valencia, Register 2011/0219).

CAP was defined as a new pulmonary infiltrate on chest radiograph on 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 wk, on cytotoxic therapy, or with acquired immunodeficiency syndrome) and active tuberculosis were exclusion criteria. We included patients from nursing homes. 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 reexamined or contacted by telephone 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: in-hospital mortality and/or need for critical care support greater than or equal to 3 days (composite outcome) (5, 8); and 30-day mortality. We defined need for critical care support as admission to an intensive care unit (ICU) or high-dependency unit (HDU).

Score Definition

We clustered the six scores in those that might facilitate the clinician's initial decision (SIRS, qSOFA, and CRB), and clinician's decision after initial management and

additional examinations (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 (*see* Table E1 in the online supplement) (5, 8, 9, 12). For SOFA score, we calculated the respiratory, hematologic, 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]) (*see* 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 nonparametric 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 logistic regression model. The baseline risk model included age, sex, 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), whereas clinicians usually apply different weights during the decision-making process (23). Decision-curve analysis is a method that depicts the predicted net benefit (NB; NB = benefit × true-positive
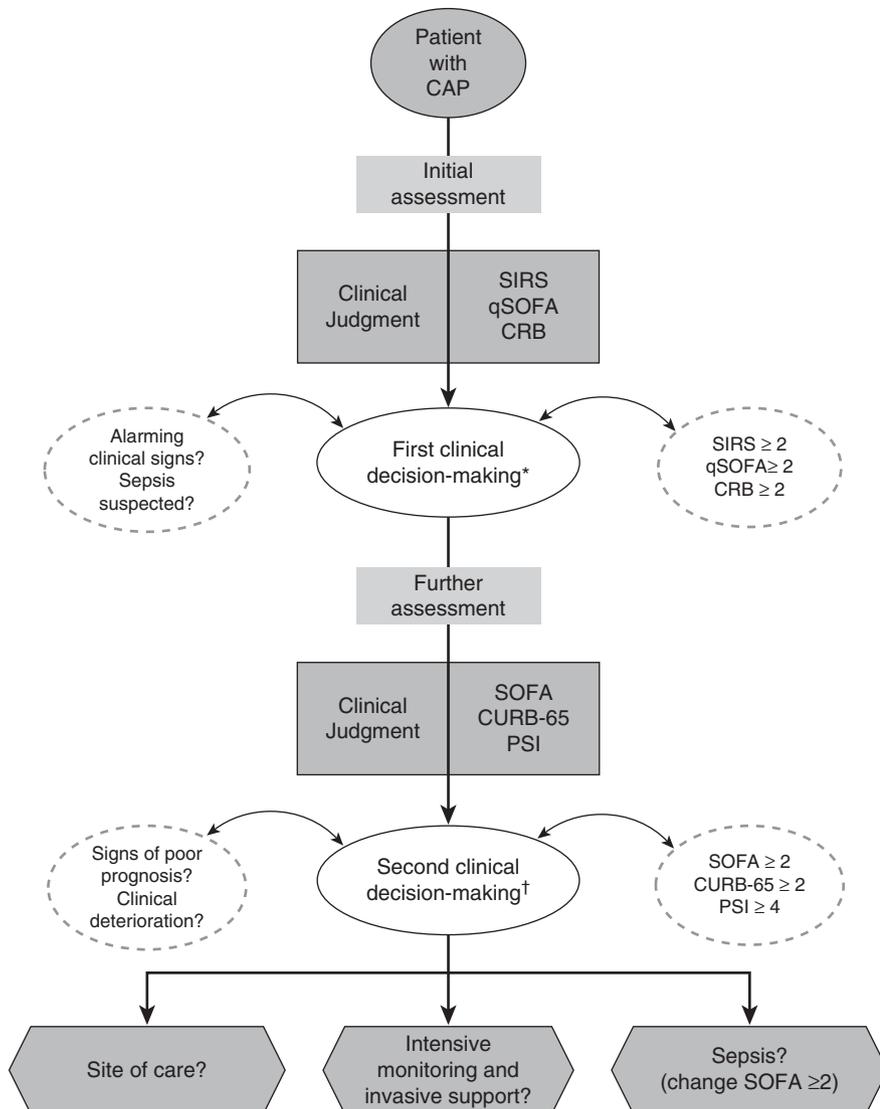
Figure 1. Flowchart about the decision-making process for community-acquired pneumonia management at the emergency department. *First clinical decision encompasses the decision to assess organ dysfunction and pneumonia severity with additional laboratory and/or invasive procedures. †Second clinical decision encompasses the decision, after full assessment, to admit the patient to the ward/intensive care unit, 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. CAP = community-acquired pneumonia; CRB = Confusion, Respiratory Rate and Blood Pressure; CURB-65 = Confusion, Urea, Respiratory Rate, Blood Pressure and Age; PSI = Pneumonia Severity Index; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS = systemic inflammatory response syndrome.

classifications − harm/cost × 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 NB 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 patients with sepsis 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 among SIRS (resembling Sepsis-2 definition), qSOFA (Sepsis-3 flowchart), and CRB with mSOFA.

Sensitivity, specificity, and positive and negative predictive values were calculated as shown elsewhere. Because 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 prespecified two subgroups, defined by age (<65, ≥65 yr) and chronic comorbidities (without chronic comorbidities, one or more 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 Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines

(32) and further information about methods and statistical analysis are in the online supplement.

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, 65.5% presented one or more chronic comorbidity, and 2,860 (41.8%) had microbiologic confirmation (Table 1). A total of 778 (11.3%) patients were discharged after a short-stay period in the ED, and 5,146 (74.9%) and 950 (13.8%) were admitted to the ward and ICU/HDU, respectively. Overall, in-hospital mortality

Table 1. Characteristics of Community-acquired Pneumonia Patients from Two Cohorts in Spain

	Values (N = 6,874)
Demographic variables	
Age, yr, mean (SD)	66.1 (19)
Age ≥65 yr, n (%)	4,170 (60.9)
Sex, male, n (%)	4,259 (62.2)
Comorbidities, 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 on presentation, mean (SD)	
Respiratory rate, per min	26 (8)
Heart rate, beats per min	98 (19)
Systolic blood pressure, mm Hg	131 (27)
Diastolic blood pressure, mm Hg	73 (14)
Mean arterial pressure, mm Hg	92 (16)
Temperature, °C	37.5 (1)
Laboratory findings on presentation	
Leukocyte, cells/mm^3, median (IQR)	12,400 (8,570–17,280)
C-reactive protein, mg/dl, median (IQR)	16.7 (7.8–26.3)
Creatinine, mg/dl, median (IQR)	1.0 (0.8–1.4)
Total bilirubin, mg/dl, median (IQR)	0.50 (0.40–0.80)
Platelets, cells/mm^3, median (IQR)	234 (187–302)
Pa$_{O_2}$/Fi$_{O_2}$ ratio, mm Hg, median (IQR)	281 (238–328)
Microbiology confirmed, n (%)	2,860 (41.8)
Scores on diagnosis	
SIRS, points, mean (SD)/median (IQR)	2.3 (1)/2 (2–3)
SIRS ≥2, n (%)	4,908 (78.8)
qSOFA, points, mean (SD)/median (IQR)	1.0 (0.7)/1 (0–1)
qSOFA ≥2, n (%)	1,260 (20.5)
CRB, points, mean (SD)/median (IQR)	0.7 (0.8)/1 (0–1)
CRB ≥2, n (%)	902 (14.6)
mSOFA, points, mean (SD)/median (IQR)	2.2 (2)/2 (1–3)
mSOFA ≥2, n (%)	4,288 (62.4)
CURB-65, points, mean (SD)/median (IQR)	1.4 (1)/1 (1–2)
CURB ≥2, n (%)	2,592 (42.0)
PSI, points, mean (SD)/median (IQR)	3.4 (1)/4 (2–4)
PSI ≥4, n (%)	3,379 (55.4)
Outcomes	
In-hospital mortality, n (%)	442 (6.4)
In-hospital mortality or 3 d of ICU stay, n (%)	716 (10.4)
30-d mortality, n/N (%)	477/6,377 (7.5)
Hospital stay, d, median (IQR)	7 (4–10)

Definition of abbreviations: CRB = Confusion, Respiratory Rate and Blood Pressure; CURB-65 = Confusion, Urea, Respiratory Rate, Blood Pressure and Age; ICU = intensive care unit; IQR = interquartile range; 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.

occurred in 442 (6.4%) patients, in-hospital mortality or greater than or equal than 3 ICU/HDU days in 716 (10.4%), and 30-day mortality in 477 of 6,377 (7.5%) (Table 1).

Score Distribution

Our complete-case analysis comprised 6,024 patients (87.6%) (*see* Tables E2 and 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 (*see* Figure E2) and for secondary outcomes (*see* Figures E3 and E4).

Very few patients who were discharged after a short ED stay had qSOFA and CRB greater than or equal to two points (4% and 2%, respectively), whereas 61% had SIRS greater than or equal to two points (Figure 3). These patients had very low 30-day mortality (3 of 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).

Score 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; *see* Figure E5). All scores presented worse discrimination for in-hospital mortality in patients greater than or equal to 65 years old. In those patients without chronic comorbidities, the discrimination of all scores improved (*see* Table E4). Regarding calibration, in general scores overestimated in-hospital mortality, and miscalibration was more evident for qSOFA, mSOFA, and CURB-65 (Table 2; *see* 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 the Barcelona and Valencia cohorts separately, but mSOFA and CURB-65 had better discrimination in the Valencia cohort (*see* Table E5). We found similar results when

analyzing the imputed data (*see* Table E6, Figure E6) and for secondary outcomes (*see* Tables E7 and E8, Figures E7 and 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 (*see* Table E9, Figure E9).

Additional Contribution of Scores 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 integrated discrimination improvement (IDI) measures (*see* Table E10, Figure E10).

### Clinical Usefulness and Decision-Curve Analysis

Among the tools for the initial assessment, SIRS greater than or equal to two presented high sensitivity and low specificity, whereas qSOFA greater than or equal to two and CRB greater than or equal to two presented moderate sensitivity and high specificity for in-hospital mortality (Table 2). Among follow-up tools, mSOFA greater than or equal to two presented high sensitivity and low specificity and CURB-65 greater than or equal to two and PSI greater than or equal to four 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 negative likelihood ratio (0.16; 95% CI, 0.12–0.23) (Table 2). We observed the same pattern in the imputed data (*see* Table E6) and for secondary outcomes (*see* Tables E7 and E8). In the subgroup analysis for in-hospital mortality, we observed similar findings except that mSOFA greater than or equal to two had higher specificity in the subgroup of patients aged less than 65 years (sensitivity, 94%; specificity, 51%) and without chronic comorbidities (sensitivity, 88%; specificity, 51%) (*see* Table E11).

The NB of qSOFA and CRB outperformed SIRS for in-hospital mortality, and SIRS showed an NB close to the "treat-all" strategy for most of the NWT values (Figure 4A). For NWT between 15 and 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 eight, 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 4B). 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 (Figures 4C–4F). The NB of the full models showed that baseline + 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, baseline + mSOFA presented the highest NB over the whole NWT range (*see* Figure E11).

Patients positive for SIRS/mSOFA (n = 3,274; 54%) had 9.0% (95% CI, 8.5–9.5) in-hospital mortality, whereas 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 (*see* Figure E12), and among patients with or without chronic comorbidities (*see* Figure E13).

Discussion

In a population of patients with CAP, 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 most scenarios. For a comprehensive
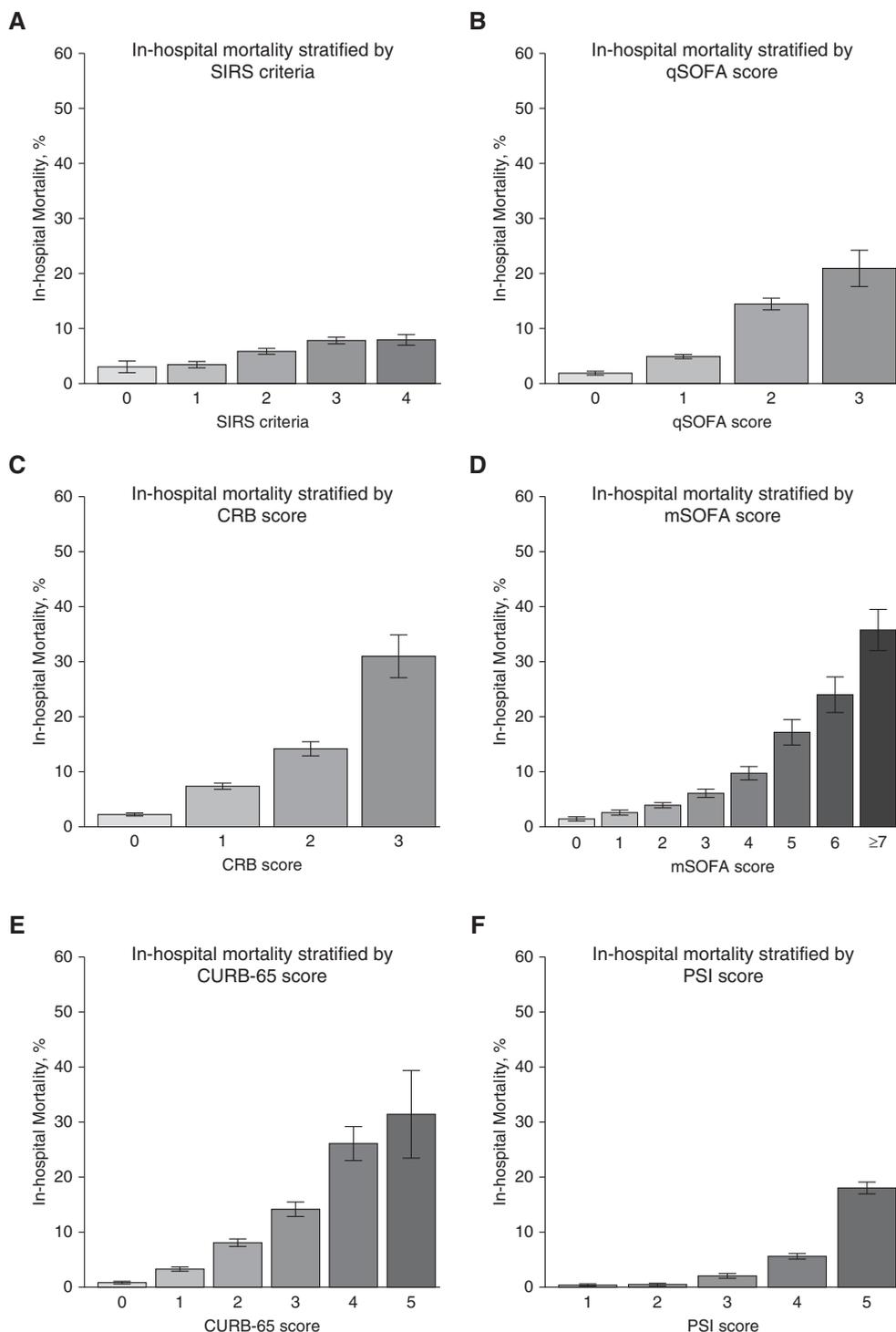
Figure 2. In-hospital mortality stratified by each score in patients with community-acquired pneumonia. (*A*) SIRS criteria. (*B*) qSOFA score. (*C*) CRB points. (*D*) mSOFA points. (*E*) CURB-65 points. (*F*) PSI risk class. *Error bars* denote 95% confidence intervals. The *x-axis* for mSOFA score was winsorized for values higher than 7 points for illustration. 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.
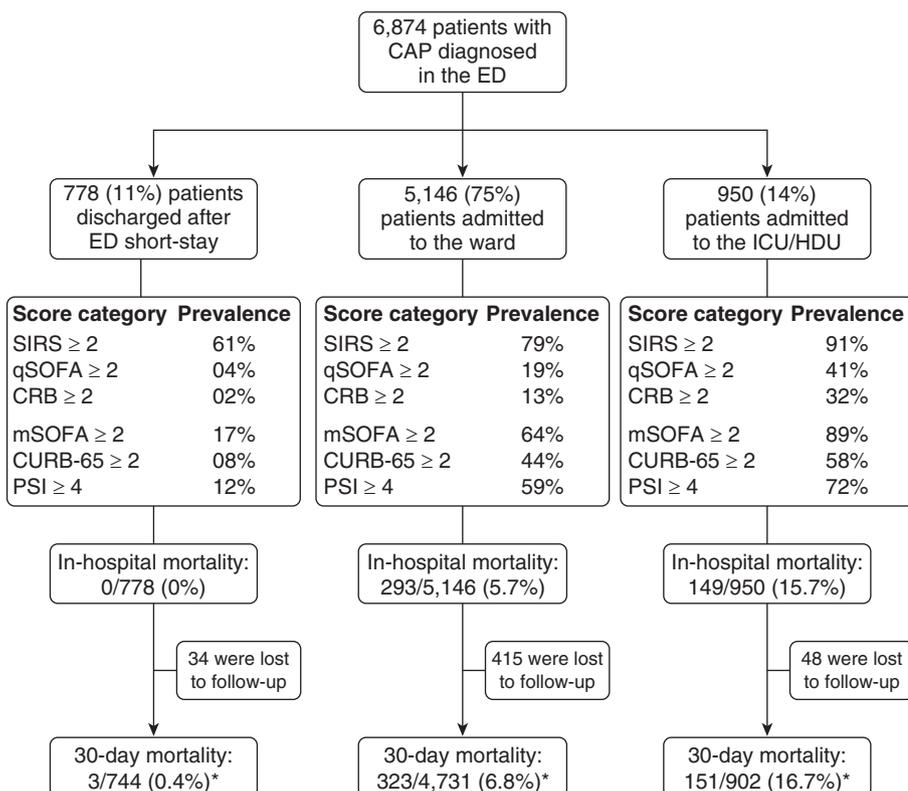
Figure 3. Distribution of SIRS, qSOFA, CRB, mSOFA, CURB-65, and PSI scores; place of treatment; and mortality for patients with community-acquired pneumonia. CAP = community-acquired pneumonia; CRB = Confusion, Respiratory Rate and Blood Pressure; CURB-65 = Confusion, Urea, Respiratory Rate, Blood Pressure and Age; ED = emergency department; HDU = high-dependency unit; ICU = intensive care unit; 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. *Patients were followed up until hospital discharge, and all survivors were reexamined or contacted by telephone 30 days after hospital discharge.

assessment of CAP, mSOFA and PSI had the best predictive performance and highest NB. 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 patients with CAP confirmed these limitations (3) and highlighted risks of overtreatment, demonstrating that the NB of SIRS is comparable with 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 SIRS or a "treat-all" strategy for NWT values lower than 40, which seems reasonable for use in the ED (5, 30), where qSOFA and CRB can be easily assessed. Given that CRB and CURB-65 were specifically developed for patients with CAP, they had better calibration and discrimination than qSOFA, and higher specificity. Thus, rather than qSOFA, physicians could consider CRB or CURB-65 for the initial risk stratification of patients with CAP.

For a comprehensive assessment of CAP, PSI had the best mortality prediction and highest NB from high NWT values down to an NWT of eight, reinforcing its pivotal role in CAP management. mSOFA seemed to be more applicable for NWT values lower than 12, mainly when considering ICU admission. This might be because PSI comprises 20 variables and has age as a main determinant for risk classification, whereas mSOFA measures acute organ dysfunctions in six 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 greater than or equal to 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 greater than or equal to two and CRB greater than or equal to two, respectively. However, SIRS greater than or equal to two was present in most of the promptly discharged patients (61%). Interestingly, 46% of patients had qSOFA less than two/mSOFA greater than or equal to two. In-hospital mortality in these patients was low (5.4%); this might indicate that patients with qSOFA less than two 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

Table 2. Score Performance and Clinical Utility for In-Hospital Mortality in 6,024 Patients with Community-acquired Pneumonia (Complete-Case Analysis)

| Score	Discrimination: AUROC (95% CI)	Calibration		Overall Performance		Score Category	Clinical Utility					
		Hosmer-Lemeshow Test	Global Calibration Test*	Scaled Brier Score (%)	Nagelkerke R-Square (%)		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)

Definition of abbreviations: 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.

AUROC curve comparisons: SIRS versus: qSOFA ($P < 0.001$), CRB ($P < 0.001$), mSOFA ($P < 0.001$), CURB-65 ($P < 0.001$), PSI ($P < 0.001$). qSOFA versus CRB ($P = 0.055$), mSOFA ($P = 0.001$), CURB-65 ($P < 0.001$), PSI ($P < 0.001$), CRB versus: mSOFA ($P = 0.035$), CURB-65 ($P = 0.001$), PSI ($P < 0.001$). mSOFA versus: CURB-65 ($P = 0.924$), PSI ($P = 0.031$). CURB-65 versus: PSI ($P = 0.002$).

*Le Cessie-van Houwelingen-Copas-Hosmer test.
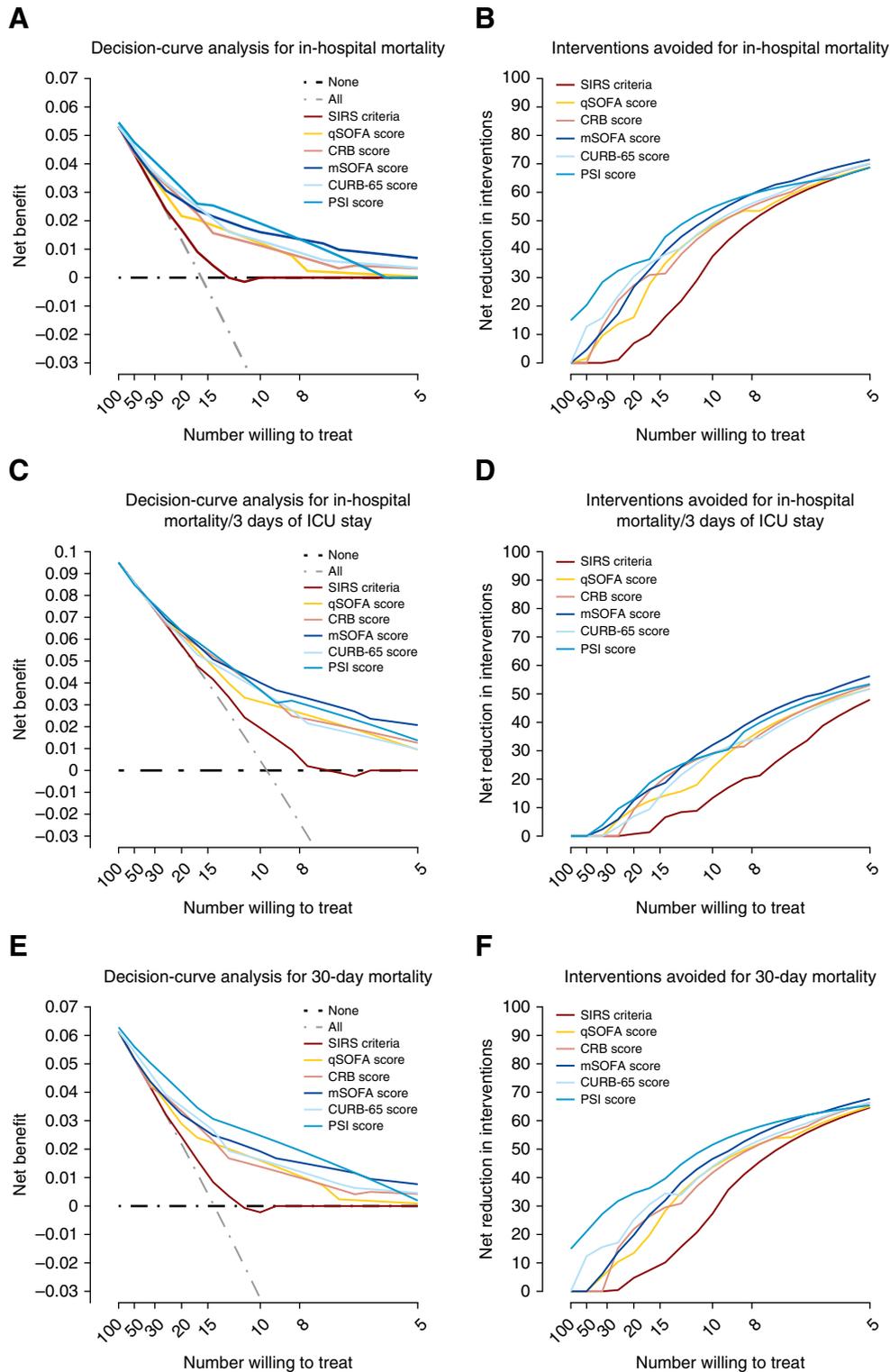
Figure 4. Decision curves showing the net benefit of SIRS criteria, qSOFA, CRB, mSOFA, CURB-65, and PSI scores in the treatment of patients with community-acquired pneumonia at risk of in-hospital mortality, composite outcome, and 30-day mortality. (*A*, *C*, and *E*) Net benefit (net benefit = benefit × true-positive classifications − harm/cost × 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). (*B*, *D*, and *F*) 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, there is a threshold probability of 5% and a NWT of 20. This choice specifies that death of a patient with community-acquired pneumonia who remained untreated is 19 times worse than the
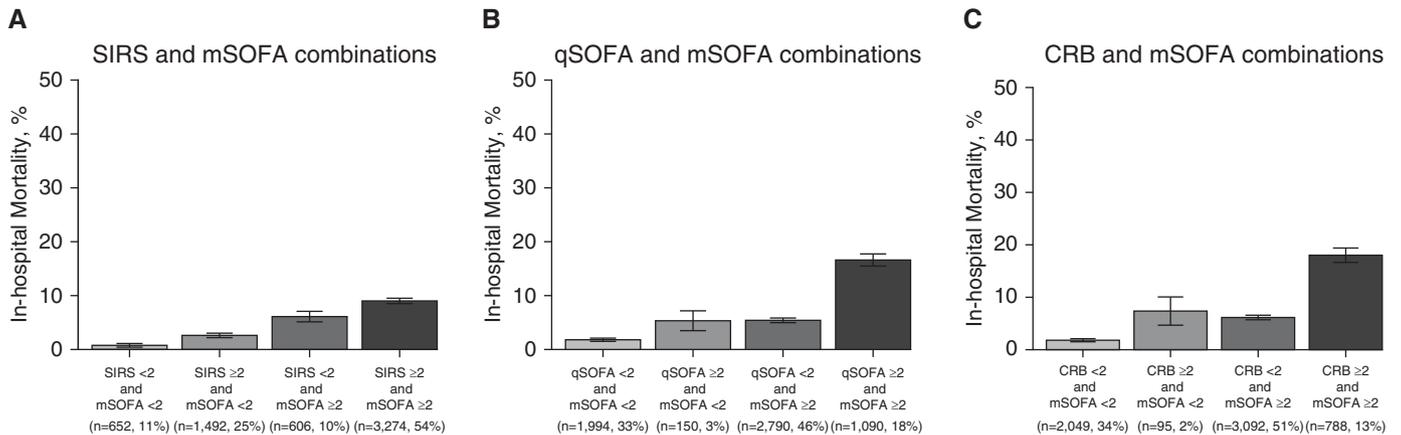
Figure 5. In-hospital mortality of 6,024 patients with community-acquired pneumonia, stratified by SIRS, qSOFA, or CRB combined with mSOFA. (*A*) In-hospital mortality stratified by SIRS criteria and mSOFA score combinations. (*B*) In-hospital mortality stratified by qSOFA and mSOFA score combinations. (*C*) In-hospital mortality stratified by CRB and mSOFA score 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; mSOFA = modified Sequential (Sepsis-related) Organ Failure Assessment; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS = systemic inflammatory response syndrome.

important to emphasize that the suggested cutoff of two 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, because of 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 neurologic 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 miscalibration, 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 patients with CAP in the ED. Moreover, CRB and CRB-65 had better predictive performance than qSOFA for initial stratification of patients with CAP in some scenarios, including higher NB for some values of NWT. For the comprehensive assessment of CAP, PSI had the best predictive performance and NB for mortality, whereas 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. ∎

Figure 4. (Continued). consequences of overtreatment of an unnecessarily treated patient. At a NWT of 19, the net benefit of the SOFA, qSOFA, and CRB scores outperforms SIRS and treat-all strategies. At the same time, at a NWT of 20, one 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; ICU = intensive care unit; PSI = Pneumonia Severity Index; qSOFA = quick Sequential (Sepsis-related) Organ Failure Assessment; SIRS = systemic inflammatory response syndrome.

References

1. Alberti C, Brun-Buisson C, Chevret S, Antonelli M, Goodman SV, Martin C, Moreno R, Ochagavia AR, Palazzo M, Werdan K, et al.; European Sepsis Study Group. 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. *Curr Opin Infect Dis* 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, et al. 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, et al. 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, et al. 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, España PP, Quintana JM, Areitio I, Gorordo I, Egurrola M, Bilbao A. Validation of a predictive rule for the management of community-acquired pneumonia. *Eur Respir J* 2006;27:151–157.
11. Bauer TT, Ewig S, Marre R, Suttorp N, Welte T; CAPNETZ Study Group. 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 Jeune I, Macfarlane JT, Read RC, Roberts HJ, Levy ML, et al.; Pneumonia Guidelines Committee of the BTS Standards of Care Committee. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009;64:iii1–iii55.
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, et al. Validation of quick SOFA in a large population of patients with community-acquired pneumonia. ESICM LIVES 2016: part one. Milan, Italy. 1–5 October 2016. *Intensive Care Med Exp* 2016;4(Suppl 1):A8.
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. Cillóniz C, Ewig S, Polverino E, Marcos MA, Esquinas C, Gabarrús 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 F Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: 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.
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, et al. 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. Vienna, Austria: R Foundation for Statistical Computing; 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;162:55–63.
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 Jr, Taniguchi LU, Azevedo LCP, Mendes PV, Zampieri FG, Park M. Sepsis-3 definitions predict ICU mortality in a low-middle-income country. *Ann 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.