Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Assessment of Clinical Criteria for Sepsis
For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)

Christopher W. Seymour, MD, MSc; Vincent X. Liu, MD, MSc; Theodore J. Iwashyna, MD, PhD; Frank M. Brunkhorst, MD; Thomas D. Rea, MD, MPH;
André Scherag, PhD; Gordon Rubenfeld, MD, MSc; Jeremy M. Kahn, MD, MSc; Manu Shankar-Hari, MD, MSc; Mervyn Singer, MD, FRCP;
Clifford S. Deutschman, MD, MS; Gabriel J. Escobar, MD; Derek C. Angus, MD, MPH

IMPORTANCE The Third International Consensus Definitions Task Force defined sepsis as "life-threatening organ dysfunction due to a dysregulated host response to infection." The performance of clinical criteria for this sepsis definition is unknown.

OBJECTIVE To evaluate the validity of clinical criteria to identify patients with suspected infection who are at risk of sepsis.

DESIGN, SETTINGS, AND POPULATION Among 1.3 million electronic health record encounters from January 1, 2010, to December 31, 2012, at 12 hospitals in southwestern Pennsylvania, we identified those with suspected infection in whom to compare criteria. Confirmatory analyses were performed in 4 data sets of 706 399 out-of-hospital and hospital encounters at 165 US and non-US hospitals ranging from January 1, 2008, until December 31, 2013.

EXPOSURES Sequential [Sepsis-related] Organ Failure Assessment (SOFA) score, systemic inflammatory response syndrome (SIRS) criteria, Logistic Organ Dysfunction System (LODS) score, and a new model derived using multivariable logistic regression in a split sample, the quick Sequential [Sepsis-related] Organ Failure Assessment (qSOFA) score (range, 0-3 points, with 1 point each for systolic hypotension [≤100 mm Hg], tachypnea [≥22/min], or altered mentation).

MAIN OUTCOMES AND MEASURES For construct validity, pairwise agreement was assessed. For predictive validity, the discrimination for outcomes (primary: in-hospital mortality; secondary: in-hospital mortality or intensive care unit [ICU] length of stay ≥3 days) more common in sepsis than uncomplicated infection was determined. Results were expressed as the fold change in outcome over deciles of baseline risk of death and area under the receiver operating characteristic curve (AUROC).

RESULTS In the primary cohort, 148 907 encounters had suspected infection (n = 74 453 derivation; n = 74 454 validation), of whom 6347 (4%) died. Among ICU encounters in the validation cohort (n = 7932 with suspected infection, of whom 1289 [16%] died), the predictive validity for in-hospital mortality was lower for SIRS (AUROC = 0.64; 95% CI, 0.62-0.66) and qSOFA (AUROC = 0.66; 95% CI, 0.64-0.68) vs SOFA (AUROC = 0.74; 95% CI, 0.73-0.76; P < .001 for both) or LODS (AUROC = 0.75; 95% CI, 0.73-0.76; P < .001 for both). Among non-ICU encounters in the validation cohort (n = 66 522 with suspected infection, of whom 1886 [3%] died), qSOFA had predictive validity (AUROC = 0.81; 95% CI, 0.80-0.82) that was greater than SOFA (AUROC = 0.79; 95% CI, 0.78-0.80; P < .001) and SIRS (AUROC = 0.76; 95% CI, 0.75-0.77; P < .001). Relative to qSOFA scores lower than 2, encounters with qSOFA scores of 2 or higher had a 3- to 14-fold increase in hospital mortality across baseline risk deciles. Findings were similar in external data sets and for the secondary outcome.

CONCLUSIONS AND RELEVANCE Among ICU encounters with suspected infection, the predictive validity for in-hospital mortality of SOFA was not significantly different than the more complex LODS but was statistically greater than SIRS and qSOFA, supporting its use in clinical criteria for sepsis. Among encounters with suspected infection outside of the ICU, the predictive validity for in-hospital mortality of qSOFA was statistically greater than SOFA and SIRS, supporting its use as a prompt to consider possible sepsis.

JAMA. 2016;315(8):762-774. doi:10.1001/jama.2016.0288

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Christopher W. Seymour, MD, MSc, Departments of Critical Care Medicine and Emergency Medicine, University of Pittsburgh School of Medicine, Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, 3550 Terrace St, Scaife Hall, Ste 639, Pittsburgh, PA 15261 (seymourcw@upmc.edu).

Although common and associated with high morbidity and mortality,[1,2] sepsis and related terms remain difficult to define. Two international consensus conferences in 1991 and 2001 used expert opinion to generate the current definitions.[3,4] However, advances in the understanding of the pathobiology and appreciation that elements of the definitions may be outdated, inaccurate, or confusing prompted the European Society of Intensive Care Medicine and the Society of Critical Care Medicine to convene a Third International Consensus Task Force to re-examine the definitions. Like many syndromes, there is no "gold standard" diagnostic test for sepsis. Therefore, the task force chose several methods to evaluate the usefulness of candidate clinical criteria, including clarity, reliability (consistency and availability), content validity (biologic rationale and face validity), construct validity (agreement between similar measures), criterion validity (correlation with established measures and outcomes), burden, and timeliness. Unlike prior efforts, the task force used systematic literature reviews and empirical data analyses to complement expert deliberations.

Based on clarity and content validity and after literature review and expert deliberation, the task force recommended elimination of the terms *sepsis syndrome*, *septicemia*, and *severe sepsis* and instead defined sepsis as "life-threatening organ dysfunction due to a dysregulated host response to infection."[5] Of note, the task force did not attempt to redefine infection. Rather, it next sought to generate recommendations for clinical criteria that could be used to identify sepsis among patients with suspected or confirmed infection. The purpose of this study was to inform this step by analyzing data from several large hospital databases to explore the construct validity and criterion validity of existing and novel criteria associated with sepsis.

EHR *electronic health record*

GCS *Glasgow Coma Scale*

ICU *intensive care unit*

LODS *Logistic Organ Dysfunction System*

qSOFA *quick Sequential [Sepsis-related] Organ Function Assessment*

SIRS *systemic inflammatory response syndrome*

SOFA *Sequential [Sepsis-related] Organ Function Assessment*

Methods

This study was approved with waiver of informed consent by the institutional review boards of the University of Pittsburgh, Kaiser Permanente Northern California (KPNC), Veterans Administration (VA) Ann Arbor Health System, Washington State Department of Health, King County Emergency Medical Services (KCEMS), University of Washington, and Jena University Hospital.

Study Design, Setting, and Population

A retrospective cohort study was performed among adult encounters (age ≥18 years) with suspected infection. The primary cohort was all hospital encounters from 2010 to 2012 at 12 community and academic hospitals in the UPMC health care system in southwestern Pennsylvania. The cohort included all medical and surgical encounters in the emergency department, hospital ward, and intensive care unit (ICU). We created a random split sample (50/50) from the UPMC cohort, the derivation cohort for developing new criteria, and the validation cohort for assessment of new and existing criteria.

We also studied 4 external data sets: (1) all inpatient encounters at 20 KPNC hospitals from 2009 to 2013; (2) all encounters in 130 hospitals in the United States' VA system from 2008 to 2010; (3) all nontrauma, nonarrest emergency medical services records from 5 advanced life support agencies from 2009-2010 transported to 14 hospitals with community infection in King County, Washington (KCEMS)[6]; and (4) all patients from 2011-2012 at 1 German hospital enrolled with hospital-acquired infection in the ALERTS prospective cohort study.[7] These cohorts were selected because they included patient encounters from different phases of acute care (out-of-hospital, emergency department, hospital ward) and countries (United States and Germany) with different types of infection (community and nosocomial). The UPMC, KPNC, and VA data were obtained from the electronic health records (EHRs) of the respective health systems; KCEMS data were obtained from the administrative out-of-hospital record; and ALERTS data were collected prospectively by research coordinators.

Defining a Cohort With Suspected Infection

For EHR data (UPMC, KPNC, and VA), the first episode of suspected infection was identified as the combination of antibiotics (oral or parenteral) and body fluid cultures (blood, urine, cerebrospinal fluid, etc). We required the combination of culture and antibiotic start time to occur within a specific time epoch. If the antibiotic was given first, the culture sampling must have been obtained within 24 hours. If the culture sampling was first, the antibiotic must have been ordered within 72 hours. The "onset" of infection was defined as the time at which the first of these 2 events occurred (eAppendix in the Supplement). For non-EHR data in ALERTS, patients were included who met US Centers for Disease Control and Prevention definitions or clinical criteria for hospital-acquired infection more than 48 hours after admission as documented by prospective screening.[7] For non-EHR data in KCEMS, administrative claims identified infection present on admission (Angus implementation of infection using *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) diagnosis codes).[6]

Determining Clinical Criteria for Sepsis Using Existing Measures

In UPMC derivation and validation data, indicators were generated for each component of the systemic inflammatory response syndrome (SIRS) criteria[4]; the Sequential [Sepsis-related] Organ Failure Assessment (SOFA) score[8]; and the Logistic Organ Dysfunction System (LODS) score,[9] a weighted organ dysfunction score (**Table 1**). We used a modified version of the LODS score that did not contain urine output (because of poor accuracy in recording on hospital ward encounters), prothrombin, or urea levels. The maximum SIRS criteria, SOFA score, and modified LODS score were calculated for the time window from 48 hours before to 24 hours after the onset of infection, as well as on each calendar day. This window was used

Table 1. Variables for Candidate Sepsis Criteria Among Encounters With Suspected Infection

Systemic Inflammatory Response Syndrome (SIRS) Criteria (Range, 0-4 Criteria)	Sequential [Sepsis-related] Organ Failure Assessment (SOFA) (Range, 0-24 Points)	Logistic Organ Dysfunction System (LODS) (Range, 0-22 Points)[a]	Quick Sequential [Sepsis-related] Organ Failure Assessment (qSOFA) (Range, 0-3 Points)
Respiratory rate, breaths per minute	Pao_2/Fio_2 ratio	Pao_2/Fio_2 ratio	Respiratory rate, breaths per minute
White blood cell count, 10^9/L	Glasgow Coma Scale score	Glasgow Coma Scale score	Glasgow Coma Scale score
Bands, %	Mean arterial pressure, mm Hg	Systolic blood pressure, mm Hg	Systolic blood pressure, mm Hg
Heart rate, beats per minute	Administration of vasopressors with type/dose/rate of infusion	Heart rate, beats per minute	
Temperature, °C	Serum creatinine, mg/dL, or urine output, mL/d	Serum creatinine, mg/dL	
Arterial carbon dioxide tension, mm Hg	Bilirubin, mg/dL	Bilirubin, mg/dL	
	Platelet count, 10^9/L	Platelet count, 10^9/L	
		White blood cell count, 10^9/L	
		Urine output, L/d	
		Serum urea, mmol/L	
		Prothrombin time, % of standard	

Abbreviation: Fio_2, fraction of inspired oxygen.

[a] Measurement units for LODS variables per original description by Le Gall et al.[9]

for candidate criteria because organ dysfunction in sepsis may occur prior to, near the moment of, or after infection is recognized by clinicians or when a patient presents for care. Moreover, the clinical documentation, reporting of laboratory values in EHRs, and trajectory of organ dysfunction are heterogeneous across encounters and health systems. In a post hoc analysis requested by the task force, a change in SOFA score was calculated of 2 points or more from up to 48 hours before to up to 24 hours after the onset of infection.

Deriving Novel Clinical Criteria for Sepsis

In the derivation cohort (UPMC), new, simple criteria were developed according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) recommendations.[10] This entailed 2 steps: (1) assessing candidate variable quality and frequency of missing data and (2) developing a parsimonious model and simple point score.[3,8,11] Because of the subjective nature and complexity of variables in existing criteria, we sought a simple model that could easily be used by a clinician at the bedside.

Based on the assumption that hospital mortality would be far more common in encounters with infected patients who have sepsis than in those who do not, all continuous variables were dichotomized by defining their optimal cutoffs using the minimum 0/1 distance on the area under the receiver operating characteristic curve (AUROC) for in-hospital mortality.[12] Cutoffs were rounded to the nearest integer, and standard single-value imputation was used, with normal value substitution if variables were missing. The latter approach is standard in clinical risk scores[8,13,14] and mirrors how clinicians would use the score at the bedside. Multiple logistic regression was used with robust standard errors and forward selection of candidate variables using the Bayesian information criterion to develop the "quick SOFA" (qSOFA) model. The Bayesian information criterion is a likelihood-based stepwise approach that retains variables that improve the model's overall ability to predict the outcome of

interest while incorporating a penalty for including too many variables. Favoring simplicity over accuracy, a point score of 1 was assigned to each variable in the final model, irrespective of the regression coefficients. Model calibration was assessed by comparing clinically relevant differences in observed vs expected outcomes, as the Hosmer-Lemeshow test may be significant due to large sample sizes.[15]

Assessments of Candidate Clinical Criteria

The test:retest or interrater reliability of individual elements was not assessed, in part because most elements have known reliability. However, the frequency of missing data was determined for each element because more common missing data for individual elements will potentially affect the reliability of integrated scores such as the SOFA score. Construct validity was determined by examining the agreement between different measures analogous to the multitrait-multimethod matrix approach of Campbell and Fiske, using the Cronbach α to measure agreement or commonality.[16,17] Confidence intervals were generated with the bootstrap method (100 replications).

Criterion validity was assessed using the predictive validity of the candidate criteria with outcomes (primary outcome: in-hospital mortality; secondary outcome: in-hospital mortality or intensive care unit [ICU] length of stay ≥3 days). These outcomes are objective, easily measured across multiple hospitals in US/non-US cohorts, and are more likely to be present in encounters with patients with sepsis than those with uncomplicated infection. To measure predictive validity, a baseline risk model was created for in-hospital mortality based on preinfection criteria using multivariable logistic regression. The baseline model included age (as a fractional polynomial), sex, race/ethnicity (black, white, or other), and the weighted Charlson comorbidity score (as fractional polynomial) as a measure of chronic comorbidities.[18,19] Race/ethnicity was derived from UPMC registration system data using fixed categories consistent with the Centers for Medicare & Medicaid Services EHR

Table 2. Summary of Data Sets

Characteristics	UPMC[a]	KPNC	VA	ALERTS	KCEMS
Years of cohort	2010-2012	2009-2013	2008-2010	2011-2012	2009-2010
No. of hospitals	12	20	130	1	14
Total No. of encounters	1 309 025	1 847 165	1 640 543	38 098	50 727
Data source and study design	Retrospective study of EHRs	Retrospective study of EHRs	Retrospective study of EHRs	Prospective cohort study	Retrospective study of administrative records
Setting	Integrated health system in southwestern Pennsylvania	Integrated health system in northern California	All hospitals in the US VA system	Single university hospital, Jena, Germany	Out-of-hospital records from integrated emergency medical services system in King County, Washington
Definition of suspected infection	Combination of body fluid culture and nonprophylactic antibiotic administration in the EHR[b]	Combination of body fluid culture and nonprophylactic antibiotic administration in the EHR[b]	Combination of body fluid culture and nonprophylactic antibiotic administration in the EHR[b]	CDC criteria for hospital-acquired infections[c]	ICD-9-CM codes for infection, with present-on-admission indicators[d]
No. with suspected infection (% of total)	148 907 (11)	321 380 (17)	377 325 (23)	1186 (3)	6508 (13)
Location at onset of infection, No. (%) infected					
Intensive care unit	15 768 (11)	7031 (2)	73 264 (19)	300 (25)	0
Outside of intensive care unit	133 139 (89)	314 349 (98)	304 061 (81)	886 (75)	6508 (100)
In-hospital mortality, No. (%) infected[e]	6347 (4)	16 092 (5)	22 593 (6)	210 (18)	700 (11)

Abbreviations: KCEMS, King County Emergency Medical Services; KPNC, Kaiser Permanente Northern California; EHR, electronic health record; *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification*; VA, Veterans Administration.

[a] Referred to as the primary cohort, further divided into derivation (n = 74 453) and validation (n = 74 454) cohorts.

[b] See the eAppendix in the Supplement for details about time windows specified between body fluid cultures and antibiotic administration.

[c] Patients were enrolled in ALERTS if the in-hospital stay was longer than 48 hours and in-person prospective screening revealed hospital-acquired infection criteria according to Centers for Disease Control and Prevention (CDC) guidelines.[7]

[d] Required Angus implementation *ICD-9-CM* code for infection accompanied by present-on-admission indicator, as previously validated.[6]

[e] Among UPMC encounters, 28 286 (19%) had in-hospital mortality plus intensive care unit length of stay of 3 days or longer.

meaningful use data set.[20] Race/ethnicity was included in the baseline model because of its described association with the incidence and outcomes of sepsis.[21]

Encounters were then divided into deciles of baseline risk. Within each decile, the rate of in-hospital mortality ± ICU length of stay of 3 days or longer was determined comparing encounters with infection with 2 or more SIRS, SOFA, LODS, and qSOFA points vs encounters with less than 2 criteria of the same score (threshold of 2 points was determined a priori). Model discrimination was assessed with the AUROC for each outcome using the continuous score(s) alone, then added to the baseline risk model. Analyses were separately performed in ICU encounters and non-ICU encounters at the onset of infection. New, simple criteria in external data sets were assessed in both ICU and non-ICU encounters.

Because serum lactate is widely used as a screening tool in sepsis,[22] how its measurement would improve predictive validity of new criteria was assessed in post hoc analyses. Evaluation included qSOFA models that did and did not include serum lactate at thresholds of 2.0, 3.0, and 4.0 mmol/L (18, 27, and 36 mg/dL) and as a continuous variable.[23] Only KPNC data were used for these analyses because an ongoing quality improvement program promoting frequent serum lactate measurement across the health system minimized confounding by indication.[24]

Several sensitivity analyses were performed to assess robustness of the findings. These included a variety of restrictions to the cohort, more rigorous definitions of suspected or presumed infection, alternative ways to measure clinical variables (such as altered mentation in the EHR), and multiple imputation analyses for missing data. There are many possible time windows for criteria around the onset of infection. A variety of windows differing from the primary analysis were tested, including (1) 3 hours before to 3 hours after; (2) 12 hours before to 12 hours after; and (3) restricting to only the 24 hours after the onset of infection. Detailed descriptions are in the Supplement.

All analyses were performed with STATA software, version 11.0 (Stata Corp). All tests of significance used a 2-sided $P \le .05$. We considered AUROCs to be poor at 0.6 to 0.7, adequate at 0.7 to 0.8, good at 0.8 to 0.9, and excellent at 0.9 or higher.[25]

Results

Cohorts and Encounter Characteristics

At 177 hospitals in 5 US and non-US data sets between 2008 and 2013 (**Table 2**), 4 885 558 encounters were studied. In the primary cohort of 1 309 025 records (UPMC derivation and validation; **Figure 1**), 148 907 encounters had suspected infection, most often presenting outside of the ICU (n = 133 139 [89%]). As shown in **Table 3**, first infection was commonly suspected within 48 hours of admission (86%), most often presenting in the emergency department (44%) compared with the ward (33%) or ICU (11%), and mortality was low (4%). The

Figure 1. Accrual of Encounters for Primary Cohort

ED indicates emergency department; ICU, intensive care unit; PACU, postanesthesia care unit.
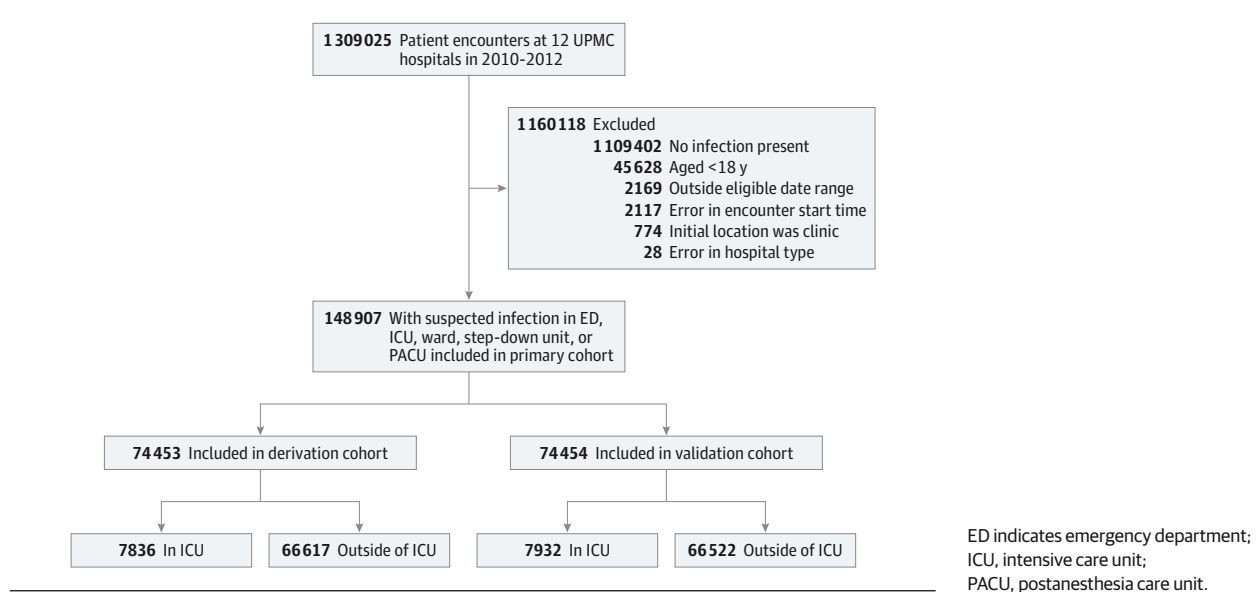
median time from the start of the encounter until the onset of suspected infection (defined as culture or antibiotics order) was 4.2 hours (interquartile range, 1.6-19.2 hours). In KPNC hospitals (eTable 1 in the Supplement), first suspected infections occurred outside the ICU (98%) with similar mortality (5%) and proportion identified within 48 hours of admission (81%). Serum lactate was measured in 57% of suspected infection encounters in KPNC hospitals compared with less than 10% in the other cohorts. In VA hospitals, encounters with suspected infection had similar mortality (6%) but were more likely to be first identified in the ICU (19%). A minority of first infection episodes occurred following surgery, and positive blood cultures were found in 5% to 19% of encounters. In the baseline risk model, using only demographics and comorbidities, there was a 10-fold variation for in-hospital mortality across deciles of baseline risk, ranging from 0.7% to 8% (eFigure 1 in the Supplement).

Frequency of Missing Data Among Clinical and Laboratory Variables

In the UPMC derivation cohort, SIRS criteria and selected laboratory tests in SOFA and LODS were variably measured in the EHR near the onset of infection (eFigure 2 in the Supplement). Tachycardia, tachypnea, and hypotension, although present in less than 50% of encounters, were the most common clinical abnormalities. Encounters in the ICU were more likely to have SIRS and SOFA variables measured and values were more likely to be abnormal. For encounters outside of the ICU, laboratory data were less available, with total bilirubin, ratio of PaO$_2$ to fraction of inspired oxygen, and platelet counts absent in 62%, 74%, and 15% of encounters, respectively.

Performance of Existing Criteria in the ICU in the UPMC Cohort

Among ICU encounters with suspected infection in the UPMC validation cohort (n = 7932 [11%]), most had 2 or more LODS

points (88%), SOFA points (91%), or SIRS criteria (84%) near the time of suspected infection, with mortality rates of 18% for all scores at this threshold (Figure 2 and eFigure 3 in the Supplement). SOFA and LODS had greater statistical agreement with each other (α = 0.87; 95% CI, 0.87-0.88) but lower with SIRS (α = 0.43 [95% CI, 0.41-0.46] for SOFA; α = 0.41 [95% CI, 0.38-0.43] for LODS) (Figure 3). Encounters in the ICU with 2 or more vs less than 2 SIRS criteria were compared within decile of baseline risk and observed a 1- to 2-fold increased rate of hospital mortality compared with a 3- to 11-fold increase in mortality comparing those with 2 or more vs less than 2 SOFA points (Figure 4). The fold change in the LODS score was even greater than that for SOFA.

In the ICU, the predictive validity for hospital mortality using SOFA (AUROC = 0.74; 95% CI, 0.73-0.76) and LODS (AUROC = 0.75; 95% CI, 0.73-0.76; P = .20) were not statistically different but were statistically greater than that of SIRS (AUROC = 0.64; 95% CI, 0.62-0.66; P < .001 for either LODS or SOFA vs SIRS) (Figure 3 and eFigure 4 and eTable 2 in the Supplement). Results for a change in SOFA of 2 points or more were significantly greater compared with SIRS (AUROC = 0.70; 95% CI, 0.68-0.71; P < .001 vs SIRS criteria). The SOFA score was 2 or more in 98% of decedents (95% CI, 97%-99%); among survivors, the SOFA score was less than 2 in 10% (95% CI, 10%-11%). These proportions were similar for a LODS threshold of 2 or 3 (eTable 3 in the Supplement). Among decedents, 2 or more SIRS criteria were present in 91% (95% CI, 89%-92%). Results were consistent for the combined outcome (eFigures 5 and 6 in the Supplement).

Performance of Existing Criteria Outside the ICU in the UPMC Cohort

For encounters with suspected infection outside of the ICU (n = 66 522 [89% of cohort]), 20 130 (30%) had no SIRS criteria, 27 560 (41%) had no SOFA points, and 29 789 (45%) had no LODS points (Figure 2). Agreement followed a pattern simi-

Table 3. Characteristics of Encounters With Suspected Infection in the Primary Cohort at 12 UPMC Hospitals From 2010 to 2012 (N = 148 907)[a]

Variables	All Encounters	Derivation Cohort		Validation Cohort	
		ICU Encounters	Encounters Outside of ICU	ICU Encounters	Encounters Outside of ICU
Total encounters with suspected infection, No.	148 907	7836	66 617	7932	66 522
Infection type, No. (%)[b]					
Presumed	112 850 (76)	7282 (93)	49 287 (74)	7351 (93)	48 930 (74)
Confirmed bacteremia	6875 (5)	646 (8)	2780 (4)	652 (8)	2797 (4)
Age, mean (SD), y	61 (19)	62 (17)	61 (20)	62 (17)	60 (20)
Male, No. (%)	63 311 (43)	4192 (54)	27 418 (41)	4255 (54)	27 446 (41)
Race/ethnicity, No. (%)					
White	113 029 (76)	5774 (74)	50 843 (76)	5881 (74)	50 531 (76)
Black	20 892 (14)	808 (10)	9552 (14)	777 (10)	9755 (15)
Other	14 986 (10)	1254 (16)	6222 (9)	1274 (16)	6236 (9)
Weighted Charlson comorbidity index, median (IQR)	1 (0-2)	1 (0-2)	1 (0-2)	1 (0-2)	1 (0-2)
Surgery prior to infection suspected, No. (%)	17 327 (12)	2153 (27)	6517 (10)	2171 (27)	6486 (10)
Onset of infection within 48 h of admission, No. (%)	128 358 (86)	6022 (77)	58 187 (87)	5993 (76)	58 156 (87)
Unit location at time infection suspected, No. (%)					
Emergency department	65 934 (44)		32 902 (50)		33 032 (50)
Ward	49 354 (33)		24 787 (37)		24 567 (37)
ICU	15 768 (11)	7836 (100)		7932 (100)	
Postacute care unit or procedure unit	1965 (1)		960 (1)		1005 (2)
Step-down unit	15 662 (11)		7855 (12)		7807 (12)
Other or missing data	224 (<1)		113 (<1)		111 (<1)
SIRS near onset of suspected infection[c]					
Mean (SD)	1.3 (1.1)	2.5 (1.0)	1.2 (1.1)	2.5 (1.0)	1.2 (1.0)
Median (IQR)	1 (0-2)	3 (2-3)	1 (0-2)	3 (2-3)	1 (0-2)
SOFA near onset of suspected infection[d]					
Mean (SD)	2.0 (2.7)	6.3 (4.0)	1.4 (1.9)	6.2 (3.9)	1.4 (2.0)
Median (IQR)	1 (0-3)	6 (3-9)	1 (0-2)	6 (3-9)	1 (0-2)
LODS near onset of suspected infection[e]					
Mean (SD)	2.0 (2.8)	6.3 (3.9)	1.5 (2.1)	6.3 (3.8)	1.5 (2.1)
Median (IQR)	1 (0-3)	6 (4-9)	1 (0-3)	6 (3-9)	1 (0-3)
Serum lactate measured on day of infection, No. (%)	13 492 (9)	3187 (41)	3611 (5)	3067 (39)	3627 (5)
Serum lactate ≥2.0 mmol/L, No. (%)	6177 (4)	1643 (21)	1444 (2)	1555 (20)	1535 (2)
ICU admission, No. (%)	37 528 (25)	7836 (100)	10 935 (16)	7932 (100)	10 825 (16)
Hospital length of stay, median (IQR), d	6 (3-10)	12 (7-20)	6 (3-9)	12 (7-19)	6 (3-9)
Hospital mortality, No. (%)	6347 (4)	1298 (17)	1874 (3)	1289 (16)	1886 (3)

SI conversion: To convert serum lactate to milligrams per deciliter, divide by 0.111.

Abbreviations: ICU, intensive care unit; IQR, interquartile range; LODS, Logistic Organ Dysfunction System; SIRS, systemic inflammatory response syndrome; SOFA, Sequential [Sepsis-related] Organ Function Assessment.

[a] Data derived from electronic health records.

[b] Presumed infection is a subset of suspected infection in which encounters received 2 or more doses of an antibiotic within 96 hours of onset of infection. Confirmed bacteremia is a subset among which blood cultures were positive during the encounter.

[c] SIRS criteria range from 0 to 4, wherein 1 point is given for perturbations of the following variables: respiratory rate, white blood cell count/bands, heart rate,

and temperature (see Table 1).[29] Maximum score is determined from 48 hours before to 24 hours after onset of infection.

[d] The SOFA score ranges from 0 to 24, where 0 to 4 points are assigned for 1 of 6 organ dysfunctions: hematologic, hepatic, respiratory, neurologic, cardiac, and renal.[8] A greater score corresponds to greater severity. Maximum score is determined from 48 hours before to 24 hours after onset of infection.

[e] The LODS score, modified for available data, ranges from 0 to 22 points, wherein points are assigned with increasing severity to hematologic, hepatic, pulmonary, neurologic, cardiovascular, and renal dysfunction.[9] Maximum score is determined from 48 hours before to 24 hours after onset of infection.

lar to that in the ICU encounters but with generally smaller Cronbach α statistics (Figure 3). Over deciles of baseline risk (Figure 4), encounters with 2 or more vs less than 2 SIRS cri-

teria had a 2- to 7-fold increase in the rate of in-hospital mortality compared with up to an 80-fold change for 2 or more vs less than 2 SOFA points.

Figure 2. Distribution of Patient Encounters Over SIRS Criteria and SOFA, LODS, and qSOFA Scores Among ICU Patients and Non-ICU Patients With Suspected Infection in the UPMC Validation Cohort (N = 74 454)
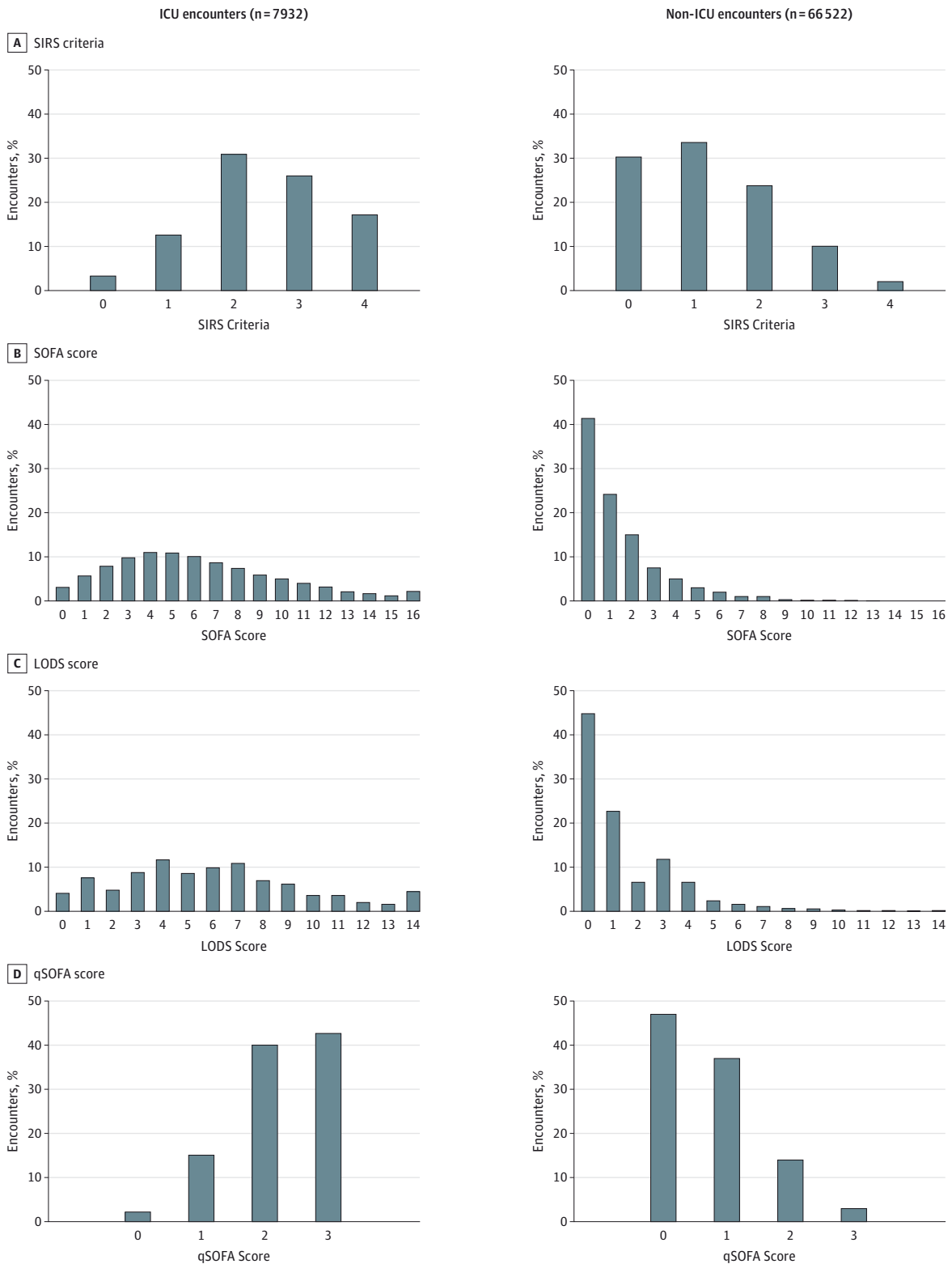
ICU indicates intensive care unit; LODS, Logistic Organ Dysfunction System; qSOFA, quick Sequential [Sepsis-related] Organ Function Assessment; SIRS, systemic inflammatory response syndrome; SOFA, Sequential [Sepsis-related]

Organ Function Assessment. The x-axis is the score range, with LODS truncated at 14 points (of 22 points) and SOFA truncated at 16 points (of 24 points) for illustration.

Figure 3. Area Under the Receiver Operating Characteristic Curve and 95% Confidence Intervals for In-Hospital Mortality of Candidate Criteria (SIRS, SOFA, LODS, and qSOFA) Among Suspected Infection Encounters in the UPMC Validation Cohort (N = 74 454)

A ICU encounters (n = 7932)

	SIRS	SOFA	LODS	qSOFA
SIRS	0.64 (0.62-0.66)	0.43 (0.41-0.46)	0.41 (0.38-0.43)	0.46 (0.43-0.48)
SOFA	<.001	0.74 (0.73-0.76)	0.87 (0.87-0.88)	0.65 (0.63-0.66)
LODS	<.001	0.20	0.75 (0.73-0.76)	0.76 (0.75-0.77)
qSOFA	.01	<.001	<.001	0.66 (0.64-0.68)

B Non-ICU encounters (n = 66 522)

	SIRS	SOFA	LODS	qSOFA
SIRS	0.76 (0.75-0.77)	0.52 (0.51-0.53)	0.43 (0.42-0.44)	0.61 (0.61-0.62)
SOFA	<.001	0.79 (0.78-0.80)	0.80 (0.80-0.81)	0.59 (0.58-0.60)
LODS	<.001	<.001	0.81 (0.80-0.82)	0.68 (0.68-0.69)
qSOFA	<.001	<.001	.72	0.81 (0.80-0.82)

ICU indicates intensive care unit; LODS, Logistic Organ Dysfunction System; qSOFA, quick Sequential [Sepsis-related] Organ Function Assessment; SIRS, systemic inflammatory response syndrome; SOFA, Sequential [Sepsis-related] Organ Function Assessment. The area under the receiver operating characteristic curve (AUROC) data in the blue-shaded diagonal cells derive from models that include baseline variables plus candidate criteria. For comparison, the AUROC of the baseline model alone is 0.58 (95% CI, 0.57-0.60) in the ICU and 0.69 (95% CI, 0.68-0.70) outside of the ICU. Below the AUROC data cells are P values for comparisons between criteria, while above the AUROC data cells are Cronbach α data (with bootstrap 95% confidence intervals), a measure of agreement.

The discrimination of hospital mortality using SOFA (AUROC = 0.79; 95% CI, 0.78-0.80), LODS (AUROC = 0.82; 95% CI, 0.81-0.83), or change in SOFA (AUROC = 0.79; 95% CI, 0.78-0.79) scores was significantly greater compared with SIRS criteria (AUROC = 0.76; 95% CI, 0.75-0.77; P < .01 for all) (Figure 3 and eFigure 4 and eTable 2 in the Supplement). Sixty-eight percent (95% CI, 66%-70%) of decedents had 2 or more SOFA points and 67% (95% CI, 66%-67%) of survivors had less than 2 SOFA points. In comparison, only 55% (95% CI, 53%-57%) of decedents had 2 or more SIRS criteria, whereas 81% of survivors had less than 2 SIRS criteria (95% CI, 81%-82%) (eTable 3 in the Supplement). Results were consistent for the combined outcome (eFigures 5 and 6 in the Supplement).

Performance of New, Simple Criteria
The final qSOFA model included Glasgow Coma Scale (GCS) score of 13 or less, systolic blood pressure of 100 mm Hg or less, and respiratory rate of 22/min or more (1 point each; score range, 0-3) (Table 4). Most encounters with infection (73%-90%) had less than 2 qSOFA points, and mortality ranged from 1% to 24% over the score range (eFigure 7 in the Supplement). Calibration plots showed similar observed vs expected proportion of deaths across qSOFA scores (eFigure 8 in the Supplement). The qSOFA agreed reasonably well with both SOFA (α = 0.73; 95% CI, 0.73-0.74) and LODS (α = 0.79; 95% CI, 0.78-0.79) and, unlike SOFA and LODS, also agreed more with SIRS (α = 0.69; 95% CI, 0.68-0.69) (Figure 3). The 24% of encounters with infection with 2 or 3 qSOFA points accounted for 70% of deaths, 70% of deaths or ICU stays of 3 days or longer.

In the ICU, the predictive validity for hospital mortality of qSOFA above baseline risk (AUROC = 0.66; 95% CI, 0.64-0.68) was statistically greater than SIRS criteria (P = .01) but significantly less than SOFA (P < .001) (Figure 3 and eFigure 4 and eTable 2 in the Supplement). Outside of the ICU, there was a 3- to 14-fold increase in the rate of hospital mortality across the entire range of baseline risk comparing those with 2 or more vs less than 2 qSOFA points (Figure 4). The predictive validity of qSOFA was good for in-hospital mortality (AUROC = 0.81; 95% CI, 0.80-0.82), was not statistically different from LODS (P = .77) and was statistically greater than SOFA or change in SOFA score (P < .001 for both) (Figure 3, Figure 4, and eFigure 4 and eTable 2 in the Supplement). Seventy percent (95% CI, 69%-72%) of decedents had 2 or more qSOFA points and 78% (95% CI, 78%-79%) of survivors had less than 2 qSOFA points (eTable 3 in the Supplement). Results were consistent for the combined outcome (eFigures 5 and 6 in the Supplement).

Among encounters with 2 or more qSOFA points, 75% also had 2 or more SOFA points (eFigure 9 in the Supplement). This proportion was greater among decedents (89%) and ICU encounters (94%) and increased as the time window for evaluation was extended to 48 hours (90%) and 72 hours (92%) after the onset of infection.

External Data Sets
The qSOFA was tested in 4 external data sets comprising 706 399 patient encounters at 165 hospitals in out-of-hospital (n = 6508), non-ICU (n = 619 137), and ICU (n = 80 595)

Figure 4. Fold Change in Rate of In-Hospital Mortality (Log Scale) Comparing Encounters With ≥2 vs <2 Criteria for Each Decile of Baseline Risk in the UPMC Validation Cohort (N = 74 454)
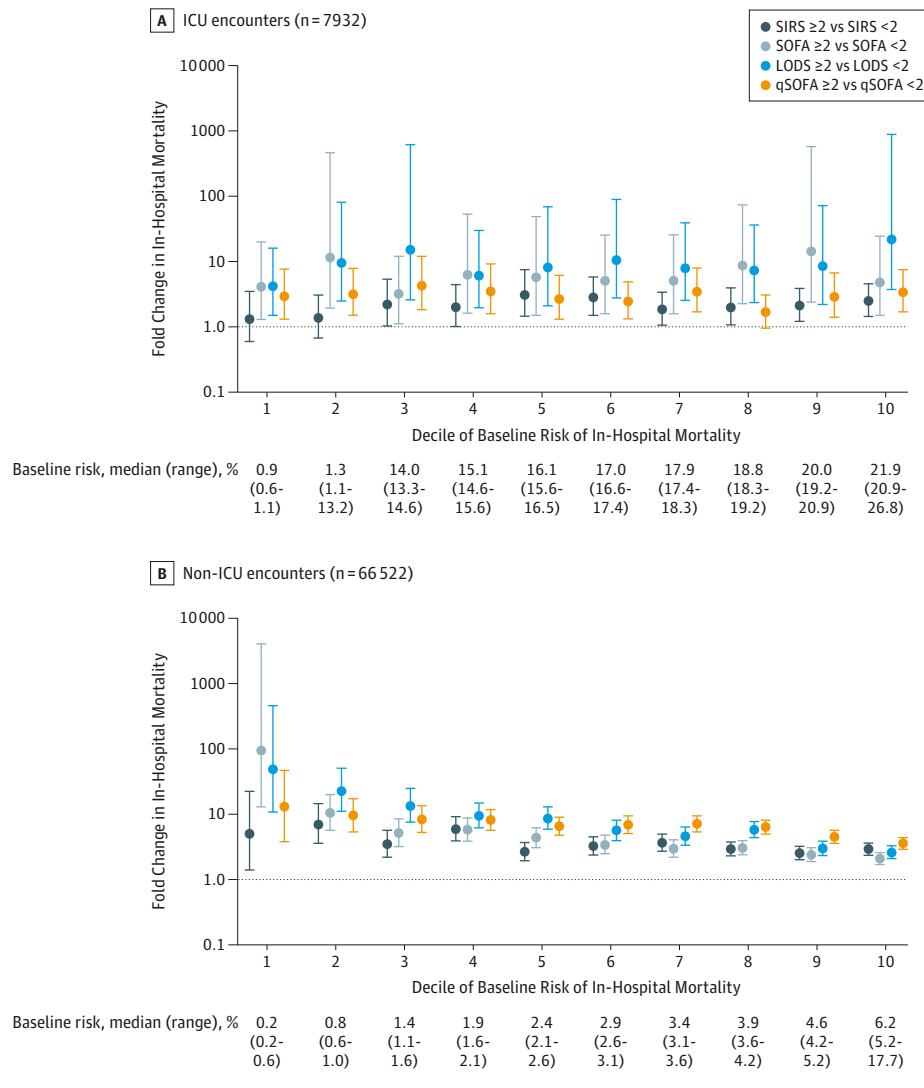
ICU indicates intensive care unit; LODS, Logistic Organ Dysfunction System; qSOFA, quick Sequential [Sepsis-related] Organ Function Assessment; SIRS, systemic inflammatory response syndrome; SOFA, Sequential [Sepsis-related] Organ Function Assessment. Panel A shows ICU encounters comparing fold change for SIRS, SOFA, LODS, and qSOFA. Panel B shows non-ICU encounters. Medians and ranges of baseline risk of in-hospital mortality within decile shown are below the x-axis.

Interpretive example: The x-axis divides the cohort into deciles of baseline risk, determined by age, sex, comorbidities, and race/ethnicity. For a young woman with no comorbidities (panel A, decile 2) admitted to the ICU with pneumonia, her chance of dying in the hospital is 10-fold greater if she has 3 SOFA points compared with 1 SOFA point. On the other hand, she has only a small increase in the chance of dying if she has 3 SIRS criteria compared with 1 SIRS criterion. For an older woman with chronic obstructive pulmonary disease admitted to the ward with pneumonia (panel B, decile 6), her chance of dying in the hospital is 7-fold higher if she has 3 qSOFA points compared with 1 qSOFA point. On the other hand, she has only a 3-fold increase in odds of dying if she has 3 SIRS criteria compared with 1 SIRS criterion.

settings (eTable 1 in the Supplement). Among encounters with community infection (KCEMS) or hospital-acquired infection (ALERTS), qSOFA had consistent predictive validity (AUROC = 0.71 and 0.75, respectively) (Table 5 and eFigure 4 in the Supplement). Results were similar in the VA data set (AUROC = 0.78), in which no GCS data were available.

Serum Lactate

During model building in UPMC data, serum lactate did not meet prespecified statistical thresholds for inclusion in qSOFA. In KPNC data, the post hoc addition of serum lactate levels of 2.0 mmol/L (18 mg/dL) or more to qSOFA (revised to a 4-point score with 1 added point for elevated serum lactate level) statistically changed the predictive validity of qSOFA (AUROC with lactate = 0.80; 95% CI, 0.79-0.81 vs AUROC without lactate = 0.79; 95% CI, 0.78-0.80; P < .001) (eFigure 10A in the Supplement). As shown in eTable 4 in the Supplement, this was consistent for higher thresholds of lactate (3.0 mmol/L

[27 mg/dL], 4.0 mmol/L [36 mg/dL]) or using a continuous distribution (P < .001). However, the clinical relevance was small as the rates of in-hospital mortality comparing encounters with 2 or more vs less than 2 points across deciles of risk were numerically similar whether or not serum lactate was included in qSOFA (eFigure 10B in the Supplement).

Among encounters with 1 qSOFA point but also a serum lactate level of 2.0 mmol/L or more, in-hospital mortality was higher than that for encounters with serum lactate levels of less than 2.0 mmol/L across the range of baseline risk. The rate of in-hospital mortality was numerically similar to that for encounters with 2 qSOFA points using the model without serum lactate (eFigure 11 in the Supplement). Because serum lactate levels are widely used for screening at many centers, the distribution of qSOFA scores over strata of serum lactate level was investigated. The qSOFA consistently identified higher-risk encounters even at varying serum lactate levels (eFigure 12 in the Supplement).

Table 4. Odds Ratios for Baseline Model and qSOFA Variables for In-Hospital Mortality in the UPMC Derivation Cohort (N = 74 453)

	Total No. With Categorical Variable	Deaths, No. (% of Total)	In-Hospital Mortality, Adjusted Odds Ratio (95% CI)
Baseline risk model[a]			
Age, y[b]			1.03 (1.03-1.03)
Charlson comorbidity index[b]			1.13 (1.11-1.15)
Race/ethnicity			
White	56 617	2470 (4)	1 [Reference]
Black	10 360	319 (3)	0.89 (0.79-1.01)
Other	7476	383 (5)	1.37 (1.22-1.53)
Male			
No	42 843	1467 (3)	1 [Reference]
Yes	31 610	1705 (5)	1.56 (1.45-1.68)
qSOFA model[c]			
Respiratory rate, /min			
<22	45 398	676 (1)	1 [Reference]
≥22	29 055	2496 (9)	3.18 (2.89-3.50)
Systolic blood pressure, mm Hg			
>100	44 669	789 (2)	1 [Reference]
≤100	29 784	2383 (8)	2.61 (2.40-2.85)
Altered mental status, Glasgow Coma Scale score			
14-15	66 879	1677 (3)	1 [Reference]
≤13	7574	1495 (20)	4.31 (3.96-4.69)

Abbreviations: qSOFA, quick Sequential [Sepsis-related] Organ Failure Assessment; UPMC, University of Pittsburgh School of Medicine.

[a] Fully parameterized using fractional polynomials in final analyses.

[b] Odds ratios correspond to a comparison between encounters separated by 1 unit change in age or Charlson comorbidity index score.

[c] Multivariable logistic regression model of qSOFA variables illustrates their association with in-hospital mortality. The odds ratios compare groups of encounters with vs without the specified criteria.

Table 5. AUROCs for In-Hospital Mortality for qSOFA in External Data Sets

Data Set and Infection Type	No. of Patients With Suspected Infection	AUROC (95% CI)	
		Baseline Model	Baseline Model + qSOFA
KPNC (all suspected infections)	321 380	0.67 (0.67-0.67)	0.78 (0.78-0.78)
ICU patients	7031	0.64 (0.62-0.66)	0.72 (0.70-0.73)
Non-ICU patients	314 349	0.68 (0.67-0.68)	0.78 (0.78-0.79)
VA (all suspected infections)[a]	377 325	0.73 (0.73-0.74)	0.78 (0.78-0.79)
ALERTS (hospital-acquired infections)	1186	0.55 (0.51-0.60)	0.73 (0.69-0.77)
KCEMS (community-acquired infections)	6508	0.59 (0.57-0.62)	0.71 (0.69-0.73)

Abbreviations: AUROC, area under the receiver operating characteristic curve; ICU, intensive care unit; KCEMS, King County Emergency Medical Services; KPNC, Kaiser Permanente Northern California; qSOFA, quick Sequential [Sepsis-related] Organ Function Assessment; VA, Veterans Administration.

[a] The VA data did not include Glasgow Coma Scale scores; the qSOFA is a modified 2-variable model (systolic blood pressure and respiratory rate only), with a range from 0 to 2 points.

Time Windows for Measuring qSOFA Variables

When qSOFA variables were measured in the time window from 3 hours before/after or 12 hours before/after the onset of infection in KPNC data (eTable 4 in the Supplement), results were not significantly different from the original model (P = .13 for 3 hours and P = .74 for 12 hours). When qSOFA variables were restricted to only the 24-hour period after the onset of infection, the predictive validity for in-hospital mortality was significantly greater (AUROC = 0.83; 95% CI, 0.83-0.84; P < .001) compared with the primary model.

Additional sensitivity analyses are shown in eTable 4 in the Supplement. The predictive validity of qSOFA was not significantly different when using more simple measures, such as any altered mentation (GCS score <15 [P = .56] compared with the model with GCS score ≤13). The predictive validity was also not

significantly different when performed after multiple imputation for missing data and in a variety of a priori subgroups.

Discussion

The Third International Consensus Definitions Task Force defined sepsis as a "life-threatening organ dysfunction due to a dysregulated host response to infection."[5] In the absence of a gold-standard test for sepsis, several domains of validity and usefulness were used to assess potential clinical criteria to operationalize this definition. Among encounters with suspected infection in the ICU (Figure 3), SOFA and LODS had statistically greater predictive validity compared with SIRS criteria. Outside of the ICU, a simple model (qSOFA) of altered menta-

tion, low systolic blood pressure, and elevated respiratory rate had statistically greater predictive validity than the SOFA score (Figure 3). The predictive validity of qSOFA was robust to evaluation under varied measurement conditions, in academic and community hospitals, in international locations of care, for community and hospital-acquired infections, and after multiple imputation for missing data. It was, however, statistically inferior compared with SOFA for encounters in the ICU and has a statistically lower content validity as a measure of multiorgan dysfunction. Thus, the task force recommended use of a SOFA score of 2 points or more in encounters with infection as criteria for sepsis and use of qSOFA in non-ICU settings to consider the possibility of sepsis.

Criteria Outside of the ICU

For infected patients outside of the ICU, there is an increasing focus on early recognition of sepsis. Potential criteria for organ dysfunction like SOFA or LODS required clinical and laboratory variables that may be missing and difficult to obtain in a timely manner. These characteristics may increase measurement burden for clinicians. In comparison, a simple model (qSOFA) uses 3 clinical variables, has no laboratory tests, and has a predictive validity outside of the ICU that is statistically greater than the SOFA score ($P < .001$). The qSOFA and SOFA scores also had acceptable agreement in the majority of encounters.

However, 3 potentially controversial issues are worth noting. First, qSOFA was derived and tested among patient encounters in which infection was already suspected. The qSOFA is not an alert that alone will differentiate patients with infection from those without infection. However, at least in many US and European hospital settings, infection is usually suspected promptly, as evidenced by rapid initiation of antibiotics.[26,27]

Second, mental status is assessed variably in different settings, which may affect the performance of the qSOFA. Although the qSOFA appeared robust in sensitivity analyses to alternative GCS cut points, further work is needed to clarify its clinical usefulness. In particular, the model evaluated only whether mental status was abnormal, not whether it had changed from baseline, which is extremely difficult to operationalize and validate, both in the EHR and as part of routine charting. An alternative to the GCS (eg, Laboratory and Acute Physiology Score, version 2, in KPNC encounters)[28] found similar results.

Third, serum lactate levels, which have been proposed as a screening tool for sepsis or septic shock, were not retained in the qSOFA during model construction. One reason may be because serum lactate levels were not measured commonly in the UPMC data set. When serum lactate levels were added to qSOFA post hoc in the KPNC health system data set, in which measurement of lactate levels was common, the predictive validity was statistically increased but with little difference in how encounters were classified. This analysis assessed only how serum lactate levels at different thresholds contributed above and beyond the qSOFA model. However, among intermediate-risk encounters (qSOFA score = 1), the addition of a serum lactate level of 2.0 mmol/L (18 mg/dL) or higher identified those with a risk profile similar to those with 2 qSOFA points. Thus, areas for further inquiry include whether serum lactate lev-

els could be used for patients with borderline qSOFA values or as a substitute for individual qSOFA variables (particularly mental status, given the inherent problems discussed above), especially in health systems in which lactate levels are reliably measured at low cost and in a timely manner.

Criteria in the ICU

Among ICU encounters, the diagnosis of sepsis may be challenging because of preexisting organ dysfunction, treatment prior to admission, and concurrent organ support. In this study, as others have reported in a distinct geographic region and health care system,[29] traditional tools such as the SIRS criteria have poor predictive validity among patients who are infected. Yet in our study, SOFA and LODS scores had superior predictive validity in the ICU and greater agreement, perhaps because more variables were likely to be measured, abnormal, and independent of ongoing interventions. These results are consistent with prior studies of SOFA and LODS in the ICU.[30,31] On average, only 2 of 100 infected decedents in the ICU had a SOFA or LODS score of less than 2. The qSOFA score had statistically worse predictive validity in the ICU, likely related to the confounding effects of ongoing organ support (eg, mechanical ventilation, vasopressors).

Advances Using EHRs

The data from these analyses provided the Third International Consensus Task Force with evidence about clinical criteria for sepsis using EHRs from 3 large health systems with both academic and community hospitals. More than 60% of US nonfederal, acute care hospitals (and all US federal hospitals) now use advanced EHRs. Adoption of EHRs has increased 8-fold since 2009 in the United States and will continue to increase.[32] The EHR may present hospitals with an opportunity to rapidly validate criteria for patients likely to have sepsis, to test prompts or alerts among infected patients with specific EHR signatures suggestive of sepsis, and to build platforms for automated surveillance.[33] In addition, criteria such as in the qSOFA can be measured quickly and easily and assessed repeatedly over time in patients at risk of sepsis, perhaps even in developing countries without EHRs.

Limitations

This investigation has several limitations. First, we studied only patients in whom infection was already suspected or documented. We did not address how to diagnose infection among those in whom life-threatening organ dysfunction was the initial presentation. Therefore, these data alone do not mandate that hospitalized patients with SOFA or qSOFA points be evaluated for the presence of infection.

Second, we chose to develop simple criteria that clinicians could quickly use at the bedside, balancing timeliness and content validity with greater criterion validity. We acknowledge that predictive validity would be improved with more complex models that include interaction terms or serial measurements over time.[3,34,35] We tested how the change in SOFA score over time would perform, and although similar to the maximum SOFA score, the optimal time windows over which change should be measured are not known.

Third, no organ dysfunction measurements evaluated in this study distinguish between chronic and acute organ dysfunction, assess whether the organ dysfunction has an explanation other than infection, or attribute dysfunction specifically to a dysregulated host response. For example, a patient with dementia with an abnormal GCS score at baseline will always have 1 qSOFA point but may not be as likely to have sepsis as a patient with a normal baseline sensorium. As such, we illustrated the predictive validity of various criteria across a full range of underlying risk determined from comorbidity and demographics.

Fourth, we chose 2 outcomes associated more commonly with sepsis than with uncomplicated infection. These outcomes have high content validity and were generalizable across data sets, but there are certainly alternative choices.[36]

Fifth, we compared predictive validity with tests of inference that may be sensitive to sample size. We found that statistically significant differences in AUROC were often present, yet these resulted in differences in classification with debatable clinical relevance. We reconciled these data by reporting the fold change in outcome comparing encounters of different scores to provide more clinical context.

Sixth, the acute, life-threatening organ dysfunction in sepsis may also occur at different times in different patients (before, during, or after infection is recognized).[37] Results were unchanged over a variety of time windows, including both long (72-hour) and short (6-hour) windows around the onset of infection. Prospective validation in other cohorts, assessment in low- to middle-income countries, repeated measurement, and the contribution of individual qSOFA elements to predictive validity are important future directions.

Conclusions

Among ICU encounters with suspected infection, the predictive validity for in-hospital mortality of SOFA was not significantly different than the more complex LODS but was statistically greater than SIRS and qSOFA, supporting its use in clinical criteria for sepsis. Among encounters with suspected infection outside of the ICU, the predictive validity for in-hospital mortality of qSOFA was statistically greater than SOFA and SIRS, supporting its use as a prompt to consider possible sepsis.

Georgia; Tom van der Poll, MD, Academisch Medisch Centrum, Amsterdam, the Netherlands; Richard S. Hotchkiss, MD, Washington University School of Medicine, St Louis, Missouri; Jean-Louis Vincent, MD, PhD, Université Libre de Bruxelles, and Department of Intensive Care, Erasme University Hospital, Brussels, Belgium; and Mitchell M. Levy, MD, Division of Pulmonary and Critical Care Medicine, Brown University School of Medicine, Providence, Rhode Island. These contributions were provided without compensation.

REFERENCES

1. Liu V, Escobar GJ, Greene JD, et al. Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA*. 2014;312(1):90-92.

2. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med*. 2001;29(7):1303-1310.

3. Levy MM, Fink MP, Marshall JC, et al; SCCM/ESICM/ACCP/ATS/SIS. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med*. 2003;31(4): 1250-1256.

4. Bone RC, Balk RA, Cerra FB, et al; ACCP/SCCM Consensus Conference Committee; American College of Chest Physicians/Society of Critical Care Medicine. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest*. 1992;101(6):1644-1655.

5. Shankar-Hari M, Phillips GS, Levy ML, et al; Sepsis Definitions Task Force. Developing a new definition and assessing new clinical criteria for septic shock: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. doi:10.1001/jama.2016.0289.

6. Seymour CW, Rea TD, Kahn JM, Walkey AJ, Yealy DM, Angus DC. Severe sepsis in pre-hospital emergency care: analysis of incidence, care, and outcome. *Am J Respir Crit Care Med*. 2012;186(12): 1264-1271.

7. Hagel S, Ludewig K, Frosinski J, et al. Effectiveness of a hospital-wide educational programme for infection control to reduce the rate of health-care associated infections and related sepsis (ALERTS)—methods and interim results [in German]. *Dtsch Med Wochenschr*. 2013; 138(34-35):1717-1722.

8. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-Related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med*. 1996;22(7):707-710.

9. Le Gall JR, Klar J, Lemeshow S, et al; ICU Scoring Group. The Logistic Organ Dysfunction System: a new way to assess organ dysfunction in the intensive care unit. *JAMA*. 1996;276(10):802-810.

10. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.

11. Rangel-Frausto MS, Pittet D, Costigan M, Hwang T, Davis CS, Wenzel RP. The natural history of the

systemic inflammatory response syndrome (SIRS): a prospective study. *JAMA*. 1995;273(2):117-123.

12. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163(7):670-675.

13. Seymour CW, Kahn JM, Cooke CR, Watkins TR, Heckbert SR, Rea TD. Prediction of critical illness during out-of-hospital emergency care. *JAMA*. 2010;304(7):747-754.

14. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE—Acute Physiology and Chronic Health Evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9(8):591-597.

15. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med*. 2007;35(9):2052-2056.

16. Bland JM, Altman DG. Cronbach's α. *BMJ*. 1997; 314(7080):572.

17. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959;56(2):81-105.

18. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in *ICD-9-CM* and *ICD-10* administrative data. *Med Care*. 2005;43 (11):1130-1139.

19. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology—with an emphasis on fractional polynomials. *Methods Inf Med*. 2005;44 (4):561-571.

20. Centers for Medicaid and Medicare Services. *Eligible Professional Meaningful Use Core Measures: Measure 7 of 13*. 2014. https://www.cms.gov /Regulations-and-Guidance/Legislation /EHRIncentivePrograms/downloads/7_Record _Demographics.pdf. Accessed January 1, 2016.

21. Barnato AE, Alexander SL, Linde-Zwirble WT, Angus DC. Racial variation in the incidence, care, and outcomes of severe sepsis: analysis of population, patient, and hospital characteristics. *Am J Respir Crit Care Med*. 2008;177(3):279-284.

22. Dellinger RP, Levy MM, Rhodes A, et al; Surviving Sepsis Campaign Guidelines Committee Including the Pediatric Subgroup. Surviving Sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*. 2013;41(2):580-637.

23. Cecconi M, De Backer D, Antonelli M, et al. Consensus on circulatory shock and hemodynamic monitoring: task force of the European Society of Intensive Care Medicine. *Intensive Care Med*. 2014; 40(12):1795-1815.

24. Liu V, Morehouse JW, Soule J, Whippy A, Escobar GJ. Fluid volume, lactate values, and mortality in sepsis patients with intermediate lactate values. *Ann Am Thorac Soc*. 2013;10(5):466-473.

25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36.

26. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med*. 2014;42(8): 1749-1755.

27. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med*. 2004;164(6):637-644.

28. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care*. 2013;51(5): 446-453.

29. 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(17):1629-1638.

30. Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. *Crit Care Med*. 1998;26(11):1793-1800.

31. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*. 2001;286(14):1754-1758.

32. Charles D, Gabriel M, Furukawa MF. *Adoption of Electronic Health Record Systems among US Non-Federal Acute Care Hospitals: 2008-2013*. Washington, DC: Office of the National Coordinator for Health Information Technology; 2014. ONC Data Brief 16.

33. Rhee C, Murphy MV, Li L, Platt R, Klompas M; Centers for Disease Control and Prevention Epicenters Program. Comparison of trends in sepsis incidence and coding using administrative claims vs objective clinical data. *Clin Infect Dis*. 2015;60(1): 88-95.

34. Shapiro NI, Wolfe RE, Moore RB, Smith E, Burdick E, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. *Crit Care Med*. 2003;31(3):670-675.

35. Opal SM. Concept of PIRO as a new conceptual framework to understand sepsis. *Pediatr Crit Care Med*. 2005;6(3)(suppl):S55-S60.

36. Iwashyna TJ, Odden A, Rohde J, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the Angus implementation of the International Consensus Conference definition of severe sepsis. *Med Care*. 2014;52(6):e39-e43.

37. Kellum JA, Kong L, Fink MP, et al; GenIMS Investigators. Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) study. *Arch Intern Med*. 2007;167(15): 1655-1663.