

OPEN

Accuracy of Quick Sequential Organ Failure Assessment Score to Predict Sepsis Mortality in 121 Studies Including 1,716,017 Individuals: A Systematic Review and Meta-Analysis

Velma Herwanto, MD^{1,2}; Amith Shetty, MD³; Marek Nalos, MD¹; Mandira Chakraborty, MD¹; Anthony McLean, MD¹; Guy D. Eslick, PhD⁴; Benjamin Tang, MD^{1,2,4}

Objectives: We performed a meta-analysis to assess whether the newly introduced quick Sequential Organ Failure Assessment score could predict sepsis outcomes and compared its performance to systematic inflammatory response syndrome, the previously widely used screening criteria for sepsis.

Data Sources: We searched multiple electronic databases including MEDLINE, the Cochrane Library, Embase, Web of Science,

¹Department of Intensive Care Medicine, Nepean Hospital, Sydney, NSW, Australia.

²Centre for Immunology and Allergy Research, the Westmead Institute for Medical Research, Sydney, NSW, Australia.

³Department of Emergency Medicine, Westmead Hospital, Sydney, NSW, Australia.

⁴Nepean Clinical School, University of Sydney, Sydney, NSW, Australia.

Drs. Herwanto, Eslick, and Tang designed the study protocol. Drs. Herwanto and Shetty collected the published articles and independently extracted the data. Dr. Herwanto performed the analyses with assistance from Dr. Eslick. She had full access to all the data in the study and had final responsibility for the decision to submit for publication. Dr. Tang conceived the study and wrote the article. All co-authors participated in the critical review of the article and approved the final draft.

The authors have disclosed that they do not have any potential conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

For information regarding this article, E-mail: vher6289@uni.sydney.edu.au

Copyright © 2019 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Crit Care Expl 2019; 1:e0043

DOI: 10.1097/CCE.0000000000000043

and Google Scholar (up to March 1, 2019) that evaluated quick Sequential Organ Failure Assessment score, systemic inflammatory response syndrome, or both (International Prospective Register of Systematic Reviews [PROSPERO]: CRD42018103327).

Study Selection: Studies were included if the outcome was mortality, organ dysfunction, admission to ICU, ventilatory support, or prolonged ICU stay and if prediction performance was reported as either area under the curve, odds ratio, sensitivity, or specificity.

Data Extraction: The criterion validity of the quick Sequential Organ Failure Assessment score and systemic inflammatory response syndrome criteria were assessed by measuring its predictive validity for primary (mortality) and secondary outcomes in pooled metrics as mentioned. The data were analyzed using random effects model, and heterogeneity was explored using prespecified subgroup analyses.

Data Synthesis: We screened 1,340 studies, of which 121 studies (including data for 1,716,017 individuals) were analyzed. For mortality prediction, the pooled area under the curve was higher for quick Sequential Organ Failure Assessment score (0.702; 95% CI, 0.685–0.718; $I^2 = 99.41\%$; $p < 0.001$) than for systemic inflammatory response syndrome (0.607; 95% CI, 0.589–0.624; $I^2 = 96.49\%$; $p < 0.001$). Quick Sequential Organ Failure Assessment score consistently outperformed systemic inflammatory response syndrome across all subgroup analyses (area under the curve of quick Sequential Organ Failure Assessment vs. area under the curve of systemic inflammatory response syndrome $p < 0.001$), including patient populations (emergency department vs ICU), study design (retrospective vs prospective), and countries (developed vs resource-limited). Quick Sequential Organ Failure Assessment score was more specific (specificity, 74.58%; 95% CI, 73.55–75.61%) than systemic inflammatory response syndrome (specificity, 35.24%; 95% CI, 22.80–47.69%) but less sensitive (56.39%; 95% CI, 50.52–62.27%) than systemic inflammatory response syndrome (78.84%; 95% CI, 74.48–83.19%).

Conclusions: Overall, quick Sequential Organ Failure Assessment score outperforms systemic inflammatory response syndrome in predicting sepsis outcome, but quick Sequential Organ Failure Assessment score has relative strengths/weaknesses (more specific but less sensitive) compared with systemic inflammatory response syndrome.

Key Words: meta-analysis; mortality; quick Sequential Organ Failure Assessment; sepsis; systemic inflammatory response syndrome

Sepsis causes an estimated 6 million deaths worldwide annually. It has superseded tuberculosis (1.29 million deaths) and HIV (1.3 million deaths) as an important cause of global disease burden (1). An effective strategy to prevent sepsis-related deaths is early recognition (2). However, early recognition of sepsis is difficult because its clinical manifestations are often nonspecific. Measuring biomarkers of host response may help, but this is not feasible in most healthcare settings, especially in low- or middle-income countries where access to laboratory facilities is either nonexistent or very limited. Because 87% of the world's population lives in low- and middle-income countries, the development of a fast, reliable, and affordable method for early recognition of sepsis will have a significant effect on resource allocation and global health outcomes.

Quick Sequential Organ Failure Assessment (qSOFA) is a low-cost, 2-minute bedside clinical tool recently introduced to facilitate early recognition of sepsis (3). The qSOFA score is marked by its extraordinary simplicity (measuring only blood pressure, respiratory rate, and mental state) and its utility in any health setting (including resource-poor countries). Earlier studies indicated that qSOFA could aid early recognition of sepsis; an abnormal qSOFA score (fulfilment of two or more criteria) predicted poor outcomes in patients with suspected sepsis (3, 4). However, recent studies have challenged this finding, with some studies showing that qSOFA accuracy was no better than systemic inflammatory response syndrome (SIRS) criteria (5–7), the widely used screening criteria for sepsis (8, 9). A number of recent meta-analyses had compared qSOFA and SIRS in their relative performance; however, these meta-analyses selected only a small number of all available studies (10–16). Currently, no meta-analyses assessed all available evidence published to date.

In this study, we performed a meta-analysis of published studies on qSOFA in order to evaluate its ability to predict sepsis outcomes (e.g., mortality) and to compare its performance to that of SIRS criteria.

METHODS

Overview

This study was performed using a prospectively developed protocol, which prescribed the search strategy, study eligibility criteria, and the methods of data extraction and analysis. The study protocol was registered with International Prospective Register of Systematic Reviews (PROSPERO) (CRD42018103327; an updated search and analysis using the same protocol were performed on March 1, 2019). The reporting of findings followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (see attached PRISMA checklist) (17). The

qSOFA score is defined by the Third International Consensus Definition for Sepsis and Septic Shock (Sepsis-3) Task Force and consists of three criteria: systolic blood pressure less than or equal to 100 mm Hg, respiratory rate greater than or equal to 22 breaths/min, and altered mental state (decrease in Glasgow Coma Scale \geq 1 point from patient's normal baseline) (3). The SIRS criteria are defined as respiratory rate greater than 20 breaths/min, or serum partial pressure of carbon dioxide less than 32 mm Hg, body temperature greater than 38°C or less than 36°C, heart rate greater than 90 beats/min, or altered white cell counts ($> 12,000/\text{mm}^3$ or $< 4,000/\text{mm}^3$ or $> 10\%$ bands) (8). For both qSOFA and SIRS, fulfilment of two or more criteria in each score is used as a threshold to indicate an increased risk of an adverse outcome (e.g., death), in accordance with international consensus definitions (3, 8).

Search Strategy and Selection Criteria

We searched relevant publications between February 23, 2016 (the first publication of the qSOFA score), and March 1, 2019, using electronic database, including MEDLINE, the Cochrane Library, Embase, Web of Science, and Google Scholar. The search strategy used the following MeSH terms: 1) "quick Sequential Organ Failure Assessment" or "qSOFA"; 2) "sepsis-3" or "sepsis" or "infection"; 3) "systemic inflammatory response syndrome" or "SIRS"; 4) "mortality" or "death"; and 5) "non-ICU" or "emergency department" or "ICU." The reference lists of each primary study were also hand-searched for additional publications. Further searches were done by manually reviewing abstract booklets, conference proceedings, and review articles. No language restriction was used, and foreign language publications were translated.

Study Eligibility and Quality Assessment

We included all studies that met the following criteria: contained data of, at least, qSOFA or SIRS; enrolled patients presenting to emergency departments (EDs), hospitals, ICUs, or other healthcare settings (e.g., prehospital); included patients 18 years old or older; used accepted definitions of SIRS and sepsis (described above), and reported outcomes such as mortality, organ dysfunction, admission to ICU, ventilatory support, or prolonged ICU stay. Studies were excluded if they were duplicated studies. We assessed the quality of each included study using the Quality Assessment of Diagnostic Accuracy Studies (18), in accordance with guidelines by the Cochrane Collaboration (19).

Methodologic Framework

We adopted a methodologic framework similar to that developed by the original authors of the qSOFA score (3, 20). In this methodologic framework, the criterion validity of the qSOFA score was assessed by measuring its predictive validity with outcomes. In the original qSOFA study, the primary outcome was hospital mortality and the secondary outcome was a composite outcome consisting of hospital mortality or ICU stay more than 3 days (3). Since the publication of the original qSOFA study, other primary outcomes (e.g., 28-d mortality) and secondary outcomes (e.g., organ dysfunction) have been reported in the literature. We have, therefore, in this meta-analysis, extended the primary outcome definition to include other measures of mortality and the secondary outcome definition to include organ dysfunction, ICU admission, prolonged ICU stay,

or ventilatory support. To measure criterion validity, area under the curve (AUC) of the receiver operator characteristic (AUROC) curves and odds ratio (OR) of the logistic regression model were used, as in the original qSOFA study (3). Because sensitivity and specificity have also been reported widely in the literature, these two metrics were also included in the current analysis.

Statistical Analysis

Summary estimates were presented, with 95% CIs, as pooled AUC/OR and pooled sensitivity/specificity. A p value of less than 0.05 was deemed statistically significant. Meta-analysis of included studies was performed using established methods, with pooled analysis of AUC using methods by Zhou et al (for calculating the weighted summary AUROC under the random effects model) and pooled analysis of OR using methods by Mantel-Haenszel (for calculating the weighted pooled OR) and DerSimonian and Laird (for calculating the summary OR under the random effects model). Because a high level of heterogeneity was expected, we performed all analyses using random effects models and we quantified, in each analysis, the variability due to heterogeneity by using I^2 statistic. We further explored potential sources of heterogeneity by performing subgroup analyses. All subgroup variables were predefined prior to the analysis; they included study design (prospective vs retrospective), geography (resource-limited vs developed countries), population (ED vs ICU), and quality of

the study (high vs low). In subgroup analysis, in order to correct for multiple comparisons across subgroups, a higher stringency level ($p < 0.005$) was adopted to determine statistical significance. Additional analyses were performed to assess the consistency and robustness of the meta-analysis findings. In the influence analysis, this was done by removing/adding one study at a time to assess individual study's effect on the pooled estimates. In the accumulative analysis, the studies were sequentially pooled, starting with the earliest studies with each successive analysis summarizing all the studies in the preceding years, thereby allowing the tracking of the pooled estimates over time. All statistical analyses were performed using Comprehensive Meta-Analysis and MedCalc.

RESULTS

Our search strategy identified 1,340 records, of which 143 eligible studies met the inclusion criteria. Of these studies, 22 studies were excluded due to duplication or missing data (**Fig. 1**). A total of 121 studies were included in the final analysis, which included 1,716,017 individuals (for a full list of included studies, see **Supplementary File 1**, Supplemental Digital Content 1, <http://links.lww.com/CCX/A94>).

The included studies were drawn from a broad range of clinical settings (**Table 1**), including ED, ICU, general wards, and prehospital setting. The patient populations were heterogeneous, including patients with different infections (e.g., pneumonia and urinary tract infection) and comorbidities (e.g., malignancy, cirrhosis, and hemodialysis). The primary outcome was reported in most studies as hospital mortality (67%) or 28- or 30-day mortality (35%). Secondary outcomes were reported in 47 out of the 121 studies (**Table 1**). Most studies had data available for both qSOFA and SIRS ($n = 68$), but some reported only qSOFA ($n = 52$) or SIRS ($n = 1$) (**Fig. 1**). In addition, subgroup analyses were also performed in both primary (**Table 3**) and secondary outcomes (**Supplementary File 2**, Supplemental Digital Content 2, <http://links.lww.com/CCX/A95>) to test the consistency of qSOFA and SIRS performance.

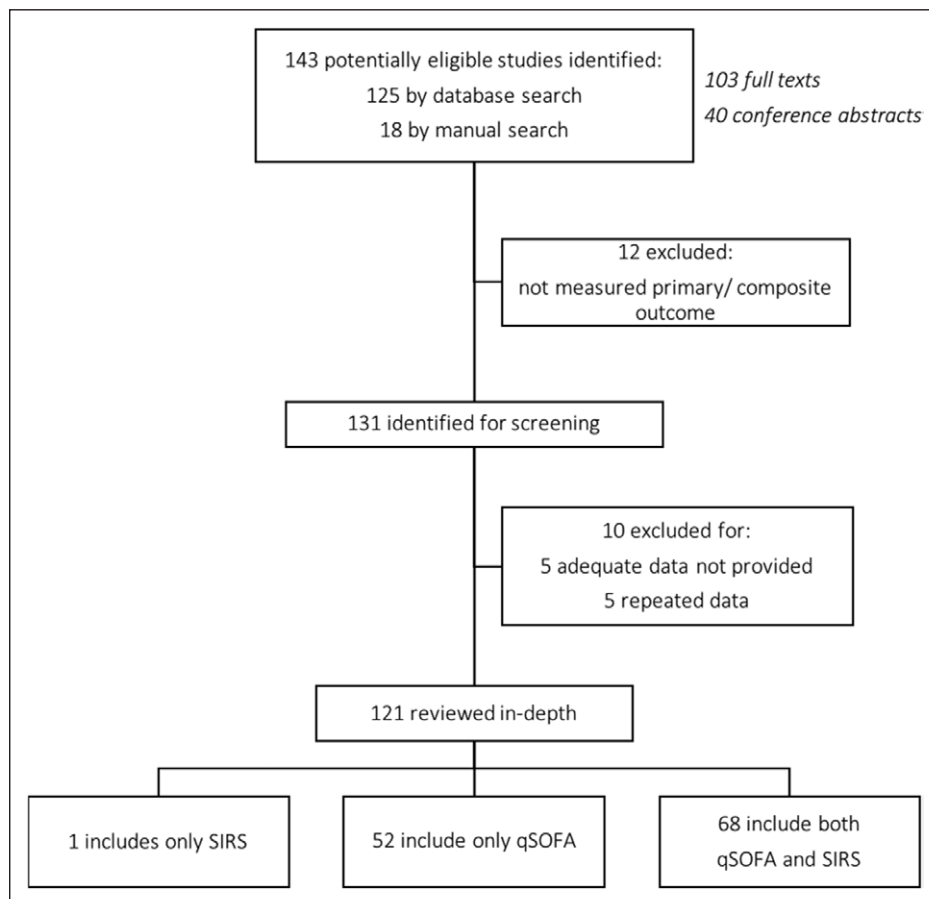


Figure 1. Study flow chart. qSOFA = quick Sequential Organ Failure Assessment, SIRS = systemic inflammatory response syndrome.

For both primary and secondary outcomes, the pooled AUROC curves showed that qSOFA outperformed SIRS (**Fig. 2** and **Table 2**). For primary outcome, the pooled AUC for qSOFA (0.702; 95% CI, 0.685–0.718; $I^2 = 99.41%$; $p < 0.001$) was higher than that for SIRS (0.607; 95% CI, 0.589–0.624; $I^2 = 96.49%$; $p < 0.001$). The AUC difference between qSOFA and SIRS was statistically significant (AUC_{qSOFA} 0.702 vs AUC_{SIRS} 0.607; $p < 0.0001$). For secondary outcomes, a statistically significant difference was also noted, which showed qSOFA outperformed SIRS (**Table 2**). In addition to AUC, OR revealed a similar trend (**Table 2**).

TABLE 1. Characteristics of Included Studies

Characteristic	Studies	Patients
Total	121	1,716,017
Study design		
Prospective	41	88,071
Retrospective	80	1,627,946
Population		
ED	67	1,333,875
Any patient (including ED and ICU)	28	176,309
ICU	13	203,479
Prehospital	6	1,174
Non-ICU	5	960
Outpatient	1	220
Case-mix		
Infection/sepsis	111	1,688,827
Respiratory tract infection	9	20,748
Urinary tract infection	1	141
Setting		
Developed countries	105	1,697,497
Resource-limited countries	16	18,520
Baseline condition/comorbidity		
Unspecified	110	1,710,980
Elderly	3	2,043
Malignancy	4	981
Cirrhosis	2	1,697
On hemodialysis	1	220
Hematology cases	1	96
Outcome measures		
Hospital mortality	81	1,422,358
28- or 30-d mortality	42	294,269
Early mortality (2–7 d)	3	73,744
Late mortality (2–6 mo)	3	2,243
Any mortality	1	37
ICU admission	27	63,535
Ventilatory support	3	33,705
Prolonged ICU stay	14	222,167
Organ dysfunction	4	16,634
30-d readmission	1	316
Publications type		
Full-text articles	93	1,573,029
Conference abstracts	28	142,988

(Continued)

TABLE 1. (Continued). Characteristics of Included Studies

Characteristic	Studies	Patients
Prediction score		
Both SIRS and qSOFA	68	1,561,080
qSOFA only	52	146,213
SIRS only	1	8,724
Inclusion criteria		
Fever	1	519
Infection	66	604,570
Sepsis-2	23	169,539
Sepsis-3	7	2,152
Either Sepsis-2 or 3	20	937,431
Positive blood culture	4	1,806

ED = emergency department, qSOFA = quick Sequential Organ Failure Assessment, SIRS = systemic inflammatory response syndrome.

For both primary and secondary outcomes, pooled OR showed that qSOFA strongly associated with sepsis outcomes, whereas SIRS did not show a statistically significant association with sepsis outcomes. For example, an individual with an increased qSOFA score (fulfilment of two or more criteria) was 3.205 times more likely to die of sepsis or 3.410 times more likely to experience secondary outcomes (e.g., organ dysfunction and ICU admission). There was no statistically significant association between SIRS and the primary outcome (OR, 1.163; 95% CI, 0.897–1.508) or between SIRS and the secondary outcome (OR, 0.796; 95% CI, 0.315–2.011).

Subgroup analyses confirmed that qSOFA outperformed SIRS across different settings and patient populations (Table 3). The difference between qSOFA and SIRS remained the same irrespective of study design (prospective vs retrospective), patient populations (ED vs ICU), countries (developed vs resource-limited), or study quality (all studies vs low-quality studies excluded) (Table 3). In these analyses, an adjusted statistical significance level ($p < 0.005$) was used to correct for multiple comparisons. Despite an increased stringency level, the difference between qSOFA and SIRS was statistically significant in most subgroups, except in one subgroup, namely, the ICU studies (Table 3).

For primary outcome, qSOFA showed a lower sensitivity (56.39%) but a higher specificity (78.84%), whereas SIRS showed a higher sensitivity (74.58%) but a lower specificity (35.24%) (Table 2). For secondary outcome, a similar trend was observed; the qSOFA showed a lower sensitivity (54.86%) but a higher specificity (74.22%), whereas SIRS showed a higher sensitivity (73.02%) but a lower specificity (39.64%). Thus, in predicting both primary and secondary outcomes, qSOFA was more specific, whereas SIRS was more sensitive.

As expected, the included studies were highly heterogenous, with I^2 over 85% (indicating high heterogeneity) in most analyses. An extensive search for the source of heterogeneity across

TABLE 2. Summary of Pooled Estimates of Quick Sequential Organ Failure Assessment and Systemic Inflammatory Response Syndrome for Primary and Secondary Outcomes

Metric	Quick Sequential Organ Failure Assessment			Systemic Inflammatory Response Syndrome			p ^a
	Pooled Estimate	Studies	Patients	Pooled Estimate	Studies	Patients	
Mortality							
AUC	0.702 (0.685–0.718)	81	1,487,113	0.607 (0.589–0.624)	47	1,397,673	< 0.0001
OR	3.205 (2.691–3.816)	34	162,372	1.163 (0.897–1.508)	10	145,010	0.0246
Sensitivity (%)	56.39 (50.52–62.27)	64	267,099	78.84 (74.48–83.19)	33	179,254	< 0.0001
Specificity (%)	74.58 (73.55–75.61)	62	265,826	35.24 (22.80–47.69)	31	178,368	< 0.0001
Secondary outcomes ^b							
AUC	0.701 (0.674–0.728)	33	295,299	0.634 (0.608–0.660)	18	236,031	0.0005
OR	3.410 (2.447–4.752)	14	27,678	0.796 (0.315–2.011)	3	3,798	0.0658
Sensitivity (%)	54.86 (47.62–62.09)	13	103,024	74.22 (63.41–85.02)	9	14,532	0.0048
Specificity (%)	73.02 (69.46–76.58)	12	102,138	39.64 (28.29–50.99)	8	13,646	0.0048

AUC = area under the curve, OR = odds ratio.

^ap values are calculated by comparing the pooled estimates between quick Sequential Organ Failure Assessment and systemic inflammatory response syndrome.

^bSecondary outcome is a composite outcome including organ dysfunction, ICU admission, ventilatory support, prolonged ICU stay, or 30-d readmission.

predefined subgroups was performed. In this analysis, pooled AUCs were compared across prospective and retrospective studies, ED patients and ICU patients, developed countries and resource-limited countries, and between all studies and low-quality studies excluded. This analysis did not identify any obvious source of heterogeneity in either primary outcome or secondary outcomes (**Supplementary File 3**, Supplemental Digital Content 3, <http://links.lww.com/CCX/A96>; and **Supplementary File 4**, Supplemental Digital Content 4, <http://links.lww.com/CCX/A97>).

Sensitivity analysis was performed to examine the robustness of the findings of the meta-analysis. Because the primary outcome was reported as hospital mortality (67%) or 28- or 30-day mortality (35%), we analyzed separately the studies with hospital mortality and 28- or 30-day mortality (Table 3). This analysis showed that the pooled estimates were likely similar among the hospital mortality (AUC_{qSOFA}, 0.704; 95% CI, 0.683–0.725), 28- or 30-day mortality (AUC_{qSOFA}, 0.694; 95% CI, 0.669–0.720), and all mortality (AUC_{qSOFA}, 0.702; 95% CI, 0.685–0.718). Because a small number of studies did not fully meet the quality assessment criteria (**Supplementary File 5**, Supplemental Digital Content 5, <http://links.lww.com/CCX/A98>), we repeated the analysis after excluding these studies. The findings of the repeated analysis did not differ significantly from the original findings (Table 3).

Two additional analyses were performed to assess the effect of potential bias on our findings. First, an influence analysis was performed on both SIRS and qSOFA, across all metrics (AUC, OR, sensitivity, and specificity) and in both primary and secondary outcomes. This analysis showed that no individual study exerted a dominant effect on the pooled estimates (**Supplementary File 6**, Supplemental Digital Content 6, <http://links.lww.com/CCX/A99>; legend: influence meta-analysis on AUC [qSOFA for mortality prediction]). Second, a cumulative analysis was performed across the same parameters (AUC, OR, sensitivity, and specificity) and in

both primary and secondary outcomes. It showed that the evidence was consistent over time, with the pooled estimates and their CIs stabilized as evidence accumulated over time. Importantly, the pool estimates remained unchanged even as more studies were added (**Supplementary File 7**, Supplemental Digital Content 7, <http://links.lww.com/CCX/A100>; legend: cumulative meta-analysis on AUC [qSOFA for mortality prediction]), suggesting that adding more studies is unlikely to change the existing body of evidence.

Finally, we compared our meta-analysis to other recently published meta-analyses (**Supplementary File 8**, Supplemental Digital Content 8, <http://links.lww.com/CCX/A101>). In terms of sensitivity and specificity, these meta-analyses showed similar findings to our findings (qSOFA is more specific, whereas SIRS is more sensitive). However, these meta-analyses did not perform analysis on AUC or OR, with the exception of one study by Song et al (11), which had a limited analysis on AUC and was performed on a smaller sample size including 23 studies ($n = 146,551$). Overall, these meta-analyses did not conclusively show whether qSOFA outperforms SIRS. On July 1, 2019, we updated our literature search and identified four additional studies that performed a head-to-head comparison of qSOFA versus SIRS. All four studies showed qSOFA outperforms SIRS (**Supplementary File 9**, Supplemental Digital Content 9, <http://links.lww.com/CCX/A102>), which is consistent with our original finding.

DISCUSSION

Sepsis is characterized by life-threatening organ dysfunction triggered by infection. It induces a myriad of host response patterns, resulting in highly varied clinical courses/manifestations in different patients. This vast heterogeneity makes it challenging to develop a prognostic tool that can reliably forecast disease progression for each and every sepsis patient. The recently introduced

TABLE 3. Subgroup Analysis of Quick Sequential Organ Failure Assessment and Systemic Inflammatory Response Syndrome For Primary Outcome^a

Subgroup	Quick Sequential Organ Failure Assessment			Systemic Inflammatory Response Syndrome			<i>p</i> ^b
	AUC (95% CI)	Studies	Patients	AUC (95% CI)	Studies	Patients	
Mortality							
Study design							
Prospective	0.720 (0.685–0.754)	23	54,705	0.595 (0.565–0.626)	12	43,009	< 0.0001
Retrospective	0.693 (0.676–0.710)	58	1,432,408	0.611 (0.589–0.632)	35	1,354,664	< 0.0001
Populations							
Emergency department	0.709 (0.687–0.731)	41	1,112,504	0.606 (0.582–0.631)	23	1,090,031	< 0.0001
ICU	0.667 (0.618–0.716)	11	210,109	0.597 (0.575–0.620)	10	209,445	0.0107
Setting							
Developed countries	0.693 (0.678–0.709)	67	1,469,198	0.610 (0.590–0.630)	39	1,382,807	< 0.0001
Resource-limited countries	0.740 (0.687–0.793)	14	17,915	0.577 (0.557–0.598)	8	14,866	< 0.0001
Study quality							
All studies ^c	0.710 (0.692–0.798)	63	1,393,916	0.613 (0.591–0.634)	35	1,413,941	< 0.0001
Low-quality studies excluded	0.703 (0.680–0.726)	38	1,320,991	0.620 (0.593–0.646)	27	1,302,987	< 0.0001
Types of mortality							
Hospital mortality	0.704 (0.683–0.725)	55	1,386,317	0.611 (0.589–0.633)	34	1,314,339	< 0.0001
28- or 30-d mortality	0.694 (0.669–0.720)	29	108,122	0.592 (0.562–0.621)	15	90,467	< 0.0001
Publication type							
Full text	0.710 (0.692–0.728)	63	1,389,574	0.613 (0.591–0.634)	36	1,344,826	< 0.0001
Conference abstract	0.666 (0.634–0.697)	18	92,960	0.586 (0.558–0.615)	12	52,847	0.0002

AUC = area under the curve.

^aThis analysis was performed on AUC of mortality prediction because it was reported in the majority of studies.

^b*p* values are calculated by comparing the summary estimates between quick Sequential Organ Failure Assessment and systemic inflammatory response syndrome. *p* < 0.005 is used as an adjusted level for statistical significance to correct for multiple comparisons across subgroups.

^cAdequate data were available for quality assessment to be performed in 93 of the 121 studies. The remaining 28 studies were in abstract format, which did not contain enough data for quality assessment. Of these 93 studies in which quality assessment was performed, 63 studies (shown in this table) reported AUC for mortality prediction. Thus, the quality assessment presented in this table is limited to the 63 studies in which sufficient data were available.

qSOFA faces such a challenge; its acceptance by clinicians is predicated on the evidence that qSOFA can predict outcome, consistently and reliably, across different patient populations and clinical settings. This analysis, based on 121 studies (*n* = 1,716,017), is the largest meta-analysis to date that evaluates the ability of qSOFA to predict sepsis outcomes. Its findings confirmed that qSOFA score predicts sepsis outcome and this predictive performance is consistent across different patient populations and clinical settings. This analysis also revealed the performance of qSOFA to be modest (e.g., AUC 0.702 for mortality); nevertheless, qSOFA consistently outperformed SIRS criteria irrespective of study design (prospective vs retrospective), patient populations (ED vs ICU), or geography (developed vs resource-limited). Collectively, these findings provide a comprehensive evidence base to inform the current understanding of the qSOFA score in predicting sepsis outcome.

This meta-analysis differs from previous meta-analyses on three major aspects. First, our meta-analysis is the largest study

conducted so far; it has included 121 studies with over 1.7 million study participants. All previous meta-analyses included much smaller sample sizes. Second, analyses of AUC and ORs provide important information regarding the predictive performance of qSOFA. This approach (using AUC and ORs) was adopted by the original qSOFA authors (3). Our meta-analysis provides the most extensive analyses, to date, on AUC and ORs of qSOFA score. In contrast, previous meta-analyses did not provide AUC/OR data except for one study (Song et al [11]), which performed a limited analysis on AUC. Third, by providing extensive analyses on AUC and ORs, our findings provide strong evidence that qSOFA outperforms SIRS. Previous meta-analyses do not provide such high-level evidence because they did not perform an extensive AUC and odd ratios analysis. In summary, our meta-analysis differs from other meta-analyses by having the largest sample size, a similar methodologic approach with the original qSOFA study and compelling evidence that qSOFA outperforms SIRS.

qSOFA was designed as a clinical prompt to assist clinicians to identify high-risk patients (3). In this aspect, qSOFA share some common characteristics with SIRS. As was recently shown by an extensive analysis of a large dataset ($n = 1,171,797$), SIRS also predicts sepsis outcomes (e.g., mortality) (9). However, qSOFA and SIRS use different parameters; qSOFA includes blood pressure, respiratory rate (> 22 breath/min), and changes in mental state, whereas SIRS includes respiratory rate (> 20 breaths/min), temperature, white cell count, and heart rate. Unsurprisingly, the two clinical scores display different sensitivity and specificity (qSOFA more specific and SIRS more sensitive), as demonstrated by our findings. Neither score is perfect because each score has its own limitations—qSOFA has a lower sensitivity (i.e., it may miss those individuals with subclinical organ dysfunction), whereas SIRS has a lower specificity (i.e., it may cause unnecessary testing in low-risk patients). In practice, the choice (qSOFA vs SIRS) is based on an optimal balance between sensitivity and specificity. The original qSOFA authors (the Sepsis-3 Task Force) had addressed this issue by calculating global performance parameters (e.g., AUC) and they compared the relative performance of qSOFA and SIRS by using these global parameters (3). In this meta-analysis, we used a similar approach—it showed that the combined analysis by AUC, OR, and sensitivity/specificity provided a more accurate evaluation of the clinical scores' performance than using sensitivity/specificity alone.

This study has several strengths, including a large sample size, an inclusion of all relevant metrics (AUC, OR, sensitivity, and specificity) and the consistency of its findings. We used an exhaustive search strategy to identify both published and unpublished studies. This results in a large sample size (121 studies), providing us with an increased statistical power to detect a difference between qSOFA and SIRS. By analyzing global metrics (such as AUC and OR), we were able to demonstrate that the overall performance of qSOFA was better than SIRS, a finding not shown by previous meta-analyses (10–16). This difference in AUC between qSOFA and SIRS seems consistent because the same difference was observed across different clinical settings, patient populations, or study design.

This study focused on criterion validity of qSOFA; it does not address other important aspects of sepsis diagnosis, such as content validity and construct validity (20, 21). As stated by the original authors, qSOFA is not intended to be used as diagnostic criteria for sepsis; rather, qSOFA was designed as a clinical prompt to alert clinicians to consider the diagnosis of sepsis (22, 23). The diagnosis of sepsis requires meeting a different set of criteria, namely the international consensus definition of sepsis (Sepsis-3), which has a higher content validity and construct validity than qSOFA score (20, 21). The more formal Sepsis-3 criteria lack flexibility in enabling early recognition of sepsis (it requires laboratory tests to be performed, which can be time consuming and costly). qSOFA addresses this limitation by possessing three desirable characteristics of a less formal bedside tool, namely, 1) low measurement burden; 2) reproducibility; and 3) timeliness (20).

Like all clinical prediction tools, there is an inherent risk in using qSOFA score in practice. Some qSOFA-negative patients do develop organ failure, and these false-negative cases can result in

serious consequence (i.e., missed treatment opportunity). Thus, a negative qSOFA requires clinicians to continue to look elsewhere for evidence of sepsis. This raises the question of whether combining qSOFA with other tools, such as SIRS, may reduce false-negative cases. Our findings suggested that qSOFA and SIRS do have complementary strengths; qSOFA is more specific, whereas SIRS is more sensitive. An intriguing next question is, therefore, whether a combined qSOFA/SIRS score may improve the overall prediction accuracy. Such questions should be addressed in future studies.

The qSOFA was designed to be used in the non-ICU setting. In this meta-analysis, a majority of the studies were performed in this setting, including 67 studies performed in ED and 12 studies performed in other non-ICU settings. The remaining studies included 28 studies performed in ED/ICU, and 13 studies performed exclusively in ICU. Overall, the relative proportions of distribution were 56% (ED), 10% (other non-ICU settings), 23% (ED/ICU), and 11% (ICU only).

As expected, heterogeneity was evident across the entire dataset. Despite an extensive search, the sources of heterogeneity could not be identified. There are several explanations for this. First, the subgroup analyses may have excluded key factors that had contributed to heterogeneity (e.g., timing of qSOFA measurement or stage of illness); however, information on these additional variables was not available in many studies, thereby precluding their analyses. Second, there were low number of studies in some subgroups, making them underpowered to detect a statistically significant difference across subgroups. Third, traditional metrics to define heterogeneity (e.g., patient populations, study design, and settings) may have been inadequate. Emerging evidence from “omics” studies has revealed that sepsis subtypes (“endotypes”) are present, but they are usually undetectable by routine clinical evaluation or conventional laboratory tests (24). These sepsis subtypes may have contributed to the heterogeneity observed in this meta-analysis.

It is expected that new qSOFA studies will continue to emerge, given the ease of qSOFA measurement and the low cost of performing such studies. A recent search in PROSPERO (a registry for meta-analyses) indicates that there are at least 12 meta-analyses on qSOFA, with some published but a large majority are still in progress. Therefore, an important question is whether adding new findings or future studies may change our findings. In our opinion, the additional studies are unlikely to change our findings for two reasons. First, our meta-analysis has a large sample size (121 studies consisting of 1,716,017 patients)—this generates a point estimate ($AUROC_{qSOFA} = 0.70$) with a very narrow CI (0.69–0.72). Thus, adding more studies to the dataset is unlikely to narrow this CI any further. Second, we find that the point estimate stabilizes over time (as shown by our cumulative meta-analysis), and thus, adding more studies is unlikely to change the final point estimate. We expect that future studies are likely to gravitate toward this final point estimate, as predicted by the well-established regression to mean principle (25).

This study has several limitations. First, it did not assess the incremental predictive validity of qSOFA. Our analyses were limited to analyzing the effect of having two or more qSOFA criteria fulfilled; the effect of having only one qSOFA criterion fulfilled remains unknown. Second, our analysis did not consider the effect

of the timing of measurement. This needs to be addressed in future studies. Third, most included studies did not provide data on the component variable of either qSOFA or SIRS. Thus, the contribution of individual component (e.g., respiratory rate) to the overall predictive performance is unclear.

In conclusion, we found that qSOFA score has a modest ability to predict sepsis outcomes, but its predictive performance is better than SIRS. The higher performance of qSOFA over SIRS is consistent in different patient populations and across a diverse range of settings. However, our findings are limited by the presence of significant heterogeneity, which cannot be adequately explained by subgroup analyses.

REFERENCES

1. Reinhart K, Daniels R, Kissoon N, et al: Recognizing sepsis as a global health priority - A WHO resolution. *N Engl J Med* 2017; 377:414–417
2. Seymour CW, Gesten F, Prescott HC, et al: Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med* 2017; 376:2235–2244
3. Seymour CW, Liu VX, Iwashyna TJ, 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
4. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810
5. Askim Å, Moser F, Gustad LT, et al: Poor performance of quick-SOFA (qSOFA) score in predicting severe sepsis and mortality - A prospective study of patients admitted with infection to the emergency department. *Scand J Trauma Resusc Emerg Med* 2017; 25:56
6. Churpek MM, Snyder A, Han X, et al: 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
7. Dorsett M, Kroll M, Smith CS, et al: qSOFA has poor sensitivity for pre-hospital identification of severe sepsis and septic shock. *Prehosp Emerg Care* 2017; 21:489–497
8. Bone RC, Balk RA, Cerra FB, et al: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101:1644–1655
9. Kaukonen KM, Bailey M, Pilcher D, et al: Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med* 2015; 372:1629–1638
10. Fernando SM, Tran A, Taljaard M, et al: Prognostic accuracy of the quick Sequential Organ Failure Assessment for mortality in patients with suspected infection: A systematic review and meta-analysis. *Ann Intern Med* 2018; 168:266–275
11. Song JU, Sin CK, Park HK, et al: Performance of the quick sequential (sepsis-related) organ failure assessment score as a prognostic tool in infected patients outside the intensive care unit: A systematic review and meta-analysis. *Crit Care* 2018; 22:28
12. Serafim R, Gomes JA, Povoia P: A Comparison of the quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality. *Chest* 2018; 153:646–655
13. Maitra S, Som A, Bhattacharjee S: Accuracy of quick Sequential Organ Failure Assessment (qSOFA) score and Systemic Inflammatory Response Syndrome (SIRS) criteria for predicting mortality in hospitalized patients with suspected infection: A meta-analysis of observational studies. *Clin Microbiol Infect* 2018; 24:1123–1129
14. Jiang J, Yang J, Mei J, et al: Head-to-head comparison of qSOFA and SIRS criteria in predicting the mortality of infected patients in the emergency department: A meta-analysis. *Scand J Trauma Resusc Emerg Med* 2018; 26:56
15. Franchini S, Scarallo L, Carlucci M, et al: SIRS or qSOFA? Is that the question? Clinical and methodological observations from a meta-analysis and critical review on the prognostication of patients with suspected sepsis outside the ICU. *Intern Emerg Med* 2019; 14:593–602
16. Liu YC, Luo YY, Zhang X, et al: Quick Sequential Organ Failure Assessment as a prognostic factor for infected patients outside the intensive care unit: A systematic review and meta-analysis. *Intern Emerg Med* 2019; 14:603–615
17. Moher D, Liberati A, Tetzlaff J, et al; PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 2009; 339:b2535
18. Whiting P, Rutjes AW, Reitsma JB, et al: The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25
19. Cochrane Collaboration: Cochrane Handbook For Systematic Reviews of Diagnostic Test Accuracy. 2009. Available at: <https://methods.cochrane.org/sdt/handbook-dta-reviews>. Accessed December 1, 2017
20. Angus DC, Seymour CW, Coopersmith CM, et al: A framework for the development and interpretation of different sepsis definitions and clinical criteria. *Crit Care Med* 2016; 44:e113–e121
21. Seymour CW, Coopersmith CM, Deutschman CS, et al: Application of a framework to assess the usefulness of alternative sepsis criteria. *Crit Care Med* 2016; 44:e122–e130
22. Singer M, Shankar-Hari M: qSOFA, cue confusion. *Ann Intern Med* 2018; 168:293–295
23. Vincent JL, Martin GS, Levy MM: qSOFA does not replace SIRS in the definition of sepsis. *Crit Care* 2016; 20:1–3
24. Scicluna BP, van Vught LA, Zwinderman AH, et al: Classification of patients with sepsis according to blood genomic endotype: A prospective cohort study. *Lancet Respir Med* 2017; 5:816–826
25. Barnett AG, van der Pols JC, Dobson AJ: Regression to the mean: What it is and how to deal with it. *Int J Epidemiol* 2005; 34:215–220