Internal and Emergency Medicine (2019) 14:593–602
https://doi.org/10.1007/s11739-018-1965-0

CE - SYSTEMATIC REVIEW

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

Stefano Franchini[1] · Luca Scarallo[2] · Michele Carlucci[1,3] · Luca Cabrini[4] · Moreno Tresoldi[5]

Received: 12 April 2018 / Accepted: 6 October 2018 / Published online: 15 October 2018

Abstract

The purpose of the study was to assess the prognostic performances, in terms of in-hospital mortality, of the quick sequential organ failure assessment (qSOFA) score and the systemic inflammatory response syndrome (SIRS) criteria applied to patients with suspected infection outside the ICU, and to critically reappraise the results and the clinical impact of the SEPSIS-3 study and of the subsequent trials. We performed bivariate meta-analysis, evaluation of the Bayesian post-test probabilities of death, and computation of the unidentified deaths for every 1000 screened cases (UDS_{1000}). The use of qSOFA for screening instead of the SIRS implies a relevant increase in the UDS_{1000}. However, this difference appears far smaller in the SEPSIS-3 study, largely due to an underestimation of SIRS sensitivity. The increment in the pre-test probability of death implied by a positive qSOFA is higher than that implied by a positivity of the SIRS. However, the included studies use highly variable definitions of "suspected sepsis" and carry very high levels of heterogeneity. SIRS overperforms qSOFA as a rule-out tool for mortality, while qSOFA shows a higher rule-in power. However, the evident lack of consistency across the published studies undermines the significance of both the meta-analytic approach and the reproducibility of the outcomes, and demands for a standardized definition of the target population.

Keywords qSOFA · SIRS · Sepsis screening · Sepsis prognosis

Introduction

In 2016, the Society of Critical-Care Medicine/European Society of Intensive Care Medicine 'SEPSIS-3' Task-Force (SEPSIS-3) redefined sepsis from an unbridled systemic inflammatory response syndrome (SIRS) caused by infection, to a life-threatening organ dysfunction secondary to a dysregulated host response to infection [1, 2]. The new SEPSIS-3 definitions dismissed the SIRS criteria, which for decades had played a fundamental screening role, and based the identification of sepsis on the Sequential [Sepsis-related] Organ Failure Assessment (SOFA) score. However, since the SOFA score requires multiple clinical and laboratory data and may prove impractical in the screening phase of suspected infection outside the intensive care unit (ICU) [3], the SEPSIS-3 Task-Force proposed the "quick SOFA" (qSOFA) [1, 4, 5], a new bedside three-point clinical score meant to be used outside the ICU to rapidly identify adult patients with suspected infection more likely to have poor outcomes [1].

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11739-018-1965-0) contains supplementary material, which is available to authorized users.

✉ Stefano Franchini
franchini.stefano@hsr.it

1 Emergency Department, Ospedale San Raffaele Scientific Institute, Pronto Soccorso, Via Olgettina 60, 20132 Milan, Italy

2 Faculty of Medicine, Vita-Salute San Raffaele University, Via Olgettina 58, 20132 Milan, Italy

3 Department of General and Emergency Surgery, Ospedale San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy

4 Department of Anaesthesia and Intensive Care, Ospedale San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy

5 Unit of General Medicine and Advanced Care, Ospedale San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy

These definitions have raised a lot of discussion and controversy [6] regarding, mainly: (i) the retrospective design of the SEPSIS-3 study that led to the identification of the qSOFA score, and to the choice of a SOFA ≥ 2 as the new defining criterion for diagnosis [7]; (ii) the fact that, by targeting greater severity, the new definitions may delay both recognition and therapeutic intervention [3, 8, 9]; (iii) the content validity of the qSOFA that does not explore 3 out of the 6 organ dysfunctions assessed by the SOFA score and, thus, its possible insufficient sensitivity and inadequate screening capacity [7, 8]; (iv) the lack of adequate validation that does not safeguard from the possible risks of abandoning a well-known and already tested system (whose application led to a significant reduction in global sepsis mortality) [9].

To validate the new definitions, several studies regarding the identification of sepsis in patients with suspected infection and their prognostic stratification have followed.

The purpose of this study is to systematically select and critically appraise the published studies regarding the application of the qSOFA score and the SIRS criteria outside the ICU, in order to explore the screening and prognostic performances of SIRS and qSOFA, and to translate them into clinically meaningful notions, and to compare them with the results obtained in the SEPSIS-3 study.

Methods

The setting selected for this review is the in-hospital evaluation of the adult patient with suspected infection or sepsis. The performances of the SIRS criteria and the qSOFA score were analyzed in terms of sensitivity (*Se*), specificity (*Sp*), and positive and negative likelihood ratios (LR+ and LR−). We chose LRs over predictive values because the latter are heavily dependent on the outcome prevalence and, thus, are hardly transferrable to settings with variable prevalence of the outcome of interest [10]. The major outcome evaluated in this review is in-hospital mortality (or, where this was not available, 30-day or 28-day mortality, as surrogate major outcomes). We performed a systematic review and meta-analysis following the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [11] and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) proposal [12].

An electronic search of the published literature was conducted in January 2018 in PubMed and EMBASE (from February 1st 2016, to January 10th 2018). We performed open queries utilizing all the expressions by which the qSOFA is mentioned in the literature ("qSOFA", "quick sofa" or "quicksofa", or "quick sepsis-related organ failure assessment", or "quick sequential organ failure assessment") linked only by the boolean operator "OR", without any keyword or MESH term identification, nor other restrictions regarding the type of publication or the database fields included in the search. We excluded all the papers that were not clinical studies. After identifying all the qSOFA-related clinical studies, we reviewed them and excluded those not including an assessment of the SIRS criteria. Thus, we included only studies that jointly compared both the SIRS criteria and the qSOFA score, excluding those that have evaluated one tool.

Our subsequent selection process was based on the following criteria. (i) We included only full-length reports published in peer-reviewed journals. (ii) We excluded all the studies that did not consider, as their outcome, in-hospital mortality (or one of the two surrogate mortality outcomes). We retained only studies (iii) performed in the in-hospital setting, (iv) on adult subjects (>16-year old), (v) where the acquisition of clinical data occurred outside the ICU, and (vi) the major inclusion criterion was suspected acute infection or sepsis. (vii) We discarded the studies where the values of "true positives", "false positive", "true negatives", and "false negatives" with respect to our outcome were not directly available or could not be extrapolated from the provided data. According to the published definitions, the qSOFA was considered positive when at least 2 variables were present, and the SIRS when at least 2 criteria were met [1, 13].

Two reviewers (S.F. and L.S.) independently evaluated all the articles identified through the electronic search and decided, by shared consensus, whether a study should be included based on the above-mentioned criteria.

To assess the methodological quality of the studies, we used the Newcastle–Ottawa Quality Assessment Scale (NOS) [14].

The aggregated prognostic measures of the selected studies were compared with the results obtained in the non-ICU subgroup of the UPMC Validation Cohort included in the original SEPSIS-3 study [4].

The performance measures and their 95% confidence intervals (95%CI) were determined for each one of the included studies (Fig. 1), for the pooled data from all the included studies and for the non-ICU subgroup of the UPMC Validation Cohort included in the original SEPSIS-3 study (Tables 1 and 2).

We also computed, for each tool: the post-test probabilities (according to the Bayes' theorem) resulting from the application of the two tools; and the number of deaths that, on average, would go unidentified at screening for every 1000 screened cases of suspected infection that we called "UDS_{1000}" (where UDS stands for "unidentified deaths at screening", and it is calculated as "FN/Nx1000", where N represents the total number of included patients and FN the number of "false negatives").

Bivariate meta-analyses were performed and heterogeneity across the studies was assessed by Q test and I^2 statistic [15].

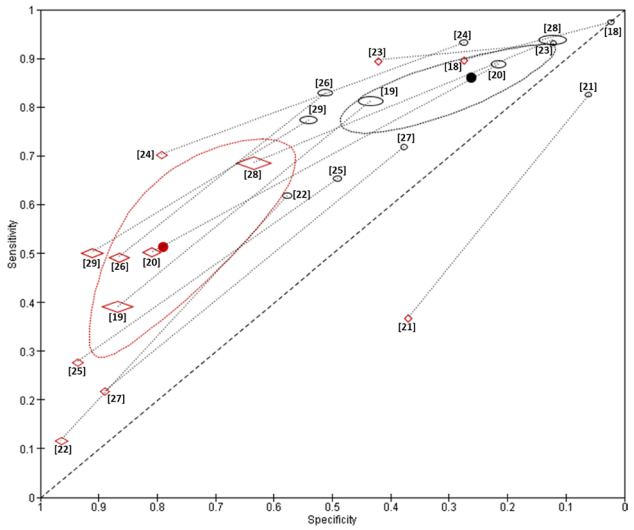

Fig. 1 Scatter plot with summary operating points for SIRS (full black dot) and qSOFA (full red dot) and their 95% confidence regions. The empty black ellipses and red diamonds represent sample size scaled points of the included studies for, respectively, SIRS and qSOFA. Gray dotted lines join the results for SIRS and qSOFA within each study (reference numbers are reported in square brackets). The 95% confidence regions drawn in the figure confirm the significant differences in sensitivity and specificity between SIRS and qSOFA

We performed LR Chi-squared statistic, considering SIRS and qSOFA as covariates, to assess changes in the −2Log likelihood when a covariate was added to the bivariate model. We estimated mean logit *Se* and *Sp*, their standard errors, 95% CI, their between-study variability, and the covariance between them. We used such parameters to determine summary operating points for SIRS and qSOFA and their 95% confidence contour ellipsoids (Fig. 1).

Data were managed and analyzed using Microsoft Excel 2013. We used the packages "meta4diag", "lme4", and "lmtest" in CRAN-R ver 3.5.0 (64-bit version) [16] to produce the meta-analysis forest plots, to perform the LR Chi-squared statistic, and to determine the fundamental bivariate model parameters. Such parameters were entered in Review Manager (RevMan Version 5.3; Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014) to produce the summary operating point scatter plot (Fig. 1). We used MetaDiSc ver 1.4 [17] to assess heterogeneity across the studies.

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Results

Figure 1 in the Supplementary material summarizes the search and selection process that we followed. During the selection process, 3 of the reviewed studies raised initial disagreement between the two evaluators, but a definitive consensus was then reached by further discussion. Twelve observational studies were finally selected for our review [18–27], of whom 5 were retrospective [18–21, 28], and 7 prospective [22–27, 29]. The risk of bias was acceptable for all the studies, as measured by the NOS quality assessment (Table 1 in the Supplementary material).

Overall 80,941 patients were included in the analysis. The definitions of infection or sepsis used by the included studies were highly heterogeneous. Table 2 in the Supplementary material details all such definitions.

The pooled global mortality was 5.2%; however, mortality ranged from 3.7% to 20.5% in the selected studies. Table 1 specifies all the general characteristics for each included study. Table 2 shows the pooled values of all the considered predictive measures with their 95% ICs, for the whole included sample and for the SEPSIS-3 study.

Both *Se* and *Sp* differ significantly when we add the clinical tool type (SIRS or qSOFA) as a covariate to our bivariate model (for *Se*: Chi square = 13.015, $P < 0.001$; for *Sp*: Chi square = 15.456, $P < 0.001$).

As for the capacity of ruling-out the risk of mortality (Figs. 1, 2, and Table 2), *Se* is significantly higher for SIRS as compared to qSOFA (respectively, 86% [95% CI 78–92%] vs 51% [95% CI 38–65%]). However, the SIRS LR− is only slightly (and not significantly) lower than qSOFA LR−.

The clinical impact of these differences in *Se* values might be represented by the changes in the UDS_{1000} (Table 3): based on the pooled screening measures, the application of the SIRS criteria at screening would result in 7 unrecognized death for every 1000 patients (UDS_{1000}) with suspected acute infection screened. If the qSOFA score is instead applied, these figures would worsen to 24 UDS_{1000}.

On the other hand, qSOFA score shows a higher rule-in power with respect to the SIRS criteria, as indicated by its approximately threefold values of *Sp* (respectively 78% [95% CI 66–88%] vs 27% [95% CI 15–41%]) and its higher LR+ (2.49 [95% CI 1.58–3.89] vs 1.17 [95% CI 1.04–1.40]) (Figs. 1, 2; Table 2).

To better appreciate the clinical relevance of these findings, we input the aggregated values of *Sp*, *Se* and LR+ that we found into a Bayesian analysis model that considers the overall mortality as its pre-test probability, in order to quantify the post-test theoretical changes in the probability of death that would be allowed by the application of the SIRS or the qSOFA (Table 3).

Table 1 Main characteristics of the studies included in the meta-analysis and of the non-ICU subgroup of the SEPSIS-3 study [4]

Selected studies Author, Year [Refs]	Type	Location	Setting	Timeframe of data acquisition	Primary outcome	Patients included in the analysis	Overall mortality	SIRS		qSOFA	
								Se	*Sp*	*Se*	*Sp*
April 2016 [18]	R	USA	ED[a]	ED stay	In-hospital mortality	214	18,2%	96% [91–99%]	3% [1–6%]	86% [76–94%]	28% [22–35%]
Askim 2017 [22]	P	Norway	ED	first assessment in the ED	30 day mortality	1535	4,4%	63% [52–73%]	58% [55–60%]	15% [8–23%]	96% [95–97%]
Churpek 2017 [28]	R	USA	ED/HW	all time-data	In-hospital mortality	30,677	5,4%	94% [93–95%]	12% [12–13%]	69% [66–71%]	64% [63–64%]
Finkelsztein 2017 [23]	P	USA	ED/HW[a]	within 8 h before ICU admission	In-hospital mortality	152	19,1%	92% [83–97%]	13% [8–19%]	83% [71–93%]	42% [34–51%]
Freund 2017 [24]	P	France, Spain, Belgium, and Switzerland	ED	ED stay	In-hospital mortality	879	8,4%	91% [84–96%]	27% [24–30%]	67% [57–77%]	79% [76–82%]
Gonzales del Castillo 2017 [25]	P	Spain	ED[b]	first assessment in the ED	30 day mortality	1071	6,7%	67% [56–77%]	49% [46–52%]	28% [19–38%]	94% [92–95%]
Henning 2017 [26]	P	USA	ED	ED stay	In-hospital mortality	6750[c]	4,1%	82% [78–87%]	51% [50–52%]	49% [43–55%]	87% [86–87%]
Moskowitz 2017 [19]	R	USA	ED	ED stay	In-hospital mortality	24,164	4,9%	81% [79–83%]	43% [43–44%]	39% [36–42%]	87% [86–87%]
Ranzani 2017 [20]	R	Spain	ED	ED stay	In-hospital mortality	6024	6,2%	89% [85–92%]	22% [20–23%]	50% [45–55%]	81% [80–82%]
Szakmany 2017 [27]	P	UK	ED/HW	within 24 h of study enrollment	30 day mortality	380	20,5%	73% [64–82%]	38% [33–43%]	25% [16–34%]	89% [86–92%]
Weigle 2016 [21]	R	USA	ED	ED stay	In-hospital mortality	224	20,5%	87% [77–81%]	8% [4–12%]	45% [32–59%]	39% [32–46%]
Williams 2017 [29]	P	Australia	ED	ED stay	30 day mortality	8871	3,7%	77% [73–81%]	54% [53–55%]	50% [44–55%]	91% [91–92%]
SEPSIS-3 [4]	R	USA	AHA	from 48 h before to 24 h after the onset of infection	In-hospital mortality	66,522[d]	2,8%	64% [62–66%]	65% [65–65%]	55% [53–57%]	84% [84–84%]

[a] evaluation occurred outside the ICU; however, only patients eventually admitted to the ICU were included. [b] only patients older than 75 years were included in this study. [c] data from the third cohort of this study were not included due to the different data acquisition process and data availability for this cohort. [d] data related to the non-ICU subgroup of the UPMC Validation Cohort included in the original Sepsis-3 study [4]. *Se* sensitivity, *Sp* specificity. 95%CI is reported in square brackets. *R* retrospective study, *P* prospective study, *ED* emergency department, *HW* hospital (non-intensive) wards

Table 2 Pooled values of the considered predictive measures for the cohorts included in this review, compared with the values reported in the non-ICU subgroup of the UPMC Validation Cohort included in the original Sepsis-3 study [4]

Cohorts	Patients included in the analysis	Mortality	SIRS				qSOFA			
			Se	*Sp*	LR+	LR−	*Se*	*Sp*	LR+	LR−
SEPSIS-3	66,522	2,8%	64% [62–66%]	65% [65–66%]	1,83 [1,76-1,89]	0,55 [0,53-0,57]	55% [53–57%]	84% [84–84%]	3,44 [3,29-3,59]	0,54 [0,51-0,56]
This review	80,941	5,2%	86% [78–92%]	27% [15–41%]	1,17 [1,04-1,40]	0,56 [0,39-0,82]	51% [38–65%]	78% [66–88%]	2,49 [1,58–3,89]	0,62 [0,45–0,79]

This analysis shows relevant differences in the pre-test death probability increments implied by a positive result. Indeed, while the positivity of the SIRS criteria would result in only a 0.9% increment of the pre-test probability of death (from 5.2 to 6.1%), the positivity of the qSOFA score would imply a much greater increase (from 5.2 to 11.3%). As described in Table 3, the differences in lowering the pre-test probability of death are more marginal. Indeed, a negative result of the two tools would imply a reduction of the mortality risk from 5.2 to 3.3% for the qSOFA score, and to 2.8% for the SIRS criteria.

The quantification of heterogeneity across the included studies, as measured by Q test and I^2 statistics, shows extremely high levels of heterogeneity in the distribution of *Se* and *Sp* across the included studies (Fig. 2; Supplementary Table 3). Two subgroup meta-analyses comparing prospective with retrospective studies and studies using an "in-hospital mortality" outcome with studies using a "30-day (or 28-day) mortality" outcome do not lead to any appreciable reduction of heterogeneity (Supplementary Table 3).

Discussion

We focused our analysis on the non-ICU setting since the qSOFA was specifically designed to be used outside the ICU.

We chose to include only studies that directly compared both tools, rather than all eligible studies that have evaluated one or both tests. This methodology is more reliable, since it allows for a direct comparison, and is less likely to be biased due to potential confounders [30].

Our analysis shows that, in the non-intensive setting, the best rule-out clinical tool for mortality is the SIRS criteria, while the qSOFA score provides the highest *Sp* and LR+ (Table 2).

Any useful tool aimed at identifying sepsis and its severity should be applied to the subject with "suspected infection or sepsis." Although we put a lot of emphasis in achieving an unambiguous definition and the highest qualitative homogeneity for the selected cohorts, we find that even the few selected studies bear a high heterogeneity in the construct of their definitions for the target population (Tables 2 and 3 in the Supplementary Material). Another major reason for inconsistency across the studies, (which is likely correlated with the heterogeneity in the definition of the target population) lies in the different clinical characteristics of the included cohorts. This is illustrated well by the variability in the mortality rates among the different studies. Further heterogeneity across the studies is added by a mix of designs, settings, timeframes of data acquisition, and outcomes considered (Table 1). This evident burden of inconsistency translates into a relevant variability in the

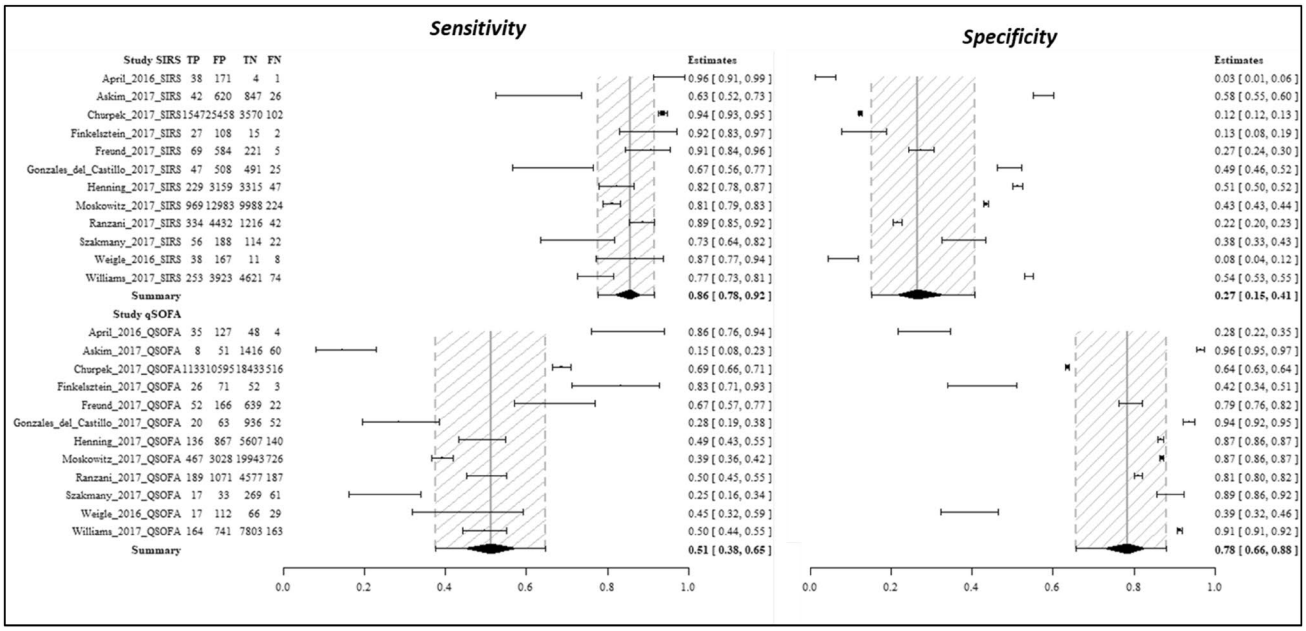

Fig. 2 Forest plots (bivariate model)

Table 3 Pre-test probabilities of death (represented by the mortality rates) and their changes, depending upon a positive (POS) or negative (NEG) result of the applied tool (according to the Bayesian analysis model), and screening performance of the SIRS criteria and the qSOFA score expressed as unidentified deaths for every 1000 screened cases (UDS_{1000}), for the cohorts included in this review and for the non-ICU subgroup of the UPMC Validation Cohort included in the original Sepsis-3 study [4]

Tool	Cohort	Mortality (pre-test prob)	Post-test prob if POS	Post-test prob if NEG	UDS_{1000}
SIRS	SEPSIS-3	2.8%	5.1%	1.6%	10
	This review	5.2%	6.1%	2.8%	7
qSOFA	SEPSIS-3	2.8%	9.1%	1.5%	13
	This review	5.2%	11.3%	3.3%	24

distribution of *Se* and *Sp* values, even within subgroups with shared study design (i.e., retrospective vs prospective; "in-hospital mortality" vs "30(or 28)-day mortality"). Indeed, we find remarkably high values of the I^2 index, which have also been reported by the three previous meta-analyses. This great heterogeneity that could jeopardize the meta-analytic approach [15] and limit the generalizability of the reported outcomes, demands, in our opinion, some considerations that might be valuable for future research. Since heterogeneity is inevitably inherent to a complex condition such sepsis, large high-quality prospective trials with well-defined target population, outcomes and subgroups are needed. These findings probably call for a standardized definition of "suspected infection or sepsis," (for both retrospective and prospective studies) that should guide the design of future trials.

We compared the results of our analysis to the results of the original SEPSIS-3 study, which led to the introduction of the qSOFA. The non-ICU cohort of the original SEPSIS-3 study by Seymour et al. [4] includes a sample size comparable to the number of patients included in our analysis (respectively 66,522 and 80,941 patients). The major differences between the fundamental prognostic measures of that cohort and those found in the present study are as follows (Table 2).

The SEPSIS-3 study reports: lower *Se* values for SIRS (64% [95% CI 62–66%] in the SESPSIS-3 study vs 86% [95% CI 78–92%] in the present study), higher *Sp* values for SIRS (65% [95% CI 65–66%] vs 27% [95% CI 15–41%]), and moderately (but not significantly) higher *Sp* values for qSOFA (84% [95% CI 84–84] vs 78% [95% CI 66–88%]). *Se* values for qSOFA are instead substantially equal (55% [95% CI 53–57%] vs 51% [95% CI 38–65%]). Thus, if we consider the data obtained by this meta-analysis as a reference for validation, the worse qSOFA rule-out performance, with respect to SIRS, highlighted by the studies that followed the SEPSIS-3 study is not due to a lower *Se* of the qSOFA *per se* but, rather, to an underestimation of SIRS *Se* in the original study. As a result, while the better *Sp* and LR+ of the qSOFA showed in the original SEPSIS-3 study are generally confirmed by our analysis (qSOFA *Sp* is only slightly overestimated in the SEPSIS-3 study), the gap in ruling-out capacity in favor of SIRS that this review documents appears

negligible in the SEPSIS-3 study. This is also confirmed by the comparison of UDS_{1000} values (Table 3): the differences between SIRS and qSOFA in terms of unidentified deaths at screening would be 3 (13 *minus* 10) for every 1000 screened patients according to the SEPSIS-3 study, while it would rise to 17 (24 *minus* 7) for every 1000 screened patients according to the pooled data of this review.

The relevant differences in the inherent cohort characteristics (well represented by their different mortality rates, reported in Table 1) and in the operative definition for suspected infection or sepsis used by the SEPSIS-3 study as compared to the studies included in this review (reported in Table 2 of the Supplementary Material), have certainly played a role in the 22% underestimation of SIRS *Se* that we noticed in the SEPSIS-3 study [4], as compared to the subsequent trials that we analyzed (Table 2). This leads to a much smaller *Se* gap between the two tools, which explains why in the SEPSIS-3 study the LR− value for qSOFA is even slightly better than that of SIRS (Table 2), and why, based on the result of that study, the post-test probability of death (Table 3) is even lower (1.5%) in case of a negative qSOFA than in case of a negative SIRS (1.6%). These findings of the original SEPSIS-3 study, in the face of the significantly better LR+ and area under the receiver operating characteristic curve (AUROC) values of the qSOFA also found in that study, might have induced the SEPSIS-3 Task Force to introduce the qSOFA, without major concerns about the dismission of the SIRS criteria.

A meta-analysis recently published by Serafim et al. [31] compares qSOFA vs SIRS performances in in-hospital mortality prediction. They use AUROC statistic and conclude that qSOFA is "slightly better" than SIRS in predicting in-hospital mortality. However, they do not consider any *predictive* measure for mortality.

Serafim's meta-analysis [31], as well as the SEPSIS-3 study, and also a large quote of the recently published studies on sepsis base their major outcome analysis on AUROC. Nonetheless, it must be considered that AUROC is a global measure of diagnostic accuracy used to assess the *discriminative* property of a test, rather than its *predictive* ability. *Discriminative* measures represent the best methods to assess the *global* performance of a diagnostic (or prognostic) test, and to minimize the test failure rates (i.e., the sum of false-negative and false-positive results) by identifying optimal cutoff points. In fact, they are greatly useful and widely used for decisions aimed at optimizing the yield of a health policy [10]. However, the AUROC *per se* tells nothing about the specific contribution of a test in ruling-in and ruling-out a diagnosis, or about the change in the outcome probability associated with the results of the test. In this instance, *predictive* measures, such as predictive values and LRs, are more meaningful than discriminative measures [10, 32]. This concept is clearly exemplified by the results reported by another meta-analysis by Song et al. [33]. Although in their study, *discrimination* for in-hospital mortality has similar AUROCs between the two tools, when they use *predictive* measures, it is clear that qSOFA has a far better rule-in power, while SIRS has a better rule-out capacity.

Any difference in *discriminative* measures (such as AUROC) found in the comparison of two different tests should be further explored by means of appropriate *predictive* measures and then judged in the light of the clinical context within which the tests are to be employed, on the basis of clinical, ethical, and economical considerations. This is especially important when dealing with life-threatening conditions, and the considered outcome is mortality [34–36].

Fernando and colleagues performed a large meta-analysis [37] finding a largely higher *Sp* for the qSOFA as compared to SIRS (80% vs 34%) with regard to in-hospital mortality of septic patients outside the ICU, while the SIRS has higher *Se* (82% for SIRS vs 51% for qSOFA). These values are remarkably similar to those found by Song JU et al. (where *Sp* is 83% for qSOFA vs 29% for SIRS, and *Se* is 86% for SIRS vs 51% for qSOFA) [33]. However, interestingly, while the former authors state that the qSOFA has only "moderate" specificity for prediction of mortality [37], Song and colleagues conclude that "A positive qSOFA score has a high specificity outside the ICU in early detection of in-hospital mortality" [33].

Why such discordance in judgement? The assessment of the adequacy of *Sp, Se* and their related *predictive* measures must take into account the real impact that these values imply in the specific clinical context they refer to. Thus, how can we effectively translate them into clinical practice?

The adequacy of the ~80% *Sp* reported by the two abovementioned meta-analyses [33, 37] for the qSOFA strongly depends on to what extent the application of the tool modifies the pre-test risk of death. This change of probability upon application of a prognostic tool is well interpreted by the Bayesian model, which is based on LR+. We calculated post-test probabilities of death according to the Bayesian analysis for both tools (Table 3) and find that the positivity of the qSOFA score would imply a relevantly higher increase in the pre-test risk of death as compared to the positivity of SIRS criteria. This analysis, which supports the message carried by *Sp* and LR+ values, provides a further and more practical quantification of the clinical impact of qSOFA, when it is used for the specific function that it was designed for (the prognostication of patients with acute infection outside the ICU).

As for *Se* values adequacy, the clinically meaningful (and simple) question is: how many deaths would cause a screening failure due to a lack of *Se*? We introduced an easily readable parameter, the UDS_{1000}, which—similarly to predictive values—is influenced by overall mortality rate in the included patients, and this must be acknowledged when

transferring the obtained results to settings with variable prevalence of the outcome of interest. Nonetheless, UDS_{1000} simply provides a direct answer to the above-mentioned question, which in our analysis confirms—but also quantifies—the most common criticism to the qSOFA, namely the inadequacy of this score as a rule-out tool for sepsis-associated mortality (Table 3).

Sepsis represents a great burden of disease. A recent meta-analysis estimates its annual global incidence at 31.5 million cases, with 19.4 million cases of severe sepsis, resulting in 5.3 million deaths [38, 39]. Sepsis contributes to as many as half of all inpatient deaths [40], and it was demonstrated that every hour of delayed treatment represents increasing mortality [41]. The life-threatening nature of sepsis and its rapid evolution potential call for a screening tool that allows for both early identification (preferably at triage), and timely selection of the most appropriate care. Based on our analysis, neither SIRS nor qSOFA fully achieves such features. There are other already existing clinical tools that might be implemented alternatively to the SIRS and qSOFA, which provide promising results, such as a National Early Warning Score (NEWS) ≥ 8, which in the study published by Churpek et al. achieves *Se* and *Sp* rates for mortality respectively of 80% and 60% [28].

However, regardless of the chosen tool, the performance of a clinical score is always dependent on the characteristics of the cohort under consideration. Thus, in a cohort of patients at possible moderate or high risk of death, the occurrence of acute infection should be treated promptly, even before the development of the diagnostic criteria for sepsis. Moreover, scores should be recalculated after the initiation of therapy, as they may give different results respect to the moment when an infection is first suspected, and no therapy has yet been initiated [28]. In this regard, a bedside score as simple as the qSOFA can be easily computed in multiple sequential assessments, enhancing its surveillance function.

The present study has some limitations. First, the high values of the I^2 index confirm relevant levels of inconsistency which, as above discussed, might jeopardize the consistency of a meta-analytic approach. This review highlights some methodological issues regarding the lack of standardization in trial design and reporting. The extreme construct variability found in the definitions reported in this review probably represents a major source of inconsistency.

Sepsis is an extraordinary complex condition and, to date, despite its hallmarks being first recognized over two millennia ago, there is still no gold standard for its diagnosis [42]. In our opinion, to overcome such complexity, we should start by tackling the heterogeneity of clinical research with high-quality prospective studies based on standardized and widely accepted definitions for the target population, and on meaningful *predictive* outcome measures.

The second important limitation is the relatively small number of selected studies which, in the face of such a great deal of diversification, prevent any in-depth analysis aimed at identifying and quantifying specific sources of heterogeneity.

Conclusions

Early prognostication of the patient with suspected infection is a major clinical challenge that we face every day in our professional life. Indeed, both the identification of sepsis among subjects presenting with suspected infection, and the determination of its severity, have crucial implications in terms of prognosis, diagnostic flow, management, and resource allocation [43]. Based on the pooled *Se, Sp* and LRs values, our analysis, which includes 80,941 subjects with suspected sepsis, confirms that, in the non-intensive setting, the SIRS criteria are a better rule-out clinical tool for mortality than the qSOFA score. Thus, it recommends against dismissing the SIRS criteria for screening outside the ICU.

Nonetheless, our analysis also strongly supports the continued use of qSOFA to identify adult patients with suspected infection more likely to have poor outcomes.

Thus, until a new and more efficient tool will be available, SIRS and qSOFA should probably not be deemed mutually exclusive. Rather, a two-step approach, based on an SIRS-based screening and repeated sequential qSOFA assessments for prognostication, would optimize the use of these two tools, taking advantage of their strong points and providing protection from their weaknesses.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Statement of human and animal rights This article does not contain any studies with human participants or animal performed by any of the authors.

Informed consent: None.

References

1. Shankar-Hari M, Phillips GS, Levy ML et al (2016) 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 315:775. https://doi.org/10.1001/jama.2016.0289
2. Peach BC (2017) Implications of the new sepsis definition on research and practice. J Crit Care 38:259–262. https://doi.org/10.1016/j.jcrc.2016.11.032

3. Marik PE, Taeb AM (2017) SIRS, qSOFA and new sepsis definition. J Thorac Dis 9:943–945. https://doi.org/10.21037/jtd.2017.03.125
4. Seymour CW, Liu VX, Iwashyna TJ et al (2016) Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 315:762. https://doi.org/10.1001/jama.2016.0288
5. Innocenti F, Tozzi C, Donnini C et al (2018) SOFA score in septic patients: incremental prognostic value over age, comorbidities, and parameters of sepsis severity. Intern Emerg Med 13:405–412. https://doi.org/10.1007/s11739-017-1629-5
6. Vincent JL, Martin GS, Levy MM (2016) qSOFA does not replace SIRS in the definition of sepsis. Crit Care. https://doi.org/10.1186/s13054-016-1389-z
7. Aublanc M, Richard J-C (2016) Assessment of clinical criteria for sepsis—was the cart put before the horse? J Thorac Dis 8:E816–E818. https://doi.org/10.21037/jtd.2016.07.51
8. Sprung CL, Schein RMH, Balk RA (2016) The new sepsis consensus definitions: the good, the bad and the ugly. Intensive Care Med 42:2024–2026. https://doi.org/10.1007/s00134-016-4604-0
9. Simpson SQ (2016) New sepsis criteria. Chest 149:1117–1118. https://doi.org/10.1016/j.chest.2016.02.653
10. Šimundić A-M (2009) Measures of diagnostic accuracy: basic definitions. EJIFCC 19:203–211
11. Liberati A, Altman DG, Tetzlaff J et al (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 62:e1–34. https://doi.org/10.1016/j.jclinepi.2009.06.006
12. Stroup DF, Berlin JA, Morton SC et al (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283:2008–2012
13. Levy MM, Fink MP, Marshall JC et al (2003) 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med 29:530–538. https://doi.org/10.1007/s00134-003-1662-x
14. Wells GA, Shea B, O'Connell D, et al The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 25 Mar 2018
15. Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560. https://doi.org/10.1136/bmj.327.7414.557
16. Guo J, Riebler A (2018) meta4diag : Bayesian bivariate meta-analysis of diagnostic test studies for routine practice. J Stat Softw. https://doi.org/10.18637/jss.v083.i01
17. Zamora J, Abraira V, Muriel A et al (2006) Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 6:31. https://doi.org/10.1186/1471-2288-6-31
18. April MD, Aguirre J, Tannenbaum LI et al (2017) Sepsis clinical criteria in emergency department patients admitted to an intensive care unit: an external validation study of quick sequential organ failure assessment. J Emerg Med 52:622–631. https://doi.org/10.1016/j.jemermed.2016.10.012
19. Moskowitz A, Patel PV, Grossestreuer AV et al (2017) Quick sequential organ failure assessment and systemic inflammatory response syndrome criteria as predictors of critical care intervention among patients with suspected infection. Crit Care Med 45:1813–1819. https://doi.org/10.1097/ccm.0000000000002622
20. Ranzani OT, Prina E, Menéndez R et al (2017) New Sepsis definition (Sepsis-3) and community-acquired pneumonia mortality. A validation and clinical decision-making study. Am J Respir Crit Care Med 196:1287–1297. https://doi.org/10.1164/rccm.201611-2262OC
21. Weigle H, Trigonis R, Bunn D et al (2016) 345 Severe Sepsis: What gets it right? A comparison of the quick sequential organ failure assessment versus systemic inflammatory response syndrome criteria. Ann Emerg Med 68:S132–S133. https://doi.org/10.1016/j.annemergmed.2016.08.361
22. Askim Å, Moser F, Gustad LT et al (2017) 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. https://doi.org/10.1186/s13049-017-0399-4
23. Finkelsztein EJ, Jones DS, Ma KC et al (2017) Comparison of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis outside the intensive care unit. Crit Care. https://doi.org/10.1186/s13054-017-1658-5
24. Freund Y, Lemachatti N, Krastinova E et al (2017) Prognostic accuracy of Sepsis-3 criteria for in-hospital mortality among patients with suspected infection presenting to the emergency department. JAMA 317:301. https://doi.org/10.1001/jama.2016.20329
25. representing the Infectious Disease Group of the Spanish Emergency Medicine Society (see addendum), González del Castillo J, Julian-Jiménez A et al (2017) Prognostic accuracy of SIRS criteria, qSOFA score and GYM score for 30-day-mortality in older non-severely dependent infected patients attended in the emergency department. Eur J Clin Microbiol Infect Dis 36:2361–2369. https://doi.org/10.1007/s10096-017-3068-7
26. Henning DJ, Puskarich MA, Self WH et al (2017) An emergency department validation of the SEP-3 sepsis and septic shock definitions and comparison with 1992 consensus definitions. Ann Emerg Med 70:544–552.e5. https://doi.org/10.1016/j.annemergmed.2017.01.008
27. Szakmany T, Pugh R, Kopczynska M et al (2018) Defining sepsis on the wards: results of a multi-centre point-prevalence study comparing two sepsis definitions. Anaesthesia 73:195–204. https://doi.org/10.1111/anae.14062
28. Churpek MM, Snyder A, Han X et al (2017) 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 195:906–911. https://doi.org/10.1164/rccm.201604-0854OC
29. Williams JM, Greenslade JH, McKenzie JV et al (2017) Systemic inflammatory response syndrome, quick sequential organ function assessment, and organ dysfunction. Chest 151:586–596. https://doi.org/10.1016/j.chest.2016.10.057
30. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0. The Cochrane Collaboration, 2010. http://srdta.cochrane.org/. Accessed 9 Aug 2018
31. Serafim R, Gomes JA, Salluh J, Póvoa P (2018) A comparison of the Quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality. Chest 153:646–655. https://doi.org/10.1016/j.chest.2017.12.015
32. Eusebi P (2013) Diagnostic accuracy measures. Cerebrovasc Dis 36:267–272. https://doi.org/10.1159/000353863
33. Song J-U, Sin CK, Park HK et al (2018) 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 22:28. https://doi.org/10.1186/s13054-018-1952-x
34. Seymour CW, Cooke CR, Heckbert SR et al (2013) Prehospital systolic blood pressure Thresholds: a community-based outcomes study. Acad Emerg Med 20:597–604. https://doi.org/10.1111/acem.12142

35. Perkins NJ, Schisterman EF (2006) The inconsistency of "Optimal" cut-points using two roc based criteria. Am J Epidemiol 163:670–675. https://doi.org/10.1093/aje/kwj063
36. Halligan S, Altman DG, Mallett S (2015) Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. Eur Radiol 25:932–939. https://doi.org/10.1007/s00330-014-3487-0
37. Fernando SM, Tran A, Taljaard M et al (2018) 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 168:266–275. https://doi.org/10.7326/M17-2820
38. Raith EP, Udy AA, Bailey M et al (2017) Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the Intensive care unit. JAMA 317:290–300. https://doi.org/10.1001/jama.2016.20328
39. Fleischmann C, Scherag A, Adhikari NKJ et al (2016) Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. Am J Respir Crit Care Med 193:259–272. https://doi.org/10.1164/rccm.201504-0781OC
40. Liu V, Escobar GJ, Greene JD et al (2014) Hospital deaths in patients with sepsis from 2 independent cohorts. JAMA 312:90–92. https://doi.org/10.1001/jama.2014.5804
41. Kumar A, Roberts D, Wood KE et al (2006) Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 34:1589–1596. https://doi.org/10.1097/01.CCM.0000217961.75225.E9
42. Churpek MM, Snyder A, Sokol S et al (2017) Investigating the impact of different suspicion of infection criteria on the accuracy of quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores. Crit Care Med 45:1805–1812. https://doi.org/10.1097/ccm.0000000000002648
43. Donnino MW, Fisher J, Fischer J (2006) Determining disease severity in severe sepsis and septic shock. Intern Emerg Med 1:219–220