RESEARCH ARTICLE

The timing of use of risk stratification tools affects their ability to predict mortality from sepsis. A meta-regression analysis. [version 1; peer review: 1 approved with reservations]

Molly Flint¹, Fergus Hamilton², David Arnold³, Edward Carlton⁴, David Hettle^{id}²

¹Medical School, University of Bristol, Bristol, Somerset, UK

²Department of Infection Sciences, North Bristol NHS Trust, Bristol, Somerset, UK

³Academic Respiratory Unit, North Bristol NHS Trust, Bristol, Somerset, UK

⁴Emergency Department, North Bristol NHS Trust, Bristol, Somerset, UK

V1 First published: 01 Dec 2021, 6:328
<https://doi.org/10.12688/wellcomeopenres.17223.1>
Latest published: 01 Dec 2021, 6:328
<https://doi.org/10.12688/wellcomeopenres.17223.1>

Abstract

Background:

Risk stratification tools (RSTs) are used in healthcare settings to identify patients at risk of sepsis and subsequent adverse outcomes. In practice RSTs are used on admission and thereafter as 'trigger' tools prompting sepsis management. However, studies investigating their performance report scores at a single timepoint which varies in relation to admission. The aim of this meta-analysis was to determine if the predictive performance of RSTs is altered by the timing of their use.

Methods:

We conducted a systematic review and meta-regression analysis of studies published from inception to 31 October 2018, using EMBASE and PubMed databases. Any cohort studies investigating the ability of an RST to predict mortality in adult sepsis patients admitted to hospital, from which a 2x2 table was available or could be constructed, were included. The diagnostic performance of RSTs in predicting mortality was the primary outcome. Sensitivity, specificity, positive predictive value, negative predictive value and area under the receiver-operating curve (AUROC) were the primary measures, enabling further meta-regression analysis.

Results:

47 studies were included, comprising 430,427 patients. Results of bivariate meta-regression analysis found tools using a first-recorded score were less sensitive than those using worst-recorded score (REML regression coefficient 0.57, 95% CI 0.07-1.08). Using worst-

Open Peer Review

Approval Status ?

1

version 1

01 Dec 2021

?

[view](#)

1. **Narani Sivayoham** ^{id}, St George's
University Hospitals NHS Foundation Trust,
London, UK

Any reports and responses or comments on the article can be found at the end of the article.

recorded score led to a large increase in sensitivity (summary sensitivity 0.76, 95% CI 0.67-0.83, for worst-recorded scores vs. 0.64 (0.57-0.71) for first-recorded scores). Scoring system type did not have a significant relationship with studies' predictive ability. The most analysed RSTs were qSOFA (n=37) and EWS (n=14). Further analysis of these RSTs also found timing of their use to be associated with predictive performance.

Conclusion:

The timing of any RST is paramount to their predictive performance. This must be reflected in their use in practice, and lead to prospective studies in future.

Keywords

Sepsis, qSOFA, Infection, EWS, Scoring, Risk stratification tool

Corresponding author: David Hettle (d.hettle@nhs.net)

Author roles: **Flint M:** Data Curation, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Hamilton F:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Writing – Review & Editing; **Arnold D:** Conceptualization, Supervision, Writing – Review & Editing; **Carlton E:** Supervision, Writing – Review & Editing; **Hettle D:** Conceptualization, Investigation, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Flint M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Flint M, Hamilton F, Arnold D *et al.* **The timing of use of risk stratification tools affects their ability to predict mortality from sepsis. A meta-regression analysis.** [version 1; peer review: 1 approved with reservations] Wellcome Open Research 2021, 6:328 <https://doi.org/10.12688/wellcomeopenres.17223.1>

First published: 01 Dec 2021, 6:328 <https://doi.org/10.12688/wellcomeopenres.17223.1>

Introduction

Background

Sepsis is a major global health issue, with 48.9 million cases per year worldwide resulting in 11 million deaths^{1–3}. In 2016, the SEPSIS-3 task force identified the need for “earlier recognition and more timely management”⁴. Early detection, coupled with rapid management improves patient outcomes^{4–8} by reducing progression to severe sepsis, the need for intensive care unit (ICU) care, and mortality⁹. Early detection also enables identification of those in whom a good outcome is likely, allowing more conservative management strategies¹⁰. Early identification is therefore a priority, but this is challenging due to variation in clinical presentations of sepsis and the lack of a single, gold-standard diagnostic test^{11,12}. A variety of risk stratification tools (RSTs) are used across multiple healthcare settings to identify patients at risk of sepsis and subsequent adverse outcomes. Using physiological and biochemical parameters RSTs aim to predict patients at risk of adverse outcomes from sepsis, a deteriorating condition¹³, or a specific infective diagnosis¹⁴. Examples include the: Early Warning Score (EWS)¹⁵, with numerous variants; Sequential Organ Failure Assessment score (SOFA)¹⁴; Quick-SOFA (qSOFA)⁴; Systemic Inflammatory Response Syndrome criteria (SIRS)¹⁶; and organ-specific scores, such as CURB-65¹⁷.

Importance

Inconsistencies in the research evidencing RSTs and their use in clinical practice potentially leads to exaggeration of their predictive ability^{18,19}. A major variation is in the timing of the use of RSTs. In practice RSTs are often used on admission to hospital and in the emergency department (ED) to guide management, and thereafter as trigger tools to monitor patient deterioration. However, studies investigating RSTs vary in their timing that the tool was performed in relation to admission, indeed a large number of studies use a worst-recorded score within 24–48 hours rather than reflecting the way RSTs are used in practice, which is likely to affect their reported performance.

Goals of this investigation

The primary objective of this meta-analytical study was to determine if the predictive performance of RSTs is altered by the timing of their use.

Methods

The protocol for this study was registered with PROSPERO (CRD42019146321).

Study selection and inclusion

A systematic search of EMBASE and PubMed databases from inception until October 2018 was undertaken. We aimed to include any prognostic tool identified by the search that was mainly based on physiological markers. RSTs to be included were not pre-defined, rather guided by the search strategy findings. The search identified primary research which investigated any RST used to predict mortality in adult patients (>18yo) admitted to hospital with sepsis, suspected sepsis or pneumonia. The search strategy combined terms for RSTs with terms for infection, sepsis and mortality.

Studies must have reported an RST score calculated on patients' hospital admission or during the initial period of observation following admission for suspected infection or sepsis. Mortality was a required outcome measure of included studies, which was primarily defined as 28-day or 30-day mortality, however other mortality measures (e.g. in-hospital mortality) were accepted. Mortality was selected as the use of RSTs have been driven by campaigns such as Surviving Sepsis⁸, advocating for early identification of septic patients to reduce mortality as far as possible. Studies in a non-ICU setting were included as several RSTs are validated for use in patients at risk of deterioration or sepsis in this setting. Studies conducted exclusively in ICU, and those investigating only trauma, paediatric, obstetric and gynaecological patients, or those with alternative diagnoses to sepsis were excluded. Language of articles was limited to English unless a translation was available. Any article that was not an observational cohort study was excluded.

Two authors (MF, DH) independently screened identified titles and abstracts using a validated web-based application, Rayyan²⁰, to produce a list of relevant articles for full-text review. Disagreements were resolved through consensus between further authors (FH, DA).

Data extraction and outcomes

The data extraction recorded: (i) RST used, (ii) timing of RST use, (iii) study design, (iv) sample size, (v) healthcare setting, (vi) geographical location, (vii) diagnosis, (viii) outcome measure for mortality, (ix) sensitivity, (x) specificity, (xi) positive predictive value, (xii) negative predictive value and (xiii) area under the receiver-operator curve (AUROC). If a study did not publish 2×2 data, this data was calculated using specificity, sensitivity, number of participants and mortality. Where studies published more than one set of 2×2 data for a single score, e.g. for multiple cut-offs, the best was taken. If identified studies investigated the use of more than one RST, all 2×2 datasets were collected and included in the analysis. Timing data was classified as ‘first recorded in ED’, ‘worst in ED’, ‘within 24 hours’, ‘within 48 hours’, or other. We then used a pragmatic approach, categorising timing further into ‘first recorded scores’, ‘worst recorded within a time point scores’ (hereafter ‘worst-recorded’) and ‘other’ for analysis. Mortality measures were coded as ‘28- or 30-day’, ‘in-hospital’ or ‘other’. As there are many Early Warning Score (EWS) variants which are all physiological scoring tools with broadly similar scoring parameters, we pragmatically cohorted these scores for analysis, allowing more powerful analysis. One author (MF) extracted data from eligible studies, which were confirmed by a second author (DH).

Risk of bias assessment

This study did not aim to include formal risk of bias assessment, as the primary objective was to quantify whether timing impacted RST performance, which itself is a major bias. All studies that used later time points would be at significant risk of bias due to the later nature of timing used (measurements occurring after diagnosis and subsequent management), and we did not feel that formally evaluating this would be helpful.

Statistical analysis

The fundamental approach taken was a meta-regression of study performance, with timing of score as the variable of interest, for all RSTs, and then for individual RSTs. All statistical analysis was performed in R 3.6.0 and 4.00, using the package *mada* (v 0.5.1)²¹.

Initial analysis generated hierarchical summary receiver-operator characteristic (HSROC) plots, allowing for comparison of the effects of all relevant covariates across studies. Using the HSROC plots we were able to visually assess heterogeneity. We then conducted bivariate meta-regression analysis to

assess the impact of the covariates on the result of the meta-analysis, the predictive ability of RSTs.

Thereafter for any RST with adequate data for analysis, bootstrapped AUC, and summary sensitivity and specificity were generated for studies investigating the effects of different timings of RST use to enable clinical comparisons. Bootstrapping was performed 1,000 times.

Results

Study characteristics

The study selection process is illustrated in [Figure 1](#). Of the 12,853 studies identified through searches, 47 were deemed

Figure 1. PRISMA Diagram of Evidence Search and Selection.

appropriate for final inclusion, including 430,427 patients. Identified studies included a total of 113 uses of RSTs, which represented 17 different tools.

Table 1 outlines the study characteristics. The earliest study was from 2007, with the most recent published in 2018. 14 studies focussed on a single RST with 33 assessing multiple tools. The distribution of the RSTs analysed is also demonstrated in Figure 1. In terms of timing, 23 studies calculated

RST score based on first recorded score, five studies reported score within 24 hours of admission, four studies reported the worst score whilst in ED, eight studies did not state when they completed a RST and seven studies used other timing scales.

Overall, 24 studies used the primarily defined outcome measure of 28- or 30-day mortality, with a further 22 using in-hospital mortality and one coded as 'other', using a definition of in-hospital mortality within 72 hours of admission. Average

Table 1. Study characteristics.

Lead author	Year	n	Mortality	Setting	Diagnostic group	Score used	Sensitivity	Specificity	AUROC
Chen ²²	2018	69115	8.16%	ED	Suspected infection	qSOFA	29%	92%	0.69
						SIRS	67%	46%	0.60
Szakmany ²³	2018	380	20.5%	ED	Suspected infection or infection	NEWS	41%	73%	0.59
						SOFA	86%	32%	0.7
						qSOFA	22%	89%	0.57
Redondo-González ²⁴	2018	349	21.8%	ED	Sepsis	SOFA	83%	42%	0.72
						qSOFA	65%	58%	0.67
						EWS	91%	41%	0.73
Aluisio ²⁵	2018	760	25.4%	ED	Infection	qSOFA	86%	43%	0.70
Geier ²⁶	2013	151	14.6%	ED	Suspected sepsis	MEWS	42.9%	74.4%	0.642
						MEDS	85.7%	68.2%	0.871
Quinten ²⁷	2018	193	3.6%	ED	Suspected infection	qSOFA	71.4%	84.4%	0.848
						PIRO	85.7%	76.9%	0.784
Canet ²⁸	2018	11205	4.5%	ED	Suspected infection	qSOFA	61%	80%	0.76
Williams ²⁹	2017	8871	3.7%	ED	Suspected infection	SIRS	72.1%	61%	0.72
						qSOFA	29.7%	96.1%	0.73
Rodriguez ³⁰	2018	3743	8.8%	ED	Suspected infection	qSOFA	64.4%	83.9%	0.788
Freund ³¹	2017	879	8%	ED	Suspected infection	SOFA	73%	70%	0.77
						qSOFA	70%	79%	0.80
						SIRS	93%	27%	0.65
Müller ³²	2017	527	13.3%	ED	Pneumonia	qSOFA	26.6%	88.3%	0.587
						SIRS	15.7%	97%	0.497
						CURB-65	19.2%	90.9%	0.65
Zhou ³³	2018	226	21.68%	ED	Pneumonia	SOFA	91.5%	81.6%	0.852
						qSOFA	72.3%	71.4%	0.724
						CURB-65	65%	87.8%	0.805
						PSI	83.1%	67.3%	0.81
Hwang ¹⁸	2018	1395	15%	ED	Severe sepsis or septic shock	qSOFA	39%	77%	0.58
						qSOFA	82%	41%	0.60
						qSOFA	91%	23%	0.57

Lead author	Year	n	Mortality	Setting	Diagnostic group	Score used	Sensitivity	Specificity	AUROC
Shu ³⁴	2019*	2292	1.4%	ED	Sepsis or infection	qSOFA	40.6%	91.9%	NR
Chen ³⁵	2011	110	43.8%	Ward	Infection	SOFA	67.3%	85.2%	0.845
						APACHE II	75.5%	73.8%	0.806
Churpek ³⁶	2017	30677	5.4%	ED	Suspected infection	NEWS	71.9%	72.2%	0.71
						MEWS	71.4%	65%	0.66
						qSOFA	68.7%	63.5%	0.63
						SIRS	77.5%	43.8%	0.69
Askim ³⁷	2017	1535	2.6%	ED	Severe sepsis	SIRS	65%	55%	0.6048
Barlow ³⁸	2007	419	19%	Ward	Pneumonia	CRB-65	73%	59%	0.73
Guirgis ³⁹	2017	3297	10.1%	ED	Sepsis	SOFA	90%	50%	0.82
						qSOFA	38%	86%	0.68
Park ⁴⁰	2017	1009	15.8%	ED	Suspected infection	qSOFA	53%	84%	0.733
Wang ⁴¹	2016	477	27.5%	ED	Infection	qSOFA	42.9%	82.6%	0.666
Cildir ⁴²	2013	230	32.2%	ED	Sepsis	MEWS	87.5%	30.4%	0.574
						MEWS	48.5%	67%	0.596
Haydar ⁴³	2017	200	11.1%	ED	Suspected sepsis	qSOFA	90.1%	45.7%	0.68
						SIRS	95.5%	5.6%	0.51
Tokioka ⁴⁴	2018	1045	4.9%	ED	Pneumonia	qSOFA	39.1%	87.8%	0.69
						CURB-65	87.5%	41%	0.75
						PSI	89.1%	42%	0.74
van der Woude ⁴⁵	2018	577	3.6%	ED	Infection	MEWS	23.8%	87%	NR
						SOFA	66.7%	79.8%	NR
						qSOFA	33.3%	96.4%	NR
						SIRS	61.9%	56.9%	NR
Henning ⁴⁶	2017	7637	4.4%	ED	Suspected infection	qSOFA	52%	86%	0.77
						SIRS	83%	50%	NR
Goulden ⁴⁷	2018	1818	15%	ED	Suspected sepsis	NEWS	74%	43%	0.6517
						qSOFA	37%	79%	0.6271
						SIRS	80%	21%	0.4891
Raymond ⁴⁸	2019*	228	11%	ED	Suspected sepsis	mSOFA	88%	67.5%	NR
Choi ⁴⁹	2019*	991	22.3%	ED	Sepsis	qSOFA	65.6%	54.8%	0.62
						SIRS	91.9%	11%	0.482
Chen ⁵⁰	2016	1641	33%	ED	Pneumonia	qSOFA	53%	75%	0.655
						CRB-65	70%	57%	0.661
						CRB	36%	81%	0.651
Gaini ⁵¹	2019*	323	7%	Ward	Infection	SOFA	100%	44%	0.83
						qSOFA	38%	89%	0.67
						SIRS	81%	28%	0.61

Lead author	Year	n	Mortality	Setting	Diagnostic group	Score used	Sensitivity	Specificity	AUROC
González Del Castillo ⁵²	2017	10776	6.5%	ED	Infection	qSOFA	27.8%	93.7%	0.69
						SIRS	65.3%	49.1%	0.65
Hifumi ⁵³	2016	171	17%	Ward	Infection	SOFA	97%	84%	0.954
Camm ⁵⁴	2018	316	7.91%	ED	Suspected infection	NEWS	44%	73.5%	NR
						qSOFA	84%	44.3%	NR
						SIRS	56%	67%	NR
Rannikko ⁵⁵	2018	481	14%	ED	Infection	qSOFA	77.3%	77.3%	0.72
de Groot ⁵⁶	2017	2280	9.5%	ED	Suspected infection	NEWS	63%	63%	0.67
						MEWS	42%	77%	0.63
						qSOFA	83%	47%	0.68
						PIRO	55%	77%	0.73
						MEDS	81%	62%	0.8
Ranzani ⁵⁷	2017	6024	6.4%	ED	Pneumonia	qSOFA	50%	81%	0.697
						SIRS	89%	22%	0.579
						CURB-65	78%	60%	0.746
						PSI	92%	47%	0.78
						CRB	40%	87%	0.716
Chen ⁵⁸	2014	680	26.2%	ED	Sepsis	CRB	57.9%	89.4%	0.74
Lee ⁵⁹	2016	36	25%	ED	Sepsis	SOFA	67%	85%	0.815
Huson ⁶⁰	2017	458	23%	Ward	Suspected infection	qSOFA	72%	68%	0.73
Jo ⁶¹	2016	533	10.8%	ED	Pneumonia	NEWS	68.3%	57.2%	0.70
						CURB-65	71.7%	52.7%	0.66
						PSI	75%	47.3%	0.68
Kim ⁶²	2017	125	10.4%	ED	Pneumonia	SOFA	69.2%	83.9%	0.83
						qSOFA	53.9%	89.3%	0.81
						CURB-65	53.9%	83.4%	0.77
						PSI	100%	49.1%	0.86
						APACHE II	69.2%	77.7%	0.85
Tirotta ⁶³	2017	526	14.8%	ED	Infection	MEWS	55%	59%	0.596
Vaittinada Ayar ⁶⁴	2018	332	27%	ED	Suspected infection	qSOFA	60%	67%	0.69
Kofoed ⁶⁵	2008	151	5.96%	ED	Infection	SOFA	44%	95%	0.80
						SAPS II	100%	68%	0.89
Lafon ⁶⁶	2018	374	13%	Ward	Sepsis	qSOFA	74%	87%	0.8
Osatnik ⁶⁷	2018	157	14%	ED	Sepsis or infection	qSOFA	63.64%	67.41%	0.65
						SIRS	81.82%	24.44%	0.53

* e-published ahead of print by October 2018

mortality of studies investigating the primary outcome of 28- or 30-day mortality was 8.57%. Extending this to all included studies with any mortality end point, mortality was 7.59%.

Study performance

There was heterogeneity in overall performance across different scoring systems across all studies, with sensitivities and specificities ranging from nearly 0 to 100, as demonstrated in [Figure 2](#). Heterogeneity was also formally assessed by Chi-square goodness of fit ($p < 0.001$). The summary AUC across all RSTs was 0.72 (0.70–0.74), with summary sensitivity of 0.66 (0.62–0.70), and summary specificity of 0.69 (0.64–0.74).

Meta-regression

In a bivariate analysis using timing data alone, scores using first-recorded timing were less sensitive than worst-recorded scores (REML regression coefficient 0.57 (0.07–1.08)), with a trend towards increased specificity. This corresponds to a large increase in sensitivity with later timing, with summary sensitivity in first-recorded RSTs of at 0.64 (0.57–0.71) compared to summary sensitivity of worst-recorded RSTs at 0.76 (0.67–0.83). [Figure 3](#) demonstrates this graphically, comparing studies using first-recorded scores than those using worst-recorded scores.

In the subsequent analysis including all relevant study level covariates only three study level factors significantly impact study performance: mortality, LMIC setting, and usage of a worst-recorded timing approach (eTable 2 in the supplement). Increasing mortality in a study was associated with increased sensitivity and reduced specificity, and this was also true for studies reporting worst-recorded timings. Studies in LMIC settings also had higher specificity. No other factor (including scoring system type) was found to have a significant relationship with studies' predictive ability.

Individual scores

For the two scores with sufficient data individually to generate meaningful outputs (qSOFA and EWS), further analysis was performed. For both these scores, a similar relationship was found between timing and study performance, although neither met statistical significance for either sensitivity or specificity, due to low study numbers.

Discussion

Summary of findings

The meta-analysis of observational cohort studies aimed to investigate whether the predictive performance of RSTs is altered by the timing of their use. Our search strategy identified 47 studies including over 430,000 patients, with qSOFA the most analysed RST. It demonstrates that the timing of performing a RST affects its predictive performance: studies utilising any RST based on first-recorded patient observations and laboratory results have a lower sensitivity than RSTs which are based on worst-recorded values in order to predict infection or sepsis-related mortality. From sub-group analysis, the effect of timing was particularly evident in studies investigating the

various EWS derivatives. Alongside timing, only LMIC setting significantly impacted study performance, and no significant differences were identified between RSTs.

Clinical context

Given the global burden of sepsis, ensuring rigorous validation of the research underpinning the tools used to allow the early identification and diagnosis of patients with suspected sepsis as advocated by SEPSIS-3⁴, is essential⁶⁸. Failure to do so may result in unnecessary sepsis-related morbidity and mortality. However, several criticisms have been made of RSTs' use in sepsis, concerning the research evidencing their use and their application in clinical practice^{10,18,36,69}. These largely focus on either a lack of specificity or sensitivity in different settings and cohorts, and it remains clear that no one tool fully achieves a gold-standard for diagnosis^{12,36,68,70,71}. In particular, Sepsis-3 and the qSOFA score have been criticised for their largely retrospective evidence base⁷², with variable predictive performance when investigated prospectively^{73–75}. Other RSTs, such as EWS, were developed to monitor patients for signs of deterioration⁷⁶ and are increasingly used as risk-stratifying tools for deterioration of any cause in hospital inpatients, endorsed by the Royal College of Physicians⁷⁷ and NICE¹⁴ despite concern into their prognostic value¹⁰.

Song *et al.*¹² note that there is lower predictive performance if tools such as qSOFA are completed at the initial suspicion for infection. Further, while Hamilton *et al.*¹⁰ have investigated the role of EWS in predicting sepsis-related mortality, we are not aware of other meta-analyses clearly demonstrating the critical role that timing plays in RSTs' predictive role.

Given these challenges in the role of RSTs and in diagnosing sepsis accurately, Franchini *et al.*⁶⁸ advocate for more prospective studies and outcome measures. Our study supports this, in recognising that the methods of investigating RSTs as indicators for management of sepsis must be prospective and consider predictive outcome measures, rather than the current body of evidence which is largely retrospective. Should we continue to rely on tools whose use is evidenced based on timings other than when they are used in clinical practice, that is the worst-recorded approach to researching RSTs, we risk overstating their predictive ability in identifying a condition which has potentially fatal consequences.

Limitations and strengths

There are some limitations to our work. There is wide heterogeneity between case definitions used in studies investigating RSTs' use as well as in the mortality outcome measure reported across studies, a reflection of the likely clinical heterogeneity in our sample. This may impact the consistency of our meta-analysis, although it is not unusual to identify large amounts of heterogeneity in systematic analysis of studies investigating diagnostic test accuracy⁷⁸, though no clear consistent source was identified which can singly explain the heterogeneity across studies. As a meta-analysis our results may be affected by publication bias or selective reporting of datasets within included studies, however the PRISMA statement on

0.05 0.29 0.52 0.76 1.00

Sensitivity

Figure 2. Forest plots summarising the sensitivity and specificity in analysed RSTs.

Figure 3. Comparison of studies using first-recorded timings versus studies using worst-recorded timings.

meta-analysis of diagnostic test accuracy studies⁷⁹ concluded that there is less risk of these biases than in reviews of primary interventional studies, as well as no adequately powered statistical test for a comprehensive assessment of bias in studies such as ours. Risk of bias assessment is challenging in this setting, as a major bias in all studies is that physiological responses change with treatment. As such, all studies that used later time points are at significant risk of bias, which is the focus of this study. Having said that the strengths of our study lie primarily in the use of a comprehensive search strategy across multiple databases, resulting in a large cohort of 430,427 patients being included in the final analysis. Using this large cohort and bivariate random effects analysis we hope to have minimised biases often present in observational cohorts, allowing the construction of HS-ROC curves. The large number of identified studies also allowed sub-group analysis of qSOFA and EWS in further detail, delivering greater insight into their use as RSTs and the impact of timing. Though there is suspicion regarding the lack of high-quality prospective studies investigating RSTs' use, that this analysis

includes a significant number of prospective studies reinforces that these findings are likely generalisable to either mode of study design.

Implications for Clinical Practice and Research

Our findings outline that clinicians must recognise that the performance of any RST is largely related to when they are used, not the individual tool. Given the low specificity of first-recorded RSTs and the impact that research driven by worst-recorded scores has had on policy design in potentially infected or septic patients, leading to the wide use of RSTs worldwide, there must be more prospective studies investigating their role and predictive value aligned with their use in practice: at the point of admission, or first suspicion of infection. In most settings, where infection and sepsis are priorities and the threat of antibiotic resistance looms large, failure to investigate the RSTs evidencing early management and antibiotic administration could lead to inaccurate identification of patients at risk of sepsis and inappropriate use of antimicrobials.

Conclusion

In summary, we must be rigorous in ensuring that the tools and scores used to predict sepsis-related mortality, and enable management and treatment decisions are used and evidenced appropriately. It remains challenging to determine how effective RSTs are in this role, as the timing of RSTs' use in the evidence base is varied⁸⁰, often reflecting a worst-recorded in a time point approach, unlike their use in clinical practice. This meta-analysis has shown that the timing of RSTs is paramount to their predictive performance. This has important implications for their use in practice and stresses the importance of prospective studies in the future.

Data availability

Underlying data

Underlying Data for "The timing of use of clinical screening tools affects their ability to predict sepsis mortality. A meta-regression analysis." DOI: <https://doi.org/10.5281/zenodo.5519552>⁸¹

This project contains the following underlying data:

- Author link.csv
- Data additions.docx

- Master_document_for_generating_figures.Rmd
- Master_final.csv
- PRISMA-2009-Checklist-MS-Word WOR.doc

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Authors' contributions

DH, FH and DA conceived the idea. MF, DH and FH undertook abstract screening and data extraction. FH, DA and EC acted as content experts in the field of sepsis and RSTs. MF and DH drafted the manuscript and all authors contributed to its editing and revision. All authors interpreted data and approved the final version of the manuscript.

Acknowledgements

We would like to thank Sarah Rudd, MSc, North Bristol NHS Trust clinical librarian for assistance in conducting the literature searches.

References

- World Health Organisation: **Sepsis**. 2018; [updated 19 April 2018]. [Reference Source](#)
- Rudd KE, Johnson SC, Agesa KM, *et al.*: **Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study**. *Lancet*. 2020; **395**(10219): 200–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Angus DC, Linde-Zwirble WT, Lidicker J, *et al.*: **Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care**. *Crit Care Med*. 2001; **29**(7): 1303–10. [PubMed Abstract](#) | [Publisher Full Text](#)
- Singer M, Deutschman CS, Seymour CW, *et al.*: **The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)**. *JAMA*. 2016; **315**(8): 801–10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rivers E, Nguyen B, Havstad S, *et al.*: **Early goal-directed therapy in the treatment of severe sepsis and septic shock**. *N Engl J Med*. 2001; **345**(19): 1368–77. [PubMed Abstract](#) | [Publisher Full Text](#)
- Nguyen HB, Corbett SW, Steele R, *et al.*: **Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality**. *Crit Care Med*. 2007; **35**(4): 1105–12. [PubMed Abstract](#) | [Publisher Full Text](#)
- 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**(23): 2235–44. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhodes A, Evans LE, Alhazzani W, *et al.*: **Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016**. *Intensive Care Med*. 2017; **43**(3): 304–77. [PubMed Abstract](#) | [Publisher Full Text](#)
- Goulden R, Hoyle MC, Monis J, *et al.*: **qSOFA, SIRS and NEWS for predicting in-hospital mortality and ICU admission in emergency admissions treated as sepsis**. *Emerg Med J*. 2018; **35**(6): 345–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Hamilton F, Arnold D, Baird A, *et al.*: **Early Warning Scores do not accurately predict mortality in sepsis: A meta-analysis and systematic review of the literature**. *J Infect*. 2018; **76**(3): 241–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Carrigan SD, Scott G, Tabrizian M: **Toward resolving the challenges of sepsis diagnosis**. *Clin Chem*. 2004; **50**(8): 1301–14. [PubMed Abstract](#) | [Publisher Full Text](#)
- 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**(1): 28. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Alam N, Hobbelink EL, van Tienhoven AJ, *et al.*: **The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review**. *Resuscitation*. 2014; **85**(5): 587–94. [PubMed Abstract](#) | [Publisher Full Text](#)
- National Institute of Clinical Excellence: **Sepsis: Recognition, Diagnosis and Early Management**. London: NICE; 2016. [Reference Source](#)
- NHS England: **National Early Warning Score (NEWS)**. 2019. [Reference Source](#)
- Levy MM, Fink MP, Marshall JC, *et al.*: **2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference**. *Intensive Care Med*. 2003; **29**(4): 530–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- National Institute of Clinical Excellence: **Pneumonia in adults: diagnosis and management**. In: Health, editor. 2014. [Reference Source](#)
- Hwang SY, Jo JJ, Lee SU, *et al.*: **Low Accuracy of Positive qSOFA Criteria for Predicting 28-Day Mortality in Critically Ill Septic Patients During the Early Period After Emergency Department Presentation**. *Ann Emerg Med*. 2018; **71**(1): 1–9.e2. [PubMed Abstract](#) | [Publisher Full Text](#)
- Moskowitz A, Andersen LW, Cocchi M, *et al.*: **The Misapplication of Severity-of-Illness Scores toward Clinical Decision Making**. *Am J Respir Crit Care Med*. 2016; **194**(3): 256–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Ouzzani M, Hammady H, Fedorowicz Z, *et al.*: **Rayyan-a web and mobile app for systematic reviews**. *Syst Rev*. 2016; **5**(1): 210. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

21. Doeblner P, Holling H: **Meta-analysis of diagnostic accuracy with mda.** [Reference Source](#)
22. Chen FC, Kung CT, Cheng HH, *et al.*: **Quick Sepsis-related Organ Failure Assessment predicts 72-h mortality in patients with suspected infection.** *Eur J Emerg Med.* 2019; **26**(5): 323–8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Szakmany T, Pugh R, Kopczynska M, *et al.*: **Defining sepsis on the wards: results of a multi-centre point-prevalence study comparing two sepsis definitions.** *Anaesthesia.* 2018; **73**(2): 195–204. [PubMed Abstract](#) | [Publisher Full Text](#)
24. Redondo-González A, Varela-Patiño M, Álvarez-Manzanares J, *et al.*: **[Assessment of the severity scores in patients included in a sepsis code in an Emergency Department].** *Rev Esp Quimioter.* 2018; **31**(4): 316–22. [PubMed Abstract](#) | [Free Full Text](#)
25. Aluisio AR, Garbern S, Wiskel T, *et al.*: **Mortality outcomes based on ED qSOFA score and HIV status in a developing low income country.** *Am J Emerg Med.* 2018; **36**(11): 2010–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Geier F, Popp S, Greve Y, *et al.*: **Severity illness scoring systems for early identification and prediction of in-hospital mortality in patients with suspected sepsis presenting to the emergency department.** *Wien Klin Wochenschr.* 2013; **125**(17–18): 508–15. [PubMed Abstract](#) | [Publisher Full Text](#)
27. Quinten VM, van Meurs M, Wolffensperger AE, *et al.*: **Sepsis patients in the emergency department: stratification using the Clinical Impression Score, Predisposition, Infection, Response and Organ dysfunction score or Quick Sequential Organ Failure Assessment score?** *Eur J Emerg Med.* 2018; **25**(5): 328–34. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Canet E, Taylor DM, Khor R, *et al.*: **qSOFA as predictor of mortality and prolonged ICU admission in Emergency Department patients with suspected infection.** *J Crit Care.* 2018; **48**: 118–23. [PubMed Abstract](#) | [Publisher Full Text](#)
29. Williams JM, Greenslade JH, McKenzie JV, *et al.*: **Systemic Inflammatory Response Syndrome, Quick Sequential Organ Function Assessment, and Organ Dysfunction: Insights From a Prospective Database of ED Patients With Infection.** *Chest.* 2017; **151**(3): 586–96. [PubMed Abstract](#) | [Publisher Full Text](#)
30. Rodriguez RM, Greenwood JC, Nuckton TJ, *et al.*: **Comparison of qSOFA with current emergency department tools for screening of patients with sepsis for critical illness.** *Emerg Med J.* 2018; **35**(6): 350–6. [PubMed Abstract](#) | [Publisher Full Text](#)
31. Freund Y, Lemachatti N, Krastinova E, *et al.*: **Prognostic Accuracy of Sepsis-3 Criteria for In-Hospital Mortality Among Patients With Suspected Infection Presenting to the Emergency Department.** *JAMA.* 2017; **317**(3): 301–8. [PubMed Abstract](#) | [Publisher Full Text](#)
32. Muller M, Guignard V, Schefold JC, *et al.*: **Utility of quick sepsis-related organ failure assessment (qSOFA) to predict outcome in patients with pneumonia.** *PLoS One.* 2017; **12**(12): e0188913. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Zhou H, Guo S, Lan T, *et al.*: **Risk stratification and prediction value of procalcitonin and clinical severity scores for community-acquired pneumonia in ED.** *Am J Emerg Med.* 2018; **36**(12): 2155–60. [PubMed Abstract](#) | [Publisher Full Text](#)
34. Shu E, Ives Tallman C, Frye W, *et al.*: **Pre-hospital qSOFA as a predictor of sepsis and mortality.** *Am J Emerg Med.* 2019; **37**(7): 1273–8. [PubMed Abstract](#) | [Publisher Full Text](#)
35. Chen SJ, Chao TF, Chiang MC, *et al.*: **Prediction of patient outcome from *Acinetobacter baumannii* bacteremia with Sequential Organ Failure Assessment (SOFA) and Acute Physiology and Chronic Health Evaluation (APACHE) II scores.** *Intern Med.* 2011; **50**(8): 871–7. [PubMed Abstract](#) | [Publisher Full Text](#)
36. 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**(7): 906–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. 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**(1): 56. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Barlow G, Nathwani D, Davey P: **The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia.** *Thorax.* 2007; **62**(3): 253–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Guirgis FW, Puskarich MA, Smotherman C, *et al.*: **Development of a Simple Sequential Organ Failure Assessment Score for Risk Assessment of Emergency Department Patients With Sepsis.** *J Intensive Care Med.* 2020; **35**(3): 270–278. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Park HK, Kim WY, Kim MC, *et al.*: **Quick sequential organ failure assessment compared to systemic inflammatory response syndrome for predicting sepsis in emergency department.** *J Crit Care.* 2017; **42**: 12–7. [PubMed Abstract](#) | [Publisher Full Text](#)
41. Wang JY, Chen YX, Guo SB, *et al.*: **Predictive performance of quick Sepsis-related Organ Failure Assessment for mortality and ICU admission in patients with infection at the ED.** *Am J Emerg Med.* 2016; **34**(9): 1788–93. [PubMed Abstract](#) | [Publisher Full Text](#)
42. Cildir E, Bulut M, Akalin H, *et al.*: **Evaluation of the modified MEDS, MEWS score and Charlson comorbidity index in patients with community acquired sepsis in the emergency department.** *Intern Emerg Med.* 2013; **8**(3): 255–60. [PubMed Abstract](#) | [Publisher Full Text](#)
43. Haydar S, Spanier M, Weems P, *et al.*: **Comparison of QSOFA score and SIRS criteria as screening mechanisms for emergency department sepsis.** *Am J Emerg Med.* 2017; **35**(11): 1730–3. [PubMed Abstract](#) | [Publisher Full Text](#)
44. Tokioka F, Okamoto H, Yamazaki A, *et al.*: **The prognostic performance of qSOFA for community-acquired pneumonia.** *J Intensive Care.* 2018; **6**: 46. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. van der Woude SW, van Doormaal FF, Hutten BA, *et al.*: **Classifying sepsis patients in the emergency department using SIRS, qSOFA or MEWS.** *Neth J Med.* 2018; **76**(4): 158–66. [PubMed Abstract](#)
46. Henning DJ, Puskarich MA, Self WH, *et al.*: **An Emergency Department Validation of the SEP-3 Sepsis and Septic Shock Definitions and Comparison With 1992 Consensus Definitions.** *Ann Emerg Med.* 2017; **70**(4): 544–52.e5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Goulden R, Hoyle MC, Monis J, *et al.*: **qSOFA, SIRS and NEWS for predicting in-hospital mortality and ICU admission in emergency admissions treated as sepsis.** *Emerg Med J.* 2018; **35**(6): 345–9. [PubMed Abstract](#) | [Publisher Full Text](#)
48. Raymond NJ, Nguyen M, Allmark S, *et al.*: **Modified Sequential Organ Failure Assessment sepsis score in an emergency department setting: Retrospective assessment of prognostic value.** *Emerg Med Australas.* 2019; **31**(3): 339–46. [PubMed Abstract](#) | [Publisher Full Text](#)
49. Choi A, Park YS, Shin TG, *et al.*: **Prognostic performance of disease severity scores in patients with septic shock presenting to the emergency department.** *Am J Emerg Med.* 2019; **37**(6): 1054–9. [PubMed Abstract](#) | [Publisher Full Text](#)
50. Chen YX, Wang JY, Guo SB: **Use of CRB-65 and quick Sepsis-related Organ Failure Assessment to predict site of care and mortality in pneumonia patients in the emergency department: a retrospective study.** *Crit Care.* 2016; **20**(1): 167. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Gaini S, Relster MM, Pedersen C, *et al.*: **Prediction of 28-days mortality with sequential organ failure assessment (SOFA), quick SOFA (qSOFA) and systemic inflammatory response syndrome (SIRS) - A retrospective study of medical patients with acute infectious disease.** *Int J Infect Dis.* 2019; **78**: 1–7. [PubMed Abstract](#) | [Publisher Full Text](#)
52. Gonzalez Del Castillo J, Julian-Jimenez A, Gonzalez-Martinez F, *et al.*: **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.* 2017; **36**(12): 2361–9. [PubMed Abstract](#) | [Publisher Full Text](#)
53. Hifumi T, Fujishima S, Abe T, *et al.*: **Prognostic factors of Streptococcus pneumoniae infection in adults.** *Am J Emerg Med.* 2016; **34**(2): 202–6. [PubMed Abstract](#) | [Publisher Full Text](#)
54. Camm CF, Hayward G, Elias TCN, *et al.*: **Sepsis recognition tools in acute ambulatory care: associations with process of care and clinical outcomes in a service evaluation of an Emergency Multidisciplinary Unit in Oxfordshire.** *BMJ Open.* 2018; **8**(4): e020497. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Rannikko J, Seiskari T, Huttunen R, *et al.*: **Plasma cell-free DNA and qSOFA score predict 7-day mortality in 481 emergency department bacteraemia patients.** *J Intern Med.* 2018; **284**(4): 418–26. [PubMed Abstract](#) | [Publisher Full Text](#)
56. de Groot B, Stolwijk F, Warmerdam M, *et al.*: **The most commonly used disease severity scores are inappropriate for risk stratification of older emergency department sepsis patients: an observational multi-centre study.** *Scand J Trauma Resusc Emerg Med.* 2017; **25**(1): 91. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Ranzani OT, Prina E, Menendez R, *et al.*: **New Sepsis Definition (Sepsis-3) and Community-acquired Pneumonia Mortality. A Validation and Clinical Decision-Making Study.** *Am J Respir Crit Care Med.* 2017; **196**(10): 1287–97. [PubMed Abstract](#) | [Publisher Full Text](#)
58. Chen YX, Li CS: **Arterial lactate improves the prognostic performance of severity score systems in septic patients in the ED.** *Am J Emerg Med.* 2014; **32**(9): 982–6. [PubMed Abstract](#) | [Publisher Full Text](#)
59. Lee WJ, Woo SH, Kim DH, *et al.*: **Are prognostic scores and biomarkers such as procalcitonin the appropriate prognostic precursors for elderly patients**

- with sepsis in the emergency department? *Aging Clin Exp Res.* 2016; **28**(5): 917–24.
[PubMed Abstract](#) | [Publisher Full Text](#)
60. Huson MAM, Katete C, Chunda L, *et al.*: Application of the qSOFA score to predict mortality in patients with suspected infection in a resource-limited setting in Malawi. *Infection.* 2017; **45**(6): 893–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 61. Jo S, Jeong T, Lee JB, *et al.*: Validation of modified early warning score using serum lactate level in community-acquired pneumonia patients. The National Early Warning Score-Lactate score. *Am J Emerg Med.* 2016; **34**(3): 536–41.
[PubMed Abstract](#) | [Publisher Full Text](#)
 62. Kim MW, Lim JY, Oh SH: Mortality prediction using serum biomarkers and various sequential organ failure assessment score among patients with community-acquired pneumonia. *Scand J Clin Lab Invest.* 2017; **77**(7): 486–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
 63. Tirodda D, Gambacorta M, La Regina M, *et al.*: Evaluation of the threshold value for the modified early warning score (MEWS) in medical septic patients: a secondary analysis of an Italian multicentric prospective cohort (SNOOPII study). *QJM.* 2017; **110**(6): 369–73.
[PubMed Abstract](#) | [Publisher Full Text](#)
 64. Vaittinada Ayar P, Delay M, Avondo A, *et al.*: Prognostic value of prehospital quick sequential organ failure assessment score among patients with suspected infection. *Eur J Emerg Med.* 2019; **26**(5): 329–33.
[PubMed Abstract](#) | [Publisher Full Text](#)
 65. Kofoed K, Eugen-Olsen J, Petersen J, *et al.*: Predicting mortality in patients with systemic inflammatory response syndrome: an evaluation of two prognostic models, two soluble receptors, and a macrophage migration inhibitory factor. *Eur J Clin Microbiol Infect Dis.* 2008; **27**(5): 375–83.
[PubMed Abstract](#) | [Publisher Full Text](#)
 66. Lafon T, Vallejo C, Organista A, *et al.*: Validation du score qSOFA au Service d'Accueil des Urgences: étude prospective monocentrique. *Annales Françaises de Médecine d'Urgence.* 2018; **8**(3): 165–71.
[Publisher Full Text](#)
 67. Osatnik J, Tort-Oribea B, Folco J, *et al.*: Predictive Performance of Quick Sequential Organ Failure Assessment Scoring in an Argentinian Hospital. *Journal of Clinical and Diagnostic Research.* 2018; **12**(10): 22–6.
 68. 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**(4): 593–602.
[PubMed Abstract](#) | [Publisher Full Text](#)
 69. Krebs ED, Hassinger TE, Guidry CA, *et al.*: Non-utility of sepsis scores for identifying infection in surgical intensive care unit patients. *Am J Surg.* 2019; **218**(2): 243–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
 70. Serafim R, Gomes JA, Salluh J, *et al.*: A Comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome Criteria for the Diagnosis of Sepsis and Prediction of Mortality: A Systematic Review and Meta-Analysis. *Chest.* 2018; **153**(3): 646–55.
[PubMed Abstract](#) | [Publisher Full Text](#)
 71. 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**(1): 56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 72. Abdullah SMOB, Sørensen RH, Dessau RBC, *et al.*: Prognostic accuracy of qSOFA in predicting 28-day mortality among infected patients in an emergency department: a prospective validation study. *Emerg Med J.* 2019; **36**(12): 722–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
 73. 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**(1): 56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 74. del Castillo JG, Julien-Jiménez A, González-Martínez F, *et al.*: 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.* 2017; **36**(12): 2361–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
 75. Raith EP, Udy AA, Bailey M, *et al.*: Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the Intensive Care Unit. *JAMA.* 2017; **317**(3): 290–300.
[PubMed Abstract](#) | [Publisher Full Text](#)
 76. Smith MEB, Chiovaro JC, O'Neil M, *et al.*: Early warning system scores for clinical deterioration in hospitalized patients: A systematic review. *Ann Am Thorac Soc.* 2014; **11**(9): 1454–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
 77. Physicians RCo: National Early Warning Score (NEWS) 2 Standardising the assessment of acute-illness severity in the NHS: Updated report of a working party. London: RCP; 2017.
[Reference Source](#)
 78. Dinnes J, Deeks J, Kirby JPR, *et al.*: A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess.* 2005; **9**(12): 1–113, iii.
[PubMed Abstract](#) | [Publisher Full Text](#)
 79. McInnes MDF, Moher D, Thombs BD, *et al.*: Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA.* 2018; **319**(4): 388–96.
[PubMed Abstract](#) | [Publisher Full Text](#)
 80. Nannan Panday RS, Minderhoud TC, Alam N, *et al.*: Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. *Eur J Intern Med.* 2017; **45**: 20–31.
[PubMed Abstract](#) | [Publisher Full Text](#)
 81. Flint M, Hamilton F, Arnold D, *et al.*: Supplementary Data for sepsis timing pre-print (1.0). *Zenodo.* 2021.
<http://www.doi.org/10.5281/zenodo.5519552>

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 13 June 2022

<https://doi.org/10.21956/wellcomeopenres.19034.r50932>

© 2022 Sivayoham N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Narani Sivayoham

Department of Emergency Medicine, St George's University Hospitals NHS Foundation Trust, London, UK

Thank you for asking me to review the paper entitled 'The timing of use of risk-stratification tools affects their ability to predict mortality from sepsis. A meta-regression analysis'. This study seeks to understand the different test characteristics of different RSTs depending on the timing of their use. The study involves a large number of patients and the authors should be commended for putting in the work commanded by this study. Overall, it is well written. However, I do have some major and minor concerns.

Major concerns:

1. The authors have concluded that 'the timing of the RSTs is paramount to their predictive performance'. In the results section, under the sub-heading meta-regression, the authors have stated the REML regression co-efficient is 0.57 (0.07-1.08), and the sensitivity analyses for the first recorded RST and worst-recorded RST were 0.64 (0.57-0.71) and 0.76 (0.67-0.83), respectively. My understanding is that as the regression co-efficient crosses 1 and because the 95% confidence intervals for the respective sensitivities overlap, it cannot be concluded that the first-recorded RSTs are less sensitive than the worst-recorded RSTs. So, whilst there may be a trend towards a difference, the analysis in this study fails to confirm any difference. A statistical review may be beneficial for this point.
2. Could the authors clarify their conclusion?
3. The authors have performed further analyses on qSOFA and EWS scores as these scores had 'sufficient data'. They state that they 'found a similar relationship to timing and study performance', but then state that neither met statistical significance for sensitivity or specificity due to 'low study numbers'.
 - Either they had sufficient data or not. Could the authors please explain which it is?
 - Also, how did they determine what 'sufficient data' was?

- In the results section of the abstract, the last sentence states that further analysis of qSOFA and EWS 'found timing of their use to be associated with predictive performance'. Yet, in the results section in the body of the manuscript, it states that the difference in sensitivity and specificity did not reach statistical significance. Could the authors please explain the discrepancy?
4. Whilst qSOFA has a consistent cut-off point of ≥ 2 to identify the high-risk group, the cut-off point for the EWS score can vary. When comparing sensitivity and specificity of the EWS score, between early and worst-recorded EWS scores, were the same cut-off points used in both groups?

Minor concerns:

1. What does REML stand for?
2. What does LMIC stand for? This should also be mentioned in the methods section.
3. The penultimate paragraph of the discussion section has 'hope to' mentioned twice.
4. Final paragraph of the discussion: 'Given the low specificity...', should this be 'sensitivity', not 'specificity'?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My area of expertise involves the risk stratification of patients with suspected sepsis in the emergency department.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.