OPEN ACCESS

University of BRISTOL

Peer reviewed version

Link to published version (if available):
10.1016/j.jinf.2018.01.002

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Accepted Manuscript
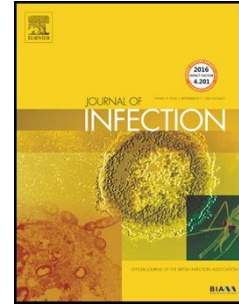
Early Warning Scores do not accurately predict mortality in sepsis: A meta-analysis and systematic review of the literature.

Hamilton F[1], Arnold D[2], Baird A[3], Albur M[4], Whiting P[5,6]

[1] Medical Trainee, Department of Medicine, Weston Area Health Trust, UK

[2] Academic Respiratory Registrar, Department of Medicine, North Bristol NHS Trust, UK

[3] Clinical Fellow, Department of Critical Care, Gloucestershire Royal Hospital, UK

[4] Consultant Microbiologist, Department of Microbiology, North Bristol NHS Trust, UK

[5] Senior Research Fellow, NIHR CHLARC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

[6] Senior Research Fellow, School of Social and Community Medicine, University of Bristol, UK

Corresponding author:

Primary email: gushamilton@gmail.com

Secondary email: Fergus.hamilton@nhs.net

Telephone: +44 7743165499

Work address:

Department of Medicine, Weston General Hospital, W-S-M, BS23 4TQ, UK

Word Count for abstract: 199

Word Count for full text: 3,737

Highlights:

- Early Warning Scores are commonly used to assess patients in hospital
- The available evidence suggests they are inaccurate at predicting mortality
- The cannot rule in (LR+ 1.79) or rule out (LR- 0.59) mortality
- The studies involved are generally at moderate-high risk of bias

Abstract

Objectives

Early Warning scores are used to evaluate patients in many hospital settings. It is not clear if these are accurate in predicting mortality in sepsis. We performed a systematic review and meta-analysis of multiple studies in sepsis. Our aim was to estimate the accuracy of EWS for mortality in this setting.

Methods

PubMED, CINAHL, Cochrane, Web of Science and EMBASE were searched to October 2016. Studies of adults with sepsis who had EWS calculated using any appropriate tool (e.g. NEWS, MEWS) were eligible for inclusion. Study quality was assessed using QUADAS-2. Summary estimates were derived using HSROC analysis.

Results

Six studies (4,298 participants) were included. Results suggest that EWS cannot be used to predict which patients with sepsis will (positive likelihood ratio 1.79, 95% CI 1.53 to 2.11) or will not die (negative likelihood ratio 0.59, 95% CI 0.45 to 0.78). Two studies were rated as

low risk of bias and one as unclear risk of bias on all domains. The other three studies were

judged at high risk of bias in one domain.

Conclusion

Early Warning Scores are not sufficiently accurate to rule in or rule out mortality in patients

with sepsis, based on the evidence available, which is generally poor quality.

Introduction

Sepsis is a major problem in emergency departments and hospitals causing significant morbidity and mortality.[1,2] A steady increase in the incidence of sepsis has been noted worldwide, with recent UK estimates of around 50 cases of severe sepsis per 100,000 people per year.[3] Collecting data on accurate mortality figures is difficult, but recent estimates suggest an estimated case mortality of between 20-35%, although rates differ widely by definition.[4,5] Timely diagnosis of patients with sepsis can be difficult, because of the differences in clinical presentations, and few uniquely identifying features, especially in the elderly.[6]

Once patients with sepsis are identified, rapid treatment has been shown to improve mortality, although nearly all evidence comes from observational studies.[7–9] Due to the significant number of patients who present with sepsis, it is critical to identify the patients at risk of deterioration, and patients requiring urgent treatment or critical care input. Alongside this, there is also a need to identify the significant number of patients who are likely to have a good outcome and hence can be managed more conservatively.

Early Warning Scores (EWS) are physiological composite scores comprising pulse rate, blood pressure, temperature, respiratory rate, mental state and oxygen saturation. Each of these physical observations are given a score, where 0 is considered normal. Simple addition of these observations allows a total score to be calculated, usually between 0 and 12.

Different versions of EWS often have minor modifications, such as the addition of points if the patient is receiving oxygen therapy, or variations in specific cut-offs.

Recently, the UK National Institute for Health and Care Excellence (NICE) has published guidelines on the recognition, diagnosis and management of sepsis.[10] These guidelines recommend considering the use of Early Warning Scores in assessing patients with suspected sepsis in acute hospital settings, and highlight this area as a key research topic. The 2015 National Confidential Enquiry on Patient Outcome and Death report on sepsis also suggested use of EWS as a principal recommendation in all care settings, specifically to assess severity of sepsis and to prioritise urgent care.[11]

Over the last ten years, EWS have been introduced into nearly all UK hospitals, and are already recommended in NICE guidance for monitoring critically ill patients in hospital and Royal College of Physician guidance for monitoring of all adult patients in acute hospital settings, .[12,13] They have recently been introduced into emergency departments, and also into many ambulances services and community settings and primary care.[14] Given these scores are often calculated on every patient, there have been concerns about the volume of workload created and the sensitivity of these scores for identifying unwell patients.[15]

EWS are often used, both informally and formally, to guide treatment decisions such as the best location for care (inpatient, outpatient, ICU), and the level of monitoring or seniority of doctor that should see the patient. In some centres, a certain score (for example, greater

than 5), will trigger a pager alert to senior medical staff or critical care outreach services.
Although there is some evidence that this method identifies sick patients, the evidence
relating to patients with sepsis is limited.[16,17]

Methods

A protocol was developed based on recommended standards for conduct of systematic
reviews.[18,19] The review was registered with PROSPERO (CRD42016047125), and the PRISMA
guidelines were followed when reporting results.

Study identification

MEDLINE, EMBASE, CINAHL, the Cochrane Library, and Web of Science were searched from
inception to October 2016. The search strategy combined terms for sepsis with terms for
EWS. The full search study is available as a web appendix. References of included studies
were screened and all first authors emailed to identify additional relevant studies.

Study Selection

Studies were eligible for inclusion if they assessed any form of Early Warning Score for the
prediction of mortality in adult patients with sepsis. Studies were required to report
sufficient data to construct a 2×2 table of test performance. Studies on a particular type of
infection (e.g. meningitis) or a particular microbial cause (e.g *Streptococcus pneumoniae)*
were excluded. Studies were limited to secondary care facilities. A diagnosis of sepsis was
considered as any diagnosis made prospectively or retrospectively by physicians, and coded

as such. Mortality data was ideally 28-day mortality, but any form of mortality reporting was accepted. Mortality was chosen as this is the most commonly reported and most important outcome in sepsis.

Two reviewers (AB, DA) independently screened titles and abstracts identified by the search. Full text articles considered potentially relevant were obtained. Inclusion assessment was performed by a third reviewer (FH); disagreements were resolved through consensus between all three.

Data extraction and quality assessment

Data on study design, number of participants, country, setting, EWS methodology, definitions of sepsis, mortality, sensitivity, specificity and area under the curve (AOC) were extracted using a predesigned form. If studies did not publish 2 x 2 data, this was calculated from the number of participants, sensitivity, specificity, and mortality. Where studies reported more than one set of 2x2 data, for example for different thresholds, then each set of 2x2 data was extracted. Study quality was assessed using the QUADAS-2 tool.[20] We modified our QUADAS assessment to look specifically at whether studies were biased by the use of EWS reducing mortality. If an EWS was in use and was acted on, this would be expected to impact on mortality and so would bias any estimate of the accuracy of NEWS in predicting mortality. This is because EWS should identify patients who are deteriorating and action should have been taken which may have reduced mortality - known as treatment paradox in the context of diagnostic test evaluation. Data extraction and quality assessment were carried out by one reviewer and checked by a second. Disagreements were resolved through discussion.

Data synthesis and analysis

We calculated sensitivity and specificity for each set of 2x2 data and plotted these in summary receiver operating characteristic (SROC) space. Where studies reported data for multiple thresholds, data for the threshold most similar to that in other studies were extracted. Only one set of 2x2 data per study contributed to the analysis. The bivariate/hierarchical summary receiver operating characteristic (HSROC) models were used to estimate summary sensitivity and specificity with 95% confidence and prediction regions around the summary points and to derive an HSROC curve. This approach allows for between-study heterogeneity in sensitivity and specificity, and for the trade-off (negative correlation) between sensitivity and specificity commonly seen in diagnostic meta-analyses. Heterogeneity was assessed visually using the SROC plot. Summary positive and negative likelihood ratios were derived from the summary estimates of sensitivity and specificity. We also estimated the summary area under the ROC (AUC) by pooling AUCs reported in individual studies using a random effects model. We used Stata 13 (StataCorp LP, College Station, Texas, USA), mainly using the *metandi* .[21]command and using MedCalc for Windows, version 15.0 (MedCalc Software, Ostend, Belgium).

We conducted sensitivity analysis to investigate the effect of type of early warning score, study quality, setting, and threshold. Analyses were restricted to studies that were considered at low risk of bias or low/unclear concerns regarding applicability across all QUADAS-2 domains, to studies conducted in the ED, and to studies that used a threshold of ≥5 to define a positive EWS result. We also conducted subgroup analyses stratified according to type of EWS.

A flow diagram was created to put these results into clinical context. This shows the results that would be obtained if EWS were used in a hypothetical population of 1000 patients presenting to the emergency department with sepsis where 150 (15%) will die. This figure was chosen as the average mortality across our six studies was around this.

Results

Figure 1 shows the flow of studies through the review process. The searches identified 620 possible studies, of which 6 were included (4,298 participants). Two of the studies were conducted in the UK [22,23], and one each in Germany[24], Israel[25], Italy[26] and Turkey[27] (**Table 1**). The Italian study was available as a conference abstract only.

Four studies were performed in emergency departments[22–24,27]; two were in internal medicine. [25,26] Three studies were relatively small in size (<400 participants),[23,24,27] one was moderate (n=535)[26] and two included more than 1000 participants.[22,25] Five studies used the MEWS system, one study used the National Early Warning Score (NEWS), a slight variation.[13,22] NEWS and MEWS both include respiration rate, oxygen saturation, supplemental oxygen, temperature, blood pressure, heart rather and level of consciousness. The NEWS score includes scoring for oxygen saturation, whereas the MEWS score includes scoring for urine output.

Three studies used the American College of Chest Physicians / Society of Critical Care Medicine (ACCP/SCCM) consensus definition as the basis of their sepsis diagnoses.[22,25,27]

One study was conducted across two sites, one site used the consensus definition, the other used a 'clinician suspicion of infection' alongside evidence that blood cultures had been drawn.[23] One study only reported that patients has "objectively diagnosed sepsis."[26] The final study concentrated on 'suspected sepsis' using some of the ACCP/SCCM definition or a 'working diagnosis that included sepsis'.[24] Two studies were retrospective case note reviews[22,23], three identified sepsis prospectively.[25–27] One study[24] used prospective identification, but allowed retrospective case note review to ensure no missed cases.

 Two studies were judged at low risk of bias for all QUADAS domains (**Table 2**). The Italian study, [26] which was available only as an abstract, was judged as unclear risk of bias for all domains except the reference standard domain for which it was judged as low risk of bias. This is because insufficient information was available to reach a judgement for the other domains. Three studies were rated at high risk for bias in terms of flow and timing. These were all prospective studies that recorded EWS in real time. This leads to the potential to them being acted upon – the 'treatment paradox', and hence lowering mortality and biasing our estimate of EWS accuracy.[24,25,27] There were some concerns regarding the applicability of studies to the review question. The Italian study was judged at unclear concerns regarding applicability for all domains due to the lack of information available. There were high concerns about the applicability of the population in one study[24]. Diagnostic criteria for sepsis were 'suspected sepsis' rather than confirmed sepsis and more than 20% of patients included had a final diagnosis that was not sepsis. The applicability of the index test (EWS) was considered as unclear in four studies as the time point at which the EWS was performed was not documented (i.e – whether on admission to hospital, or during the stay in the ED).[23,24,26,27] Given that EWS are based on physiology, and hence change continually and

with treatment, there was the potential for this to have affected the applicability of results

if EWS had not been recorded on admission. One study was considered at high concern

regarding applicability.[24] The definition of mortality in this study was in hospital mortality

compared to our outcome of interest of 30 day mortality. The accuracy of EWS for

prediction of mortality at this time point may be different and so this was flagged as a high

concern regarding applicability.

Estimates of sensitivity and specificity showed considerable heterogeneity across studies.

Specificity ranged from 34 to 89%, and specificity from 30 to 84% (Table 1). The summary

ROC plot (**figure 2)** shows individual study estimates of sensitivity and specificity, the

summary estimate across studies and the HSROC curve. Individual estimates of sensitivity

and specificity were close to the HSROC curve, suggesting that much of the variation across

studies was a result of the trade-off between sensitivity and specificity. The area under the

curve (AUC) was more similar across studies, ranging from 0.59 to 0.72 with a summary AUC

of 0.68 (95% CI 0.65, 0.71; 5 studies). Summary sensitivity was 66% (95% CI 46, 81%), and

summary specificity was 62% (95% CI 45, 76%) based on all six studies. These estimates

should be interpreted with some caution due to the observed heterogeneity. There was

less heterogeneity in likelihood ratios with positive likelihood ratios ranging from 1.27 to

2.06 and negative likelihood ratios ranging from 0.33 to 0.79. The summary positive

likelihood ratio was 1.79 (95% CI 1.53, 2.11) and summary negative likelihood ratio was 0.59

(95% CI 0.45, 0.78). These suggest that EWS cannot be used to rule in or rule out mortality.

We investigated the effect of study quality, type of EWS, threshold and setting on results

(**Table 3**). We restricted analysis to studies at low risk of bias (n=2), studies with low

concern regarding applicability (n=2), studies with low or unclear concerns regarding

applicability (n=5), studies conducted in emergency departments (n=4), studies that

evaluated the MEWS score (n=5) and studies that defined a score of 5 or more as a positive

EWS result (4 studies). There was evidence that accuracy varied according to type of EWS,

all other sensitivity analysis showed similar results. Analyses stratified on type of EWS

suggested that MEWS is slightly better for predicting mortality (LR+ 1.90, 1.66 to 2.18; 5

studies, **Figure 3**) than NEWS (LR+ 1.27, 1.22 to 1.34; 1 study) but worse at ruling out

mortality (LR- 0.66, 0.55 to 0.80 compared to 0.33, 0.24 to 0.47 for NEWS). The accuracy

was still not sufficient to predict or rule out mortality. Only one study evaluated NEWS[22] but

this was the largest study included in the review and was considered at low risk of bias.

To put the above figures in context, **Figure 4** shows that in theory, if EWS were to be used in

a group of 1000 patients presenting to the emergency department where 150 (15%) will die,

an estimated 422 will have an EWS predicting mortality but only 99 (24%) will actually die.

Of the 578 people with a result suggesting they will not die, 51 (9%) will die.

Discussion

Summary of findings

We identified six relevant studies for inclusion in our review. Results suggest that EWS

cannot be used to predict who will (positive likelihood ratio 1.79, 95% CI 1.53, 2.11) or will

not (negative likelihood ratio 0.59, 95% CI 0.45, 0.78) die in patients with sepsis. Sensitivity

analyses suggested that results were similar when restricted to studies at low risk of bias,

studies with low concern regarding applicability, studies with low or unclear concerns

regarding applicability, studies conducted in emergency departments, and studies that defined a score of 5 or more as a positive EWS result.   There was some evidence that MEWS may be slightly better at predicting who will die but worse at predicting who will not die than the NEWS score. However, results from sensitivity analyses should be interpreted with extreme caution due to the small number of studies included in the review.

Study quality was variable, with two studies rated as low risk of bias on all domain and, three studies rated as high risk of bias as there was it was possible that the EWS result was acted on prior and so could have affected the outcome. One study was available only as a conference abstract, and as such was judged at unclear risk of bias. There were some concerns regarding applicability because of diagnosis of sepsis in one study, lack of information on when the EWS was performed in four studies, and in-hospital mortality rather than 30-day mortality as an outcome in one study.

Strengths and weaknesses of this review

Strengths of this review include a sensitive search strategy across multiple databases. We therefore consider it unlikely that we have missed relevant studies. We included a search of conference abstracts to minimise the likelihood of publication bias and contacted authors of included studies to ask them if they were aware of any additional studies. It was not possible to formally assess publication bias as methods commonly used to detect this are not suitable for use in test accuracy reviews.[19] However, publication bias may be less of a problem for studies of diagnostic test than for reviews of interventions. A recent study found that studies reporting higher accuracy were no more likely to be reported in full.[28]

Two reviewers were involved in all stages of the review process minimising the likelihood that bias or errors were introduced during the review process. We used statistically robust hierarchical models to estimate summary sensitivity and specificity and to derive summary received operating characteristic plots. These allow for both between-study variability in sensitivity and specificity, and the negative correlation between them.[21] We used the validated QUADAS-2 criteria to assess study quality. We incorporated the results of the quality assessment into our synthesis by restricting analyses to studies judged as low risk of bias or applicability.

There are many variations of EWS using different scoring systems. We included studies of any EWS in our review but only identified studies evaluating MEWS or NEWS. Our analysis restricted to studies that had evaluated MEWS showed it was slightly better at predicting mortality than NEWS, but worse at ruling mortality out. However, as only one study evaluated NEWS there were not sufficient data to allow firm conclusions regarding whether there is a difference in accuracy of these two scores. As we did not identify studies of other EWS it is not possible to determine whether these might have differing accuracies from those EWS included in our review.

Two new diagnostic criteria for sepsis have been published this year – US consensus guidelines (SEPSIS-3) and new guidance from NICE[10,29]. The SEPSIS-3 guidelines in particular differ quite significantly from the ACCP/SCCM definition used by the studies included in our review. Given the change in diagnostic criteria, the group categorised as having sepsis may well differ significantly from previous studies. The accuracy of EWS for predicting mortality

in patients diagnosed according to new criteria may therefore not be the same as those diagnosed using previous diagnostic criteria.

The treatment paradox – where the observed mortality may be significantly lower in studies where EWS are acted upon, is a high potential for bias in prospective studies introducing EWS. Three of our six studies were potentially at risk of this, although mortality in each of these studies was not substantially different from the other three studies (mean mortality 22.7% vs 17.5%).

Finally, and importantly, EWS were not designed to be used as single time point diagnostic or predictive tools. They are intended to be used to 'track and trigger' individual patients, and hence the evaluation of 'one off' EWS may not reflect day to day clinical practice. Estimates of accuracy from these studies therefore only show how accurate a single time point assessment of EWS is in predicting mortality. Accuracy may be different if multiple time points were considered. However, we were unable to find any studies in sepsis evaluating EWS in any other context except 'one off' readings. Alongside this, the timings of the particular EWS measurement were not clear in many of the studies (i.e. if they were the first or the worst set of observations taken). Given the dynamic nature of sepsis, and heterogeneous nature of EWS usage in these studies, this is an important consideration.

Comparisons with previous literature

We are unaware of any other systematic review that have evaluated EWS in patients with sepsis. Previous literature around EWS in other patient populations has been studied more

extensively. A review conducted in 2014[30] found that EWS can predict cardiac arrest or death in patients on hospital wards, estimating an area under the curve of between 0.88-0.93 based on eight studies, but noted significant methodological limitations in the studies. Another review of EWS for predicting mortality, also published in 2014, was unable to draw conclusions due to substantial heterogeneity of EWS systems alongside poor methodology[31]. One study, by Quan et al was identified as a conference abstract but did not have enough extractable data to be included in the formal SR. This was a retrospective study of ED patients with sepsis, that showed an area under the curve of 0.61 (0.52-0.70) for predicting mortality, but we were unable to gather any further information (despite emailing the lead author)[32].

We are aware of two further primary studies on early warning scores in specific infections rather than sepsis. One study has been published on EWS in Gram-negative bacteraemia[33]. This study collected EWS data for multiple days prior to and after the onset of bacteraemia, and found cumulative EWS scores (over multiple days) to be highly predictive of mortality (AUC 0.90, for a 7 day average of EWS after onset of bacteraemia). Another study looked at standardised early warning scores (SEWS) in patients with pneumonia[34]. This study included 419 patients with pneumonia, and found a sensitivity of 52% and specificity of 67% for predicting mortality, similar to our study

Implications for practice

This review suggests that EWS cannot be used to predict who will or will not die in patients with sepsis, although the quality of studies was poor. This is important for two reasons. Firstly, we should not be using EWS alone to prognosticate in patients with sepsis. They are

not accurate enough to guide treatment decisions alone (on this data), such as whether or not patients should be admitted to ITU or remain on wards.

Secondly, EWS is now commonly used in the triage of patients presenting to healthcare facilities at different levels such as primary care or ED in the secondary / tertiary care hospitals. This approach is now endorsed by NICE for patients with 'suspected sepsis', based on the assumption that these scores predict mortality and other outcomes, and was a principal recommendation in the 2015 NCEPOD report on sepsis. Although this review did not look at the predictive value of EWS in the setting of 'suspected sepsis', it would seem likely that EWS will have a poor performance in this group too, given that EWS does not predict mortality in sepsis.

Therefore, EWS is not likely to be sensitive or specific enough to perform a triaging role alone, and is likely to lead to a significant increase in resources spent on low risk patients with suspected sepsis (if a low cut off is used), or missing of patients with a high mortality (if a high cut off is used).

Implications for Research

As noted above, there have been two new consensus guidelines on diagnosis of sepsis published this year[10,29]. It will be important to re-assess tools such as EWS in patients diagnosed with sepsis based on these new criteria, as they may perform quite differently. Secondly, there has been a gradual introduction of NEWS in the ambulance services and the community in the U.K, and research might be best to concentrate on NEWS in particular, alongside research comparing this to other EWS.

Hence, large scale, multi-centre studies are needed to further assess the role of EWS in sepsis in the community and in ambulance or similar services. Further research would also be useful in patients with suspected sepsis, and perhaps in using multiple recordings of EWS in an emergency department stay to assess response to early treatment.

Conclusions

EWS have poor prognostic value in predicting sepsis mortality. Based on the existing data, which is of poor quality, EWS should not be used on their own to guide prognosis in patients with sepsis, and are unlikely to be reliable alone in identifying patients at risk or death. Further work is needed to assess the role of EWS in patients with suspected sepsis.

References

1.   Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med*. 2001;29(7):1303-1310. http://www.ncbi.nlm.nih.gov/pubmed/11445675. Accessed February 24, 2015.

2.   Harrison DA, Welch CA, Eddleston JM. The epidemiology of severe sepsis in England, Wales and Northern Ireland, 1996 to 2004: secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database. *Crit Care*. 2006;10(2):R42. doi:10.1186/cc4854.

3.   Harrison DA, Welch CA, Eddleston JM. The epidemiology of severe sepsis in England,

Wales and Northern Ireland, 1996 to 2004: secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database. *Crit Care*. 2006;10(2):R42. doi:10.1186/cc4854.

4. Martin GS, Mannino DM, Eaton S, Moss M. The Epidemiology of Sepsis in the United States from 1979 through 2000. *N Engl J Med*. 2003;348(16):1546-1554. doi:10.1056/NEJMoa022139.

5. Kaukonen K-M, Bailey M, Suzuki S, Pilcher D, Bellomo R. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. *JAMA*. 2014;311(13):1308-1316. doi:10.1001/jama.2014.2637.

6. Carrigan SD, Scott G, Tabrizian M. Toward resolving the challenges of sepsis diagnosis. *Clin Chem*. 2004;50(8):1301-1314. doi:10.1373/clinchem.2004.032144.

7. Chamberlain DJ, Willis EM, Bersten AB. The severe sepsis bundles as processes of care: a meta-analysis. *Aust Crit Care*. 2011;24(4):229-243. doi:10.1016/j.aucc.2011.01.003.

8. Daniels R, Nutbeam T, McNamara G, Galvin C. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. *Emerg Med J*. 2011;28(6):507-512. doi:10.1136/emj.2010.095067.

9. Barochia A V, Cui X, Vitberg D, et al. Bundled care for septic shock: an analysis of clinical trials. *Crit Care Med*. 2010;38(2):668-678. doi:10.1097/CCM.0b013e3181cb0ddf.

10. NICE. *Sepsis: Recognition, Diagnosis and Early Management | Recommendations | Guidance and Guidelines | NICE*. London: NICE; 2016. https://www.nice.org.uk/guidance/ng51/chapter/Recommendations. Accessed July 21, 2016.

11. National Confidential Enquiry into Patient Outcome and Death. *NCEPOD - Sepsis: Just Say Sepsis! (2015)*.; 2015. http://www.ncepod.org.uk/2015sepsis.html. Accessed March 11, 2017.

12. NICE. *Acutely Ill Adults in Hospital: Recognising and Responding to Deterioration*. London: NICE; 2007. https://www.nice.org.uk/guidance/cg50. Accessed November 13, 2016.

13. RCP Acute Medicine Task Force. National Early Warning Score (NEWS). RCP. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news. Published 2015. Accessed September 22, 2016.

14. NasMeD. *Future National Clinical Priorities for Ambulance Services in England*.; 2014. http://aace.org.uk/wp-content/uploads/2014/05/Future-national-clinical-priorities-for-ambulance-services-in-England-FINAL-2.pdf. Accessed November 13, 2016.

15. Jarvis S, Kovacs C, Briggs J, et al. Aggregate National Early Warning Score (NEWS) values are more important than high scores for a single vital signs parameter for discriminating the risk of adverse outcomes. *Resuscitation*. 2014;87C:75-80. doi:10.1016/j.resuscitation.2014.11.014.

16. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl*. 2006;88(6):571-575. doi:10.1308/003588406X130615.

17. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM*. 2001;94(10):521-526. http://www.ncbi.nlm.nih.gov/pubmed/11588210. Accessed March 18, 2016.

18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339(jul21_1):b2535.

doi:10.1136/bmj.b2535.

19.   Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Cochrane Handbook

        for Systematic Reviews of Diagnostic Test Accuracy Chapter 11 Interpreting results

        and drawing conclusions. 2013. http://srdta.cochrane.org/. Accessed July 21, 2016.

20.   Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of

        QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included

        in systematic reviews. *BMC Med Res Methodol*. 2003;3(1):25. doi:10.1186/1471-

        2288-3-25.

21.   Harbord R. METANDI: Stata module to perform meta-analysis of diagnostic accuracy.

        Stat Softw Components. 2008.

22.   Corfield AR, Lees F, Zealley I, et al. Utility of a single early warning score in patients

        with sepsis in the emergency department. *Emerg Med J*. 2014;31(6):482-487.

        doi:10.1136/emermed-2012-202186.

23.   Vorwerk C, Loryman B, Coats TJ, et al. Prediction of mortality in adult emergency

        department patients with sepsis. *Emerg Med J*. 2009;26(4):254-258.

        doi:10.1136/emj.2007.053298.

24.   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):508-515.

        doi:10.1007/s00508-013-0407-2.

25.   Ghanem-Zoubi NO, Vardi M, Laor A, Weber G, Bitterman H. Assessment of disease-

        severity scoring systems for patients with sepsis in general internal medicine

        departments. *Crit Care*. 2011;15(2):R95. doi:10.1186/cc10102.

26.   Regina La M, Dentali F, Bonardi G, et al. Accuracy of MEWS score in predicting

mortality in septic patients admitted to Internal Medicine units. In: *XIX Congresso Nazionale Della Società Scientifica FADOI*. Bologna; 2014:69.

27. Çıldır E, Bulut M, Akalın H, Kocabaş E, Ocakoğlu G, Aydın ŞA. 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):255-260. doi:10.1007/s11739-012-0890-x.

28. Korevaar DA, Cohen JF, Spijker R, et al. Reported estimates of diagnostic accuracy in ophthalmology conference abstracts were not associated with full-text publication. *J Clin Epidemiol*. 2016;79:96-103. doi:10.1016/j.jclinepi.2016.06.002.

29. 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. doi:10.1001/jama.2016.0287.

30. 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):1454-1465. doi:10.1513/AnnalsATS.201403-102OC.

31. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PWB. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation*. 2014;85(5):587-594. doi:10.1016/j.resuscitation.2014.01.013.

32. Quan G, Lebovic G, McGowan M, Gray S. Can Early Warning scores predict mortality of ED Septic patients? A comparison of VIEWS-L and NEWS-L. In: *Critical Care Medicine*. Vol 43. ; 2015:256. doi:10.1097/01.ccm.0000474848.31895.a7.

33. Albur M, Hamilton F, MacGowan AP. Early warning score: a dynamic marker of severity and prognosis in patients with Gram-negative bacteraemia and sepsis. *Ann*

Clin Microbiol Antimicrob. 2016;15(1):23. doi:10.1186/s12941-016-0139-z.

34. 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):253-259. doi:10.1136/thx.2006.067371.

Role of the funding source

Competing interests

Authors' contributions

FH and MA conceived the idea for the review. FH and PW drafted the article. .FH, DA, AB and MA served as a content experts in the field of EWS. PW provided methodological support. FH, DA and AB undertook screening and data extraction. FH and DA performed the risk of bias assessment. All authors contributed to the interpretation of results, commented on draft manuscripts and have given their approval for publication. FH (lead author and manuscript's guarantor) affirms that the manuscript is an honest, accurate, and

transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Ethics approval

As this was a systematic review no ethics approval was needed.

Data sharing statement

The full dataset is available in tables included in the review.

Acknowledgements

Transparency:

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Patient involvement

Patients were not involved in the design or conduct of this study.

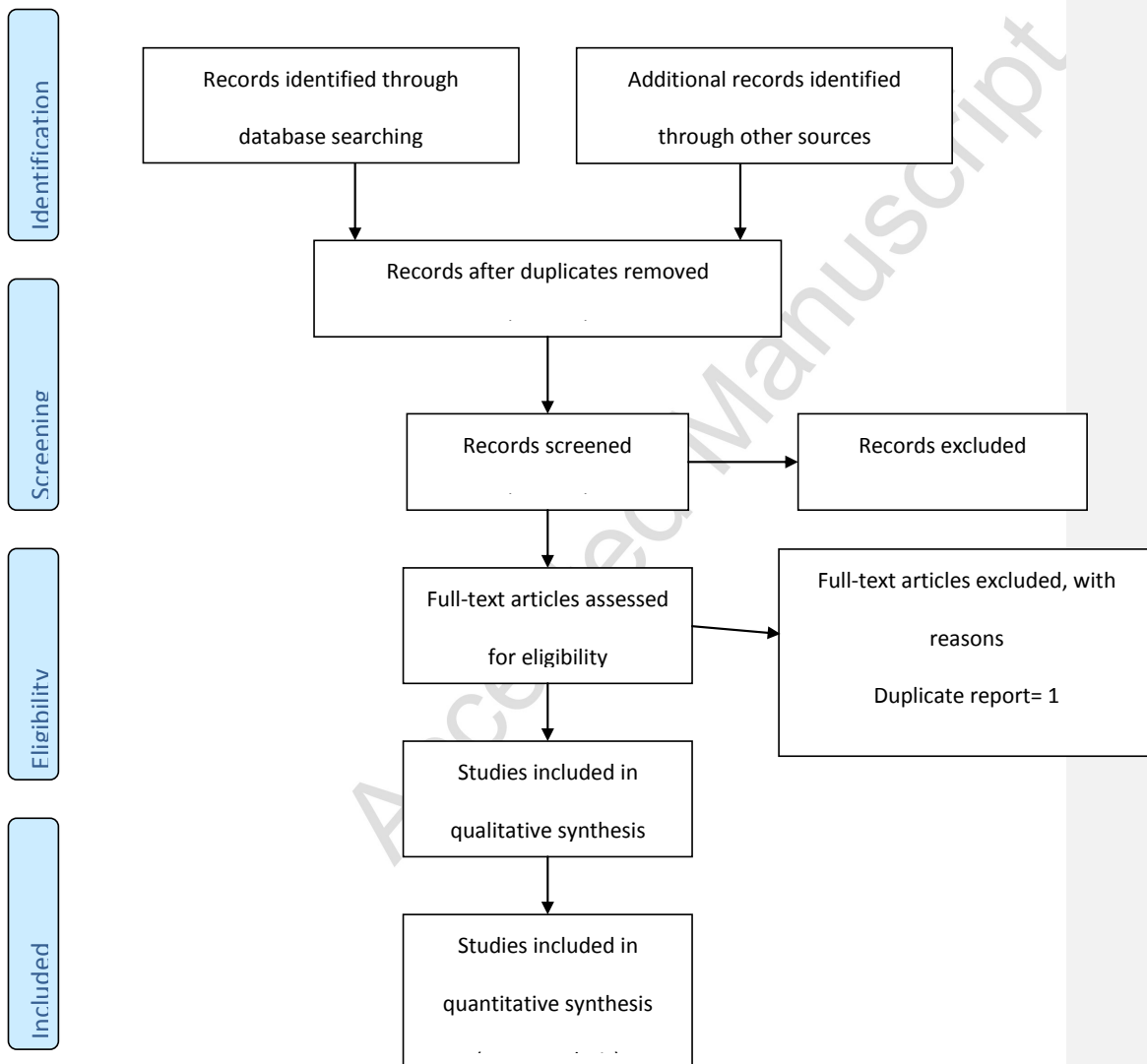Figure 1: Flow of studies through the review process

Figure 2: Summary receiver operating characteristic plot for all studies

Circles indicate individual study estimates; the green line shows the hierarchical summary receiver operating characteristic curve, the black square denotes the summary estimate and the dashed line its 95% confidence region, while the dotted line shows a 95% prediction region
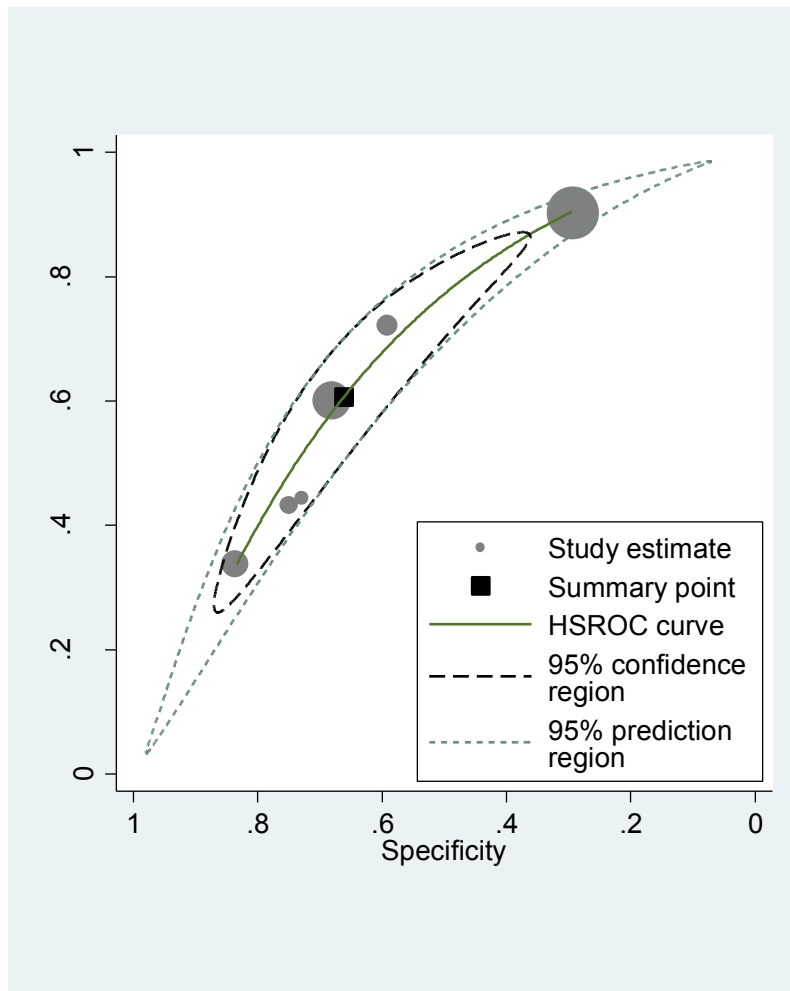
Figure 3: Summary receiver operating characteristic plot for studies that evaluated MEWS

Circles indicate individual study estimates; the green line shows the hierarchical summary receiver

operating characteristic curve, the black square denotes the summary estimate and the dashed line

its 95% confidence region, while the dotted line shows a 95% prediction region
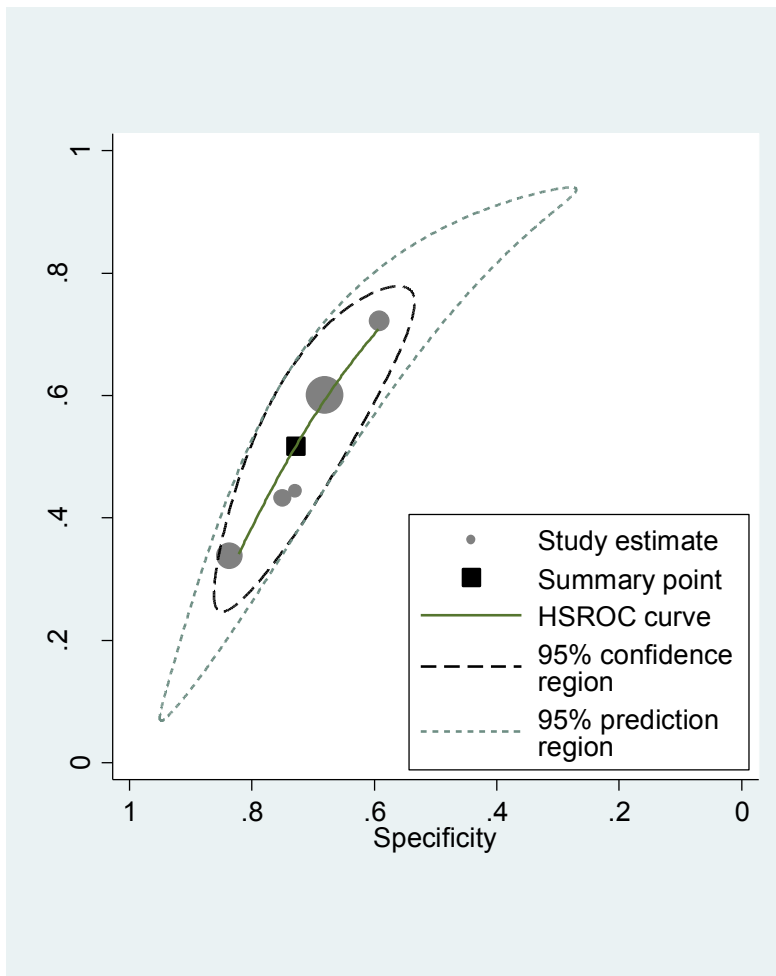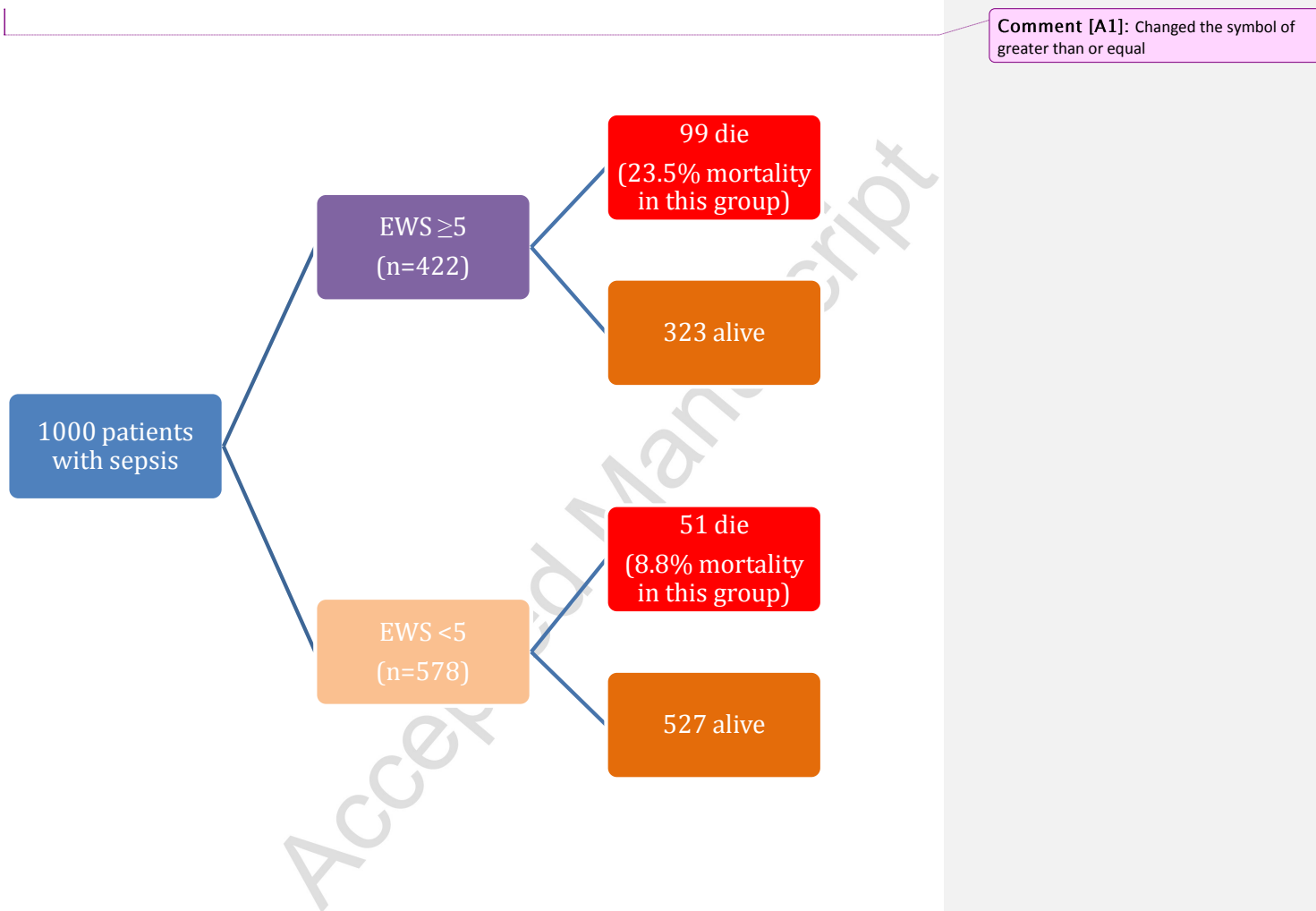
Figure 4: Flow diagram showing the results that would be obtained if EWS were used in a hypothetical population of 1000 patients presenting with sepsis where 150 (15%) will die.

Table 1 Study details and individual study results

Study and Country	How sepsis diagnosed?	EWS	Sepsis definition	Threshold	Setting	n	Mean/ median age	% male	Mortality prevalence (%)	AUC (95% CI)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	LR+ (95% CI)	LR- (95% CI)
Cildir (2013) [27] Turkey	Prospective, clinician identification	MEWS	ACCP/SCCM	≥6	ED	230	NR	NP	32.2	0.61 (NR)	43 (32, 55)	75 (67, 81)	1.73 (1.19, 2.52)	0.76 (0.62, 0.93)
Corfield (2014) [22] Scotland	Retrospective case note review	NEWS	ACCP/SCCM	≥5	ED	2003	72 (IQR 59, 81)	47%	14.8	0.70 (0.67, 0.74)	89 (85, 92)	30 (28, 32)	1.27 (1.22, 1.34)	0.33 (0.24, 0.47)
Geier (2013) [24] Germany	Prospective clinician identification and retrospective case note review.	MEWS	Clinical diagnosis	≥5	ED	151	68.3 (SD 18)	54	14.6	0.64 (0.55, 0.73)	43 (22, 67)	74 (63, 81)	1.64 (0.90, 3.00)	0.77 (0.50, 1.15)
Ghanem, Zoubi (2011) [25] Israel	Prospective Electronic identification	MEWS	ACCP/SCCM	≥5	IM	1072	74.7 (SD 16.1)	54%	19.4	0.67 (0.63, 0.71)	60 (51, 69)	68 (64, 71)	1.88 (1.58, 2.23)	0.59 (0.47, 0.73)

30

La Regina (2014) [26] Italy	Unknown	MEWS	Unknown	≥4	IM	535	73	49%	14.4	0.59 (0.51, 0.66)	34 (27, 45)	84 (80, 87)	2.06 (1.42, 3.0)	0.79 (0.67, 0.93)
Vorwerk (2008) [23] UK	Retrospective, case note review.	MEWS	Clinical + blood culture	≥5	ED	307	69.7 (IQR 67.5, 71.8)	51%	23.4	0.72 (0.67, 0.77)	72 (60, 82)	59 (53, 65)	1.76 (1.43, 2.18)	0.47 (0.32, 0.69)

(n= number of participants, AUC = area under curve, Prevalence = prevalence of mortality, NR = not reported, ED= emergency department, IM= internal medicine, MEWS = modified early warning score, NEWS = National early warning score)

31

Table 2 Results of the QUADAS-2 assessment

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Cilder (2013) [27]	☺	☺	☺	☹	☺	?	☺
Corfield (2014) [22]	☺	☺	☺	☺	☺	☺	☺
Geier (2013) [24]	☺	☺	☺	☹	☹	?	☹
Ghanem-Zoubi (2011) [25]	☺	☺	☺	☹	☺	☺	☺
La Regina (2014) [26]	?	?	☺	?	?	?	?
Vorwerk (2008) [23]	☺	☺	☺	☺	☺	?	☺

☺ Low Risk   ☹ High Risk   ? Unclear Risk

32

Table 3 Summary estimates of sensitivity, specificity, and likelihood ratios (LR+ and LR-) for the primary analysis and sensitivity analyses

Subgroup	Number of studies	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)
All studies	6	62 (45, 76)	66 (50, 79)	1.79 (1.53, 2.11)	0.59 (0.45, 0.78).
Low risk of bias	2	72 (60, 82) and 89 (85, 92)	59 (53, 65) and 30 (28, 32)	1.76 (1.43, 2.18) & 1.27 (1.22, 1.34)	0.47 (0.32, 0.69) 0.33 (0.24, 0.47)
Low/unclear applicability concerns	5	63 (41, 81)	65 (46, 80)	1.78 (1.47, 2.14)	0.57 (0.42, 0.78)
ED only	4	67 (42, 85)	60 (40, 77)	1.65 (1.36, 2.00)	0.56 (0.38, 0.81)
Threshold = 5	4	71 (51, 85)	58 (40, 74)	1.68 (1.37, 2.05)	0.50 (0.36, 0.69)
MEWS only	5	52 (39, 65)	73 (39, 65)	1.90 (1.66, 2.18)	0.66 (0.55, 0.80)
NEWS only	1	89 (85, 92)	30 (28, 32)	1.27 (1.22, 1.34)	0.33 (0.24, 0.47)

33

34