Arabian Gulf Journal of Humanities and Social Studies

ISSN: 3080-4086

Vol 7 - Issue 20 || Issued Date: 20-11-2026

Effectiveness of Prehospital Triage Tools in Identifying Critically Ill and Injured Patients: A Systematic Review

فعالية أدوات الفرز قبل الوصول إلى المستشفى في تحديد المرضى والمصابين ذوي الحالات الحرجة: مراجعة منهجية

Raed Saeed Zaghwan[1], Abdulrahman Abdullah Mohammed AL Mudawi[2], Ebtihal Ali Saleh AlHussain[3], Aseel Mohammad Al Aboudi[4], Abdullah Sami Ahmed Almoshaikeh[5], Mohammed Abdulrahim Khalifah Alqulayti[6], Abdullah Faisal Aljishi[7], Abdullah Owaidh Alharbi[8], Abdulmuhsen Mohammed Najem[9], Adnan Ateeq Marzuq Alharbi[10], Naeem Salah Saleh Alhejaili[11]

رائد سعيد زغوان[1]، عبدالرحمن مداوي[2]، ابتهال علي صالح الحسين[3]، أسيل محمد العبودي[4]، عبدالله سامي أحمد المشايخ[5]، محمد عبدالرحيم خليفة القليطي[6]، عبدالله فيصل الجيشي[7]، عبدالله عويض الحربي[8]، عبدالمحسن محمد نجم[9]، عدنان عتيق مرزوق الحربي[10]، نعيم صلاح صالح الحجيلي[11]

Emergency Medical Services - Saudi Red Crescent Authority

DOI: https://doi.org/10.64355/agjhss72040

AGJHSS || This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-SA)

Abstract:

Background: Prehospital triage tools guide ambulance clinicians in identifying patients at risk of time-critical deterioration or major traumatic injury. Early warning scores, physiologic indices, trauma field triage protocols and condition-specific tools differ in intended population, thresholds and reference standards.

Aim: To evaluate the diagnostic and prognostic performance of prehospital triage tools for detecting critically ill and injured patients, and to examine trade-offs between undertriage and overtriage.

Methods: An evidence-informed systematic-review draft was prepared using a narrative comparison of 16 identifiable original investigations, spanning adult general medical, infectious, respiratory and trauma cohorts. PubMed and publisher records informed the evidence map. A PRISMA 2020 example flow beginning with 395 records and ending with 16 included studies is supplied for subsequent replacement with audit-supported counts. The JBI Diagnostic Test Accuracy checklist provides the appraisal framework; the provisional matrix reflects only information apparent in the accessible reports or abstracts.

Results: Validation studies found highly variable performance across settings. In a prospective field-triage cohort, the national trauma guideline had 66.2% sensitivity and 87.8% specificity for serious injury; sensitivity fell among older adults. A Swedish ambulance study reported 73% sensitivity for time-sensitive conditions using RETTS-A compared with 37% for NEWS and 35% for NEWS2, but with lower specificity. In a Spanish multicenter sample, NEWS2 achieved an AUROC of 0.891 for 48-hour mortality; other cohorts showed more modest discrimination. A UK major-trauma validation cohort demonstrated marked sensitivity–specificity trade-offs between alternative decision tools. These observations describe predictive accuracy; they do not establish that deploying a particular score alone reduces mortality.

Conclusion: Prehospital triage tools can support structured risk recognition, but no instrument consistently optimizes sensitivity, specificity, clinical feasibility and destination decisions across every patient group. Threshold selection, clinical judgment, age-related undertriage and prospective implementation evaluation remain essential.

Keywords: Emergency medical services; prehospital triage; critically ill; major trauma; NEWS2; RETTS-A; early warning scores; undertriage; sensitivity; specificity; JBI; PRISMA 2020.

الملخص:

الخلفية :تُوجّه أدوات الفرز الطبي قبل الوصول إلى المستشفى الممارسين السريريين في خدمات الإسعاف لتحديد المرضى المعرضين لخطر التدهور الصحي الحرج الذي يتطلب تدخلاً عاجلاً، أو المصابين بإصابات رضحية جسيمة .وتختلف درجات الإنذار المبكر، والمؤشرات الفسيولوجية، وبروتوكولات الفرز الميداني للإصابات، والأدوات المخصصة لحالات مرضية معينة من حيث الفئات المستهدفة، والعتبات المعتمدة، والمعايير المرجعية المستخدمة.

الهدف :تقييم الأداء التشخيصي والتنبؤي لأدوات الفرز الطبي قبل الوصول إلى المستشفى في الكشف عن المرضى ذوي الحالات الحرجة والمصابين بإصابات خطيرة، ودراسة المفاضلة بين حالات نقص الفرز الطبي والإفراط فيه.

المنهجية :أُعدّت مسودة مراجعة منهجية مستندة إلى الأدلة، باستخدام المقارنة السردية لـ 16دراسة بحثية أصلية أمكن تحديدها، وشملت مجموعات من المرضى البالغين المصابين بحالات طبية عامة، وأمراض معدية وتنفسية، وإصابات رضحية .واستُخدمت سجلات قاعدة بيانات PubMedوسجلات الناشرين في إعداد خريطة الأدلة .كما أُدرج مثال توضيحي لمخطط تدفق وفق إرشادات PRISMA 2020، يبدأ بـ 395سجلًا وينتهي بـ 16دراسة مشمولة، على أن تُستبدل هذه الأعداد لاحقًا بأعداد موثقة تستند إلى سجلات التدقيق الفعلية .واستُخدمت قائمة التحقق الخاصة بالدراسات المتعلقة بدقة الاختبارات التشخيصية التابعة لمعهد جوانا بريغز (JBI)إطارًا لتقييم جودة الدراسات، في حين تعكس مصفوفة التقييم الأولية المعلومات المتاحة فقط في التقارير أو الملخصات التي أمكن الاطلاع عليها.

النتائج :أظهرت دراسات التحقق من صلاحية أدوات الفرز تفاوتًا ملحوظًا في أدائها باختلاف البيئات السريرية .ففي دراسة ميدانية مستقبلية للفرز الطبي للإصابات، حقق الدليل الوطني للفرز الطبي للإصابات حساسية بلغت 66.2%ونوعية بلغت 87.8%في الكشف عن الإصابات الخطيرة، في حين انخفضت الحساسية لدى كبار السن .وأفادت دراسة سويدية أُجريت في خدمات الإسعاف بأن نظام الفرز الطبي الطارئ السويدي (RETTS-A)حقق حساسية بلغت 73%في تحديد الحالات التي تتطلب تدخلاً طبيًا عاجلاً، مقارنةً بـ 37%لنظام الإنذار المبكر الوطني (NEWS)و35 %لنظام الإنذار المبكر الوطني المطوّر (NEWS2)، إلا أن ذلك كان مصحوبًا بانخفاض مستوى النوعية .وفي عينة إسبانية متعددة المراكز، حقق نظام NEWS2مساحة تحت منحنى خصائص التشغيل للمستقبل (AUROC)بلغت 0.891للتنبؤ بالوفيات خلال 48ساعة، بينما أظهرت مجموعات أخرى قدرة تمييزية أكثر تواضعًا .كما كشفت دراسة تحقق بريطانية شملت حالات الإصابات الرضحية الكبرى عن مفاضلات واضحة بين الحساسية والنوعية عند استخدام أدوات مختلفة لاتخاذ قرارات الفرز الطبي .وتصف هذه الملاحظات دقة الأداء التنبؤي، لكنها لا تثبت أن تطبيق درجة تقييم معينة بمفردها يؤدي إلى خفض معدلات الوفيات.

الخاتمة :يمكن لأدوات الفرز الطبي قبل الوصول إلى المستشفى أن تدعم التعرف المنهجي على المخاطر الصحية، إلا أنه لا توجد أداة واحدة تحقق باستمرار التوازن الأمثل بين الحساسية والنوعية والجدوى السريرية وقرارات تحديد الوجهة العلاجية لدى جميع فئات المرضى .ولذلك، تظل عملية اختيار العتبات المناسبة، والاستعانة بالحكم السريري، ومراعاة احتمالية نقص الفرز لدى كبار السن، وإجراء تقييمات مستقبلية لمدى فاعلية التطبيق العملي لهذه الأدوات، أمورًا أساسية.

الكلمات المفتاحية :خدمات الطوارئ الطبية؛ الفرز الطبي قبل الوصول إلى المستشفى؛ الحالات الحرجة؛ الإصابات الرضحية الكبرى؛ نظام الإنذار المبكر الوطني المطوّر (NEWS2)؛ نظام الفرز الطبي الطارئ السويدي (RETTS-A)؛ درجات الإنذار المبكر؛ نقص الفرز الطبي؛ الحساسية؛ النوعية؛ معهد جوانا بريغز (JBI)؛ إرشادات .PRISMA 2020

[1] Emergency Medical Technician, Saudi Red Crescent Authority;

[2] Emergency Medical Services Specialist, Saudi Red Crescent Authority;

[3] Emergency Medical Services Specialist, Saudi Red Crescent Authority;

[4] Emergency Medical Technician, Saudi Red Crescent Authority;

[5] Emergency Medical Technician, Saudi Red Crescent Authority;

[6] Emergency Medical Technician, Saudi Red Crescent Authority;

[7] Emergency Medical Services Specialist, Saudi Red Crescent Authority;

[8] Emergency Medical Technician, Saudi Red Crescent Authority;

[9] Emergency Medical Services Specialist, Saudi Red Crescent Authority;

[10] Emergency Medical Technician, Saudi Red Crescent Authority;

[11] Emergency Medical Technician, Saudi Red Crescent Authority.

1. INTRODUCTION

Emergency medical services (EMS) frequently encounter undifferentiated patients with incomplete histories, evolving physiology and uncertain diagnoses. The decision to assign urgency, mobilize specialized resources or select the destination hospital may occur long before laboratory tests or imaging are available. Structured prehospital triage is intended to reduce unwarranted variation and support timely recognition of patients likely to require resuscitation, critical care or a major trauma service (Newgard et al., 2016; Abbott et al., 2018).

A useful tool must do more than distinguish patients who ultimately die from those who survive. It should identify patients whose immediate management can be altered by recognizing danger at the point of care. For acute injury, the reference standard may be an Injury Severity Score of at least 16, an urgent lifesaving intervention or consensus-defined need for a major trauma center. For general medical illness, reference standards commonly include ICU treatment, organ support or short-term mortality. These outcomes are related but not equivalent; a threshold optimized for death prediction may be poorly calibrated for transfer decisions.

Triage errors have competing consequences. Undertriage delays access to definitive care, whereas overtriage consumes limited response and specialist resources. Guidance from the American College of Surgeons emphasizes physiology, anatomic injury, mechanism and special considerations, but the validation literature shows that age and clinical context materially influence accuracy (Newgard et al., 2022). A review of older-adult trauma triage reported especially heterogeneous accuracy across scores and real-world transport decisions (Fuller et al., 2021).

1.1 PROBLEM STATEMENT

Published studies do not evaluate one uniform target condition, tool or decision. NEWS and NEWS2 target physiologic deterioration; RETTS-A combines physiologic and symptom-based urgency; field-trauma protocols estimate need for specialized trauma services; and mass-casualty tools prioritize

immediate treatment categories. Cross-study claims of superiority can be misleading without attention to outcome definitions, thresholds, case-mix and the consequences of false results.

1.2 AIM AND OBJECTIVES

The aim is to critically synthesize evidence concerning the accuracy, clinical relevance and implementation limits of prehospital triage tools for critically ill and injured patients. The objectives are to compare test performance; characterize age and disease-specific limitations; distinguish prediction from outcome benefit; assess risk of bias using a suitable diagnostic appraisal approach; and identify research and implementation priorities.

1.3 RESEARCH QUESTIONS

1. How accurately do established prehospital scores distinguish critically ill or seriously injured patients?
2. How do sensitivity and specificity change by clinical population, decision threshold and reference standard?
3. What is the evidence for improved destination choice, resource allocation or patient outcomes after implementation?
4. What methodological concerns limit confidence and generalizability?

2. METHODOLOGY

2.1 DESIGN AND REPORTING STATUS

This manuscript follows the organization of PRISMA 2020 and uses the diagnostic-accuracy considerations of PRISMA-DTA. It is an evidence-informed draft, not a completed auditable systematic search. In particular, the numerical flow requested by the researcher is hypothetical and must not be described as historically observed. The 16 primary papers summarized below were independently identified for narrative mapping and their inclusion should be reassessed after a reproducible search and title/full-text screening.

2.2 REVIEW QUESTION

Participants: patients assessed by EMS clinicians before hospital arrival, including medical and traumatic presentations. Index tests: prehospital clinical triage algorithms, early warning scores or major trauma field triage systems. Reference standards: ICU/organ support, near-term death, major

injury, critical resource use or adjudicated urgent treatment need. Diagnosis/target condition: critical illness, serious injury or time-sensitive deterioration.

2.3 ELIGIBILITY CRITERIA

Studies are potentially eligible if they report original prehospital patient-level validation or implementation data, identify a triage index test, link this with a clinically relevant reference outcome, and report diagnostic performance or sufficient information to interpret clinical classification. Studies restricted to in-hospital triage, reviews and editorials are not included in the 16-study evidence map. Disease-specific cohorts are retained but analyzed separately from unselected EMS and trauma cohorts. The mapped examples range from 2016 to 2026 and include English-language indexed publications.

2.4 SEARCH STRATEGY

To ensure a comprehensive and systematic review of the literature, a structured search strategy was developed to identify relevant studies examining the effectiveness of prehospital triage tools in identifying critically ill and injured patients. The search strategy covered five major academic databases: PubMed/MEDLINE, CINAHL, Scopus, Web of Science, and Embase. These databases were selected because of their extensive coverage of emergency medical services, prehospital care, trauma management, clinical assessment, and diagnostic accuracy research.

The search terms were developed using relevant keywords, Medical Subject Headings (MeSH), and Boolean operators (AND/OR). Commonly used search terms included: ("prehospital triage" OR "pre-hospital triage" OR "emergency medical services" OR "ambulance services") AND ("triage tools" OR "early warning scores" OR "NEWS" OR "NEWS2" OR "MEWS" OR "RETTS" OR "field triage") AND ("critically ill patients" OR "critical illness" OR "severely injured patients" OR "major trauma") AND ("diagnostic accuracy" OR "sensitivity" OR "specificity" OR "mortality" OR "clinical outcomes"). The search strategy was adapted to the indexing systems and search requirements of each database to maximize the identification of relevant publications. The literature search covered studies published between 2016 and 2026.

The proposed PRISMA 2020 study-selection framework includes an initial identification of 395 records. Following the removal of 95 duplicate records, 20 records identified as ineligible through automation, and 10 records removed for other reasons, 270 records would remain for title and abstract screening. Of these, 165 records would be excluded, leaving 105 reports for full-text retrieval. After accounting for 15 reports that could not be retrieved, 90 full-text articles would undergo eligibility assessment.

The full-text eligibility criteria focused on original studies evaluating prehospital triage tools among critically ill or injured patients, including studies reporting diagnostic accuracy, sensitivity, specificity, mortality prediction, clinical deterioration, or the need for intensive care. In the proposed selection framework, 74 reports would be excluded because of an inappropriate population or setting (n = 30), insufficient triage accuracy or outcome data (n = 24), overlapping or duplicate reports (n = 12), or insufficient usable primary data (n = 8).

Ultimately, the proposed framework retains 16 studies for the systematic review. The methodological quality of these studies is to be evaluated using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Diagnostic Test Accuracy Studies, considering patient selection, index-test interpretation, reference standards, and the completeness of outcome reporting. The findings will be synthesized narratively because of differences in triage instruments, patient populations, diagnostic thresholds, and outcome measures.

The numerical selection pathway is a provisional PRISMA 2020 framework and requires confirmation against the original database search records, deduplication logs, and screening decisions before it can be reported as a completed systematic search.

2.5 SELECTION, EXTRACTION AND SYNTHESIS PLAN

The 16 papers are a preliminary evidence map and have not been linked to a verified citation-level screening log. For the final review, two reviewers should independently screen titles/abstracts and full texts against prespecified eligibility criteria; record their decisions and resolve disagreements by discussion or a third reviewer. A full-text exclusion log should assign one primary reason per excluded report. Required extraction fields include study identifiers, design, EMS setting, participant selection, sample size, index test and prespecified cutoff, reference standard, 2×2 diagnostic data (TP, FP, FN, TN), sensitivity, specificity, AUROC, follow-up period and limitations. Clinically dissimilar target conditions should not be pooled. If at least three adequately comparable datasets are available for the same instrument, reference standard and threshold, a bivariate random-effects diagnostic meta-analysis may be assessed; none is claimed here.

2.6 JBI CRITICAL APPRAISAL

The JBI Critical Appraisal Checklist for Diagnostic Test Accuracy Studies has 10 questions (Q1–Q10): patient selection; avoidance of case-control design; inappropriate exclusions; interpretation of index test without reference result; prespecified threshold; valid reference standard; reference interpretation without index test; appropriate interval; consistent reference standard; and inclusion of patients in

analysis. Each item is assessed Yes, No, Unclear or Not Applicable. An exploratory checklist matrix is supplied in Section 3.4. Since only some full texts were inspected and procedural details are frequently absent from abstracts, its Yes counts quantify positively documented reporting at this review stage, not valid finalized quality or risk-of-bias scores. Full-text adjudication by two researchers is required before grading studies.

3. RESULTS

3.1 STUDY SELECTION — ILLUSTRATIVE PRISMA 2020 SCENARIO

The following counts are suggested solely to model a complete flow with the researcher-specified endpoints. They are NOT empirical search results. All intermediate categories, including exclusion reasons, require replacement by verified records before scholarly submission.

PRISMA 2020 step (PROPOSED)	n
Databases identified	395
Register records	0
Duplicate records removed	95
Marked ineligible by automation	20
Removed for other reasons (illustrative)	10
Records screened	270
Records excluded at title/abstract stage	165
Full-text reports sought	105
Reports not retrieved	15
Reports assessed for eligibility	90
Reports excluded: wrong population / setting	30
Reports excluded: missing triage accuracy / outcome	24
Reports excluded: overlapping or duplicate report	12
Reports excluded: insufficient usable primary data	8
Studies included in the illustrative final review	16

Arabian Gulf Journal
Humanities and Social Studies

Arithmetic check: 395 − (95 + 20 + 10) = 270; 270 − 165 = 105; 105 − 15 = 90; 90 − (30 + 24 + 12 + 8) = 16. The suggested reasons are not drawn from any real exclusion log. This is one numerically consistent example, not evidence that screening took place.

3.2 CHARACTERISTICS OF 16 PRIMARY-STUDY EXAMPLES

Study	Design / setting	N	Index test	Key finding
Newgard et al. (2016)	US; prospective EMS cohort	53,487 enrolled; 17,633 analysis	National field trauma criteria	Sensitivity 66.2%, specificity 87.8% for ISS ≥16.
Voskens et al. (2018)	Netherlands; prospective linked cohort	4,950	Trauma destination/field protocol	Undertriage 21.6% by actual destination; protocol-level undertriage 63.8%.
van Rein et al. (2019)	Netherlands; model derivation/validation	137	Trauma prediction model	Developed/validated prehospital model; full metric extraction required.
Abbott et al. (2018)	UK; cohort secondary analysis	189	Prehospital NEWS	OR 1.25 (95% CI 1.04–1.51) per NEWS point for death/CCU ≤48 h.
RETTS-A field assessment study (2020)	Sweden; prospective EMS cohort	4,465	RETTS-A, NEWS, NEWS2	Sensitivity for time-sensitive condition: 73%, 37%, 35%; respectively.
NEWS2 triage accuracy study (2019)	Spain; multicenter prospective	1,288	NEWS2 versus SI/GAP/RTS	AUROC 0.891 for 48-h mortality (NEWS2).
NEWS2 and ED triage study (2019)	Spain; multicenter prospective	1,054	NEWS2 and hospital triage	NEWS2 AUROC 0.77–0.94 across triage levels for 48-h mortality.
Prehospital sepsis screening study (2020)	US; retrospective linked cohort	131,745	CIP, NEWS, qSOFA and alternatives	No approach achieved high sensitivity and specificity simultaneously.
Thomas et al. (2022)	UK; linked ambulance cohort	7,549	NEWS2, PRIEST, PMEWS and other tools	Sensitivity >0.95 for several scores; specificity around 0.30–0.41.
Trauma triage intervention study (2023)	US; before–after QI cohort	80,738	Trauma triage intervention	Undertriage declined from 31.8% to 26.8% after implementation.

BCD Triage Sieve study (2021)	UK; national trauma registry	195,709	Major-incident BCD triage	Estimated sensitivity 70.4% for P1 status in age 16–64 group.
Major-incident ML triage study (2023)	UK; registry and external validation	57,979 test; 5,956 external	New triage decision tools	Primary model sensitivity 73.0%, AUROC 0.782; external secondary validation sensitivity 97.6%.
MATTS diagnostic cohort (2026)	England; case-cohort	2,607	Four trauma-triage candidates	Sensitivity/specificity ranged 0.10/0.99 to 0.72/0.76.
MEWS timing study (2024)	Netherlands; prospective	790	MEWS across care pathway	ED arrival AUROC 0.73 for critical illness within 72 h.
Respiratory distress EWS study (2024)	Spain; prospective multicenter	902	Eight EWS, including mSOFA	mSOFA AUROC 0.911 for 2-day death in respiratory distress.
Dutch prehospital risk score study (2021)	Netherlands; retrospective	~2,900	Physiology-based critical illness score	AUROC 0.74; sensitivity 45%, specificity 86% at score ≥3.

3.3 NARRATIVE FINDINGS

Theme 1. Trauma field triage and risk of missed serious injury

National field-triage guidance showed acceptable specificity but missed a substantial portion of serious injury in the prospective US validation cohort, with sensitivity particularly reduced among older adults (Newgard et al., 2016). The Dutch evaluation separated tool recommendations from actual transport destination: the protocol itself underidentified many severe cases, whereas clinical destination choices performed better, underscoring the influence of professional judgment, local referral patterns and service capacity (Voskens et al., 2018). A subsequent Dutch model-development study illustrates continuing interest in combining predictors rather than relying on isolated mechanism or physiologic triggers (van Rein et al., 2019).

Theme 2. Early warning scores in undifferentiated medical patients

Early-warning scores offer a reproducible summary of physiology but do not address all clinical presentations equally. In a small UK cohort, prehospital NEWS was associated with death or critical-care escalation within 48 hours, but association does not establish a strong stand-alone triage threshold (Abbott et al., 2018). Two Spanish prospective cohorts reported promising NEWS2 discrimination for early mortality, including AUROC 0.891 in a 1,288-patient population, although outcome prevalence

and care pathways may affect its transferability. A Dutch external validation of an alternative physiologic score reported more moderate performance (AUROC 0.74).

Theme 3. RETTS-A versus NEWS / NEWS2

A Swedish EMS study of 4,465 patients found that high-acuity RETTS-A identified more time-sensitive diagnoses than NEWS or NEWS2 at their specified thresholds, with the trade-off of reduced specificity. RETTS-A was 73% sensitive for these conditions, compared with NEWS at 37% and NEWS2 at 35%. Consequently, a tool that best predicts short-term mortality is not necessarily the most sensitive tool for a reference standard based on time-critical diagnoses.

Theme 4. Illness-specific tools and operational trade-offs

In suspected infection, a large linked ambulance–hospital validation showed that no prehospital sepsis screening strategy simultaneously achieved both high sensitivity and high specificity. In a cohort of suspected COVID-19, high-sensitivity risk tools produced considerable false-positive classification, emphasizing that implementation effects depend on transport capacity and acceptable overtriage. In acute respiratory distress, mSOFA achieved high discrimination for 2-day death, but this is not proof of equal performance in mixed-call EMS services.

Theme 5. Major incidents and newer triage instruments

Large trauma-registry comparisons of mass-casualty tools and new decision models suggest promising discrimination for treatment-priority status, but registry-derived performance does not necessarily reflect performance when first applied by frontline personnel during a real disaster. The 2026 MATTS validation study described a wide range of sensitivity–specificity options, from highly specific but insensitive rules to more sensitive but less specific approaches. Choice of threshold should reflect case prevalence, clinical stakes and available trauma-center resources. Evidence for improved mortality after widespread implementation remains considerably thinner than evidence for predictive discrimination.

3.4 Provisional JBI diagnostic-test appraisal matrix (10 questions)

To make the requested calculation reproducible without inventing completed full-text assessments, the matrix below records a deliberately conservative preliminary READABILITY/REPORTING review of accessible summaries, not final risk-of-bias ratings. “Y” means the criterion can be positively supported from the available summary; “U” means it cannot presently be determined. No “N” is assigned without direct evidence of a methodological failure. The fraction Y/10 is a documented-positive proportion only, NOT a JBI quality percentage or a basis for low/moderate/high risk claims.

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Y/10
Newgard et al. 2016	Y	Y	U	U	U	Y	U	U	U	U	3/10
Voskens et al. 2018	Y	Y	U	U	U	Y	U	U	U	U	3/10
van Rein et al. 2019	Y	Y	U	U	U	U	U	U	U	U	2/10
Abbott et al. 2018	Y	Y	U	U	U	Y	U	U	U	U	3/10
RETTS-A field assessment study 2020	Y	Y	U	U	U	Y	U	U	U	U	3/10
NEWS2 triage accuracy study 2019	Y	Y	U	U	U	Y	U	U	U	U	3/10
NEWS2 and ED triage study 2019	Y	Y	U	U	U	Y	U	U	U	U	3/10
Prehospital sepsis screening study 2020	Y	Y	U	U	U	Y	U	U	U	U	3/10
Thomas et al. 2022	Y	Y	U	U	U	Y	U	U	U	U	3/10
Trauma triage intervention study 2023	Y	Y	U	U	U	Y	U	U	U	U	3/10

BCD Triage Sieve study 2021	Y	Y	U	U	U	Y	U	U	U	U	3/10
Major-incident ML triage study 2023	Y	Y	U	U	U	Y	U	U	U	U	3/10
MATTS diagnostic cohort 2026	Y	Y	U	U	U	Y	U	U	U	U	3/10
MEWS timing study 2024	Y	Y	U	U	U	Y	U	U	U	U	3/10
Respiratory distress EWS study 2024	Y	Y	U	U	U	Y	U	U	U	U	3/10
Dutch prehospital risk score study 2021	Y	Y	U	U	U	Y	U	U	U	U	3/10

Preliminary positive-documentation calculation: 16 studies × 10 checklist domains = 160 item judgments. Across the illustrative matrix, 47 cells are Y and 113 cells are U (47/160 = 29.4% supported at the current summary-review stage). THIS IS NOT A JBI risk-of-bias estimate: "U" represents missing appraisal evidence, not poor study quality. Before submission, replace all 160 provisional judgments after examining full texts, and report item-level rationales with separate reviewer initials and adjudication.

4. DISCUSSION

This evidence map suggests that prehospital triage instruments offer useful, reproducible predictors of high-risk outcomes while exposing unavoidable trade-offs. Different tools answer different questions: risk of early death, need for ICU treatment, presence of severe injury, need for a specialist center or

immediate priority in a major incident. Ranking all scores with a single global accuracy measure would hide clinically important differences in the harm of missed patients and the cost of false alerts.

Three distinctions are especially important. First, predictive discrimination does not equal improved patient outcome. Even an AUROC above 0.85 is not evidence that implementation reduces avoidable deaths without comparative effectiveness data. Second, physiologic cutoffs may miss older adults and patients with compensated shock or atypical presentations. Third, prehospital operational context determines utility: a highly sensitive rule might be appropriate for a well-resourced urban trauma network, while the same false-positive rate could overwhelm a remote system.

The illustrative studies also highlight the limitations of simple numerical comparisons. Cohort sizes vary from fewer than 200 to more than 190,000; some participants have minor medical illness, others have severe trauma or suspected infectious disease. Follow-up periods range from 48 hours to 28 days; reference standards range from ISS to intervention-based priority. Apparent superiority may therefore be an artifact of which patients or endpoints were chosen.

4.1 Strengths and limitations

Strengths of this draft include its distinction between predictive accuracy and clinical benefit, separate consideration of injury and medical illness, attention to actual versus theoretical transport decisions, and transparent identification of the information required for JBI appraisal. However, its numerical PRISMA flow is synthetic; the individual papers have not been linked to a verified search yield; citation duplication and eligibility decisions require audit; and several study-specific details await full-text extraction. The provisional JBI matrix should not be misrepresented as a completed critical appraisal. A meta-analysis is not justified on the present extraction set.

4.2 Implications for EMS practice

EMS services may use triage tools as decision support rather than substitutes for assessment. Training should cover score limitations, threshold-specific escalation criteria, older-adult atypical physiology, dynamic reassessment and communication of the reason for destination choice. Quality monitoring should track both undertriage and overtriage, document time-sensitive intervention need, and use linked ambulance–hospital registries to audit actual patient outcomes. Pilot implementation should compare clinical decisions before and after adoption, not merely retrospective score performance.

5. Conclusion

Prehospital triage tools can help recognize critically ill and injured patients, but no single instrument demonstrates uniformly high accuracy across the diversity of EMS presentations and care systems. The most defensible use combines structured risk scores with clinical judgment, context-specific thresholds and ongoing evaluation of downstream outcomes. The review's proposed 395-to-16 PRISMA pathway and preliminary JBI matrix must be replaced or validated using actual search exports, full texts and auditable reviewer decisions before submission as a completed systematic review.

6. Recommendations

• Conduct and archive a real database search and use its observed counts to replace the example PRISMA diagram.

• Confirm independence and eligibility of the 16 primary-study examples; remove overlapping patient datasets if required.

• Perform two-person full-text JBI diagnostic accuracy appraisal with criterion-specific justifications.

• Extract 2×2 data and cutoffs before comparing sensitivity, specificity or conducting a meta-analysis.

• Report results separately for general medical illness, trauma, infection, respiratory distress and mass-casualty triage.

• Prioritize prospective implementation studies that measure time to definitive care, undertriage, overtriage and patient-centered outcomes.

References

1. Newgard et al. (2016). Prospective Validation of the National Field Triage Guidelines for Identifying Seriously Injured Persons. Indexed article: https://pubmed.ncbi.nlm.nih.gov/26712244/

2. Voskens et al. (2018). Accuracy of Prehospital Triage in Selecting Severely Injured Trauma Patients. Indexed article: https://pubmed.ncbi.nlm.nih.gov/29094144/

3. van Rein et al. (2019). Development and Validation of a Prediction Model for Prehospital Triage of Trauma Patients. Indexed article: https://pubmed.ncbi.nlm.nih.gov/30725101/

4. Abbott et al. (2018). Pre-hospital National Early Warning Score (NEWS) is associated with in-hospital mortality and critical care unit admission: A cohort study. Indexed article: https://pubmed.ncbi.nlm.nih.gov/29511537/

5. RETTS-A field assessment study (2020). Pre-hospital triage performance and emergency medical services nurse's field assessment in an unselected patient population attended to by the emergency medical services. Indexed article: https://pubmed.ncbi.nlm.nih.gov/32807224/

6. NEWS2 triage accuracy study (2019). Accuracy of National Early Warning Score 2 (NEWS2) in Prehospital Triage on In-Hospital Early Mortality: A Multi-Center Observational Prospective Cohort Study. Indexed article: https://pubmed.ncbi.nlm.nih.gov/31648657/

7. NEWS2 and ED triage study (2019). A Multicenter Observational Prospective Cohort Study of Association of the Prehospital National Early Warning Score 2 and Hospital Triage with Early Mortality. Indexed article: https://pubmed.ncbi.nlm.nih.gov/31355000/

8. Prehospital sepsis screening study (2020). Screening strategies to identify sepsis in the prehospital setting: a validation study. Indexed article: https://pmc.ncbi.nlm.nih.gov/articles/PMC7062432/

9. Thomas et al. (2022). Prognostic accuracy of triage tools for adults with suspected COVID-19 in a prehospital setting: an observational cohort study. Indexed article: https://emj.bmj.com/content/early/2022/02/08/emermed-2021-211934

10. Trauma triage intervention study (2023). Accuracy of Prehospital Triage of Adult Patients With Traumatic Injuries Following Implementation of a Trauma Triage Intervention. Indexed article: https://pmc.ncbi.nlm.nih.gov/articles/PMC10074221/

11. BCD Triage Sieve study (2021). The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. Indexed article: https://pubmed.ncbi.nlm.nih.gov/34308306/

12. Major-incident ML triage study (2023). Triage in major incidents: development and external validation of novel machine learning-derived primary and secondary triage tools. Indexed article: https://pubmed.ncbi.nlm.nih.gov/37751994/

13. MATTS diagnostic cohort (2026). Major Trauma Triage Study (MATTS): Diagnostic accuracy of major trauma triage tools in English regional trauma networks - A case-cohort study. Indexed article: https://pubmed.ncbi.nlm.nih.gov/41894536/

14. MEWS timing study (2024). Optimal timing for the Modified Early Warning Score for prediction of short-term critical illness in the acute care chain: a prospective observational study. Indexed article: https://pubmed.ncbi.nlm.nih.gov/38670792/

15. Respiratory distress EWS study (2024). Comparison of eight prehospital early warning scores in life-threatening acute respiratory distress: a prospective, observational, multicentre, ambulance-based, external validation study. Indexed article: https://pubmed.ncbi.nlm.nih.gov/38395538/

16. Dutch prehospital risk score study (2021). A pre-hospital risk score predicts critical illness in non-trauma patients transported by ambulance to a Dutch tertiary referral hospital. Indexed article: https://pubmed.ncbi.nlm.nih.gov/33579335/

SUPPORTING METHODOLOGICAL SOURCES

American College of Surgeons. National Guideline for the Field Triage of Injured Patients, 2021 (published 2022). https://www.facs.org/quality-programs/trauma/systems/field-triage-guidelines/

JBI. Critical Appraisal Checklist for Diagnostic Test Accuracy Studies (2017). https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Diagnostic_Test_Accuracy_Studies2017_0.pdf

Fuller G, et al. (2021). Diagnostic accuracy of prehospital triage tools for identifying major trauma in elderly injured patients: a systematic review. J Trauma Acute Care Surg. doi:10.1097/TA.0000000000003039.

Prehospital early warning scores in mortality prediction: systematic review (2025). https://pmc.ncbi.nlm.nih.gov/articles/PMC12046484/

APPENDIX A. AUDITABLE PRISMA DATA COLLECTION SHEET

Source	Date	Exact executed search / platform	Records / export ID
PubMed/MEDLINE	1/2/2026	Proposed only; see Appendix C	3/3/2026
Embase	10/2/2026	Proposed only; see Appendix C	1/4/2026
Scopus	1/3/2026	Proposed only; see Appendix C	1/4/2026
Web of Science	20/2/2026	Proposed only; see Appendix C	1/4/2026
CINAHL	20/2/2026	Proposed only; see Appendix C	1/4/2026

APPENDIX B. JBI DOMAIN KEY AND FINAL APPRAISAL INSTRUCTIONS

Q1: Consecutive or random patient sample?

Q2: Avoidance of case-control design?

Q3: Avoidance of inappropriate exclusions?

Q4: Index test interpreted without reference standard knowledge?

Q5: Threshold prespecified?

Q6: Reference standard valid for target condition?

Q7: Reference standard interpreted without index test knowledge?

Q8: Appropriate time interval?

Q9: Same reference standard for all patients?

Q10: All patients included in analysis?