Triage Performance in Emergency Medicine: A Systematic Review

Jeremiah S. Hinson, MD, PhD*; Diego A. Martinez, PhD; Stephanie Cabral, BS; Kevin George, BS; Madeleine Whalen, MSN, MPH; Bhakti Hansoti, MBChB, PhD; Scott Levin, PhD, MS

*Corresponding Author. E-mail: hinson@jhmi.edu, Twitter: @Hinson_EM.

Study objective: Rapid growth in emergency department (ED) triage literature has been accompanied by diversity in study design, methodology, and outcome assessment. We aim to synthesize existing ED triage literature by using a framework that enables performance comparisons and benchmarking across triage systems, with respect to clinical outcomes and reliability.

Methods: PubMed, EMBASE, Scopus, and Web of Science were systematically searched for studies of adult ED triage systems through 2016. Studies evaluating triage systems with evidence of widespread adoption (Australian Triage Scale, Canadian Triage and Acuity Scale, Emergency Severity Index, Manchester Triage Scale, and South African Triage Scale) were cataloged and compared for performance in identifying patients at risk for mortality, critical illness and hospitalization, and interrater reliability. This study was performed and reported in adherence to Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines.

Results: A total of 6,160 publications were identified, with 182 meeting eligibility criteria and 50 with sufficient data for inclusion in comparative analysis. The Canadian Triage and Acuity Scale (32 studies), Emergency Severity Index (43), and Manchester Triage Scale (38) were the most frequently studied triage scales, and all demonstrated similar performance. Most studies (6 of 8) reported high sensitivity (>90%) of triage scales for identifying patients with ED mortality as high acuity at triage. However, sensitivity was low (<80%) for identification of patients who had critical illness outcomes and those who died within days of the ED visit or during the index hospitalization. Sensitivity varied by critical illness and was lower for severe sepsis (36% to 74%), pulmonary embolism (54%), and non–ST-segment elevation myocardial infarction (44% to 85%) compared with ST-segment elevation myocardial infarction (56% to 92%) and general outcomes of ICU admission (58% to 100%) and lifesaving intervention (77% to 98%). Some proportion of hospitalized patients (3% to 45%) were triaged to low acuity (level 4 to 5) in all studies. Reliability measures (κ) were variable across evaluations, with only a minority (11 of 42) reporting κ above 0.8.

Conclusion: We found that a substantial proportion of ED patients who die postencounter or are critically ill are not designated as high acuity at triage. Opportunity to improve interrater reliability and triage performance in identifying patients at risk of adverse outcome exists. [Ann Emerg Med. 2018;■:1-13.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2018.09.022

INTRODUCTION

Triage, a concept developed and refined on the battlefield, has been central to the practice of emergency medicine for more than half a century.[1] As emergency departments (EDs) face escalating patient volumes,[2] persistent crowding,[3] and patient populations with more complex disease,[4] the need for accurate and reliable triage has intensified. This has spurred the development and rapid adoption of a set of standardized triage systems designed to structure the triage decisionmaking process. Triage systems with published evidence of widespread adoption include the Australasian Triage Scale (ATS),[5] Canadian Triage and Acuity Scale (CTAS),[6] Emergency Severity Index (ESI),[7] Manchester Triage Scale (MTS),[8] and South African Triage Scale (SATS).[9]

These standardized triage systems share core elements (Table 1): their objective is to identify and prioritize patients with critical time-sensitive care needs; all deploy a 5-level classification scheme, a general practice endorsed by the American College of Emergency Physicians (ACEP) and Emergency Nursing Association[10]; all set targets for timeliness to physician contact per triage level; all were developed through provider group consensus; and they universally rely on some level of subjective judgment by trained triage providers to execute.

Despite their shared characteristics, there is substantial divergence in approach. The degree of reliance on provider judgment varies considerably by triage system. For example, the CTAS and MTS are on the prescriptive end of the spectrum, translating detailed clinical

Editor's Capsule Summary

What is already known on this topic

Numerous emergency department (ED) triage scales exist to prioritize the immediacy with which patients should be treated.

What question this study addressed

This study examined the reliability, as well as the sensitivity and specificity of predicting mortality (ED, inhospital, 1 day, and 7 days after the ED visit), critical illness, and hospitalization for 5 ED triage scales.

What this study adds to our knowledge

Not enough studies with similar methodology exist to compare the scales' validity for the outcomes identified. None of the scales demonstrated consistently high reliability.

How this is relevant to clinical practice

More studies are needed to compare the performance of ED triage scales for clinically important outcomes and within different ED patient populations.

discriminators and vital sign combinations to specific triage levels. In comparison, the user-friendly ESI provides higher-level guidance with some vital sign–based recommendations, lending itself to heavier dependence on triage provider intuition. The ESI is also the only major triage system to incorporate projected resource use into triage decisionmaking with patient flow in mind. Furthermore, the ATS (Australia), CTAS (Canada), ESI (United States), and MTS (United Kingdom) were developed for use in high-resource settings, whereas only the SATS (South Africa) considers resource-limited environments.[9] These triage system differences and variability in published evaluations (evaluation design, outcomes, and analytic methods) significantly limit the collective value of a large and rapidly increasing body of emergency medicine triage literature. Although several reviews on the topic have been performed, to our knowledge none have explored this variability or proposed a framework to facilitate standardization of study design and reporting.

Goal of This Investigation

The primary objective of this systematic review was to characterize the most commonly studied ED triage systems across the globe and to evaluate their performance with respect to identifying clinical outcomes and reliability. The goal was to develop and apply a framework to compare scientific evaluations of ED triage to establish a foundation to benchmark assessments of triage for both research and quality improvement across triage systems, and to illuminate evidence and gaps in performance of specific triage systems generally, and among particular demographic and clinical subpopulations treated in the ED.

Table 1. Triage system characteristics.

Triage System	CTAS	ESI	MTS	ATS	SATS
Stated objective	Provide patients with timely care	Prioritize patients by immediacy of care needs and resource	Rapidly assess a patient and assign a priority based on clinical need	Ensure patients are treated in order of clinical urgency and allocate patients to the most appropriate treatment area	Prioritize patients based on medical urgency in contexts where there is a mismatch between demand and capacity
Recommended time to physician contact, min	1: immediate 2: ≤15 3: ≤30 4: ≤60 5: ≤120	1: immediate 2: ≤15 3: none 4: none 5: none	Red: immediate Orange: ≤10 Yellow: ≤60 Green: ≤120 Blue: ≤240	1: immediate 2: ≤15 3: ≤30 4: ≤60 5: ≤120	Red: immediate Orange: ≤10 Yellow: ≤60 Green: ≤240 Blue: ≤120
Discriminators					
Clinical	Yes	No	Yes	Yes	Yes
Vital signs	Yes	Yes	Yes	Yes	Yes
Pain score	Yes (10-point)	Yes (visual analog scale)	Yes (3-point)	No	Yes (4-point)
Resource use	No	Yes	No	No	No
Pediatrics	Separate version	Separate vital sign differentiators	Considered within algorithm	Considered within algorithm	Separate flowchart

MATERIALS AND METHODS

Selection of Participants

Published articles were identified by a systematic search of PubMed, EMBASE, Scopus, and Web of Science, including all articles through December 31, 2016, using the search terms detailed in Appendix E1 (available online at http://www.annemergmed.com). Abstracts for original research were screened for study inclusion with the following eligibility criteria: the triage system studied was designed for use in an ED, and the triage system assessed was intended for application in a clinically undifferentiated adult population. Guideline and review publications were excluded, as were publications unavailable in English or full text. Studies of triage systems designed for outside the ED (ie, out-of-hospital) or intended for application to a specific clinical cohort (eg, trauma patients) or demographic group were also excluded. However, studies that used a tool developed for an undifferentiated adult ED population but were evaluated in specific clinical or demographic cohorts (eg, pediatrics) were included. Studies were screened for inclusion by 3 authors independently (S.C., K.G., and S.L.), with interrater reliability assessed on a test set of 100 articles, using the Cohen κ statistic. This study was performed and reported in adherence to Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines.[11]

Data Collection and Processing and Primary Data Analysis

Studies that met inclusion criteria underwent full-text review. Standardized data extraction forms (Appendix E2, available online at http://www.annemergmed.com) were used to record the triage study evaluation design, geographic setting, ED characteristics, study locations (eg, multisite versus single site), and population size (patients and triage providers). Patient outcome measures used to assess validity and evaluate triage system performance were also extracted, as were reliability measures to assess the consistency of triage system application across providers and compared with a criterion standard. When multiple criterion standard comparator groups existed, we reported the highest reliability comparison. Studies that focused on any of the major 5-level triage systems (ATS, CTAS, ESI, MTS, and SATS) were subject to cataloging and analysis of outcome and reliability measures to facilitate comparisons of the standardized triage systems most widely used in practice.

Triage studies were found to be highly variable in terms of approach to measurement, data analysis, and reporting of results. This stimulated the development of a cross-cutting framework used to organize and compare triage system performance, validity, and reliability. For studies assessing triage system validity we focused on the detection of clinical outcomes and their stratification across triage levels. Clinical outcomes were grouped into measures of mortality (eg, ED, inhospital, 7-day outcome), critical illness (eg, ST-segment elevation myocardial infarction [STEMI], pulmonary embolism), and hospitalization. The ability to identify patients at triage with these outcomes was determined with the principal summary measures of sensitivity and specificity. Sensitivity and specificity were extracted directly from primary articles when available and calculated from raw data with standard 2×2 tables when not. For the high-severity outcome groups (mortality and critical illness), we computed the sensitivity and specificity of designating the patients high acuity: treated in less than 15 minutes (level 1 or 2) (Table 1). For the hospitalization outcome, we computed sensitivity and specificity for designating patients at least moderate acuity (levels 1 to 3). Reliability measures (κ) extracted from studies were grouped into clusters based on the study design (retrospective versus prospective), rater comparison approach (between-rater versus criterion standard), patient encounters versus paper-based scenarios, and κ method (weighted versus unweighted). Studies were included in each of these analyses if they focused on nurses using a standardized triage system (ATS, CTAS, ESI, MTS, or SATS) and reported or included sufficient data to calculate our outcome-based validity or reliability performance metrics. For studies in which these data were measured but not reported in detail, we made attempts to obtain primary data by contacting corresponding authors.

We elected not to combine data to perform meta-analysis. Although results could have been combined for a small subset of included studies, the majority were divergent in terms of design, population of focus, triage scale examined, and outcome measured. Ultimately, the small number of studies that could be included in each meta-analysis and the many studies that would not be included limited the value of such analyses, and meta-analysis was determined nonbeneficial to the overall interpretation of systematic review results.

For all included studies that performed outcomes-based validity assessment of triage scales, we applied the Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS-2) tool. QUADAS-2 defines quality based on 2 constructs: risk of bias and applicability of the study to the reviewer's research question.[12] Risk of bias may be introduced through the domains of (1) patient selection, (2) index test execution and interpretation, (3) reference standard collection and interpretation, (4) and flow and timing between the index test and reference standard. Concerns about applicability may also exist for domains 1 to 3, degrading the quality of a study. QUADAS-2 domain measures were evaluated by 2 authors (J.S.H. and S.L.) independently, with disagreements resolved collectively. QUADAS-2 was applied to each study at the outcome level, and QUADAS-2 results were aggregated by the overarching outcome themes of mortality, critical illness, and hospitalization.

RESULTS

The database search strategy yielded 6,160 unique publications for screening, with 182 (3%) meeting overarching eligibility criteria (Figure 1). Interrater agreement between the 3 screening authors was good (κ=0.77; 95% confidence interval 0.74 to 0.80). Of the 182 eligible studies, a major 5-level triage system was evaluated in 73% of them: ATS (14 studies), CTAS (32), ESI (43), MTS (38), and SATS (12). The countries where triage evaluation was most frequently performed included the United States (34), Canada (23), Netherlands (13), Portugal (12), and Australia (11). The growth in volume of triage studies published over time is shown in Figure 2. Overall, 50 studies were included in the major triage system comparative performance analysis. The major attributes of the 5-level standardized triage systems of focus are shown in Table 1. Outcomes-based validity was assessed by 32 studies and reliability by 21. A summary of included studies is shown in Table 2.

Clinical outcomes used to assess the validity of the triage system within each study context were grouped by measures of mortality (17 evaluations), critical illness (22 evaluations), and inpatient hospitalization (20 evaluations). The defined clinical outcome, population studied, population size, triage system evaluated, and sensitivity and specificity of outcome detection at triage are reported in Table 2 and Figures 3 and 4.

Mortality outcomes included ED (8 evaluations), inhospital (7), 1-day (1), and 7-day mortality (1)

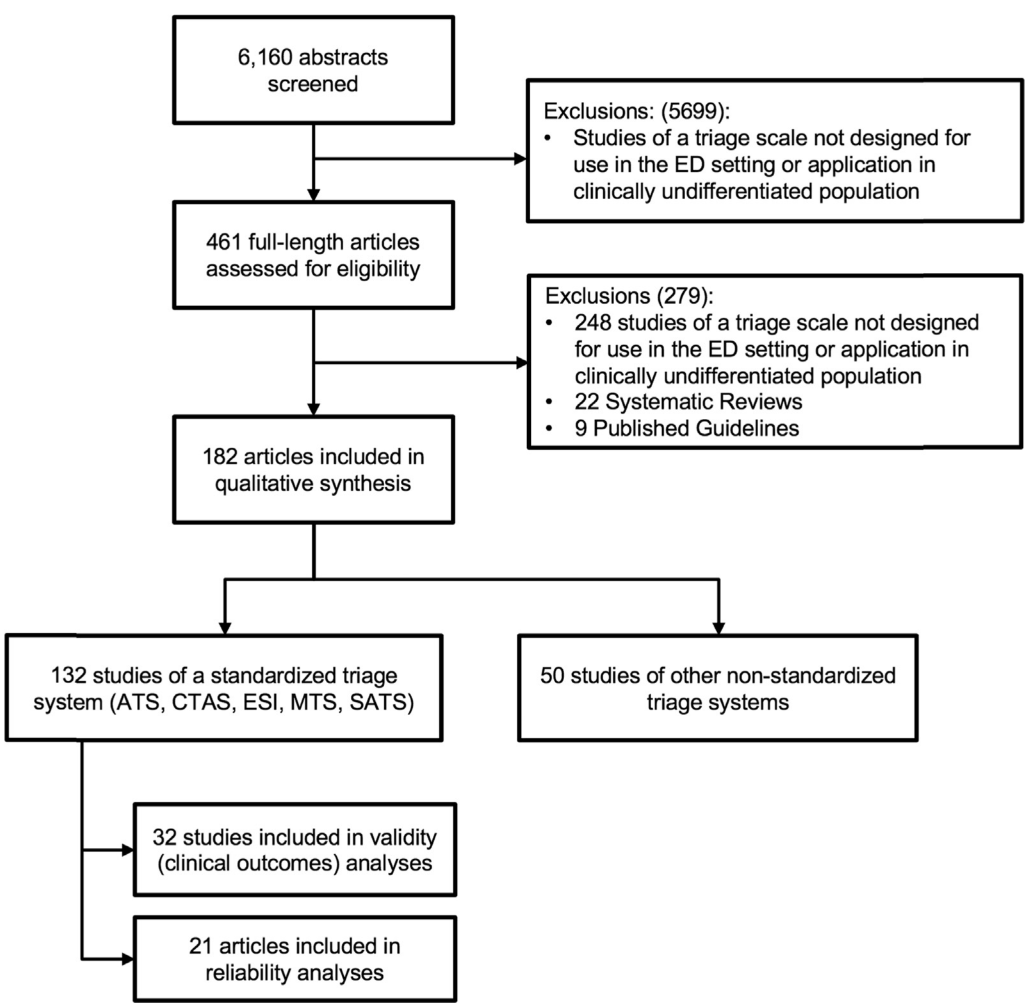


Figure 1. Study inclusion flowchart.

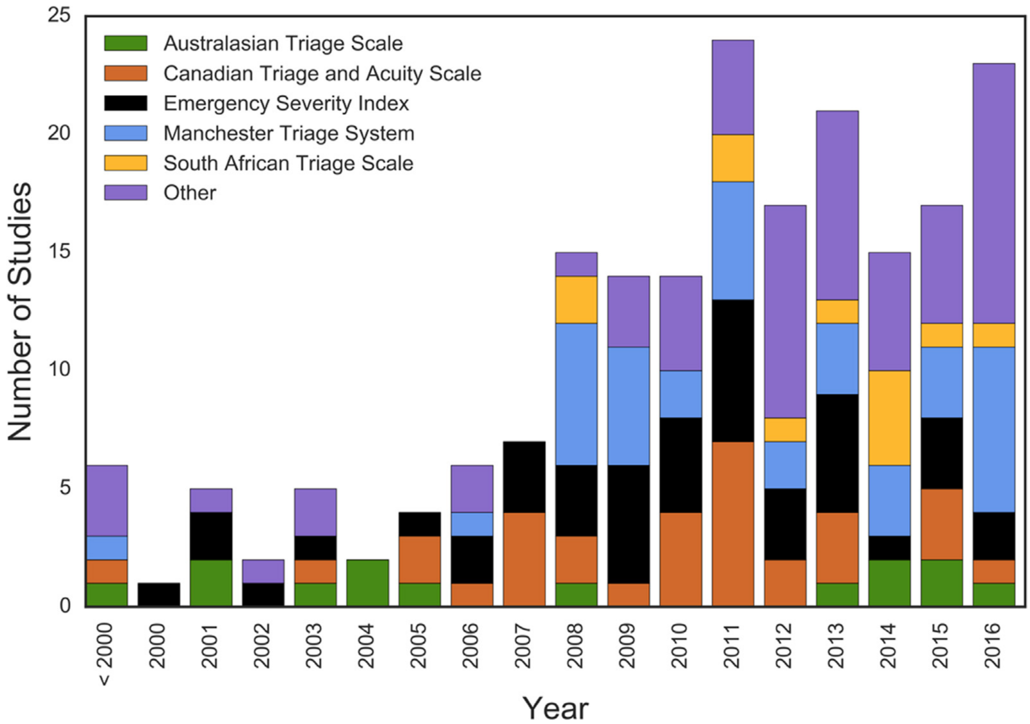


Figure 2. Triage studies over time, stratified by triage system.

(Figure 3). These outcomes were evaluated in ED populations as a whole and specific demographic (elderly and pediatric) and clinical (eg, heart failure, pulmonary embolism, sepsis) subgroups. Both sensitivity and specificity measures were improved for mortality occurring in the ED compared with postemergency care. All 9 studies found that greater than 20% of patients who died after emergency care were not designated as high acuity (<15 minutes; level 1 or 2).

Critical illness outcomes included ED diagnosis of STEMI (4 studies), non-STEMI (3), acute coronary syndrome (1), severe sepsis (3), or pulmonary embolism (1), as well as direct transfer from the ED to the ICU (7) or requirement for an immediate lifesaving intervention in the ED, as defined by the ESI implementation handbook.[7] Sensitivity of high-acuity detection of critical illness outcomes varied (Figure 4) and was on average lower for severe sepsis (36% to 74%), pulmonary embolism (54%), and non-STEMI (44% to 85%) compared with STEMI (56% to 92%) and general outcomes of ICU admission (58% to 100%) and lifesaving intervention (77% to 98%).

Although the majority of outcomes were assessed in general populations, ICU admission was also assessed in pediatric and heart failure cohorts, and requirement for immediate lifesaving intervention was assessed only in the elderly. Sensitivity for critical illness outcomes was poor for multiple scales. For example, 15% to 23% of elderly patients in need of immediate lifesaving interventions were not triaged high acuity with the ESI, and less than half were identified as requiring immediate care (ESI level 1) on arrival. Although the CTAS was highly sensitive (98%) for high-acuity detection, it also identified less than half of these patients requiring immediate care (CTAS level 1).

A total of 20 studies captured sensitivity and specificity data for triage-level designations for the outcome of inpatient hospitalization (Figure 5). The majority focused on general adult or pediatric populations, with 2 isolated studies assessing performance in elderly patients (≥65 years). The sensitivity of designating hospitalized patients at least midacuity (levels 1 to 3) was relatively high, with only 3 of 20 studies reporting less than 70%. However, every study did report hospital admissions occurring for low-acuity patients (levels 4 and 5).

Overall, risk of bias and applicability concerns for outcomes-based validity assessments were low. Results of QUADAS-2 assessment, aggregated by overarching outcome themes of mortality, critical illness, and hospitalization, are shown in Figure 6. Outcome-level assessments for individual studies are shown in Appendix E3, available online at http://www.annemergmed.com. Potential bias was introduced for 15 analyses by patient selection strategy (eg, exclusion of large numbers of patients with missing data) and for 3 by reliance on administrative diagnosis codes to define reference standard. Applicability was also affected by patient selection strategy (eg, focus on an overly differentiated cohort) for 17 studies.

Table 2. Summary characteristics of included studies.

Author and Year	N	Country	Teaching Status	Environment	Triage System	Study Design	Measure
Alquraini et al, 2015[30]	1,600	Saudi Arabia	Academic	Urban	CTAS	Paper-based scenarios	Reliability
Atzema et al, 2009[31]	1,479	Canada	Both	All	CTAS	Retrospective observational	Validity
Atzema et al, 2010[32]	1,418	Canada	Both	All	CTAS	Retrospective observational	Validity
Atzema et al, 2011[33]	6,605	Canada	Both	All	CTAS	Retrospective observational	Validity
Baumann et al, 2007[34]	929	US	Academic	Urban	ESI	Retrospective observational	Validity
Beverridge et al, 1999[35]	1,000	Canada	Not reported	Not reported	CTAS	Paper-based scenarios	Reliability
Chamberlain et al, 2015[36]	995	Australia	Academic	Not reported	ATS	Prospective observational	Validity
Considine et al, 2004[37]	4,614	Australia	Both	All	ATS	Paper-based scenarios	Reliability
Dalwai et al, 2014[38]	780	Pakistan	Community	Rural	SATS	Paper-based scenarios	Reliability
Doherty et al, 2003[39]	84,802	Australia	Both	Rural	ATS	Retrospective observational	Validity
Dong et al, 2006[14]	569	Canada	Academic	Urban	CTAS	Prospective observational	Reliability
Dong et al, 2007[15]	1,068	Canada	Academic	Urban	CTAS	Paper-based scenarios	Reliability
Eitel et al, 2003[40]	25,622	US	Both	All	ESI	Prospective observational	Reliability
Fernandes et al, 2013[41]	780	Canada	Academic	Urban	CTAS	Paper-based scenarios	Reliability
Gerdtz et al, 2008[42]	237	Australia	Community	Urban	ATS	Paper-based scenarios	Reliability
Gomez Jimenez et al, 2003[43]	32,261	Andorra	Community	Urban	CTAS	Retrospective observational	Validity
Göransson et al, 2005[44]	7,550	Sweden	Both	All	CTAS	Paper-based scenarios	Reliability
Graff et al, 2014[45]	45,469	Germany	Academic	Urban	MTS	Retrospective observational	Both
Graff et al, 2017[46]	20,836	Germany	Academic	Urban	MTS	Retrospective observational	Validity
Gravel et al, 2012[47]	1,464	Canada	Academic	All	CTAS	Paper-based scenarios	Validity
Gravel et al, 2013[48]	549,351	Canada	Academic	Suburban	CTAS	Retrospective observational	Validity
Green et al, 2012[49]	780	US	Not reported	Suburban	ESI	Retrospective observational	Validity
Grossmann et al, 2012[25]	519	Switzerland	Academic	Urban	ESI	Retrospective observational	Validity
Grossmann et al, 2011[50]	2,114	Switzerland	Academic	Urban	ESI	Retrospective observational	Reliability
Grouse et al, 2008[51]	50	Australia	Academic	Suburban	MTS	Paper-based scenarios	Reliability
Hernandez Ruiperez et al, 2015[13]	410	Spain	Academic	Rural	ESI	Retrospective observational	Validity
Jafari-Rouhi et al, 2013[52]	1,104	Iran	Academic	Urban	ESI	Prospective observational	Validity
Lee et al, 2011[24]	1,903	South Korea	Academic	Urban	CTAS	Retrospective observational	Validity
Martins et al, 2009[53]	316,622	Portugal	Community	Urban	MTS	Retrospective observational	Validity
Ng et al, 2010[54]	1,851	Taiwan	Academic	Urban	CTAS	Prospective observational	Validity
Olofsson et al, 2009[55]	1,027	Sweden	Community	Rural	MTS	Prospective observational	Reliability
Paiva et al, 2012[56]	176	Portugal	Academic	Urban	MTS	Retrospective observational	Validity
Pavlovic et al, 2008[57]	944	US	Academic	Urban	ESI	Retrospective observational	Validity
Peralta Santos et al, 2014[58]	23,615	Portugal	Community	Urban	MTS	Retrospective observational	Validity
Pinto et al, 2010[59]	218	Portugal	Academic	Urban	MTS	Retrospective observational	Validity
Platts-Mills et al, 2010[26]	773	US	Both	All	ESI	Retrospective observational	Validity
Roukema et al, 2006[60]	1,065	Netherlands	Academic	Urban	MTS	Retrospective observational	Validity
Steiner et al, 2016[61]	2,407	Switzerland	Both	All	MTS	Prospective observational	Validity
Storm-Versloot et al, 2009[16]	594	Netherlands	Academic	Urban	ESI	Paper-based scenarios	Reliability
Travers et al, 2002[62]	31,348	US	Both	All	ESI	Retrospective observational	Reliability
Trigo et al, 2008[63]	278	Portugal	Community	Suburban	MTS	Retrospective observational	Validity
Twomey et al, 2011[64]	50	South Africa	Community	Suburban	SATS	Prospective observational	Reliability
van der Wulp et al, 2008[65]	50	Netherlands	Community	Urban	MTS	Paper-based scenarios	Both
van der Wulp et al, 2009[66]	72,232	Netherlands	Community	Urban	ESI, MTS	Retrospective observational	Validity
van Spall et al, 2011[67]	68,380	Canada	Both	All	CTAS	Retrospective observational	Validity
Wuerz et al, 2000[68]	493	US	Community	Urban	ESI	Prospective observational	Both

Table 2. Continued.

Author and Year	N	Country	Teaching Status	Environment	Triage System	Study Design	Measure
Wuerz et al, 2001[69]	8,251	US	Academic	Urban	ESI	Retrospective observational	Reliability
Yergens et al, 2015[70]	1,770	Canada	Academic	Urban	CTAS	Retrospective observational	Validity
Zachariasse et al, 2016[71]	50,062	Netherlands	Academic	Urban	MTS	Retrospective observational	Validity
Zook et al, 2016[72]	54,505	US	Community	Urban	ESI	Retrospective observational	Validity

US, United States.

A total of 21 reliability studies for the major 5-level triage systems yielded 44 independent κ measurements. A schema for comparing triage reliability evaluations is shown in Table 3; these κ data are conceptually grouped by κ methodology (unweighted versus weighted [linear, quadratic, or unknown]), method of comparison (between-rater [ie, interrater] versus criterion standard), and evaluation design (patient encounters versus paper-based scenarios).

More than half of reliability evaluations were performed with paper scenarios (27), and a smaller proportion with patient encounters (16). Studies were almost evenly split between performing between rater comparisons (23) and rater versus criterion standard (20). Weighted κ was reported for a majority of evaluations (29), with unweighted κ reported for a minority (14). Not surprisingly, weighted analysis produced substantially higher estimates of reliability than unweighted analysis, a trend most easily appreciated by comparing weighted and unweighted κ for studies that performed both evaluations.[13-16] Similarly, weighted evaluations that applied quadratic weighting (14) reported higher reliability estimates compared with linear weighing (6). Eight evaluations reported weighted κ measurements that were not specified as linear or quadradic. Although the high variability in study design and analytic approach in this group of studies precluded statistical comparison, there did not appear to be major differences in outcome depending on method (between-rater versus criterion standard or patient encounters versus paper-based scenarios) across all triage systems. Similarly, there was no appreciable difference in aggregate κ measures between different triage systems (ATS, CTAS, ESI, MTS, or SATS) when measurements were made with comparable methods and analytic approach.

LIMITATIONS

There were several limitations to our systematic review. First, only studies available through our search engines and criteria were available to review. This excluded any data, write-ups, or presentations not reported in a scientific journal article and introduced potential for publication bias. This also excluded all non-English literature, biasing the geography of study sites to English-speaking countries. However, all 5 of the widely adopted triage systems were developed in English, and the majority of published ED triage literature is in English. In addition, using QUADAS-2, we identified risk of bias related to use of

Outcome	Population Type	Citation Number	Study Author and Year	Triage System	N	High Acuity Level 1 or 2 Sensitivity (95% CI)	Specificity (95% CI)
ED Mortality	General		Gomez Jimenez, et al. 2003	CTAS	23,099	0.78 (0.40, 0.97)	0.96 (0.96, 0.96)
			van der Wulp, et al. 2009	ESI	37,974	0.93 (0.76, 0.99)	0.88 (0.88, 0.89)
			Martins, et al. 2008	MTS	316,622	0.98 (0.95, 0.99)	0.74 (0.74, 0.74)
			van der Wulp, et al. 2009	MTS	34,258	0.97 (0.82, 1.00)	0.83 (0.83, 0.83)
	Pediatric		Gomez Jimenez, et al. 2003	CTAS	9,162	1.00 (0.03, 1.00)	0.99 (0.98, 0.99)
	Elderly		Lee, et al. 2011	CTAS	1,903	1.00 (0.59, 1.00)	0.85 (0.84, 0.87)
			Grossman, et al. 2012	ESI	519	1.00 (0.16, 1.00)	0.80 (0.77, 0.84)
	Severe Sepsis		Chamberlain, et al. 2015	ATS	995	0.75 (0.69, 0.80)	0.59 (0.55, 0.62)
In-Hospital Mortality	General		Doherty, et al. 2003	ATS	84,911	0.34 (0.30, 0.39)	0.94 (0.94, 0.94)
			Graff, et al. 2014	MTS	45,469	0.69 (0.66, 0.73)	0.82 (0.81, 0.82)
			Peralta Santos, et al. 2014	MTS	23,615	0.57 (0.51, 0.63)	0.85 (0.85, 0.86)
			Steiner, et al. 2016	MTS	2,407	0.39 (0.28, 0.51)	0.74 (0.72, 0.75)
	Elderly		Lee, et al. 2011	CTAS	1,903	0.64 (0.35, 0.87)	0.81 (0.78, 0.85)
			Grossman, et al. 2012	ESI	519	0.64 (0.35, 0.87)	0.81 (0.78, 0.85)
	Pulmonary Embolism		Paiva, et al. 2012	MTS	176	0.73 (0.54, 0.88)	0.50 (0.42, 0.58)
1-Day Mortality	Heart Failure		Van Spall, et al. 2011	CTAS	68,380	0.70 (0.67, 0.72)	0.57 (0.57, 0.57)
7-Day Mortality	Heart Failure		Van Spall, et al. 2011	CTAS	68,380	0.59 (0.57, 0.61)	0.57 (0.57, 0.58)

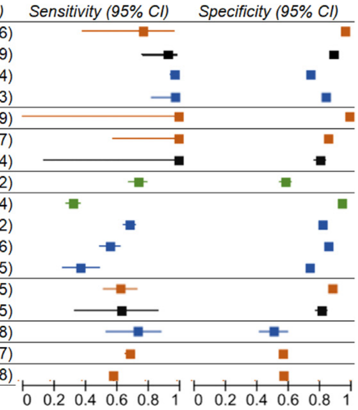


Figure 3. Performance of triage systems for mortality outcomes.

Outcome	Population Type	Citation Number	Study Author and Year	Triage System	N	High Acuity Level 1 or 2 Sensitivity (95% CI)	High Acuity Level 1 or 2 Specificity (95% CI)
STEMI	General		Atzema, et al. 2009	CTAS	1,479	0.56 (0.54, 0.59)	Not estimable
			Atzema, et al. 2010	CTAS	1,418	0.57 (0.54, 0.59)	Not estimable
			Atzema, et al. 2011	CTAS	2,095	0.74 (0.72. 0.76)	Not estimable
			Trigo, et a. 2008	MTS	278	0.79 (0.74, 0.84)	Not estimable
			Pinto, et al. 2010	MTS	78	0.92 (0.84, 0.97)	Not estimable
NSTEMI	General		Atzema, et al. 2009	CTAS	1,607	0.44 (0.41, 0.46)	Not estimable
			Atzema, et al. 2010	CTAS	1,563	0.44 (0.41, 0.46)	Not estimable
			Atzema, et al. 2011	CTAS	4,511	0.63 (0.62, 0.65)	Not estimable
			Pinto, et al. 2010	MTS	140	0.85 (0.78, 0.90)	Not estimable
Severe Sepsis	General		Chamberlain, et al. 2015	ATS	995	0.49 (0.46, 0.53)	Not estimable
			Yergens, et al. 2015	CTAS	1,770	0.74 (0.71, 0.78)	0.36 (0.33, 0.39)
			Graff, et al. 2016	MTS	20,836	0.36 (0.30, 0.42)	0.83 (0.82, 0.83)
Pulmonary Embolism	General		Paiva, et al. 2012	MTS	176	0.54 (0.46, 0.62)	Not estimable
ICU Admission	General		Graff, et al. 2014	MTS	45,469	0.82 (0.80, 0.83)	0.84 (0.84, 0.85)
			Steiner, et al. 2016	MTS	2,407	0.58 (0.47, 0.68)	0.75 (0.73, 0.76)
	Pediatric		Gravel, et al. 2013	CTAS	549,351	0.79 (0.77, 0.81)	0.88 (0.88, 0.88)
			Jafari-Rouhi, et al. 2013	ESI	1,105	1.00 (0.69, 1.00)	0.71 (0.68, 0.74)
			Zachariasse, et al. 2016	MTS	50,062	0.71 (0.68, 0.74)	0.83 (0.82, 0.83)
	Elderly		Lee, et al. 2011	CTAS	1,903	0.80 (0.74, 0.85)	0.93 (0.91, 0.94)
	Heart Failure		Van Spall, et al. 2011	CTAS	68,380	0.63 (0.62, 0.64)	0.59 (0.58, 0.59)
Immediate Life-Saving Intervention	Elderly		Lee, et al. 2011	CTAS	1,903	0.98 (0.93, 1.00)	0.89 (0.88, 0.91)
			Platss-Mills, et al. 2010	ESI	773	0.77 (0.56, 0.91)	0.77 (0.74, 0.80)
			Grossman, et al. 2012	ESI	519	0.85 (0.55, 0.98)	0.82 (0.78, 0.85)

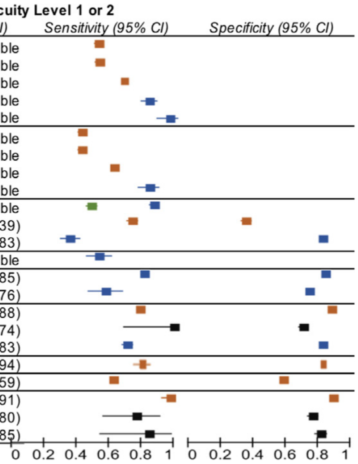


Figure 4. Performance of triage systems for critical illness.

administrative diagnostic codes as reference standard in several of the retrospective studies from which sensitivity and specificity of ED triage designation for critical illness outcomes (eg, non-STEMI, severe sepsis) were derived. Although diagnosis code data have shown variable reliability, we do not suspect systematic bias across these studies that would affect any cumulative interpretation of results.

The diversity in outcome measurements and evaluation design across the studies reviewed served as a major limitation to comparison and to our ability to draw conclusions about individual triage scales. This similarly limits the aggregate knowledge created by this body of literature that our systematic review addressed. Although our purpose was to connect these triage studies through a common evaluative framework, any summative results and conclusions should be interpreted with this variability in mind.

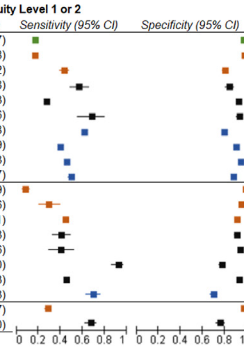
DISCUSSION

Triage remains a central process for safe management of patients under circumstances of excess demand common in many EDs. Increases in patient volume in developed countries and new evolving emergency care systems in developing countries, particularly in population epicenters, stress the need for more accurate and reliable triage. As a result, the body of scientific literature evaluating ED triage has increased (Figure 2).

The purpose of this systematic review was to introduce a framework for comparison of triage system performance with respect to clinical outcomes and reliability. We used this framework to catalog studies based on types of outcomes evaluated (mortality, critical illness, and hospitalization) and the sensitivity and specificity of their detection. It also highlights the core study design factors that may be used to group analyses of rater reliability based on κ methodology, rater comparison method, and evaluation design. This framework was proposed as a means to conceptually highlight patterns and catalog diversity in triage literature.

Applying this framework, we found a majority of studies reporting low sensitivity (<80%) in identifying patients who had critical illness outcomes or died during the index hospitalization, including those who died within days of

Population Type	Citation Number	Study Author and Year	Triage System	N	High Acuity Level 1 or 2 Sensitivity (95% CI)	High Acuity Level 1 or 2 Specificity (95% CI)	Mid-Acuity Level 1 through 3 Sensitivity (95% CI)	Mid-Acuity Level 1 through 3 Specificity (95% CI)
General		Doherty, et al. 2003	ATS	84,911	0.18 (0.18, 0.19)	0.97 (0.97, 0.97)	0.58 (0.57, 0.59)	0.81 (0.81, 0.82)
		Gomez Jimenez, et al. 2003	CTAS	23,099	0.18 (0.17, 0.19)	0.98 (0.98, 0.98)	0.69 (0.68, 0.71)	0.80 (0.79, 0.80)
		Ng, et al. 2010	CTAS	1,851	0.44 (0.40, 0.48)	0.80 (0.78, 0.82)	0.85 (0.82, 0.88)	0.34 (0.32, 0.37)
		Wuerz, et al. 2000	ESI	493	0.57 (0.49, 0.65)	0.84 (0.80, 0.88)	0.92 (0.87, 0.96)	0.54 (0.49, 0.60)
		van der Wulp, et al. 2009	ESI	37,974	0.28 (0.27, 0.29)	0.93 (0.93, 0.93)	0.86 (0.85, 0.87)	0.67 (0.66, 0.67)
		Hernandez Ruiperez, et al. 2015	ESI	410	0.68 (0.55, 0.80)	0.93 (0.90, 0.96)	0.97 (0.88, 1.00)	0.55 (0.50, 0.61)
		Martins et al. 2009	MTS	316,622	0.59 (0.59, 0.60)	0.78 (0.78, 0.78)	0.96 (0.96, 0.96)	0.25 (0.24, 0.25)
		van der Wulp, et al. 2009	MTS	34,258	0.37 (0.36, 0.39)	0.88 (0.88, 0.89)	0.86 (0.86, 0.87)	0.54 (0.53, 0.54)
		Graff, et al. 2014	MTS	45,469	0.43 (0.43, 0.44)	0.93 (0.93, 0.93)	0.87 (0.86, 0.87)	0.45 (0.44, 0.45)
		Peralta Santos, et al. 2014	MTS	23,615	0.48 (0.44, 0.51)	0.86 (0.86, 0.87)	0.85 (0.83, 0.88)	0.52 (0.51, 0.53)
Pediatric		Gomez Jimenez, et al. 2003	CTAS	9,162	0.09 (0.06, 0.13)	0.99 (0.99, 0.99)	0.55 (0.50, 0.61)	0.86 (0.85, 0.87)
		Gravel, et al. 2012	CTAS	1,464	0.30 (0.21, 0.40)	0.95 (0.94, 0.96)	0.84 (0.75, 0.90)	0.52 (0.49, 0.54)
		Gravel, et al. 2013	CTAS	549,351	0.45 (0.45, 0.46)	0.91 (0.91, 0.91)	0.91 (0.91, 0.91)	0.54 (0.54, 0.54)
		Pavlovic, et al. 2008	ESI	944	0.42 (0.34, 0.50)	0.91 (0.89, 0.93)	0.82 (0.76, 0.88)	0.68 (0.65, 0.71)
		Green, et al. 2012	ESI	780	0.41 (0.30, 0.53)	0.94 (0.92, 0.96)	0.95 (0.88, 0.99)	0.59 (0.55, 0.63)
		Jafari-Rouhi, et al. 2013	ESI	1,105	0.93 (0.87, 0.97)	0.78 (0.75, 0.80)	0.95 (0.89, 0.98)	0.66 (0.63, 0.69)
		Zook, et al. 2016	ESI	54,505	0.46 (0.45, 0.47)	0.93 (0.93, 0.93)	0.91 (0.90, 0.91)	0.66 (0.66, 0.67)
		Roumeka, et al. 2006	MTS	1,065	0.70 (0.64, 0.76)	0.70 (0.67, 0.73)	0.91 (0.87, 0.95)	0.44 (0.40, 0.47)
Elderly		Lee, et al. 2011	CTAS	1,903	0.28 (0.25, 0.31)	0.96 (0.94, 0.97)	0.92 (0.90, 0.94(	0.38 (0.35, 0.42)
		Baumann, et al. 2007	ESI	929	0.67 (0.62, 0.72)	0.76 (0.72, 0.80)	0.98 (0.97, 0.99)	0.30 (0.26, 0.34)



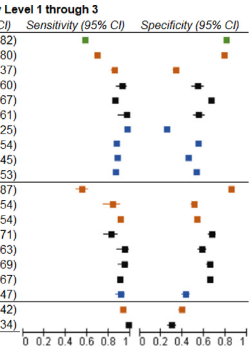


Figure 5. Performance of triage systems for hospitalization.

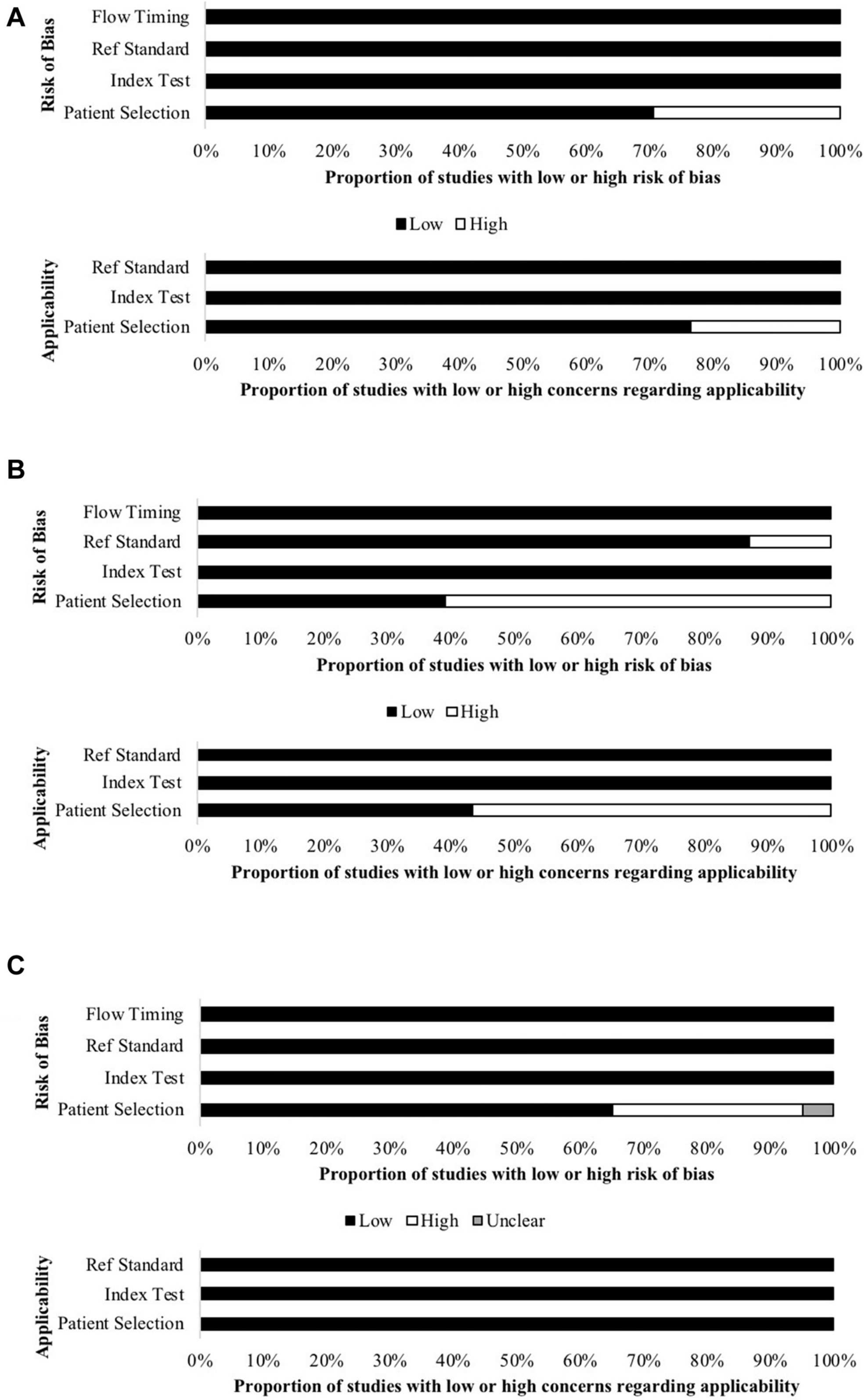


Figure 6. Aggregate risk of bias and applicability concerns as determined by QUADAS-2 among outcomes-based validity assessments. Assessments are grouped by focus on outcomes related to mortality (n =17) (*A*), critical illness (n=23) (*B*), and hospitalization (n=20) (*C*). QUADAS-2 grading for individual studies assessed at the outcome level are shown in Appendix E3, available online at http://www.annemergmed.com.

Table 3. Summary of ED triage scale reliability evaluations by analytic approach and study design.

Measure Type	Comparison Type	Evaluation Type	Study Author and Year	Triage System	Provider Type	Weight Type	Reliability
Unweighted κ	Between rater	Patient encounters	Dong et al, 2006[14]	CTAS	Nurse		0.40
			Dong et al, 2007[15]	CTAS	Nurse		0.40
		Paper scenarios	Gerdtz et al, 2008[42]	ATS	Nurse		0.41
			Storm-Versloot et al, 2009[16]	ESI	Nurse		0.46
			Storm-Versloot et al, 2009[16]	MTS	Nurse		0.76
			Dalwai et al, 2014[38]	SATS	Nurse		0.55
	Criterion standard	Patient encounters	Wuerz et al, 2000[68]	ESI	Nurse		0.70
			Hernandez et al, 2015[13]	ESI	Nurse		0.77
		Paper scenarios	Considine et al, 2004[37]	ATS	Nurse		0.43
			Göransson et al, 2005[44]	CTAS	Nurse		0.46
			Storm-Versloot et al, 2009[16]	ESI	Nurse		0.43
			Van Der Wulp et al, 2009[16]	MTS	Nurse		0.48
			Olofsson et al, 2009[55]	MTS	Nurse		0.61
			Storm-Versloot et al, 2009[16]	MTS	Nurse		0.84
Weighted κ	Between rater	Patient encounters	Dong et al, 2006[14]	CTAS	Nurse	Linear	0.52
			Dong et al, 2006[14]	CTAS	Nurse	Quadratic	0.66
			Dong et al, 2007[15]	CTAS	Nurse	Linear	0.52
			Dong et al, 2007[15]	CTAS	Nurse	Quadratic	0.65
			Eitel et al, 2003[40]	ESI	Nurse	Unknown	0.78
			Graff et al, 2014[45]	MTS	Nurse	Unknown	0.95
		Paper scenarios	Beverridge et al, 1999[35]	CTAS	Nurse	Quadratic	0.84
			Fernandes et al, 2013[41]	CTAS	Nurse	Linear	0.70
			Fernandes et al, 2013[41]	CTAS	Nurse	Quadratic	0.79
			Alquraini et al, 2015[30]	CTAS	Nurse	Quadratic	0.87
			Eitel et al, 2003[40]	ESI	Nurse	Unknown	0.76
			Storm-Versloot et al, 2009[16]	ESI	Nurse	Quadratic	0.73
			Wuerz et al, 2001[69]	ESI	Nurse	Unknown	0.80
			Grouse et al, 2008[51]	MTS	Nurse	Unknown	0.60
			Storm-Versloot et al, 2009[16]	MTS	Nurse	Quadratic	0.82
			Dalwai et al, 2014[38]	SATS	Nurse	Linear	0.65
			Dalwai et al, 2014[38]	SATS	Nurse	Quadratic	0.77
	Criterion standard	Patient encounters	Wuerz et al, 2000[68]	ESI	Nurse	Linear	0.80
			Grossmann et al, 2011[50]	ESI	Nurse	Unknown	0.99
			Travers et al, 2002[62]	ESI	Nurse	Unknown	0.68
			Wuerz et al, 2001[69]	ESI	Nurse	Unknown	0.73
			Hernandez et al, 2015[13]	ESI	Nurse	Quadratic	0.81
			Twomey et al, 2011[64]	SATS	Nurse	Quadratic	0.92
			Hernandez et al, 2015[13]	ESI	Nurse	Quadratic	0.86
		Paper scenarios	Göransson et al, 2005[44]	CTAS	Nurse	Unknown	0.71
			Storm-Versloot et al, 2009[16]	ESI	Nurse	Quadratic	0.71
			van der Wulp et al, 2008[65]	MTS	Nurse	Quadratic	0.62
			Olofsson et al, 2009[55]	MTS	Nurse	Linear	0.71
			Olofsson et al, 2009[55]	MTS	Nurse	Quadratic	0.81
			Storm-Versloot et al, 2009[16]	MTS	Nurse	Quadratic	0.87

their ED visit, as high acuity at triage. The link between the timeliness of ED care and post-ED mortality (eg, inhospital mortality) may be amorphous. However, the time-sensitive nature of critical illness outcomes (Figure 5) is definitive and heavily affected by triage decisionmaking. Many studies have established a relationship between the interval from ED arrival to treatment endpoints and health outcomes of morbidity and mortality. Examples include door-to-balloon time for acute coronary syndrome,[17,18] time to fluid resuscitation and antibiotic administration for sepsis,[19-21] time to thrombolytics for pulmonary embolism,[22] and time to ICU admission and transfer for critically ill patients.[23] Three studies that evaluated sensitivity of triage scales for identifying elderly patients with immediate life-threatening events reported more alarming results.[24-26] Between 11% (CTAS) and 23% (ESI) of elderly patients with these events were triaged to moderate acuity (level 3). In a busy ED, such patients may wait hours before evaluation by an emergency physician. Variable performance in detecting all these high-risk outcomes may be attributed to the variability in rater reliability exhibited across studies. Consistently high rates of outcome detection require commensurate rates of rater reliability; these performance measures are intertwined.

Although we had hypothesized that a single triage system would emerge as superior, we found similar performance trends across all triage scales, and weaknesses identified were common to all systems. This highlighted the general opportunity to improve triage performance and reduce untoward variability at triage as a whole. Advancements in electronic health records and decision support have opportunity to better control variability and enhance triage performance.[27-29] The framework proposed in this systematic review may provide a common measuring stick to draw meaningful connections between previous research and future triage evaluations.

Supervising editor: Melissa L. McCarthy, ScD, MS

Author affiliations: From the Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, MD (Hinson, Martinez, Whalen, Hansoti, Levin); the Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD (Cabral); and the Whiting School of Engineering, Johns Hopkins University, Baltimore, MD (George, Levin).

Author contributions: SL conceived and designed the study. JSH, SC, KG, and SL performed literature review and data extraction. MW and BH provided content expertise and advice on the study design. DAM performed data analysis and data visualization. JSH and SL drafted the article, and all authors contributed to its revision. JSH takes responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). This work was funded by grant R21HS023641 from the Agency for Healthcare Research and Quality. Dr. Levin is an engineer and the owner of a start-up company, StoCastic, LLC, that has developed several machine-learning and electronic health record-based tools to improve patient safety and throughput throughout the hospital, including a tool that aims to improve reliability of ED triage. Dr. Hinson serves as a consultant for StoCastic, LLC.

Publication dates: Received for publication June 4, 2018. Revisions received August 8, 2018, and September 11, 2018. Accepted for publication September 21, 2018.

REFERENCES

1. Iserson KV, Moskop JC. Triage in medicine, part I: concept, history, and types. *Ann Emerg Med*. 2007;49:275-281.
2. National Center for Health Statistics. National Hospital Ambulatory Care Survey. 2014. Available at: https://www.cdc.gov/nchs/data/nhamcs/web_tables/2014_ed_web_tables.pdf. Accessed November 5, 2018.
3. Pines JM. Emergency care at the crossroads: emergency department crowding, payment reform, and one potential future. *Ann Emerg Med*. 2015;66:493-495.
4. Pitts SR, Pines JM, Handrigan MT, et al. National trends in emergency department occupancy, 2001 to 2008: effect of inpatient admissions versus emergency department practice intensity. *Ann Emerg Med*. 2012;60:679-686.e3.
5. Australasian College for Emergency Medicine. Guidelines on the implementation of the Australasian Triage Scale in Emergency Department. Available at: https://acem.org.au/getmedia/51dc74f7-9ff0-42ce-872a-0437f3db640a/G24_04_Guidelines_on_Implementation_of_ATS_Jul-16.aspx. Accessed November 5, 2018.
6. Beveridge R, Clarke B, Janes L, et al. Implementation guidelines for the Canadian Emergency Department Triage and Acuity Scale (CTAS)—endorsed by the Canadian Association of Emergency Physicians, the National Emergency Nurses Affiliation of Canada, and l'association des medecins d'urgence du. 1998. Available at: https://caep.ca/resources/ctas/implementation-guidelines. Accessed November 5, 2018.
7. Gilboy N, Tanabe P, Travers D, et al, eds; Ahrq Q. *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care.* 4th ed. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
8. Mackway-Jones K, ed. *Emergency Triage*. Malden, MA: Blackwell Publishing; 1996.
9. Twomey M, ed. *South African Triage Scale Training Manual 2012*. Capetown, South Africa: Western Cape Government; 2012. Available at: https://emssa.org.za/wp-content/uploads/2011/04/SATS-Manual-A5-LR-spreads.pdf. Accessed May 30, 2018.
10. Fernandes CMB, Tanabe P, Gilboy N, et al. Five-level triage: a report from the ACEP/ENA Five-Level Triage Task Force. *J Emerg Nurs*. 2005;31:39-50.
11. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.
12. Whiting PF, Rutjes AWS, Westwood ME, et al. Research and reporting methods accuracy studies. *Ann Intern Med*. 2011;155:529-536.
13. Hernandez Ruiperez T, Costa CL, Martinez MDA, et al. Evidence of the validity of the Emergency Severity Index for triage in a general hospital emergency department. *Emergencias*. 2015;27:301-306.
14. Dong SL, Bullard MJ, Meurer DP, et al. Reliability of computerized emergency triage. *Acad Emerg Med*. 2006;13:269-275.
15. Dong SL, Bullard MJ, Meurer DP, et al. The effect of training on nurse agreement using an electronic triage system. *CJEM*. 2007;9:260-266.
16. Storm-Versloot MN, Ubbink DT, Chin A, et al. Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study. *Emerg Med J*. 2009;26:556-560.
17. Berger PB, Ellis SG, Holmes DR, et al. Relationship between delay in performing direct coronary angioplasty and early clinical outcome in patients with acute myocardial infarction: results from the Global Use of Strategies to Open Occluded Arteries in Acute Coronary Syndromes (GUSTO-IIb) trial. *Circulation*. 1999;100:14-20.
18. De Luca G, Suryapranata H, Ottervanger JP, et al. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. *Circulation*. 2004;109:1223-1225.
19. Kumar A, Haery C, Paladugu B, et al. The duration of hypotension before the initiation of antibiotic treatment is a critical determinant of survival in a murine model of *Escherichia coli* septic shock:

association with serum lactate and inflammatory cytokine levels. *J Infect Dis*. 2006;193:251-258.
20. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med*. 2014;42: 1749-1755.
21. Seymour CW, Gesten F, Prescott HC, et al. Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med*. 2017;376:2235-2244.
22. Beydilli İ, Yılmaz F, Sönmez BM, et al. Thrombolytic therapy delay is independent predictor of mortality in acute pulmonary embolism at emergency service. *Kaohsiung J Med Sci*. 2016;32:572-578.
23. Cardoso LTQ, Grion CMC, Matsuo T, et al. Impact of delayed admission to intensive care units on mortality of critically ill patients: a cohort study. *Crit Care*. 2011;15:R28.
24. Lee J, Oh S, Peck E, et al. The validity of the Canadian Triage and Acuity Scale in predicting resource utilization and the need for immediate life-saving interventions in elderly emergency department patients. *Scand J Trauma Resusc Emerg Med*. 2011;19:68.
25. Grossmann FF, Zumbrunn T, Frauchiger A, et al. At risk of undertriage? testing the performance and accuracy of the Emergency Severity Index in older emergency department patients. *Ann Emerg Med*. 2012;60:317-325.e3.
26. Platts-Mills TF, Travers D, Biese K, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. *Acad Emerg Med*. 2010;17:238-243.
27. Mistry B, Stewart De Ramirez S, Kelen G, et al. Accuracy and reliability of emergency department triage using the Emergency Severity Index: an international multicenter assessment. *Ann Emerg Med*. 2018;71: 581-587.e3.
28. Levin S, Toerper M, Hamrock E, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. *Ann Emerg Med*. 2018;71:565-574.e2.
29. Dugas AF, Kirsch TD, Toerper M, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. *J Emerg Med*. 2016;50:910-918.
30. Alquraini M, Awad E, Hijazi R. Reliability of Canadian Emergency Department Triage and Acuity Scale (CTAS) in Saudi Arabia. *Int J Emerg Med*. 2015;8:29.
31. Atzema CL, Austin PC, Tu JV, et al. Emergency department triage of acute myocardial infarction patients and the effect on outcomes. *Ann Emerg Med*. 2009;53:736-745.
32. Atzema CL, Austin PC, Tu JV, et al. ED triage of patients with acute myocardial infarction: predictors of low acuity triage. *Am J Emerg Med*. 2010;28:694-702.
33. Atzema CL, Schull MJ, Austin PC, et al. Temporal changes in emergency department triage of patients with acute myocardial infarction and the effect on outcomes. *Am Heart J*. 2011;162:451-459.
34. Baumann MR, Strout TD. Triage of geriatric patients in the emergency department: validity and survival with the Emergency Severity Index. *Ann Emerg Med*. 2007;49:234-240.
35. Beveridge R, Ducharme J, Janes L, et al. Reliability of the Canadian Emergency Department Triage and Acuity Scale: interrater agreement. *Ann Emerg Med*. 1999;34:155-159.
36. Chamberlain DJ, Willis E, Clark R, et al. Identification of the severe sepsis patient at triage: a prospective analysis of the Australasian Triage Scale. *Emerg Med J*. 2015;32:690-697.
37. Considine J, LeVasseur SA, Villanueva E. The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Ann Emerg Med*. 2004;44: 516-523.
38. Dalwai MK, Twomey M, Maikere J, et al. Reliability and accuracy of the South African Triage Scale when used by nurses in the emergency department of Timergara Hospital, Pakistan. *S Afr Med J*. 2014;104:372-375.
39. Doherty SR, Hore CT, Curran SW. Inpatient mortality as related to triage category in three New South Wales regional base hospitals. *Emerg Med*. 2003;15:334-340.
40. Eitel DR, Travers DA, Rosenau AM, et al. The Emergency Severity Index triage algorithm version 2 is reliable and valid. *Acad Emerg Med*. 2003;10:1070-1080.
41. Fernandes CMB, McLeod S, Krause J, et al. Reliability of the Canadian Triage and Acuity Scale: interrater and intrarater agreement from a community and an academic emergency department. *CJEM*. 2013;15:227-232.
42. Gerdtz MF, Collins M, Chu M, et al. Optimizing triage consistency in Australian emergency departments: the Emergency Triage Education Kit. *Emerg Med Australas*. 2008;20:250-259.
43. Gomez Jimenez J, Murray MJ, Beveridge R, et al. Implementation of the Canadian Emergency Department Triage and Acuity Scale (CTAS) in the principality of Andorra: can triage parapeters serve as emergency department quality indicators? *CJEM*. 2003;5:315-322.
44. Göransson K, Ehrenberg A, Marklund B, et al. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci*. 2005;19:432-438.
45. Graff I, Goldschmidt B, Glien P, et al. The German Version of the Manchester Triage System and its quality criteria—first assessment of validity and reliability. *PLoS One*. 2014;9:e88995.
46. Gräff I, Goldschmidt B, Glien P, et al. Validity of the Manchester Triage System in patients with sepsis presenting at the ED: a first assessment. *Emerg Med J*. 2017;34:212-218.
47. Gravel J, Gouin S, Goldman RD, et al. The Canadian Triage and Acuity Scale for children: a prospective multicenter evaluation. *Ann Emerg Med*. 2012;60:71-77.
48. Gravel J, Fitzpatrick E, Gouin S, et al. Performance of the Canadian Triage and Acuity Scale for children: a multicenter database study. *Ann Emerg Med*. 2013;61:27-32.e3.
49. Green NA, Durani Y, Brecher D, et al. Emergency Severity Index version 4: a valid and reliable tool in pediatric emergency department triage. *Pediatr Emerg Care*. 2012;28:753-757.
50. Grossmann FF, Nickel CH, Christ M, et al. Transporting clinical tools to new settings: cultural adaptation and validation of the Emergency Severity Index in German. *Ann Emerg Med*. 2011;57:257-264.
51. Grouse AI, Bishop RO, Bannon AM. The Manchester Triage System provides good reliability in an Australian emergency department. *Emerg Med J*. 2009;26:484-486.
52. Jafari-Rouhi AH, Sardashti S, Taghizadieh A, et al. The Emergency Severity Index, version 4, for pediatric triage: a reliability study in Tabriz Children's Hospital, Tabriz, Iran. *Int J Emerg Med*. 2013;6:36.
53. Martins HMG, De Castro Dominguez Cuña LM, Freitas P. Is Manchester (MTS) more than a triage system? a study of its association with mortality and admission to a large Portuguese hospital. *Emerg Med J*. 2009;26:183-186.
54. Ng CJ, Hsu KH, Kuan JT, et al. Comparison between Canadian Triage and Acuity Scale and Taiwan Triage System in emergency departments. *J Formos Med Assoc*. 2010;109:828-837.
55. Olofsson P, Gellerstedt M, Carlström ED. Manchester Triage in Sweden—interrater reliability and accuracy. *Int Emerg Nurs*. 2009;17: 143-148.
56. Paiva LV, Providencia R, Faustino A, et al. Manchester triage in acute pulmonary embolism: can it unmask the grand impersonator? *Emerg Med J*. 2012;29:1-7.
57. Pavlovic S, Suresh S, Knazik SR. Does the Emergency Severity Index predict pediatric emergency department use of resources, admission rate and length of stay? *Pediatr Emerg Care*. 2010;26:726-727.
58. Peralta Santos A, Freitas P, Martins HMG. Manchester Triage System version II and resource utilisation in the emergency department. *Emerg Med J*. 2014;31:148-152.

59. Pinto D, Lunet N, Azevedo A. Sensitivity and specificity of the Manchester Triage System for patients with acute coronary syndrome. *Rev Port Cardiol*. 2010;29:961-987.
60. Roukema J, Steyerberg EW, van Meurs A, et al. Validity of the Manchester Triage System in paediatric emergency care. *Emerg Med J*. 2006;23:906-910.
61. Steiner D, Renetseder F, Kutz A, et al. Performance of the Manchester Triage System in adult medical emergency patients: a prospective cohort study. *J Emerg Med*. 2016;50:678-689.
62. Travers DA, Waller AE, Bowling JM, et al. Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs*. 2002;28:395-400.
63. Trigo J, Gago P, Mimoso J, et al. In-hospital delay in ST-segment-elevation myocardial infarction after Manchester Triage. *Rev Port Cardiol*. 2008;27:1251-1259.
64. Twomey M, De Sá A, Wallis LA, et al. Inter-rater reliability of the South African Triage Scale: assessing two different cadres of health care workers in a real time environment. *Afr J Emerg Med*. 2011;1:113-118.
65. van der Wulp I, van Baar ME, Schrijvers AJP. Reliability and validity of the Manchester Triage System a general emergency department patient population in the Netherlands: results of a simulation study. *Emerg Med J*. 2008;25:431-434.
66. Van Der Wulp I, Schrijvers AJP, Van Stel HF. Predicting admission and mortality with the Emergency Severity Index and the Manchester Triage System: a retrospective observational study. *Emerg Med J*. 2009;26:506-509.
67. van Spall HGC, Atzema C, Schull MJ, et al. Prediction of emergent heart failure death by semi-quantitative triage risk stratification. *PLoS One*. 2011;6:1-8.
68. Wuerz RC, Milne LW, Eitel DR, et al. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med*. 2000;7:236-242.
69. Wuerz RC, Travers D, Gilboy N, et al. Implementation and refinement of the Emergency Severity Index. *Acad Emerg Med*. 2001;8:170-176.
70. Yergens DW, Ghali WA, Faris PD, et al. Assessing the association between occupancy and outcome in critically ill hospitalized patients with sepsis. *BMC Emerg Med*. 2015;15:1-8.
71. Zachariasse JM, Kuiper JW, de Hoog M, et al. Safety of the Manchester Triage System to detect critically ill children at the emergency department. *J Pediatr*. 2016;177:232-237.e1.
72. Zook HG, Kharbanda AB, Flood A, et al. Racial differences in pediatric emergency department triage scores. *J Emerg Med*. 2016;50:720-726.