Five-level emergency triage systems: variation in assessment of validity

Akira Kuriyama,[1,2] Seigo Urushidani,[3] Takeo Nakayama[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/emermed-2016-206295).

[1]Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan
[2]Department of General Medicine, Kurashiki Central Hospital, Kurashiki, Okayama, Japan
[3]Department of Emergency Medicine, Kurashiki Central Hospital, Kurashiki, Okayama, Japan

Correspondence to
Dr Akira Kuriyama, Department of Health Informatics, Kyoto University School of Public Health, Yoshida-konoe-cho Sakyo-ku Kyoto 606-8501 Japan;
akira.kuriyama.jpn@gmail.com

Received 18 August 2016
Revised 19 April 2017
Accepted 5 May 2017
Published Online First 27 July 2017

ABSTRACT

Introduction Triage systems are scales developed to rate the degree of urgency among patients who arrive at EDs. A number of different scales are in use; however, the way in which they have been validated is inconsistent. Also, it is difficult to define a surrogate that accurately predicts urgency. This systematic review described reference standards and measures used in previous validation studies of five-level triage systems.

Methods We searched PubMed, EMBASE and CINAHL to identify studies that had assessed the validity of five-level triage systems and described the reference standards and measures applied in these studies. Studies were divided into those using criterion validity (reference standards developed by expert panels or triage systems already in use) and those using construct validity (prognosis, costs and resource use).

Results A total of 57 studies examined criterion and construct validity of 14 five-level triage systems. Criterion validity was examined by evaluating (1) agreement between the assigned degree of urgency with objective standard criteria (12 studies), (2) overtriage and undertriage (9 studies) and (3) sensitivity and specificity of triage systems (7 studies). Construct validity was examined by looking at (4) the associations between the assigned degree of urgency and measures gauged in EDs (48 studies) and (5) the associations between the assigned degree of urgency and measures gauged after hospitalisation (13 studies). Particularly, among 46 validation studies of the most commonly used triages (Canadian Triage and Acuity Scale, Emergency Severity Index and Manchester Triage System), 13 and 39 studies examined criterion and construct validity, respectively.

Conclusion Previous studies applied various reference standards and measures to validate five-level triage systems. They either created their own reference standard or used a combination of severity/resource measures.

INTRODUCTION

Triage at EDs is a decision-making process that is applied to identify patients who require immediate attention to achieve optimal outcomes.[1] Triage systems are scales developed primarily to categorise patients who do and who do not need immediate intervention by urgency, and optimise resources in the ED to apply them to those who need immediate care.

Validating triage systems is essential because they can impact the outcomes of patients in need of immediate care. A challenge is to determine the appropriate reference standard and measure for validation that discriminate patients into a 'true' category of urgency. In validation studies of triage systems, criterion and construct validities have been mainly discussed. Criterion validity looks at the correlation of a scale with some external criterion of the disorder under study (reference standard).[2] This is often the 'gold standard'. Construct validity looks for the correlation in assessments obtained from several scales purported to measure the same construct (measures).[2 3] In validating triage systems, criterion validity would be 'true' urgency of patients, most likely determined by experts, while construct validity represents severity-related variables such as mortality, admission and resource and time spent on patients. Criterion validity should be preferred in validation of scales, but given the lack of a convenient gold standard for 'urgency', Moll has suggested that the best proxy reference standard comprises prognostic markers, disease severity and case complexity.[4] Furthermore, there is no consensus as to what are acceptable reference standards or measures in validating triage systems.

Therefore, we systematically reviewed the reference standards and measures used in published validation studies of triage systems to provide an understanding of the basis on which triage scales have been validated.

METHODS

This study proceeded according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement for reporting systematic reviews.[5] The protocol for this systematic review is registered in PROSPERO (CRD42015027653).

Search and selection of studies

The American College of Emergency Physicians / Emergency Nurses Association Five-Level Triage Task Force recommended the use of five-level triage scales as they generally showed better reliability compared with three-level or four-level ones.[6] Travers *et al* showed that a five-level triage system was more reliable and discriminative than a three-level one.[7] A growing trend to use the five-level triage systems was also noted.[8] Among them, the Canadian Triage and Acuity Scale (CTAS), the Emergency Severity Index (ESI), the Manchester Triage System (MTS) and the Australasian Triage System (ATS) are the five-level triage systems that were most frequently studied, and we first focused on these four. We also examined other five-level triage systems separately. We required these studies to meet the following conditions: (1) they included patients who presented at EDs from all categories of triage scales, unless a pragmatic or ethical need was required to exclude some

► http://dx.doi.org/10.1136/emermed-2016-206058
► http://dx.doi.org/10.1136/emermed-2017-206973

To cite: Kuriyama A, Urushidani S, Nakayama T. *Emerg Med J* 2017;**34**:703–710.

from certain urgency categories; (2) they were of any design and (3) were published in peer-reviewed journals, and (4) they were explicitly testing the validity of one or more triage scales. Studies from any age groups as well as ambulatory and transferred patients were included. Studies were excluded if they used these scales as characteristics of the participants or explanatory variables, or if they focused on a limited spectrum of diseases and symptoms or populations classified in certain urgency categories without clear rationale. We also excluded reviews, editorials, letters, conference proceedings or abstracts and studies that focused solely on inter-rater reliability.

We searched PubMed, EMBASE and CINAHL for potentially eligible studies. We designed a sensitive search strategy as follows: 'Canadian Triage and Acuity Scale' OR 'Emergency Severity Index' OR 'Manchester Triage Scale' OR 'Australasian Triage Scale'. Next, we searched PubMed for studies on other five-level triage systems with the following search terms: 'Emergency Service, Hospital'[Mesh] AND 'Triage'[Mesh]. There were no language restrictions. The last search date was 29 February 2016.

Two authors (AK and SU) independently screened abstracts to identify potentially eligible studies. The same authors then retrieved the full texts, independently assessed the eligibility of these studies, and screened their included reference lists. Any uncertainty about the eligibility of a study was resolved through discussion with the third author (TN).

Data extraction

Two authors (AK and SU) independently extracted the study characteristics (study design, country, number of study sites and triage scale applied), participant demographics (patient category by age and sample size) and the reference standards and measures that were used to evaluate the validity of the triage scales.

Analysis

The reference standards and measures to validate the triage systems and how they were used in evaluating the validity are described. As the goal was to describe the range of standards and measures used to validate triage systems, and not evaluate the accuracy of that validation, assessment of risk for bias in each original study was waived.

RESULTS

Description of studies

Our search for studies on the four most studied triage systems yielded 998 articles (figure 1), of which 46 met inclusion criteria[9–54] (table 1). Among them, 21 assessed ESI, 14 CTAS and 14 MTS.

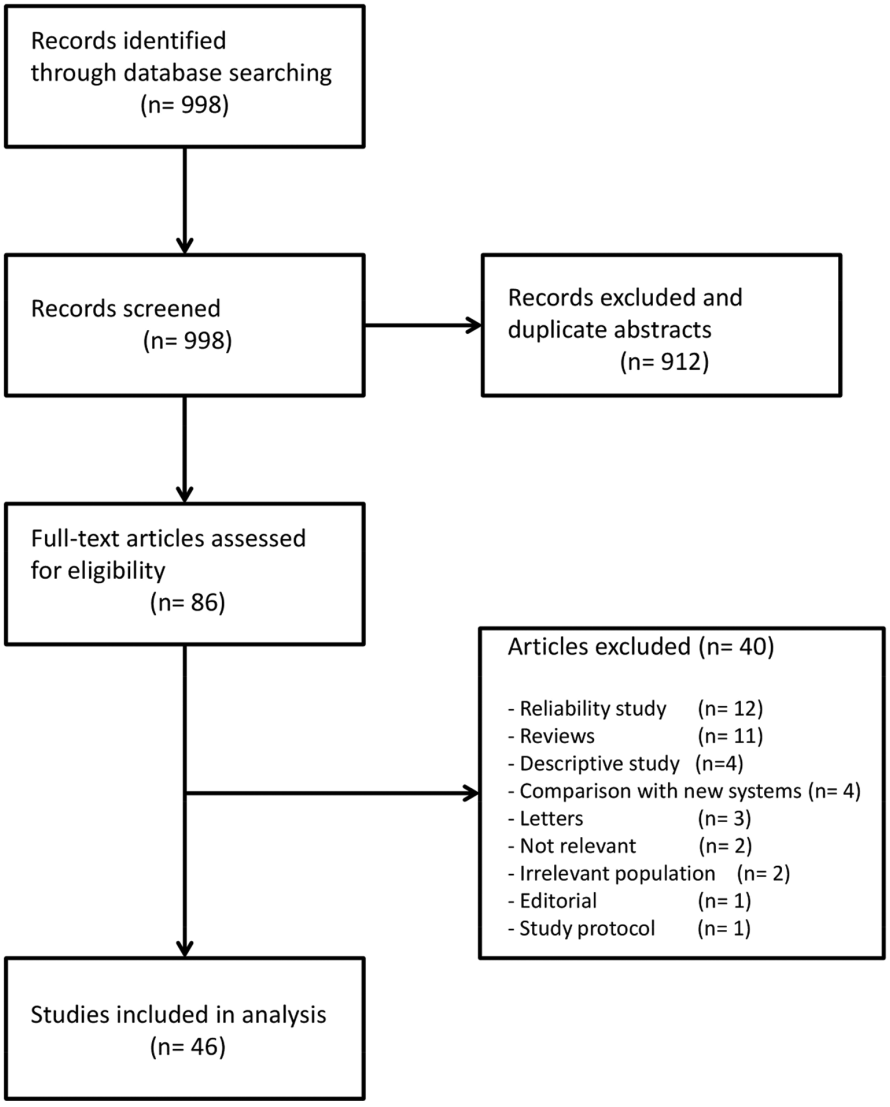

Figure 1 Study selection.

None evaluated the validity of ATS. Seven studies compared two more scales: MTS and ESI,[38] MTS, ESI and an informally structured triage system (ISS),[34] CTAS and the Taiwan Triage and Acuity Scale (TTS),[28] ESI and TTS,[48] CTAS and ESI,[41] and MTS V.1 and 2,[33] MTS and a modified version of MTS.[49] Twelve studies were from the USA,[9 11 12 15 23 31 35–37 42 43 53] nine from Canada,[10 13 18 20–22 41 47] seven from the Netherlands,[32 34 38–40 49 54] three from Switzerland,[24 25 50] two from Brazil,[30 52] Portugal[44 45] and Taiwan,[28 48] respectively, and one each was from Andorra, Germany, Kuwait, Norway, Saudi Arabia, South Korea, Sweden and Spain. One was an international multicentre study that examined MTS.[33] The median sample size was 1042 (range 50–550 940) patients, and a median of 1 site (range 1–12 sites) was studied. Adult, paediatric and geriatric populations were examined in 17, 20 and 4 articles, respectively. Six studies evaluated validity in both adult and paediatric populations, and 11 studies did not specify the populations. Eighteen studies were retrospective and 28 prospective observational.

Table 1 Characteristics of included studies on Australasian Triage System, Canadian Acuity and Triage Scale (CTAS), Emergency Severity Index (ESI) and Manchester Triage System (MTS)

Author	Year	Country	Target triage systems	Populations	Sample size	Sites (n)	Design
Wuerz[42]	2000	USA	ESI	Adult	493	2	Prospective
Wuerz[43]	2001	USA	ESI	Adult	8855	2	Retrospective
Wuerz[53]	2001	USA	ESI	Adult	202	1	Retrospective
Jiménez[26]	2003	Andorra	CTAS	Adult /paediatric	32 261	1	Retrospective
Eitel[15]	2003	USA	ESI V.2	Adult/paediatric	1042	7	Prospective
Bergeron[47]	2004	Canada	CTAS	Paediatric	55	1	Prospective
Tanabe[35]	2004	USA	ESI V.3	NS	403	1	Retrospective
Tanabe[36]	2004	USA	ESI V.3	NS	403	1	Retrospective
Gouin[18]	2005	Canada	CTAS	Paediatric	537	1	Prospective
Baumann[11]	2005	USA	ESI V.3	Paediatric	510	1	Retrospective
Dong[14]	2005	Canada	CTAS	NS	722	1	Prospective
Chi[48]	2006	Taiwan	ESI, TTS	Adult	3172	1	Prospective
Roukema[32]	2006	Netherlands	MTS	Paediatric	1065	1	Retrospective
Worster[41]	2006	Canada	CTAS, ESI V.3	NS	486	2	Prospective
Baumann[12]	2007	USA	ESI V.3	Geriatric	929	1	Retrospective
Dong[13]	2007	Canada	CTAS	NS	29 346	1	Prospective
Elshove-Bolk[17]	2007	Norway	ESI	Adult	1832	1	Prospective
van Veen[40]	2008	Netherlands	MTS	Paediatric	16 735	2	Prospective
van der Wulp[39]	2008	Netherlands	MTS	NS	50	2	Prospective
Ma[10]	2008	Canada	CTAS	Paediatric	1618	1	Retrospective
van der Wulp[38]	2009	Netherlands	ESI, MTS	NS	37 974/34 258	4	Retrospective
Travers[37]	2009	USA	ESI V.4	Paediatric	1000	5	Prospective
Olofsson[29]	2009	Sweden	MTS	Adult/paediatric	1027	7	Prospective
Gravel[22]	2009	Canada	CTAS	Paediatric	58 529	1	Retrospective
Martins[45]	2009	Portugal	MTS	Adult	316 622	1	Retrospective
Ng[28]	2010	Taiwan	CTAS, TTS	Adult	1851	3	Prospective
Platts-Mills[31]	2010	USA	ESI	Geriatric	782	1	Retrospective
van der Wulp[54]	2010	2010	ESI	Adult	584	1	Prospective
Storm-Versloot[34]	2011	Netherlands	MTS, ESI, ISS	Adult/paediatric	900	1	Prospective
Grossmann[24]	2011	Switzerland	ESI V.4	NS	2114	1	Prospective
Elkum[16]	2011	Kuwait	CTAS	Adult/paediatric	1206	1	Retrospective
Pinto[30]	2012	Brazil	MTS	Adult/paediatric	300	1	Prospective
Grossmann[25]	2012	Switzerland	ESI	Geriatric	819	1	Prospective
Lee[27]	2012	South Korea	CTAS	Geriatric	1903	1	Retrospective
Gravel[21]	2012	Canada	CTAS	Paediatric	1564	9	Prospective
Green[23]	2012	USA	ESI V.4	Paediatric	780	1	Retrospective
van Veen[49]	2012	Netherlands	MTS, modified MTS	Paediatric	11 210	2	Prospective
Gravel[20]	2013	Canada	CTAS	Paediatric	550 940	12	Retrospective
Santos[44]	2013	Portugal	MTS	NS	24 721	1	Prospective
Al-Hindi[51]	2014	Saudi Arabia	CTAS	Paediatric	3014	1	Prospective
Graff[19]	2014	Germany	MTS	NS	45 469	1	Retrospective
Seiger[33]	2014	Netherlands, UK, Portugal	MTS V.1 and 2	Paediatric	60 735	4	Prospective
Guedes[52]	2015	Brazil	MTS	Adult	577	1	Prospective
Hong[9]	2015	USA	ESI	Adult	310	1	Prospective
Ruipérez[46]	2015	Spain	ESI	NS	410	1	Prospective
Steiner[50]	2016	Switzerland	MTS	Adult	2407	1	Prospective

ISS, informally structured triage system; NS, not stated; TTS, Taiwan Triage and Acuity Scale.

Table 2 Characteristics of other five-level triage systems

Author	Year	Country	Triage systems	Populations	Sample size	Sites (n)	Design
Maningas[64]	2006	USA	Soterion Rapid Triage System	Adult /paediatric	33 850	1	Retrospective
Maningas[65]	2006	USA	Soterion Rapid Triage System	Paediatric	7077	1	Retrospective
Taboulet[57]	2009	France	FRench Emergency Nurses Classification in Hospital scale	Adult	941	1	Retrospective
Wildgren[59]	2011	Sweden	Medical Emergency Triage and Treatment System	Adult	8695	1	Retrospective
van Ierland[60]	2011	Netherlands	Netherlands Triage System	Adult/paediatric	3207	1	Prospective
Ozüçelik[58]	2013	Turkey	Hacettepe Emergency Triage System	Paediatric	308	1	Prospective
Chang[61]	2013	Taiwan	Paediatric Taiwan Triage System/ Pediatric Triage and Acuity Scale	Paediatric	42 548/42 346	1	Retrospective
Jobé[56]	2014	Belgium	Echelle Liégeoise d'Index de Sévérité à l'Admission	Adult	544	1	Retrospective
Elias[55]	2015	USA	Clinical GPS	NS	73	1	Prospective
Perez[63]	2016	Denmark	Rapid Emergency Triage and Treatment System- Hospital Unit West	Adult/paediatric	4680	1	Retrospective
Betz[62]	2016	Canada	Rapid Triage Score	NS	496	1	Prospective

NS, not stated.

Our search for other five-level triage systems yielded 3227 articles. Ten triage systems from 11 articles were finally included for analysis: Clinical GPS,[55] Echelle Liégeoise d'Index de Sévérité à l'Admission,[56] FRench Emergency Nurses Classification in Hospital scale,[57] Hacettepe Emergency Triage System,[58] Medical Emergency Triage and Treatment System,[59] Netherlands Triage System,[60] Pediatric Triage and Acuity Scale,[61] Rapid Triage Score,[62] Rapid Emergency Triage and Treatment System-Hospital Unit West[63] and Soterion Rapid Triage System.[64 65] All these systems were examined in single-centre studies, and 4 out of 11 were prospectively conducted[55 58 60 62] (table 2).

Reference standards and measures

For studies using criterion validity, reference standards included degree of urgency with objective standard criteria developed by expert panels for their studies a priori, and other triage systems that had already been in use as a means of validating a triage system. For studies of construct validity, measures included patient prognosis, costs and resource use that were gauged during the ED stay and after hospitalisation (box).

We identified five main outcomes for the validation of triage systems (table 3). For criterion validity, outcomes included (1) agreement of assigned degree of urgency with objective standard criteria set by expert panels for their studies a priori, (2) overtriage and undertriage and (3) sensitivity and specificity for defined reference standards. Outcomes for studies using construct validity were (1) associations between assigned degree of patient urgency, their prognosis, cost and resource use in EDs; and (2) associations between assigned degree of urgency and patient prognosis and cost after hospitalisation were examined.

For studies on criterion validity, nine studies used a reference standard that was developed by investigators.[2 14 29 32 34 39 40 46 49] Six of these were studies of MTS, and one each was of CTAS and ESI, respectively. One study evaluated MTS, ESI and ISS. The details of reference standards are shown in table 4.

Overtriage and undertriage of patients in ED were measured in eight studies (three examining MTS, two ESI, one CTAS, one comparing MTS and a modified version of MTS, and one comparing MTS, ESI and ISS).[29 34 37 39 46 47 49 50] Overtriage and undertriage were defined in four simulation studies of vignettes and three prospective studies as the degree of urgency assigned by nurses being above or below that assigned by experts, respectively.[29 34 39 46 47 49 50] Travers *et al* defined overtriage as patients who were rated ESI 1, 2 or 3 (acuity) but required <2 resources or those who were assessed as ESI 1 but who were not hospitalised, and undertriage as those who were assessed as ESI 4 or 5 (non-acuity) but received ≥2 resources or were hospitalised.[37]

The sensitivity and specificity of some reference standards were measured in seven studies (three respectively examining MTS and ESI, and one comparing MTS, ESI and ISS).[9 31 32 34 39 40 46 47] The reference standards comprised patients receiving immediate life-saving intervention,[9 31] each of five degrees of urgency,[34] and high and low degrees of urgency[32 39 40 46] that were predefined by investigators.

For studies on construct validity, association between assigned degree of patient urgency and one or more measures gauged during the ED stay and after hospitalisation was examined. Admissions from the ED were most commonly measured, with 33 studies using this outcome.[9 11–15 17–28 32–34 36–38 42–46 48 51]

Box Reference standards and measures used in the validation studies of five-level triage systems

Reference standards (criterion validity)

- Objective standard criteria/urgency set by expert panels for their studies a priori
- Existing emergency triage systems
- Immediate life-saving interventions

Measures (construct validity)

- Overall admissions
- Admissions to intensive care or monitored units
- ED length of stay
- Costs in EDs
- Number of resource used in EDs
- Mortality in EDs
- Leaving without being seen
- Waiting times before examinations by physicians in EDs
- Referrals to outpatients after the discharge from EDs
- In-hospital mortality
- Hospital length of stay
- Costs after hospitalisation
- Six-month survival
- Sixty-day mortality
- Thirty-day mortality
- Ninety-day mortality

Table 3 Approaches with reference standards and measures in the validation of Australasian Triage System, Canadian Triage and Acuity System (CTAS), Emergency Severity Index (ESI) and Manchester Triage Systems (MTS)

Approaches	All	CTAS (14 studies)	ESI (21 studies)	MTS (14 studies)
1. Agreement of assigned urgency levels with objective standard criteria set for studies (criterion validity)	9	1	3	7
2. Overtriage and undertriage (criterion validity)	8	1	3	5
3. Sensitivity and specificity for certain outcomes (criterion validity)	7	–	4	4
4. Associations between actual urgency levels of patients and measures in EDs (construct validity)				
Overall admissions	33	11	17	8
ED length of stay	19	6	12	1
Number of resource used in EDs	17	4	13	2
Mortality	9	2	4	4
Admissions to intensive care or monitored units	8	4	2	1
Costs	7	3	3	1
Leaving without being seen	3	2	1	–
Waiting times before examinations by physicians	1	1	–	–
Referrals to outpatients after the discharge from EDs	1	–	1	–
5. Associations between actual urgency levels of patients and outcomes after hospitalisation (construct validity)				
In-hospital mortality	8	2	5	2
Hospital length of stay	6	1	2	3
Costs after hospitalisation	1	1	–	–
Six-month survival	1	–	1	–
Thirty-day mortality	1	–	–	1

Table 4 Reference standards developed by investigators for validating triage systems

Author	Year	Triage systems	Design	Validated sample	Details of reference standards
Dong	2005	CTAS	Prospective	Patients	An expert panel determined triage levels of randomly selected patients through consensus after chart reviews, while they were kept blinded from the bedside assessment, investigation, management and patient outcomes.
Roukema	2006	MTS	Retrospective	Patients	An expert panel defined a five-level reference classification for true urgency. The classification comprised a combination of vital signs, a possible life-threatening condition, resource use (diagnostic investigation and treatment) and disposition from the ED.
van Veen	2008	MTS	Prospective	Patients	Paediatricians and a paediatric surgeon defined a five-level reference standard based on the literature and expert opinions. This was based on a combination of vital signs, diagnosis, diagnostic and therapeutic interventions, hospital admission and follow-up.
van Veen	2012	MTS and modified MTS	Prospective	Patients	Paediatricians and a paediatric surgeon created a reference urgency category based on a combination of vital signs, diagnosis, diagnostic and therapeutic interventions and hospitalisation/follow-up.
Storm-Versloot	2014	MTS, ESI, ISS	Prospective	Patients	An expert panel comprising seven experienced ED physicians set a five-level urgency classification. Each physician independently reviewed each case record and determined a category based on ED data, results of diagnostic tests, final diagnosis and patient prognosis.
Steiner	2016	MTS	Prospective	Patients	Two independent physicians determined an urgency level based on the clinical data, diagnostic tests and final diagnosis.
van der Wulp	2008	MTS	Prospective	Vignettes	Two experts independently evaluated vignettes of 50 actual patients and assigned degrees of urgency through consensus.
Olofsson	2009	MTS	Prospective	Vignettes	An expert panel rated and determined degrees of urgency in case scenarios through consensus.
Ruipérez	2015	ESI	Prospective	Patients	Triage nurses and a triage expert independently evaluated patients, and the expert's evaluation was considered as the reference standard.[46]

CTAS, Canadian Triage and Acuity Scale; ESI, Emergency Severity Index; ISS, informally structured triage system; MTS, Manchester Triage System.

Other measures gauged during the ED stay included admissions to intensive care or monitored units,[19 20 22 24 26 27 35 50] mortality in the ED (n=9),[26 27 36 38 42–45 50] patients leaving without being attended,[16 20 46] duration of wait before being examined by a physician, referrals to outpatients after discharge from the ED, Worthing Physiological Scoring System scores,[54] length of stay (LOS) in the ED[11–13 15 16 19 20 22–26 28 35 37 42 43 46 48] and costs consumed in EDs. The measure in 17 studies was the amount of resources used in EDs.[9 11 13 15 17 21–25 32 34 35 37 41 42 46 51] A few of these studies specifically counted specialist consultations, necessity for monitoring, diagnostic procedures (electrocardiography, laboratory examinations, diagnostic imaging, blood cultures and invasive diagnostic tests)[44] and treatment (intravenous fluids, transfusions, mechanical ventilation, inhalers and life-saving interventions). Measures gauged after admission included in-hospital mortality rates,[19 24 25 27 36 41 52] hospital LOS,[19 24 27 35 50 52] 30-day[50] and 6-month survival,[53] and costs consumed after admission.

Table 5 Approaches with reference standards and measures in the validation of other five-level triage systems

Approaches	All (11 studies)
1. Agreement of assigned urgency levels with existing triage systems (criterion validity)	3
2. Overtriage and undertriage (criterion validity)	1
3. Sensitivity and specificity for certain outcomes (criterion validity)	–
4. Associations between actual urgency levels of patients and measures in EDs (construct validity)	
Overall admissions	7
ED length of stay	2
Number of resource used in EDs	3
Mortality	1
Admissions to intensive care or monitored units	2
Costs	3
Referrals to outpatients after the discharge from EDs	1
5. Associations between actual urgency levels of patients and outcomes after hospitalisation (construct validity)	
In-hospital mortality	2
Hospital length of stay	2
Thirty-day mortality	1
Sixty-day mortality	1
Ninety-day mortality	1

Other five-level triage systems

Reference standards and measures used to validate triage systems were similar to those for the four triage systems (table 5). Three and nine studies used criterion and construct validity, respectively. Of note, some studies used other triage systems (CTAS and ESI) that had already been in use, as the reference standards.[55 58 62]

DISCUSSION

We found that, of 57 validation studies, a variety of reference standards and measures were used. Overall, construct validity (51 studies) was more frequently examined than criterion validity (14 studies). Particularly, validation studies of ESI (5 of 21 studies) and MTS (7 of 14 studies) more frequently used a form of criterion validity compared with CTAS (2 of 14 studies). Validation studies of these three triages commonly used some construct validity; CTAS (13 studies), ESI (20 studies) and MTS (10 studies).

Criterion validity should be preferred in validation of scales. In validating triage systems, true 'urgency' of patients should serve as reference standards, because triage systems rank the speed of care for a patient, or namely, urgency. However, our study found that many validation studies focused on severity or construct validity, likely due to the lack of established criterion validity in triage system research.[66] For criterion validity, reference standards were mostly the studies' own criteria, which were either urgency alone or a combination of urgency and severity, determined by investigators. Other triage systems that are already in use could also serve as reference standards for a newly introduced triage system, which reduces the variability of reference standards.[28 34]

Triage scales could be considered a type of decision rule, in which the goal of the rule is to predict the need for immediate care. However, unlike decision rules, they are consensus-based and lack the typical multistep data gathering and statistical processes resulting in derivation, and then prospective validation on a unique population. Moll suggested a four-step approach to validate a triage system[4]: consensus-based derivation of decision rules for different degrees of urgency, validation of a system with a reference standard as the best proxy for prognosis in a single setting (internal validation), modification of triage rules and validation in various emergency care settings (external validation). For studies evaluating criterion validity, most reference standards or criteria developed by the investigators for their validation studies followed Moll's framework of reference standards or the best 'proxy' based on the information of urgency and severity. We speculate that the investigators of studies assessing criterion validity needed to establish reference standards based on both urgency and severity because urgency is hard to determine or predict based on a limited amount of information gained during the triage.

All validation studies are currently subject to the limitations described above due to the absence of perfect standard references of urgency. Clinicians need to know that the validity of triage systems has not been perfectly determined and their weaknesses remain obscure. Bearing this in mind, clinicians need to triage patients using the available triage systems.

The present study has some limitations. First, it was designed simply to review and describe the methodologies used in validation studies of emergency triage systems without the intent to suggest the most appropriate reference standard or measure for a validation study. Second, our study focused on five-level triage systems. We might have missed other methodology as well as reference standards and measures used to examine other triage systems. Despite these limitations, we summarised the reference standards and measures used in validation studies of triage systems, and described their drawbacks and advantages. This information should provide an important rationale for future validation studies of triage systems.

CONCLUSIONS

The most commonly used triage systems have been validated using both criterion and construct validity of emergency triage systems. The difficulty in defining a surrogate for urgency means that studies must either create their own reference standard (often an expert panel) or use a combination of severity/resource measures which approximate but are not the same as urgency. Given that the limitations of validation studies are not completely understood and given the potential flaws of triage systems, future studies should attempt to elucidate the weaknesses of triage systems in terms of presenting signs and symptoms and the characteristics of patients.

Contributors AK and TN conceived the study design and interpreted the data. AK and SU acquired the data. AK analysed the data and drafted the manuscript. All authors critically revised and approved the submission of the manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 FitzGerald G, Jelinek GA, Scott D, *et al*. Emergency department triage revisited. *Emerg Med J* 2010;27:86–92.
2 Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford; New York: Oxford University Press, 2003.
3 Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Annu Rev Clin Psychol* 2009;5:1–25.

4 Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epidemiol* 2010;63:384–8.
5 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W–94.
6 Fernandes CM, 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.
7 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.
8 McHugh M, Tanabe P, McClelland M, *et al*. More patients are triaged using the emergency Severity Index than any other triage acuity system in the United States. *Acad Emerg Med* 2012;19:106–9.
9 Hong R, Sexton R, Sweet B, *et al*. Comparison of START triage categories to emergency department triage levels to determine need for urgent care and to predict hospitalization. *Am J Disaster Med* 2015;10:13–21.
10 Ma W, Gafni A, Goldman RD. Correlation of the Canadian pediatric emergency triage and acuity scale to ED resource utilization. *Am J Emerg Med* 2008;26:893–7.
11 Baumann MR, Strout TD. Evaluation of the emergency severity Index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med* 2005;12:219–24.
12 Baumann MR, Strout TD. Triage of geriatric patients in the emergency department: validity and survival with the emergency severity Index. *Ann Emerg Med* 2007;49:234–40.
13 Dong SL, Bullard MJ, Meurer DP, *et al*. Predictive validity of a computerized emergency triage tool. *Acad Emerg Med* 2007;14:16–21.
14 Dong SL, Bullard MJ, Meurer DP, *et al*. Emergency triage: comparing a novel computer triage program with standard triage. *Acad Emerg Med* 2005;12:502–7.
15 Eitel DR, Travers DA, Rosenau AM, *et al*. The emergency severity index triage algorithm version 2 is reliable and valid. *Acad Emerg Med* 2003;10:1070–80.
16 Elkum NB, Barrett C, Al-Omran H. Canadian emergency department triage and acuity scale: implementation in a tertiary care center in Saudi Arabia. *BMC Emerg Med* 2011;11:3.
17 Elshove-Bolk J, Mencl F, van Rijswijck BT, *et al*. Validation of the emergency severity Index (ESI) in self-referred patients in a european emergency department. *Emerg Med J* 2007;24:170–4.
18 Gouin S, Gravel J, Amre DK, *et al*. Evaluation of the paediatric canadian triage and acuity Scale in a pediatric ED. *Am J Emerg Med* 2005;23:243–7.
19 Gräff 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.
20 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.
21 Gravel J, Gouin S, Goldman RD, *et al*. The canadian triage and acuity Scale for children: a prospective multicenter evaluation. *Ann Emerg Med* 2012;60:71–7.
22 Gravel J, Manzano S, Arsenault M. Validity of the Canadian paediatric triage and acuity scale in a tertiary care hospital. *CJEM* 2009;11:23–8.
23 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;28:753–7.
24 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;57:257–64.
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;60:317–25.
26 Jiménez JG, Murray MJ, Beveridge R, *et al*. Implementation of the Canadian emergency department triage and acuity Scale (CTAS) in the principality of Andorra: can triage parameters serve as emergency department quality indicators? *CJEM* 2003;5:315–22.
27 Lee JY, Oh SH, Peck EH, *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.
28 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;109:828–37.
29 Olofsson P, Gellerstedt M, Carlström ED. Manchester triage in Sweden-interrater reliability and accuracy. *Int Emerg Nurs* 2009;17:143–8.
30 Pinto D, Salgado PO, Chianca TC. Predictive validity of the manchester triage system: evaluation of outcomes of patients admitted to an emergency department. *Rev Lat Am Enfermagem* 2012;20:1041–7.
31 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;17:238–43.
32 Roukema J, Steyerberg EW, van Meurs A, *et al*. Validity of the manchester triage system in paediatric emergency care. *Emerg Med J* 2006;23:906–10.
33 Seiger N, van Veen M, Almeida H, *et al*. Improving the manchester triage system for pediatric emergency care: an international multicenter study. *PLoS One* 2014;9:e83267.
34 Storm-Versloot MN, Ubbink DT, Kappelhof J, *et al*. Comparison of an informally structured triage system, the emergency severity index, and the manchester triage system to distinguish patient priority in the emergency department. *Acad Emerg Med* 2011;18:822–9.
35 Tanabe P, Gimbel R, Yarnold PR, *et al*. The emergency Severity Index (version 3) 5-level triage system scores predict ED resource consumption. *J Emerg Nurs* 2004;30:22–9.
36 Tanabe P, Gimbel R, Yarnold PR, *et al*. Reliability and validity of scores on the emergency Severity Index version 3. *Acad Emerg Med* 2004;11:59–65.
37 Travers DA, Waller AE, Katznelson J, *et al*. Reliability and validity of the emergency severity index for pediatric triage. *Acad Emerg Med* 2009;16:843–9.
38 van der Wulp I, Schrijvers AJ, 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;26:506–9.
39 van der Wulp I, van Baar ME, Schrijvers AJ. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. *Emerg Med J* 2008;25:431–4.
40 van Veen M, Steyerberg EW, Ruige M, *et al*. Manchester triage system in paediatric emergency care: prospective observational study. *BMJ* 2008;337:a1501.
41 Worster A, Fernandes CM, Eva K, *et al*. Predictive validity comparison of two five-level triage acuity scales. *Eur J Emerg Med* 2007;14:188–92.
42 Wuerz RC, Milne LW, Eitel DR, *et al*. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000;7:236–42.
43 Wuerz RC, Travers D, Gilboy N, *et al*. Implementation and refinement of the emergency severity index. *Acad Emerg Med* 2001;8:170–6.
44 Santos AP, Freitas P, Martins HM. Manchester Triage System version II and resource utilisation in the emergency department. *Emerg Med J* 2014;31:148–52.
45 Martins HM, Cuña LM, Freitas P, Manchester I. 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;26:183–6.
46 Ruipérez TH, Costa CL, Martínez MGA, *et al*. Evidencias de validez del sistema de triaje emergency Severity Index en un servicio de urgencias de un hospital general. *Emergencias* 2015;27:301–6.
47 Bergeron S, Gouin S, Bailey B, *et al*. Agreement among pediatric health care professionals with the pediatric canadian triage and acuity scale guidelines. *Pediatr Emerg Care* 2004;20:514–8.
48 Chi CH, Huang CM. Comparison of the emergency severity index (ESI) and the Taiwan triage system in predicting resource utilization. *J Formos Med Assoc* 2006;105:617–25.
49 van Veen M, Steyerberg EW, Van't Klooster M, *et al*. The manchester triage system: improvements for paediatric emergency care. *Emerg Med J* 2012;29:654–9.
50 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;50:678–89.
51 Al-Hindi AA, Al-Akhfash AA, Fareed AM, *et al*. Efficacy of implementation of a 5 scale pediatric triage and acuity scale in pediatric emergency, Saudi Arabia. *Saudi Med J* 2014;35:999–1004.
52 Guedes HM, Martins JC, Chianca TC. Predictive value of the manchester triage system: evaluation of patients' clinical outcomes. *Rev Bras Enferm* 2015;68.4045–51.
53 Wuerz R. Emergency severity index triage category is associated with six-month survival. ESI Triage Study Group. *Acad Emerg Med* 2001;8:61–4.
54 van der Wulp I, Rullmann HA, Leenen LP, *et al*. Associations of the emergency severity index triage categories with patients' vital signs at triage: a prospective observational study. *Emerg Med J* 2011;28:1032–5.
55 Elias P, Damle A, Casale M, *et al*. A web-based tool for patient triage in emergency department settings: validation using the emergency severity index. *JMIR Med Inform* 2015;3:e23.
56 Jobé J, Ghuysen A, Gérard P, *et al*. Reliability and validity of a new French-language triage algorithm: the ELISA scale. *Emerg Med J* 2014;31:115–20.
57 Taboulet P, Moreira V, Haas L, *et al*. Triage with the french emergency nurses classification in Hospital scale: reliability and validity. *Eur J Emerg Med* 2009;16:61–7.
58 Özüçelik DN, Kunt MM, Karaca MA, *et al*. A model of complaint based for overcrowding emergency department: five-level Hacettepe emergency triage system. *Ulus Travma Acil Cerrahi Derg* 2013;19:205–14.
59 Widgren BR, Jourak M. Medical emergency triage and treatment system (METTS): a new protocol in primary triage and secondary priority decision in emergency medicine. *J Emerg Med* 2011;40:623–8.
60 van Ierland Y, van Veen M, Huibers L, *et al*. Validity of telephone and physical triage in emergency care: the Netherlands triage system. *Fam Pract* 2011;28:334–41.
61 Chang YC, Ng CJ, Wu CT, Cj N, Ct W, *et al*. Effectiveness of a five-level iaediatric triage system: an analysis of resource utilisation in the emergency department in Taiwan. *Emerg Med J* 2013;30:735–9.

62 Betz M, Stempien J, Wilde A, *et al*. A comparison of a formal triage scoring system and a quick-look triage approach. *Eur J Emerg Med* 2016;23:185–9.
63 Pérez N, Nissen L, Nielsen RF, *et al*. The predictive validity of RETTS-HEV as an acuity triage tool in the emergency department of a danish Regional Hospital. *Eur J Emerg Med* 2016;23:33–7.
64 Maningas PA, Hime DA, Parker DE, *et al*. The soterion rapid triage system: evaluation of inter-rater reliability and validity. *J Emerg Med* 2006;30:461–9.
65 Maningas PA, Hime DA, Parker DE. The use of the soterion rapid triage system in children presenting to the emergency department. *J Emerg Med* 2006;31:353–9.
66 Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. *Emerg Med J* 2007;24:477–9.