



OPEN ACCESS

One-two-triage: validation and reliability of a novel triage system for low-resource settings

Ayesha Khan,¹ S V Mahadevan,¹ Andrea Dreyfuss,² James Quinn,¹ Joan Woods,³ Koy Somontha,³ Matthew Strehlow⁴

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/emmermed-2015-205430>).

¹Division of Emergency Medicine, Stanford University, Stanford, California, USA

²Department of Emergency Medicine, Highland General Hospital, Oakland, California, USA

³University Research Co. Centre for Human Services, Phnom Penh, Cambodia

⁴Division of Emergency Medicine, Stanford University, Stanford, California, USA

Correspondence to

Dr Ayesha Khan, 300 Pasteur Drive, Alway Bldg., Room M121, Stanford, CA 94305, USA; akhanx@stanford.edu

Received 1 October 2015

Revised 22 June 2016

Accepted 27 June 2016

Published Online First

27 July 2016

ABSTRACT

Objectives To validate and assess reliability of a novel triage system, one-two-triage (OTT), that can be applied by inexperienced providers in low-resource settings.

Methods This study was a two-phase prospective, comparative study conducted at three hospitals. Phase I assessed criterion validity of OTT on all patients arriving at an American university hospital by comparing agreement among three methods of triage: OTT, Emergency Severity Index (ESI) and physician-defined acuity (the gold standard). Agreement was reported in normalised and raw-weighted Cohen κ using two different scales for weighting, Expert-weighted and triage-weighted κ . Phase II tested reliability, reported in Fleiss κ , of OTT using standardised cases among three groups of providers at an urban and rural Cambodian hospital and the American university hospital.

Results Normalised for prevalence of patients in each category, OTT and ESI performed similarly well for expert-weighted κ (OTT $\kappa=0.58$, 95% CI 0.52 to 0.65; ESI $\kappa=0.47$, 95% CI 0.40 to 0.53) and triage-weighted κ ($\kappa=0.54$, 95% CI 0.48 to 0.61; ESI $\kappa=0.57$, 95% CI 0.51 to 0.64). Without normalising, agreement with gold standard was less for both systems but performance of OTT and ESI remained similar, expert-weighted (OTT $\kappa=0.57$, 95% CI 0.52 to 0.62; ESI $\kappa=0.6$, 95% CI 0.58 to 0.66) and triage-weighted (OTT $\kappa=0.31$, 95% CI 0.25 to 0.38; ESI $\kappa=0.41$, 95% CI 0.35 to 0.4). In the reliability phase, all triagers showed fair inter-rater agreement, Fleiss κ ($\kappa=0.308$).

Conclusions OTT can be reliably applied and performs as well as ESI compared with gold standard, but requires fewer resources and less experience.

INTRODUCTION

EDs require a systematic approach to prioritise patient care depending on acuity.¹ ED triage scales standardise initial patient assessment and reduce wait times for patients with life-threatening conditions, thereby minimising unnecessary deaths and disability.² However, current triage scales are designed for healthcare settings in high-income countries (HICs), requiring experienced medical providers or specialised algorithms. Due to a dearth of experienced emergency healthcare providers and limited resources, many hospitals in low- and middle-income countries (LMICs) have forgone triage altogether; others have adapted established triage scales with varying success.^{3, 4} Thus, although current triage systems theoretically apply to health professionals in various contexts, they are ineffective in many developing nations.

Key messages

What is already known on this subject?

Current triage scales are designed for healthcare settings in high-income countries, requiring experienced medical providers or specialised algorithms. Due to a dearth of experienced emergency healthcare providers and limited resources, many hospitals in low- and middle-income countries (LMICs) have forgone triage altogether or adapted established triage scales with varying success. Despite the clear need for a triage system designed specifically for LMICs that can be 'universally and broadly implemented', no comprehensive, reliable system has been established.

What might this study add?

Our study shows that one-two-triage, a triage scale created for low-resource settings, can be reliably applied by inexperienced providers in LMICs and is valid in determining acuity of patients presenting to the ED.

Prior to this study, most government hospitals in Cambodia lacked a standardised system for sorting patients based on acuity. Instead, patients typically self-triaged to the hospital department deemed most appropriate (eg, ED, intensive care unit (ICU), inpatient wards) and were seen in the order of arrival. As the number of Cambodian patients with acute medical emergencies has sharply risen,⁵ there is an immediate need to implement a validated triage system for adult and paediatric patients designed specifically for LMICs.

A panel of experts from Stanford University developed a triage system to be implemented within Cambodia's current healthcare infrastructure: one-two-triage (OTT). Named to highlight the two-stage process of collating patients by severity—the isolation of critical/emergent from non-critical patients followed by the separation of urgent from non-urgent patients—OTT was designed specifically for use in training-limited and resource-limited settings. OTT demands minimal medical knowledge for the initial information-gathering stage, can be applied to both adult and paediatric patients, and does not require a full set of vital signs to identify the most critical cases. At the request of the Cambodian Ministry of Health, Stanford University and University Research



CrossMark

To cite: Khan A, Mahadevan SV, Dreyfuss A, et al. *Emerg Med J* 2016;**33**:709–715.

Consortium partnered to implement OTT in government hospitals across Cambodia. The current study aimed to validate OTT and test its reliability following implementation in two Cambodian government hospitals.

METHODS

Study design and setting

This two-phase prospective, comparative study evaluated the validity and reliability of a novel triage system. Phase I evaluated the validity of OTT at Stanford University Hospital (SUH). Phase II tested the reliability of OTT at SUH, and at Siem Reap Provincial Hospital (SRPH) and Mehmot Referral Hospital (MRH) in Cambodia.

The three study hospitals are demographically distinct. SUH is an academic, adult and paediatric trauma centre with an annual volume of 60 000 patients. SRPH is a public, referral hospital providing adult emergency services to nearly 75 000 patients annually. MRH is a district-level facility that receives 7000 adult and paediatric patients per year. SUH employs the Emergency Severity Index (ESI) as a means of triaging patients, while neither SRPH or MRH used a formal triage process prior to OTT.

Methods and measurements

Phase I included patients presenting to the SUH ED during 36 eight-hour shifts between August and December 2013. Enrolment periods were split among day, night, weekday and weekend shifts. Twelve Stanford undergraduate students with no prior medical training enrolled patients and collected data after receiving a 1-day live training course on OTT (see online supplementary appendix E). The medical inexperience of the undergraduates was comparable to that of Cambodian nurses using the triage system for the first time. For each enrolled patient, the students assigned OTT acuity and recorded both ESI acuity (assigned by a trained ED triage nurse) and physician-defined acuity (assigned by the board-certified emergency medicine (EM) physician on shift) (figure 1A).

To assess validity, OTT-assigned acuity was compared with physician-defined acuity. We designated physician-defined acuity as the gold standard based on previous studies demonstrating that physician intuition better predicts patient mortality than exogenous physiological scoring systems and that years of experience generate accurate acuity assessment.^{6–9} For this study, physician-defined acuity was determined by board-certified EM-trained physicians at levels 1–4 corresponding to an ideal time to assessment of 0, 10, 30 and 120 min,

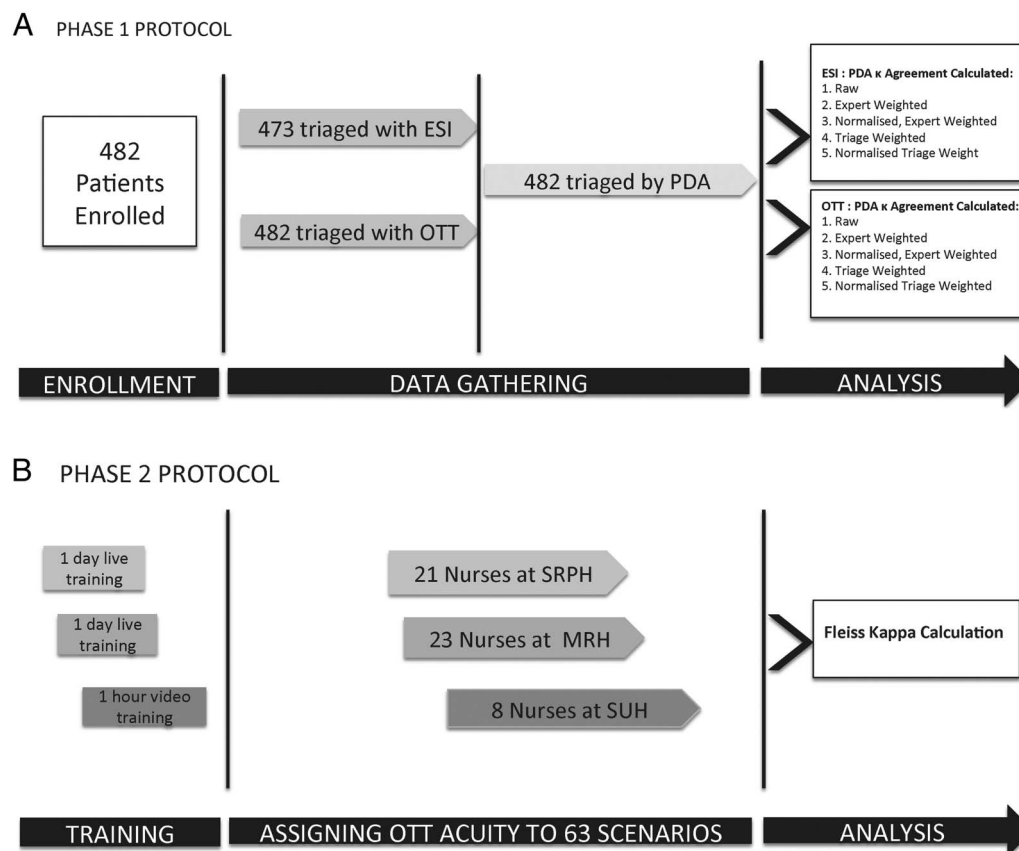


Figure 1 (A) Phase I: patients enrolled and triaged at Stanford. In phase I of the protocol, validity was assessed at Stanford University Hospital (SUH). Patients were triaged using Emergency Severity Index (ESI) by a nurse in the usual course of patient care, and they were assigned acuity using one-two triage (OTT), in parallel, during their triage assessment by an undergraduate student. The student then asked the attending physician responsible for the patient to assign physician-defined acuity (PDA) before the medical evaluation of the patient. Agreement between ESI and PDA and agreement between OTT and PDA were analysed using raw and weighted κ . (B) Phase II: 63 standardised scenarios assigned OTT acuity. In phase II of the protocol, reliability was assessed by three different groups of providers assigning OTT acuity to 63 written cases. The three groups included nurses from Siem Reap Provincial Hospital (SRPH), Mehmot Referral Hospital (MRH) and Stanford University Hospital (SUH). Training in OTT was provided to the Cambodian nurses via a 1-day live workshop and to the Stanford nurses via a 1 hour training video. Fleiss κ was used to analyse agreement.

respectively. Physicians were blinded to the acuity levels assigned by the alternative systems. As experienced board-certified EM-trained physicians are unavailable in Cambodia, SUH was chosen as the study site for phase I.

We also compared OTT with an established triage system, ESI. Rather than use a direct comparison, we measured both systems against physician-defined acuity, thus also testing the agreement between ESI and physician-defined acuity. Two a priori decisions were made to allow comparison between the two systems: (1) ESI and OTT triage scales differ in the absolute number of acuity levels, ESI levels 4 and 5 were combined, as these levels differ in resources used but not acuity, which is our primary concern. (2) Major and minor trauma activations were assigned ESI levels 1 and 2, respectively, based on analogous criteria for acuity assignment between scales (see online supplementary appendix A).

For phase I, sample size for unweighted κ was calculated using the R package kappaSize. The 'Power4Cats' function described therein was used since this was a four-category κ . The null hypothesis is that ratings were assigned by all observers randomly according to the usual ESI breakdown (0.15, 0.25, 0.25 and 0.35). We estimated a sample size of 120 patients for 90% power; unweighted κ calculation considered the usual ESI level breakdown at SUH ED (level 1 (15%), level 2 (25%), level 3 (25%) and level 4 (35%)). We estimated a sample size of 335 patients for 90% power to allow for the most extreme scenario of unevenly distributed triage (level 1 (5%), level 2 (5%), level 3 (85%) and level 4 (5%)), variability among groups of raters and weighting of the κ .

Phase II assessed the reliability of OTT through inter-rater agreement among 52 healthcare providers, each evaluating the complete set of 63 standardised, written EM cases. The participating providers comprised three groups: 8 nurses from SUH, 21 nurses and medical assistants from SRPH and 23 nurses and medical assistants from MRH. SUH nurses were trained to use OTT via a 1 hour, online training module. The Cambodian providers at SRPH and MRH received a 1-day, live training course taught in their native language (Khmer) (figure 1B). The difference in training duration reflects the existing familiarity with triage among SUH staff; the SUH module highlighted OTT-specific differences in triage compared with the more comprehensive module for Cambodian providers. The written EM case scenarios were based on cases commonly encountered in Cambodian EDs; they were devised

and tested for agreement by a team of eight board-certified EM physicians from Stanford following consultation by a doctorate-level education specialist within the SUH Department of Surgery. The 35 adult and 28 paediatric cases were distributed across chief complaint categories and OTT acuity, as shown in table 1, with red being critical, orange emergent, yellow urgent and green non-urgent.

Outcomes

Our primary study outcome for phase I was agreement between physician-defined acuity (gold standard) and assigned OTT triage level. Agreement was measured using a normalised, asymmetrical Cohen's weighted κ as it adjusts for degrees of discordance within ordinal scales. κ was weighted asymmetrically to reflect the greater clinical consequences of under-triaging versus over-triaging patients. Under-triaging—assigning a lower acuity than the patient's presentation warrants—can delay care with serious clinical consequences, whereas over-triaging—assigning a greater acuity than warranted—carries a lower risk of direct patient harm. Two different asymmetrically weighted scales were used: *triage-weighted* κ , introduced by Van der Wulp *et al*,¹⁰ and *expert-weighted* κ , a scale dependent on the consensus opinion of EM physicians at SUH ED (see online supplementary appendix B). To prevent signal loss from the relatively few high-acuity cases, the number of patients in each category was normalised to equal one another (see online supplementary appendix C). Secondary outcomes included raw agreement between OTT and physician-defined acuity (without weighting or normalisation), agreement between ESI and physician-defined acuity and comparative performance of OTT and ESI.

Our study assessed the criterion validity of OTT using emergency physician-defined acuity as the *gold standard* for correlation rather than the more commonly used correlation with surrogate markers, such as mortality, admission, length of hospital stay and recidivism.^{11–14} Since these outcomes depend on external factors, including delay in treatment, inadequate treatment or patient comorbid conditions, they may not accurately validate the system particularly in an LMIC where surrogate markers are heavily influenced by resource availability, adjunct therapies and provider inexperience.^{14–18} Conversely, criterion validity using our gold standard more directly assesses the triage tool. For further analysis, we examined parity between the criterion validity of OTT and the criterion validity of a well-

Table 1 Phase II: breakdown of standardised scenarios

Complaint type	Acuity				Age		Frequency	Percentage
	Red	Orange	Yellow	Green	Adult	Ped		
Cardiothoracic	3	1	3	3	6	4	10	15.9
ENT/eye	1	0	4	2	4	3	7	11.1
Fever	0	1	3	2	3	3	6	9.5
Genitourinary	0	1	4	2	4	3	7	11.1
Gastrointestinal	0	1	4	2	3	4	7	11.1
Neurology	0	2	4	1	4	3	7	11.1
Trauma/musculoskeletal	1	2	4	2	6	3	9	14.3
Toxicology/skin	0	0	5	1	3	3	6	9.5
General orange	0	2	0	0	1	1	2	3.2
General red	2	0	0	0	1	1	2	3.2
Total	7	10	31	15	35	28	63	100

ENT, ear, nose, throat; Ped, Paediatrics.

Table 2 Phase I: breakdown of patients' chief complaint based on organ system

Organ system	Frequency	Percentage
Gastrointestinal	100	20.8
Cardiothoracic	88	18.3
Musculoskeletal	66	13.7
Neurological	46	9.6
Fever	40	8.3
Trauma	36	7.5
Skin and soft tissue	31	6.4
Otolaryngology	25	5.2
Psychiatric	19	4.0
Genitourinary	17	3.5
Eyes	9	1.9
Toxic	4	0.8
Total	481	100

established, reputable triage scale, ESI, with previously proven correlation to surrogate markers.¹⁵

During phase II, we assessed reliability for interobserver agreement using Fleiss κ to account for multiple raters. While Fleiss is typically used for nominal, non-ordinal data, the typology of triage categories is not strictly ordinal in nature as various criteria are applied and intersected to determine a patient's number. For example, ESI is not simply ordinal because there are essentially three categories based on acuity (1, 2 and the remainder) and three categories based on anticipated resource utilisation (3, 4 and 5). A patient may be higher in the resource category but lower in the acuity category and an aggregate of these ordinal scales is applied to determine the overall category, which does not then remain strictly ordinal. Similarly, OTT acuity—like ESI—is an aggregation of various ordinal scales and thus lends itself to Fleiss analysis.¹⁹

Data analysis

Data from the validity portion were analysed with Python and STATA. We assessed homogeneity within groups using one-way analysis of variance, treating each triage group as one unit for Cohen-weighted κ .

Calculations were carried out according to formulae described by Fleiss in *Measuring nominal scale agreement among many raters* using Palantir Metropolis software V4.3.2.1. Equations for the κ and variance (which is the square of the SE) can be found in online supplementary appendix D.²⁰ Final assignment of significance of agreement was based on the ratio of absolute κ divided by its SE, where a ratio >1.96 means $p < 0.05$.¹⁵

RESULTS

Study subjects

During phase I, 482 patients were enrolled and triaged. ESI acuity was collected on 473 of 482 patients; physician acuity

was assigned for all 482 patients based on their initial assessment. Of the patients, 231 were male and 251 female, ranging from 7 days to 97 years old with a median age of 36 years. The most common organ systems associated with patient chief complaints were gastrointestinal (20.8%), cardiothoracic (18.3%) and musculoskeletal (13.7%) (table 2).

Phase I

The OTT triage system showed moderate agreement with physician-defined acuity using both normalised *expert-weighted* κ ($\kappa=0.58$, 95% CI 0.52 to 0.65) and normalised *triage-weighted* κ ($\kappa=0.54$, 95% CI 0.48 to 0.61). Without normalisation, OTT's agreement with the gold standard remained similar using *expert-weighted* κ ($\kappa=0.57$, 95% CI 0.52 to 0.62) but was weaker using *triage-weighted* κ ($\kappa=0.31$, 95% CI 0.25 to 0.38). Similar to OTT, the ESI triage system moderately agreed with physician-defined acuity. Without normalisation, agreement with the gold standard was less for both triage systems but performance of OTT and ESI remained similar. Table 3 shows the results of asymmetrically weighting the agreement using *expert-weighted* and *triage-weighted* κ .

The number of patients assigned to each triage level by the respective triage method is shown in figure 2. OTT characterised slightly more patients as higher acuity (level 1 or 2) than ESI or physician-defined acuity, whereas the physicians characterised more patients as the lowest acuity (level 4). Overall, OTT over-triaged in 34% of cases and ESI in 27%; OTT and ESI both under-triaged 13% of patients.

Phase II

In the reliability portion (phase II), all healthcare providers using OTT showed fair inter-rater agreement across questions (Fleiss κ ; $\kappa=0.308$). By individual site, SUH nurses showed substantial agreement ($\kappa=0.717$), whereas MRH and SRPH providers exhibited fair agreement ($\kappa=0.334$ and $\kappa=0.269$, respectively). When considering adult-specific and paediatric-specific questions, reliability remained consistent (figure 3). Finally, when accounting for SE, all Fleiss κ s demonstrated agreement in the assignment of the four triage categories significantly greater than by chance (figure 4).

DISCUSSION

In this study, we developed a triage tool for LMIC and tested it for validity and reliability. Despite the clear need for a triage system designed specifically for LMICs that can be 'universally and broadly implemented', no comprehensive, reliable system has been established. OTT addresses this need as it was designed to function in low-resource settings with limited provider training and little access to expensive technologies. The current study demonstrated that OTT can be reliably applied after a short training module for triagers in LMICs. OTT demonstrated good agreement in determining patient acuity compared with expert EM physician assessment. Importantly, OTT performed similar to ESI, an internationally recognised and used triage system, and outperformed ESI for the most critically ill patients.

Table 3 Phase I: weighted agreement by triage method

Rating system	N		Unweighted	Expert-weighted	Normalised expert-weighted	Triage-weighted	Normalised triage-weighted
OTT	482	κ (95% CI)	0.27 (0.22 to 0.33)	0.57 (0.52 to 0.62)	0.58 (0.52 to 0.65)	0.31 (0.25 to 0.38)	0.54 (0.48 to 0.61)
ESI	473	κ (95% CI)	0.34 (0.28 to 0.39)	0.6 (0.58 to 0.66)	0.47 (0.40 to 0.53)	0.41 (0.35 to 0.48)	0.57 (0.51 to 0.64)

ESI, Emergency Severity Index; OTT, one-two-triage.

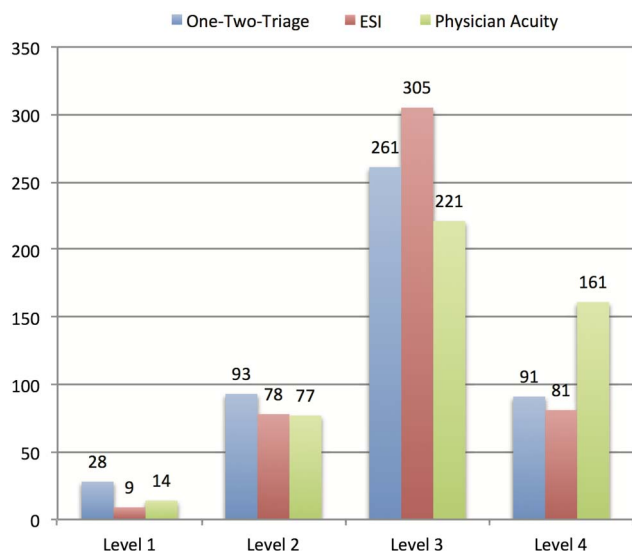
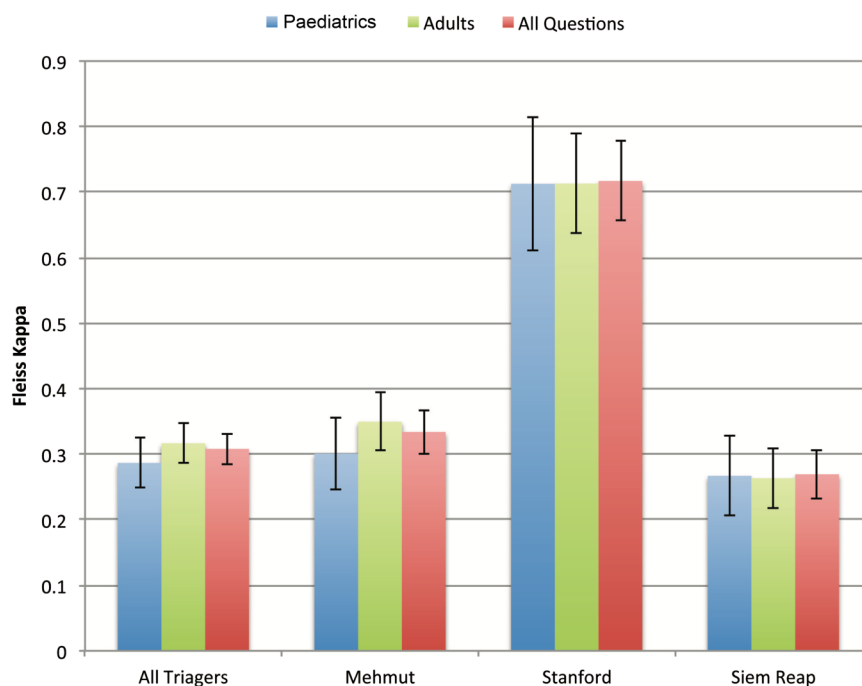


Figure 2 Phase I: number of patients assigned to each acuity level by triage method: one-two-triage (OTT), Emergency Severity Index (ESI) and physician-defined acuity (PDA). The group of patients enrolled in phase I showed a similar distribution across acuity levels when triaged by OTT, ESI and PDA. OTT characterised slightly more patients as high acuity and PDA characterised slightly more patients as the lowest acuity.

Prior to OTT, the only widely accepted triage scales developed for LMICs were the emergency triage assessment and treatment (ETAT) and the South African Triage Scale (SATS). ETAT is limited to paediatrics and has been unreliable in LMICs, with large variations in assessment of priority signs even after 20 hours of training.⁴ Conversely, OTT was reliably applied by Cambodian nurses after only 8 hours of training. A strength of OTT is that it greatly limits the inclusion of signs and symptoms that demand subjective clinical decision-making by the triaging nurse. Stage 2 requires some clinical subjectivity; reflected in the greater agreement among US-trained nurses,

Figure 3 Phase II: agreement between triagers on standardised questions by site. Reliability was assessed by three different groups of providers assigning one-two-triage acuity to 63 written cases. The three groups included nurses from Siem Reap Provincial Hospital (SRPH), Mehmot Referral Hospital (MRH) and Stanford University Hospital (SUH). Using Fleiss κ , The SUH nurses showed substantial agreement and the MRH and SRPH nurses showed fair agreement across all scenarios. When considering adult and paediatric scenarios, agreement remained the same.



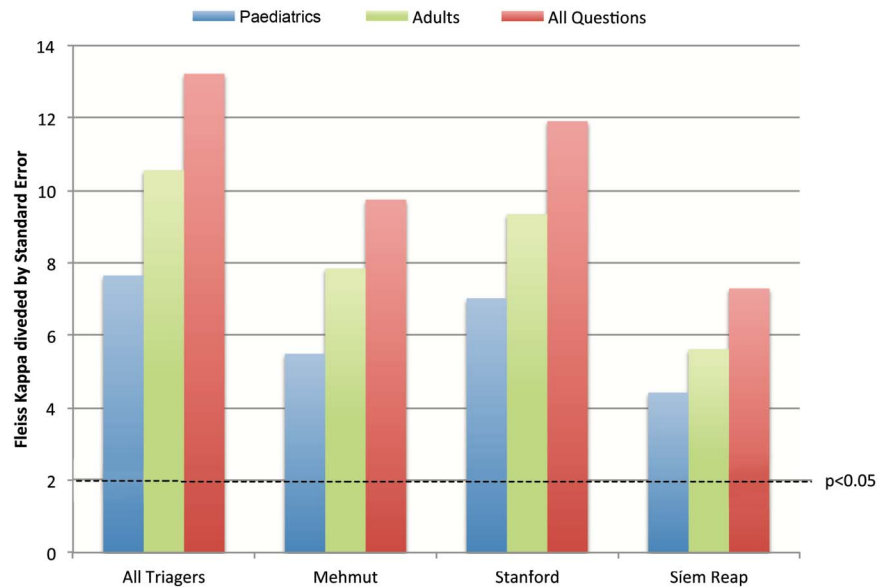
who have more standardised education and clinical experience than the Cambodian nurses.

SATS is a relatively new triage tool created for low-resource settings with some similarities to OTT; it is comprehensive for adults and paediatrics and can be applied by nursing assistants.²¹ However, unlike SATS, OTT does not require a full set of vital signs for complete triage, which necessitates additional time and staff training. The only two vital signs necessary for OTT are HR and oxygen saturation, both rapidly obtained by portable pulse oximeters, which are relatively low cost (~US\$20), durable and require low maintenance. Abnormal oxygen saturation was previously shown to predict the need for ICU admission.²² Additionally, we validated OTT prospectively with authentic emergency patients, whereas SATS was validated solely using clinical vignettes.²¹ While we found OTT to slightly over-triage, SATS has a greater tendency to under-triage higher acuity patients.²³ Over-triaging siphons ‘less-critical’ patients into the ED, potentially overwhelming medical providers or distracting care away from more acute cases; however, under-triaging critical patients has greater immediate implications for patient outcome.²¹ The underlying goal of triage is to provide the greatest good for the greatest number, so the sickest patients should be seen first, even if a few patients are occasionally judged as more acute.

At a recent International Federation of Emergency Medicine meeting held on triage for LMICs, it was noted that the development of new triage scales for LMICs may be ‘too daunting’ given the amount of work invested in existing triage scales.²⁴ In part, research and expansion of triage scales in LMICs have been hindered by the lack of a universal method for assessing triage system performance.²⁵ Our study addresses two important factors in the assessment of triage systems: (1) the gold standard of acuity against which system performance is validated and (2) the relative importance of under-triage versus over-triage.

Triage validity is challenging to determine because an effective triage scale measures the *potential* for deterioration of a patient; however, once a patient is triaged, interventions are implemented to prevent this deterioration. Therefore, we believe real-time

Figure 4 Phase II: test of significance for Fleiss κ . Reliability in phase II of the study was assessed by Fleiss κ . When accounting for SE, all Fleiss κ s demonstrated agreement in the assignment of the four triage categories significantly greater than by chance.



assessment by an experienced EM-trained provider better indicates acuity than patient outcomes.

Our study captures the relative importance of under-triage versus over-triage through an asymmetrically weighted agreement scale. The exact risk of over-triaging and under-triaging is challenging to define as it depends on local factors such as patient volume, facility resources and cultural practices.²¹ Development of a standardised asymmetric weighting scale based on hospital resources by the EM community may result in more robust validation studies in the future.

LIMITATIONS

Our study is limited by (1) the lack of a controlled environment to study triage and (2) the absence of a standard statistic to weigh agreement and determine the clinical relevance of over-triaging versus under-triaging.

Although this study used an accepted gold standard to assess clinical acuity, the Delphi method rather than a single provider may give a stronger consensus opinion, as suggested by Twomey *et al.*²⁶ Further, we assessed validity in the USA rather than in Cambodia due to a lack of available EM experts in Cambodia. While differences in epidemiology and patient characteristics might theoretically impact the applicability of OTT in LMICs, this novel triage system was designed by a panel of EM experts with clinical experience in LMICs for use in such settings. Assessing reliability using written cases is limiting in that all relevant information is provided to the nurses and not dependent on the nurse eliciting information. Further, visual cues that prompt experienced providers to assign acuity based on gestalt are absent. However, our system is created for low-resource settings, where providers may not have experience, and we aggressively attempted to minimise subjective clinical assessment.

Normalising the prevalence of triage levels introduces an artificial bias since clinically equal numbers of patients across all triage levels are not expected. With additional time and resources for sampling, choosing a larger and equal number of patients from each category would be ideal. Normalisation inflates importance of a small sample of patients beyond what it may represent; this inflation favours performance in rare, critically ill patients over a scale that performs best for the larger portion of less sick patients. However, the disproportionate

importance of correctly triaging level 1 patients may offset this limitation.

Lastly, a future study will be needed to assess reliability after implementation in other LMICs.

CONCLUSIONS

In conclusion, OTT is a valid, reliable and cost-effective alternative to ESI in Cambodia. Notwithstanding difficulties in assessment, OTT selected patients needing emergent and urgent care as effectively as ESI, a universally accepted triage scale, while requiring fewer resources. OTT outperformed ESI when dealing with the most critical patients. Moreover, OTT was applied reliably even when used by inexperienced providers in a low-resource setting. With OTT operational in 22 hospitals and pending implementation in 21 more throughout Cambodia, further research on the operational aspects of the scale, such as time-to-triage and adherence to triaging, is needed and forthcoming. While our study provides the first steps of validity and reliability testing for OTT in Cambodia, it will be important to establish that it can be replicated in other LMICs.

Acknowledgements The authors thank Dr Anne Tecklenberg Strehlow for assistance with editing.

Contributors AK, MS, AD and JQ conceived the study and designed the trial. AK and AD supervised the conduct of the trial and the data collection. JW and KS aided in data acquisition. AK provided statistical advice and AD analysed the data. AK drafted the manuscript and MS and SM contributed substantially to its revision. AK takes responsibility for the paper as a whole.

Funding Use of REDCap was supported by grant UL1 RR025744. Translation and local expenses were funded by the US Agency for International Development (USAID) under Better Health Services project, Cooperative Agreement No. 442-A-00-09-00007-00.

Disclaimer The findings of this study are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the US government.

Competing interests None declared.

Ethics approval Stanford University Hospitals Institutional Review Board, Cambodian Ministry of Health. The SUH Institutional Review Board approved both phases of the study (protocol 24500), and the Ministry of Health in Cambodia approved the triage training.

Patient consent Written consent was obtained from all participants in their native language, and Khmer-speaking and English-speaking staff were available for questions.

Data sharing statement Study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at SUH.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- 1 Quereshi N. Triage Systems: a review of the literature with reference to Saudia Arabia. *East Mediterr Health J* 2010;16:690–8.
- 2 Farrohknia N, Castren M, Ehrenberg A, et al. Emergency department triage scales and their components: a systematic review of the scientific evidence. *Scand J Trauma Resusc Emerg Med* 2011;19:42.
- 3 Jacobs B, Ir P, Bigdeli M, et al. Addressing access barriers to health services: an analytical framework for selecting appropriate interventions in low-income Asian countries. *Health Policy Plan* 2012;27:288–300.
- 4 Molyneux E, Ahmad S, Robertson A. Improved triage and emergency care for children reduces inpatient mortality in a resource-constrained setting. *Bull World Health Organ* 2006;84:314–19.
- 5 Cambodia Demographic and Health Survey 2014. 2015; Phnom Penh, Cambodia, 2015.
- 6 Sinuff T, Adhikari NK, Cook DJ, et al. Mortality predictions in the intensive care unit: comparing physicians with scoring systems. *Crit Care Med* 2006;34:878–85.
- 7 Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;41:1140–5.
- 8 Meadow W, Frain L, Ren Y, et al. Serial assessment of mortality in the neonatal intensive care unit by algorithm and intuition: certainty, uncertainty, and informed consent. *Pediatrics* 2002;109:878–86.
- 9 Knaus WA, Harrell FE Jr., Lynn J, et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. *Ann Intern Med* 1995;122:191–203.
- 10 van der Wulp I vSHF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol* 2009;62:1196–201.
- 11 Bergeron S, Gouin S, Bailey B, et al. Comparison of triage assessments among pediatric registered nurses and pediatric emergency physicians. *Acad Emerg Med* 2002;9:1397–401.
- 12 Considine J, Ung L, Thomas S. Clinical decisions using the National Triage Scale: how important is postgraduate education? *Accid Emerg Nurs* 2001;9:101–8.
- 13 Goransson K, Ehrenberg A, Marklund B, et al. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19:432–8.
- 14 Wuerz RC, Travers D, Gilboy N, et al. Implementation and refinement of the emergency severity index. *Acad Emerg Med* 2001;8:170–6.
- 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 Elshove-Bolk J, Mend F, van Rijswijk 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.
- 17 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.
- 18 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.
- 19 Velleman PF, Wilkinson L. Nominal, ordinal, interval, and ratio typologies are misleading. *Am Statistician* 1993;47:1:65–72.
- 20 Fleiss J. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971;76:378–82.
- 21 Twomey M, Wallis LA, Thompson ML, et al. The South African Triage Scale (adult version) provides reliable acuity ratings. *Int Emerg Nurs* 2012;20:142–50.
- 22 Bazerbashi H, Merriman KW, Toale KM, et al. Low tissue oxygen saturation at emergency center triage is predictive of intensive care unit admission. *J Crit Care* 2014;29:775–9.
- 23 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;104:372–5.
- 24 Bullard M. International Federation of Emergency Medicine Triage Special Interest Group Meeting, Hong Kong; 2014.
- 25 Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epidemiol* 2010;63:384–8.
- 26 Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. *Emerg Med J* 2007;24:477–9.