Validity and reliability of the triage scale in older people in a regional emergency department in Hong Kong

Kai Yeung Cheung1 and Ling Pong Leung2

Abstract
Background: Older people (≥65 years) present a unique challenge in emergency department triage. Hong Kong’s Hospital Authority adopts a five-level emergency department triage system, with no special considerations for older people. We evaluated the validity and reliability of this triage scale in older people in a regional Hong Kong emergency department.

Methods: In total, 295 cases stratified by triage category were randomly selected for review from November 2016 to January 2017. Validity was established by comparing the real emergency department patients’ triage category against (1) that of an expert panel and (2) the need for life-saving intervention. Triage notes were extracted to make case scenarios to evaluate inter- and intra-rater reliabilities. Emergency department nurses (n = 8) were randomly selected and grouped into <5 and ≥5 years emergency department experience. All nurses independently rated all 295 scenarios, blinded to clinical outcomes.

Results: The percentage agreement between the real emergency department patients’ triage category and the expert panel’s assignment was 68.5%, with 16.3% and 15.3% over-triage and under-triage, respectively. Quadratic weighting kappa for agreement with the expert panel was 0.72 (95% confidence interval: 0.53–0.91). The sensitivity, specificity and positive likelihood ratio for the need for life-saving interventions were 75.0% (95% confidence interval: 47.6%–92.7%), 97.1% (95% confidence interval: 94.4%–98.8%) and 26.2 (95% confidence interval: 12.5%–54.8%), respectively. The Fleiss kappa value for inter-rater reliability was 0.50 (95% confidence interval: 0.47–0.54) for junior and senior nurse groups, respectively.

Conclusion: The current triage scale demonstrates reasonable validity and reliability for use in our older people. Considerations highlighting the unique characteristics of older people emergency department presentations are recommended.

Keywords
Older people, geriatric, nursing, patient safety, triage

Introduction
Triage is a prioritisation process to identify people who are at risk of adverse clinical outcomes and assign an appropriate level of care.1 In Hong Kong, a five-level emergency department (ED) triage scale is implemented (Table 1).2 Trained nurses perform triage, based on the patients’ presenting conditions and the criteria suggested in triage scales. A valid and reliable triage scale is essential to ensure safety and optimal resource utilisation. Validity of

1 Accident and Emergency Department, United Christian Hospital, Kowloon, Hong Kong
2 Emergency Medicine Unit, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong

Corresponding author:
Kai Yeung Cheung, Accident and Emergency Department, United Christian Hospital, 130 Hip Wo Street, Kwun Tong, Kowloon, Hong Kong.
Email: terranceky@netvigator.com

Creative Commons Non Commercial CC BY-NC. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Table 1. The ED triage scale.

| Triage category | 1 | 2 | 3 | 4 | 5 |
|-----------------|---|---|---|---|---|
| Response time   | Immediate | <15 min | <30 min | Nil | Nil |
| Patient condition Example | Life threatening | Unstable | Major and stable vital signs | Acute and stable vital signs | Minor and stable vital signs |
|                  | Cardiac arrest | Sepsis | Fracture of upper limb | Sprain ankle | Simple respiratory tract infection |

ED: emergency department.
Response time: Recommended time to be assessed by a physician.

an ED triage scale refers to whether it accurately measures patient acuity. Inter-rater and intra-rater reliabilities address the consistency of rating the triage acuity of the same patient by different raters or by the same rater at different time points, respectively. Since there are growing concerns, which suggest that commonly used triage scales fail to appreciate the unique characteristics of older patients, including atypical clinical presentations, a different threshold to define vital sign abnormalities and multiple comorbidities, the triage scale should be valid and reliable for use in older patients.

In Hong Kong, older people represent one of the fastest growing ED populations and account for 14% to 24% of ED visits; the validity and reliability of the triage scale must be addressed when applied to older people. The objective of this study was to evaluate the validity and reliability of the current triage scale (2016 version) in an older people (aged ≥65 years) in a regional ED in Hong Kong.

Methods

Study design and setting

This study was conducted in one regional ED in Hong Kong with more than 1000 in-patient beds. The hospital served the region with a population of more than 600,000, with 17.2% aged ≥65 years in 2016. The annual ED census was 179,251 in 2016 to 2017. About 35% of ED patients were aged ≥65 years. Patients aged ≥65 years have a higher triage acuity (triage categories 1–3) when compared to age <65 years (61.8% vs. 32.7%).

The study was conducted in two phases. In the first phase, a review of real ED patients’ records was conducted to (1) construct a set of paper-based triage scenario and (2) assess the validity of the scale. To construct a set of paper-based triage scenario, all medical records for older people (≥65 years old) presented to the ED between 1 November 2016 and 31 January 2017 for care were retrieved from the electronic databases of the Hospital Authority, Hong Kong. This period was chosen as the triage guideline was updated in August 2016, which allowed a 3-month interval for ED nurses to be familiar with this guideline before the study began. We excluded patients who (1) had a scheduled follow-up, (2) discharged against medical advice, (3) left without being evaluated, (4) died before arrival and (5) had incomplete/no records available. To account for the uneven prevalence among each triage category and avoid overrepresentation of semi-urgent and non-urgent triage categories in our sample, we used a proportional stratified sampling method. The proportional sample sizes selected were triage category 1 (n = 6, 2.0%), category 2 (n = 14, 5.0%), category 3 (n = 165, 56.0%), category 4 (n = 102, 35.0%) and category 5 (n = 8, 3.0%). All eligible patients who met initial criteria were then chosen randomly until the pre-specified sample size was achieved in each triage category. Given that the \( z_{\alpha} = 1.96 \) at a 95% confidence interval (CI) and the width of CI = 0.1, we determined that a minimum of 275 patients were required in total. This was based on a previous local study, which validated the 2012 version ED triage guideline in Hong Kong. In that study, the observed agreement and chance agreement between the expert panel and the real ED patients’ triage assignment (i.e. triage category that the patient was assigned for that indexed ED visit) for all triage levels were 0.78 and 0.51, respectively. A standard data collection form was designed to collect patient-related data. It included patient characteristics, presentation-related variables and outcome variables. Each term and the coding methods were defined. The exact triage descriptions were used to construct paper-based triage scenarios without modification. If patients were re-triaged, the findings from the re-triage were used. A research assistant, blinded to the study objective, was responsible for data extraction. The researcher extracted 10% of patient records independently to double-check the correctness with the research assistant and had a percentage agreement of 96.7% (551 of 570 items were in agreement). Discrepancies were resolved by discussion and agreement.

To evaluate the validity, first, we compared the real ED patients’ triage assignment with the ratings assigned by an expert panel. The expert panel consisted of an ED physician specialist and an ED advanced practice nurse. They were invited because they had more than 5 years of triage practice review and experience in contributing to the annual ED triage audit for the Hospital Authority. They independently rated each case based on the information provided in
the written scenario. They were also blinded to other clinical information such as mortality and the real ED patients’ triage assignment. Finally, they discussed until a final triage rating was agreed upon. This triage rating became the reference standard. Second, the admission rate, ED and inpatient mortality in each real ED patients’ triage category were analysed. Third, we determined the ability of the triage scale to identify patients in categories 1 and 2 and whether they received any life-saving interventions. We referred to the Agency for Healthcare Research and Quality definition of life-saving interventions. Examples include intubation and haemodynamic support. We also considered high-concentration oxygen therapy and administration of a thrombolytic agent for acute ST segment-elevation myocardial infarction and acute ischaemic stroke as life-saving interventions, and we defined the time to receive any of these interventions as within 1 h of ED arrival.

The aims of the second phase were to evaluate the inter-rater and intra-rater reliability of the scale. Except the nursing managers and the advanced practice nurse serving as a member of the expert panel, all ED nurses who received formal triage training were eligible for participation. They had a unique number and were first stratified into senior and junior ED nurse groups, using 5 years of ED working experience as a cut-off. In each group, four participants were selected randomly. A total of eight triage nurse participants were required. These eight nurse participants were then invited to participate in the study, with the provision of written information regarding the study objectives and procedures. Written informed consent was obtained from them. Their professional qualifications and work experience details were collected, which included total nursing working experience since registration, triage experience, job title and specialty-related training. The same set of paper-based triage scenarios was then given to them to rate each patient’s triage category individually. Similarly, they were blinded to other clinical information and the results of others. For practical reason, one participant from each group was invited to rate the same set of scenarios after a 1-month interval. Inter-rater and intra-rater reliabilities were evaluated using agreement statistics. We followed the Guidelines for Reporting Reliability and Agreement Studies in reporting this work. The kappa statistic (and the strength of agreement) was described as follows: 0.81–1.00 (almost perfect), 0.61–0.80 (substantial), 0.41–0.60 (moderate), 0.21–0.40 (fair), 0.00–0.20 (slight) and <0.00 (poor). All data were analysed using Medcalc (https://www.medcalc.org/index.php), VassarStats (http://vassarstats.net) or StatTools, Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong & New Territories East Cluster (http://www.obg.cuhk.edu.hk/ResearchSupport/StatTools/ResourceIndex_Subjects.php).

### Results

Figure 1 showed a diagram of the patient selection process. Table 2 summarised the patients’ characteristics, showing that the majority of patients (>90%) were assigned to triage categories 3 and 4, and 50% of the patients were admitted. The total nursing working experience since registration and ED triage experiences were 13.5 years (standard deviation (SD) = 7.3) and 10.3 years (SD = 8.3) in the senior nurse group and 5.3 years (SD = 1.0) and 3.5 years (SD = 1.0) in the junior nurse group. There was 1 Nursing Officer in the senior nurse group. They also received specialty training in emergency nursing. All participants in the junior nurse group were Registered Nurses and only one had undergone...
specialty training in emergency nursing. In both groups, none had received structured, specialised training in geriatric emergency care.

In this cohort, all patients assigned to triage categories 1 and 2 were admitted; 15.0% (3/20) were admitted to critical care unit. The admission rate decreased to 69.0% in triage category 3 and 14.0% in triage category 4. No patients were admitted from triage category 5. The overall in-patient mortality was 4.1% (12/295), which included 16.7% (1/6), 21.4% (3/14) and 4.8% (8/165) in triage categories 1 to 3, respectively. No patients required re-triage, and there were no deaths in the ED.

Table 3 shows a comparison of the real ED patients’ triage categories versus the expert panel triage categories. The percentage agreement between the real ED patients’ triage category and the expert panel opinion was 68.5%. Over-triage was defined as the triage ratings in the real ED patients’ triage categories that were higher than the expert panel and was 16.3%. The quadratic weighted kappa between the real ED patients’ triage assignment and the expert panel for agreement was 0.72 (95% CI: 0.53–0.91), reflecting a substantial agreement.

Table 4 shows the number of patients who did and did not receive life-saving interventions based on triage category. The metrics of the ability to identify the need to receive life-saving interventions were (1) sensitivity (75.0%, 95% CI: 47.6%–92.7%), (2) specificity (97.1%, 95% CI: 94.4%–98.8%), (3) positive predictive value (60.0%, 95% CI: 41.7%–75.9%), (4) negative predictive value (98.6%, 95% CI: 96.7%–99.4%), (5) positive likelihood ratio (26.2, 95% CI: 12.5–54.8) and (6) negative likelihood ratio (0.26, 95% CI: 0.11–0.60). The Fleiss kappa values for inter-rater reliability were 0.51 (95% CI: 0.48–0.50) and 0.50 (95% CI: 0.47–0.54) for the junior and senior nurse groups, respectively, reflecting a moderate agreement; this was independent of their ED experience. The intra-rater reliability was substantial (quadratic weighted k = 0.76, 95% CI: 0.58–0.95) and almost perfect (quadratic weighted k = 0.96, 95% CI: 0.87–1.00) in the senior and junior nurse groups, respectively.

Discussion
The results of this study demonstrate that the current ED triage scale has a reasonable validity and reliability for use
in the older people. At present, different strategies exist to evaluate the validity and reliability of triage scales, and each has its strengths and limitations. The assessment of validity by establishing the association between triage categories and objective outcomes such as disposition status, resource utilisation and mortality is commonly used. Studies have examined the Emergency Severity Index (ESI) and the Canadian Triage Acuity Scale (CTAS) and found a significant association between a higher triage category (i.e. more urgent) and these measures in older people, which partly reflected the severity of the patients’ condition. A similar pattern was observed in our study. In particular, the ability of a triage scale to recognise a need for life-saving interventions is a surrogate marker to reflect both clinical urgency and how resources are utilised at a critical moment. In published studies, the sensitivity ranged from 42% to 97.9% and the specificity ranged from 89.2% to 99.8%. Our sensitivity (75.0%, 95% CI: 47.6%–92.7%) is modest by comparison and fairly sensitive, but we modified the list of life-saving interventions to include more critical interventions in our clinical practice, which may affect the sensitivity to some degree. Also, the wide CI indicates our result is imprecise and is likely due to small sample size. It should be interpreted with caution. Our specificity (97.1%, 95% CI: 94.4%–98.8%) and negative predictive value (98.6%, 95% CI: 96.7%–99.4%) were high. It suggested that our scale has a good property to identify who do not need life-saving interventions. However, using the need of life-saving intervention as an outcome measure can be confounded by clinical indications, institutional practice, patients’ and clinicians’ preferences and the availability of resources.

We also reported the results of the agreement between real ED patients’ triage assignment and expert opinions. The kappa value ranged from 0.61 to 0.76 in published studies, which is similar to the results of the present study. While some authors considered use of an expert panel as a preferred assessment for acuity rating, there are no universally accepted recommendations regarding the composition and optimal number of experts in the panel. For our expert panel, we recruited one ED physician specialist and one advanced practice nurse with experience in triage practice and audit. This composition was used in a previous study. However, it is worth mentioning that their agreement was moderate with a wide CI (quadratic weighted kappa 0.69, 95% CI: 0.47–0.91) before they achieved consensus. This suggests that a considerable variation exists, and that triage can be challenging even for experienced clinicians. Given that geriatric emergency care is an emerging ED specialty and many clinicians may not have specialised training in this field, an inter-disciplinary approach to scale development and validation with geriatric clinicians’ input may be beneficial and warrant further investigation. This may provide a more comprehensive view on the focused assessment and interpretation of findings from a geriatric perspective.

Intra-rater reliability has not been addressed in previous triage studies specific to the older people, but results from our study suggested a very good intra-rater reliability. Compared to junior nurse group, senior nurse group had a lower intra-rater reliability. Also, inter-rater reliabilities were only moderate in both triage nurse groups. It has been suggested that triage raters may modify the application of triage criteria or have different interpretations of triage criteria as they become more confident in their use, resulting in a lower reliability.

Our findings showed similar validity and reliability results as a previous local study. In a retrospective review of 100 patients from a general population, Fan and Leung reported that the inter-rater reliabilities ranged from 0.51 to 0.54 (unweighted kappa, re-calculation based on published data) and the quadratic weighted kappa was 0.77 for validity. Although there are substantial differences in the study population, setting and expert panel composition, the findings are consistent and suggest that the scale has a similar performance in older people. The over-triage rate is higher in our study than in theirs (7%), which may result in the distribution of scarce ED resources to those who do not need it. One possible interpretation for this phenomenon is related to the common practice of nurses to err ‘on the safe side’, thus over-triaging older people. The consequences of this have not been addressed frequently in published studies and should not be ignored. However, since the findings were not corrected by chance from a statistical perspective, the results should be interpreted with caution.

**Limitations**

This study has some limitations. First, triage ratings were based on reviewing medical records, in which the interactions and other non-verbal cues between participants (triage nurses) and patients are lacking. The format of triage assessment (based on a collection of paper-based triage scenarios) is unfamiliar for some nurses, which might have been challenging for them. The results of this study are limited to a single setting, where institutional practices may have potential influences on how clinicians assess and manage patients. Since sample size calculation was based on the agreement between the real ED patients’ triage assignment and the expert panel rather than the ability of the triage scale to identify the need of receiving life-saving interventions, a very small sample size was obtained with regard to this evaluation. This is characterised by a very wide CI. Therefore, results should be viewed as preliminary. Despite this, we reported detailed descriptions of the patients’ and nurses’ characteristics to allow readers a thoughtful assessment of the findings and their relevance to individual settings.
Future studies may systematically explore the characteristics of patients that were incorrectly triaged. This will increase clinicians’ awareness about the potential limitations in triage application. Also, in the context of trauma, low-energy trauma (e.g. fall <2 m), Glasgow Coma Scale and blood pressure may not adequately reflect the severity of injury in older people.\textsuperscript{22–24} These history findings and physical parameters may be incorporated in future triage scales to guide decision making.

**Conclusion**

This study suggested that the current triage scale demonstrates reasonable validity and reliability for use in the older people. Further work to improve its application in this population is recommended.

**Acknowledgements**

I would like to thank all doctors and nurses of the Accident and Emergency Department, United Christian Hospital, for their assistance. Finally, I would like to extend my heartfelt gratitude to Miss Lui Po Yee for her dedication, time and trust throughout the study.

**Author contributions**

CKY/LPL helped in study design. CKY helped in data collection and writing up the manuscript.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Availability of data and materials**

Data will be provided upon written request and availability.

**Ethical approval**

The study was approved by the Hospital Authority Research Ethics Committee (KC/KE-17-0037/ER-2).

**Informed consent**

Written consent was sought from nursing participants.

**Human rights**

The principles of autonomy and confidentiality were followed.

**ORCID iDs**

Kai Yeung Cheung \(\text{https://orcid.org/0000-0002-4386-1682}\)

Ling Pong Leung \(\text{https://orcid.org/0000-0003-0219-3814}\)

**References**

1. 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(1): 39–50; quiz118.
2. Task Group on Best Practice, COC (A&E) Nursing Development Subcommittee Hospital Authority. *A&E triage guidelines*. Hong Kong: Hospital Authority, 2016.
3. Twomey M, Wallis LA and Myers JE. Limitations in validating emergency department triage scales. *Emerg Med J* 2007; 24(7): 477–479.
4. Grossmann FF and Nickel C. Triage of older ED patients. In: Nickel C, Bellou A and Conroy S (eds) *Geriatric emergency medicine*. Basel: Springer, 2018, pp.17–22.
5. Lamantia MA, Stewart PW, Platts-Mills TF, et al. Predictive value of initial triage vital signs for critically ill older adults. *West J Emerg Med* 2013; 14(5): 453–460.
6. Yip WL, Fan KL, Lui CT, et al. Utilization of the accident & emergency departments by Chinese elderly in Hong Kong. *World J. Emerg. Med* 2015; 6: 283–288.
7. Yim VWT, Graham CA and Rainer TH. A comparison of emergency department utilization by elderly and younger adult patients presenting to three hospitals in Hong Kong. *Int J Emerg Med* 2009; 2(1): 19–24.
8. Census and Statistics Department. Population By-census, 2016, https://www.bycensus2016.gov.hk/en/copyright.html?disclaimer
9. Hospital Authority. 2016-2017 hospital authority statistical report. Hong Kong: Hospital Authority, 2018.
10. Khan KS and Chien PFW. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001; 108: 562–567.
11. Norman GR and Streiner DL. *Biostatistics: the bare essentials*. 4th ed. Shelton, CT: People’s Medical Publishing House, 2014.
12. Fan MMW and Leung LP. Validation of the Hong Kong accident and emergency triage guidelines. *Hong Kong Med J* 2013; 19(3): 198–202.
13. Gilboy N, Tanabe P, Travers DA, et al. Emergency Severity Index, Version 4: implementation handbook. AHRQ Publication No. 05-0046-2, May 2005. Rockville, MD: Agency for Healthcare Research and Quality, https://www.ahrq.gov/research/esi/ (accessed 20 July 2018).
14. Platts-Mills TF, Travers D, Biese K, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elderly emergency department patients receiving an immediate Life-saving intervention. *Acad Emerg Med* 2010; 17(3): 238–243.
15. Kottner J, Audigé L, Borson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96–106.
16. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159–174.
17. Lentz BA, Jenson A, Hinson JS, et al. Validity of ED: addressing heterogeneous definitions of over-triage and under-triage. *Am J Emerg Med* 2017; 35(7): 1023–1025.
18. 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(3): 317–325.
19. 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.

20. Grossmann FF, Bingisser R and Nickel CH. Comment on the validity of emergency department triage tools. Am J Emerg Med 2017; 35(9): 1376.

21. Dallaire C, Poitras J, Aubin K, et al. Emergency department triage: do experienced nurses agree on triage scores? J Emerg Med 2012; 42: 736–740.

22. Kehoe A, Rennie S and Smith JE. Glasgow Coma Scale is unreliable for the prediction of severe head injury in elderly trauma patients. EMJ 2015; 32: 613–615.

23. Kehoe A, Smith JE, Edwards A, et al. The changing face of major trauma in the UK. EMJ 2015; 32: 911–915.

24. Brown JB, Gestring ML, Forsythe RM, et al. Systolic blood pressure criteria in the National trauma triage protocol for geriatric trauma: 110 is the new 90. J Trauma Acute Care Surg 2015; 78(2): 352–359.