Validation of the Hong Kong Accident and Emergency Triage Guidelines

Mandy MW Fan 范文慧
LP Leung 梁令邦

Objective   To validate the Hong Kong Accident and Emergency Triage guidelines.

Design   Retrospective chart review.

Setting   The Accident and Emergency Department of a tertiary hospital in Hong Kong.

Participants   Patients who attended the Accident and Emergency Department on one day in February 2012.

Main outcome measures   The inter-rater reliability in two pairs of nurses grouped according to experience and validity as compared with an expert panel.

Results   Of the 100 patients recruited and triaged into levels 1 to 5, the weighted kappa coefficient (inter-rater reliability) for the two pairs of nurses was 0.699 and 0.717, respectively. The weighted kappa coefficient for validity was 0.766. When only patients in triage levels 3 and 4 were included, the weighted kappa coefficient for reliability dropped to 0.632 and 0.585, respectively. The weighted kappa coefficient for validity also decreased to 0.558.

Conclusions   The overall inter-rater reliability and validity of the Guidelines appeared acceptable. Further revision of the Guidelines on triaging patients to levels 3 or 4 is probably necessary.

New knowledge added by this study

- The overall inter-rater reliability and validity of the Hong Kong Accident and Emergency Triage Guidelines are substantial.
- The inter-rater reliability of the Guidelines is independent of the experience of the triage nurses.
- On applying to triage level 3 and 4 patients, who make up the main bulk of attendance to the Accident and Emergency Department, the Guidelines are less reliable and the validity also reduced.

Implications for clinical practice or policy

- Further revision of the Guidelines targeting triage level 3 and 4 patients is required.
- A well-written guideline is more important than the experience of the triage nurses.

Introduction

Triage is an integral component of an emergency department (ED). It is established to tackle the mismatch between limited resources and potentially unlimited demand for medical services. This is especially important as hospital ED overcrowding has become a burden worldwide.[1]

Different triage systems have emerged over the years, from the initial 3-level and 4-level systems, to the 5-level scales that are now widely accepted. The Emergency Severity Index (ESI) in the United States, the Canadian Triage and Acuity Scale (CTAS) in Canada, the Manchester Triage System (MTS) in the United Kingdom, Portugal and the Netherlands, and the Australian Triage Scale (ATS) in Australia are all 5-level triage scales currently being employed in established EDs. The Hong Kong Accident and Emergency Triage Guidelines (HKAETG) were first introduced in 1999. After multiple revisions by the Accident & Emergency (A&E) nursing development subcommittee of the Hospital Authority (HA), the fourth version evolved in 2011 and was implemented from January 2012. These guidelines also rely on a 5-level triage scale that aims to determine the priority, intensity,

Accident and Emergency Department, Queen Mary Hospital, Pokfulam, Hong Kong
MMW Fan, MB, BS, MCEM
LP Leung, MB, BS, FHKAM (Emergency Medicine)

Correspondence to: Dr MMW Fan
Email: mandymwfan@gmail.com

and place of care.[2] The target physician response time is linked to the first three levels of triage that serves as a performance pledge by the HA (Table 1). In Hong Kong, triage is performed by A&E registered nurses. Eligibility to carry out this task includes a minimum of 1 year's A&E experience and having attended structured triage training. To which triage level a patient belongs depends on the triage nurse's assessment of the severity of patient's presenting condition and the stability of vital signs (Table 2). This kind of assessment is not symptom-specific and its accuracy relies on the triage nurse's diagnostic skills. The expected time interval from A&E registration of a patient to being seen by the triage nurse is less than 10 minutes. Therefore, depending on the patient load, there is often a limit to the time spent on the triage process.

Evaluation of a triage scale involves assessing its reliability and validity. Most studies on this topic were about the ESI, CTAS, MTS, or ATS. However, a systematic review on ED triage scales revealed that most were not supported by strong scientific evidence with regard to inter-rater agreement and ability to predict hospitalisation and in-hospital mortality.[3] For the HKAETG, there have been no published reports on its reliability and validity despite its being in use for 13 years. Thus, this study aimed at assessing its reliability and validity when applied to the local population, with a view to establishing a reference for its future refinement and revision.

驗證香港急症室分流服務的指引

目的	驗證香港急症室分流服務的指引。
設計	回顧審查。
安排	香港一所提供第三層醫療服務醫院的急症室。
參與者	於2012年2月其中一天使用急症室的病人。
主要結果測量	按經驗分為兩組護士,然後比較這兩組護士的評分者間信度;再比較兩組護士與一專家小組的一致性效度。
結果	共100名病人被納入研究範圍,他們均獲分流至第一至第五級別,兩組護士的加權kappa值(即評分者間信度)為0.699及0.717,而一致性效度的加權kappa值為0.766。如果只包括第三至第四級別的病人,評分者間信度即降至0.632及0.585,而一致性效度的加權kappa值則下跌至0.558。
結論	分流服務指引的總評分者間信度和效度尚可接受,而把病人分流至第三或第四級別的指引可能有需要進一步修訂。

Methods

This study was a retrospective chart review of the triage records of a random sample of patients attending the A&E of a tertiary hospital and teaching institution. About 300 to 400 patients attend this A&E on a daily basis. Each patient is triaged by a designated

TABLE 1. The Hong Kong Accident and Emergency Triage Guidelines[2]

Triage category	Priority of care (% of patients required to fulfil the target response time)	Intensity of care	Place of care
1 (Critical)	Immediate medical attention (100%)	Team approach by physicians and nurses	Resuscitation area
2 (Emergent)	Medical attention within 15 minutes (95%)	Immediate continuous close monitoring	Resuscitation area or treatment cubicle
3 (Urgent)	Medical attention within 30 minutes (90%)	Nurse-initiated intervention as necessary	Cubicle
4 (Semi-urgent)	-	-	Cubicle or walk-in clinic
5 (Non-urgent)	-	-	Walk-in clinic

TABLE 2. Definition of the triage categories[2]

Triage category	Definition
1 (Critical)	A life-threatening condition caused by a major event Unstable vital signs
2 (Emergent)	A potentially life-threatening condition Borderline vital signs but with potential risk of rapid deterioration
3 (Urgent)	A major condition with potential risk of deterioration Vital signs stable
4 (Semi-urgent)	An acute but stable condition with minimal immediate risk of serious complication Vital signs stable
5 (Non-urgent)	A minor and stable condition with no serious complication expected Vital signs stable

nurse. Patients were recruited by proportional random sampling from all patients attending the A&E from 00:01 to 23:59 on one day in February 2012 using a random number table. All patients were triaged before random sampling and numbered according to the A&E number. The proportion was based on the patient distribution of each triage category on that day. Proportional random sampling was adopted because each triage category could be considered an individual non-overlapping stratum and patient distribution in each category was uneven, with patients in triage level 4 being the majority. This method of sampling is better than simple random sampling in that it reduces the possibility of over-representation of triage level 4 patients in the studied sample. The triage records of recruited patients were presented in a standardised abstract format. Each abstract contained the patient's age, gender, presenting complaints, and the data of vital signs routinely measured by the triage nurse. The length of each abstract was restricted to 50 words or less. All abstracts were prepared by one of the investigators.

For evaluating reliability, four nurses, who had been trained on the triage guidelines, were recruited. They were grouped in two pairs according to their A&E experience (nurse A and B: 5-10 years, and nurse C and D: 15-20 years). They were presented with the same set of abstracts and asked to assign a triage level to each patient according to the data provided. The time for the decision process of each patient was restricted to 2 minutes or less. All four nurses were blinded to each other's assessment. Their decisions on triage assignments were then compared within the groups. For evaluating validity, the consensus of an expert panel on each patient's triage category was taken as the gold standard. The panel was made up of two senior nurses. Both had over 20 years of clinical and administrative A&E experience and had contributed to the production of the triage guidelines. They reviewed the same set of abstracts, read them together, and arrived at a conclusion together. They were blinded to the actual triage assignment of each patient. Their consensus was then compared with the original triage assignment.

Statistical analysis

The demographic characteristics of the recruited patients were analysed with descriptive statistics. The inter-rater reliability of the two pairs of nurses was expressed as the quadratically weighted kappa, which is better than the normal kappa when the level of agreement in a scale containing different categories is measured.[4] Validity was measured as the level of agreement between the expert panel and the original triage record and was expressed as a quadratically weighted kappa. The use of kappa statistics as a measure of validity is recommended if

the measurement of the test and reference standard are ranked.[5] As for the sample size, in comparing the agreement between two observers, with a kappa of >0.4 and $P<0.05$ as statistically significant, a minimum of 54 subjects was required.[6]

Results

One hundred patients were recruited from the 420 attendances on 15 February 2012. Their ages ranged from 24 days to 103 years (mean, 51.9 years). Gender distribution was approximately equal (male:female ratio=49:51). The distribution by triage category was level 1: 2%, level 2: 2%, level 3: 26%, level 4: 65%, and level 5: 5%. For distribution by specialty of the presenting complaints, 44% belonged to Internal Medicine, 16% to Orthopaedics, 14% to Surgery, 9% to Paediatrics, 8% to Gynaecology, 3% to Ear, Nose and Throat, 3% to Psychiatry, 2% to Neurosurgery, and 1% to Ophthalmology.

The weighted kappa coefficient (inter-rater reliability) for the two pairs of nurses was 0.699 and 0.717, respectively. The strength of agreement, according to the classification by Landis and Koch,[7] was substantial. There was, however, no significant difference between the two groups as their 95% confidence intervals overlapped (Tables 3a and 4a).

TABLE 3. Inter-rater agreement between nurse A and B (a) for all triage levels and (b) for triage levels 3 and 4

(a) Triage levels 1-5

Triage level		Nurse A					
		1	2	3	4	5	Total
Nurse B	1	0	1	0	0	0	1
	2	0	1	0	0	0	1
	3	1	3	22	6	0	32
	4	0	1	8	42	1	52
	5	0	0	0	7	7	14
	Total	1	6	30	55	8	100

Quadratically weighted kappa = 0.699

95% Confidence interval = 0.591-0.807

(b) Triage levels 3 and 4 only

Triage level		Nurse A					
		1	2	3	4	5	Total
Nurse B	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
	3	0	2	22	6	0	30
	4	0	1	8	41	1	51
	5	0	0	0	6	4	10
	Total	0	3	30	53	5	91

Quadratically weighted kappa = 0.632

95% Confidence interval = 0.495-0.769

TABLE 4. Inter-rater agreement between nurse C and D (a) for all triage levels and (b) for triage levels 3 and 4

(a) Triage levels 1-5

Triage level		Nurse C					
		1	2	3	4	5	Total
Nurse D	1	1	1	0	0	0	2
	2	0	1	0	1	0	2
	3	0	0	16	7	0	23
	4	0	0	5	49	9	63
	5	0	0	0	4	6	10
	Total	1	2	21	61	15	100

Quadratically weighted kappa = 0.717

95% Confidence interval = 0.579-0.855

(b) Triage levels 3 and 4 only

Triage level		Nurse C					
		1	2	3	4	5	Total
Nurse D	1	0	0	0	0	0	0
	2	0	0	0	1	0	1
	3	0	0	16	7	0	23
	4	0	0	4	48	7	59
	5	0	0	0	4	4	8
	Total	0	0	20	60	11	91

Quadratically weighted kappa = 0.585

95% Confidence interval = 0.431-0.738

TABLE 5. Agreement between the expert panel and actual triage assignment (a) for all triage levels and (b) for triage levels 3 and 4

(a) Triage levels 1-5

Triage level		Expert panel					
		1	2	3	4	5	Total
Actual	1	2	0	0	0	0	2
	2	1	1	0	0	0	2
	3	0	0	19	7	0	26
	4	0	0	10	55	0	65
	5	0	0	0	4	1	5
	Total	3	1	29	66	1	100

Quadratically weighted kappa = 0.766

95% Confidence interval = 0.632-0.899

(b) Triage levels 3 and 4 only

Triage level		Expert panel					
		1	2	3	4	5	Total
Actual	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
	3	0	0	19	7	0	26
	4	0	0	10	55	0	65
	5	0	0	0	0	0	0
	Total	0	0	29	62	0	91

Quadratically weighted kappa = 0.558

95% Confidence interval = 0.372-0.743

The weighted kappa coefficient for validity was 0.766 (Table 5a). The strength of agreement was substantial.[7] When only patients originally triaged to level 3 and 4 were measured, the weighted kappa coefficient for the inter-rater reliability of nurse A, B and nurse C, D dropped to 0.632 and 0.585, respectively (Tables 3b and 4b). The weighted kappa coefficient for validity also decreased to 0.558 (Table 5b). The overall accuracy was 78%, with 15% being under-triaged and 7% of being over-triaged.

Discussion

Reliability is an important attribute of a triage tool. As a patient's triage category determines the priority and intensity of care the patient receives, it has a direct bearing on patient safety. Ideally, different triage nurses should assign the same triage level to the same patient. This study showed that the inter-rater reliability of the HKAETG was substantial, irrespective of the experience of the nurses. Reliability studies of other triage scales have shown variable results. According to one review, the weighted kappa coefficient ranged from 0.34 to 0.84.[8] However, direct comparison between different triage scales is impossible, because most of these scales lack external validation. To be reliable, a triage

scale should be simple to use and unambiguous to interpret, while diverse enough to match the wide spectrum of patient presentation in the ED. The HKAETG is basically a two-step process. Primary triage is a rapid screening for conditions requiring resuscitation by assessing the chief complaint and the airway, breathing, and circulation. Secondary triage focuses on assessing an apparently stable patient by the history, vital signs, and simple bedside investigations like oxygen saturation. Obviously, triaging an ill patient with overt instability to level 1 or 2 should not be difficult. Triaging a patient to level 5 is also straightforward. Problems may arise during the identification of a potentially critical patient, who should belong to level 3, from other stable patients as both groups have stable vital signs. This is reflected by a fall in the degree of agreement when only level 3 and 4 patients were studied. The fall is particularly obvious for nurse C and D where the agreement dropped from substantial to moderate.[7] This has important clinical implications as the triage to physician interval for triage level 3 patients should not exceed 30 minutes, whereas it may be hours for the level 4 patients.

Assessing the validity of a triage scale is difficult. An ideal scale should measure the true patient acuity.

This is not possible, however, because of a multitude of factors influencing a patient from presentation to discharge.[9] Certain surrogate measures, eg admission rate, resource utilisation and mortality rates, have been used. All of these have limitations. For instance, 29% of acute admissions were judged inappropriate in a local study.[10] With such a degree of inappropriateness, the admission rate cannot be considered a valid measure of patient acuity. In this study, expert consensus was taken as the gold standard. This can be viewed as a kind of construct validity and eliminates the many confounders to the surrogate measures already mentioned.[11] The HKAETG demonstrated substantial agreement between the expert panel and triage nurses. This suggests that in general, the HKAETG is a valid instrument for measuring patient acuity in the local setting. However, if the measurement is restricted to patients of level 3 and 4, who make up the major proportion of A&E attendances, the agreement is only moderate. This would appear to cast doubt on its actual ability to reflect patient acuity.

This study showed that 15% and 7% of patients were under-triaged and over-triaged, respectively. Neither under-triage nor over-triage is desirable. The former can delay necessary treatment, while the latter may subject a patient to unnecessary interventions.

Regarding limitations of this study, how patient data are presented may affect the performance of the triage nurses. At best, recourse to written cases can only partly mirror the actual patient encounter in the ED. A study of the CTAS showed that inter-rater reliability improved with the use of interactive computerised vignettes.[12] For validity, the use of expert consensus as the reference also has its drawbacks. It lacks objectivity because there are no explicit criteria of validity. The recruitment of different experts may give different results for validity. Another limitation is that findings from this study are only applicable to the local population. Validation of the HKAETG is not complete without external validation. This study was done in a single institution. A territory-wide study may give more representative results. As this study was done on patients attending the ED on a single day, it might overlook the influence on triage performance by factors like patient spectrum, workload, and the presence of an epidemic. Finally, ways to improve accuracy have to be found. Factors leading to reduced reliability and validity when managing level 3 and 4 patients also warrant further investigation.

Conclusions

The overall inter-rater reliability and validity of the HKAETG are acceptable. On the other hand, for the level 3 and 4 patients, who constitute almost 75% of attendances of the local A&E,[13] the inter-rater reliability and validity are only moderate. It suggests that further refinement and revision of the scale is necessary.

References

1. Hoot NR, Aronsky D. Systemic review of emergency department crowding: causes, effects and solutions. Ann Emerg Med 2008;52:126-36. **cross**ref

2. Task Group on Best Practice, A&E Nursing Development Subcommittee. A&E Triage Guidelines Version 4. Jun 2011.

3. Farrohknia N, Castrén 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. **cross**ref

4. van der Wulp I, van Stel HF. Calculating kappas from adjusted data improved the comparability of the reliability of triage systems: a comparative study. J Clin Epidemiol 2010;63:1256-63. **cross**ref

5. Chien PF, Khan KS. Evaluation of a clinical test II: assessment of validity. BJOG 2001;108:568-72. **cross**ref

6. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation. Stat Med 1992;11:1511-9. **cross**ref

7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74. **cross**ref

8. FitzGerald G, Jelinek GA, Scott D, Gerdtz MF. Emergency department triage revisited. Emerg Med J 2010;27:86-92. **cross**ref

9. 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. **cross**ref

10. Leung LP, Cheng YW, Fan KL. Evaluation of the appropriateness of acute hospitalizations in Hong Kong. Hong Kong J Emerg Med 2011;18:277-81.

11. Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. Emerg Med J 2007;24:477-9. **cross**ref

12. Dong SL, Bullard MJ, Meurer DP, et al. Reliability of computerized emergency triage. Acad Emerg Med 2006;13:269-75. **cross**ref

13. Hospital Authority. Annual plan 2012-2013. Hong Kong: Hospital Authority; 2012.