J Clin Res Gov 3 (2014) 124-133

Swedish Science Pioneers
Developing World Journal Series
Journal of Clinical Research & Governance
www.jcrg.sciencepub.se

Review Article

Kappa Agreement of Emergency Department Triage Scales: A Systematic Review and Meta-Analysis

Faramarz Pourasghar[a], Jafar Sadegh Tabrizi[b], Parvin Sarbakhsh[c], Amin Daemi[d]

a: Road Traffic Injury Research Center, Department of Medical Informatics, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

b: Tabriz Health Services Management Research Center and Department of Health Services Management, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

c: Department of Statistics and Epidemiology, School of Health, Tabriz University of Medical Sciences, Tabriz, Iran

d: Department of Health Services Management and Student Research Committee, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

Correspondence

Amin Daemi, Department of Health Services Management, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran.
Tell: +984113352291
Email: Amin_daemi@yahoo.com

Keywords:

Emergency Department
Triage
Inter-rater Reliability
Inter-observer Agreement
Kappa

Received: 2013-11-10
Accepted: 2014-04-03

DOI: 10.13183/jcrg.v3i2.78

Abstract

Purpose: The aim of this review was to compare the reliability of existing triage scales. The research question was that, using kappa value as reliability measure, in adult ED patients to what extent did clinicians' triage ratings agree with eachother or with a gold standard?

Methods: The search was limited to studies on adult patients (≥15 years) visiting EDs for somatic reasons and to English and Persian studies that have reported kappa. Studies solely on geriatric patients, computerized triage, telephone triage, and specialty triages were excluded. PubMed, Scopus, ISI web of knowledge, Mosby's Nursing Index, Magiran and SID, were systematically searched through August 2013. 13 articles for kappa (n=38829), 11 articles for linear weighted kappa (LWK) (n= 22697) and 12 articles for quadratic weighted kappa (QWK) (n=10915) met our inclusion criteria. Heterogeneity among studies and publication bias were checked. All pooled analyses of kappa scores for each triage scale were calculated based on random-effects model and then the scales were compared.

Results: Findings suggest that five-level scales are more reliable in triaging patients in the emergency department than others (pooled kappa: 0.53, 95% CI (0.48, 0.57), LWK= 0.76, 95% CI (0.72, 0.81) and QWK = 0.74, 95% CI (0.69, 0.79)). The Canadian Triage and Acuity Scale (CTAS) and the Emergency Severity Index (ESI) have been studied more than the other scales.

Conclusion: It seems necessary to set a consensus method to assess and compare the reliability of triage scales. This review suggests more specific studies on reliability of triage scales, especially on intra-rater agreement.

Introduction

Triage as the first step in an Emergency Department (ED) visit, sorts and prioritizes the patients. This sorting and prioritizing process aims at optimizing the patients' waiting time according to severity of their medical condition and reducing the negative impact of delayed care by giving priority to those who cannot wait safely.

Several triage scales have been introduced since the early 1990s. Of those, are the Australian Triage Scale (ATS), the Canadian Triage and Acuity Scale (CTAS), the Manchester Triage System (MTS) and the Emergency Severity Index (ESI).

Numerous studies have been conducted reviewing ED triage including the reliability of triage scales. It is ideal that triage ratings be reproducible and repeatable regardless of rater and the setting. In other words, different triage personnel must be able to reach the same or similar triage scores when assessing similar patients by the same tools.

At national level, some governmental or non-governmental agencies may need to compare EDs with each other or some hospitals may ask for a benchmark from other hospitals. Reliability of the triage scale used in these instances is an important pre-assumption.

Many EDs have adopted a triage system and others are going to do so. To choose a triage scale, reliability of the triage scale should be considered as an essential factor. Some studies reported low inter-observer agreement [1] while others reported very high kappa values [2]. It seems necessary to reach a census on reliability of triage scales to help clinicians when

adopting a system for their EDs.
The aim of this review was to compare reliability of different triage systems around the world by using kappa values. The research question was "In adult ED patients, using Kappa statistics, to what extent did clinicians' triage ratings agree with each other or with a gold standard?"

Methods

Systematically search of national and international literature including PubMed, Scopus, ISI Web of knowledge, Mosby's Nursing Index, Magiran and SID with no time-limit through August 2013 yielded 1342 articles.
To assess and report the quality and risk of bias of included studies, once all possible relevant studies were identified, each study was assessed for risks of bias such as database bias, outcome reporting bias, selection bias, attrition, performance, and other biases. This information was used in article selection and data synthesis so that the studies with high or unclear risk of bias were not used in the synthesis of the evidence.
Heterogeneity in meta-analysis refers to the variation in study outcomes between studies. When the studies' results only differ by the sampling error (homogeneous case) a fixed-effects model can be applied to obtain an average effect size. By contrast, if the study results differ by more than the sampling error (heterogeneous case), then the meta-analyst can assume a random-effects model, in order to take into account both within- and between-studies variability. Statistical heterogeneity was tested by Cochrane's Q test and the random-effects model was used to adjust for the combined effect size in meta-analysis.
Publication bias occurs when primary researchers search among elasticity estimates and select those with statistically significant coefficients, "correct" signs, and more elastic values. Publication bias was addressed using funnel plot, and in case of existing this bias, trim-and-fill method was used to adjust for this bias in meta-analysis.
Inclusion was limited to studies reporting Kappa values on adult patients (≥15 years) who visited EDs for somatic reasons and studies based on real patients triaged at EDs or fictitious patient scenarios.
Studies solely on geriatric patients, computerized triage, telephone triage, specialty triages such as ophthalmic or obstetric triage, studies on a special symptom or syndrome, and studies published in languages other than English and Persian were excluded.

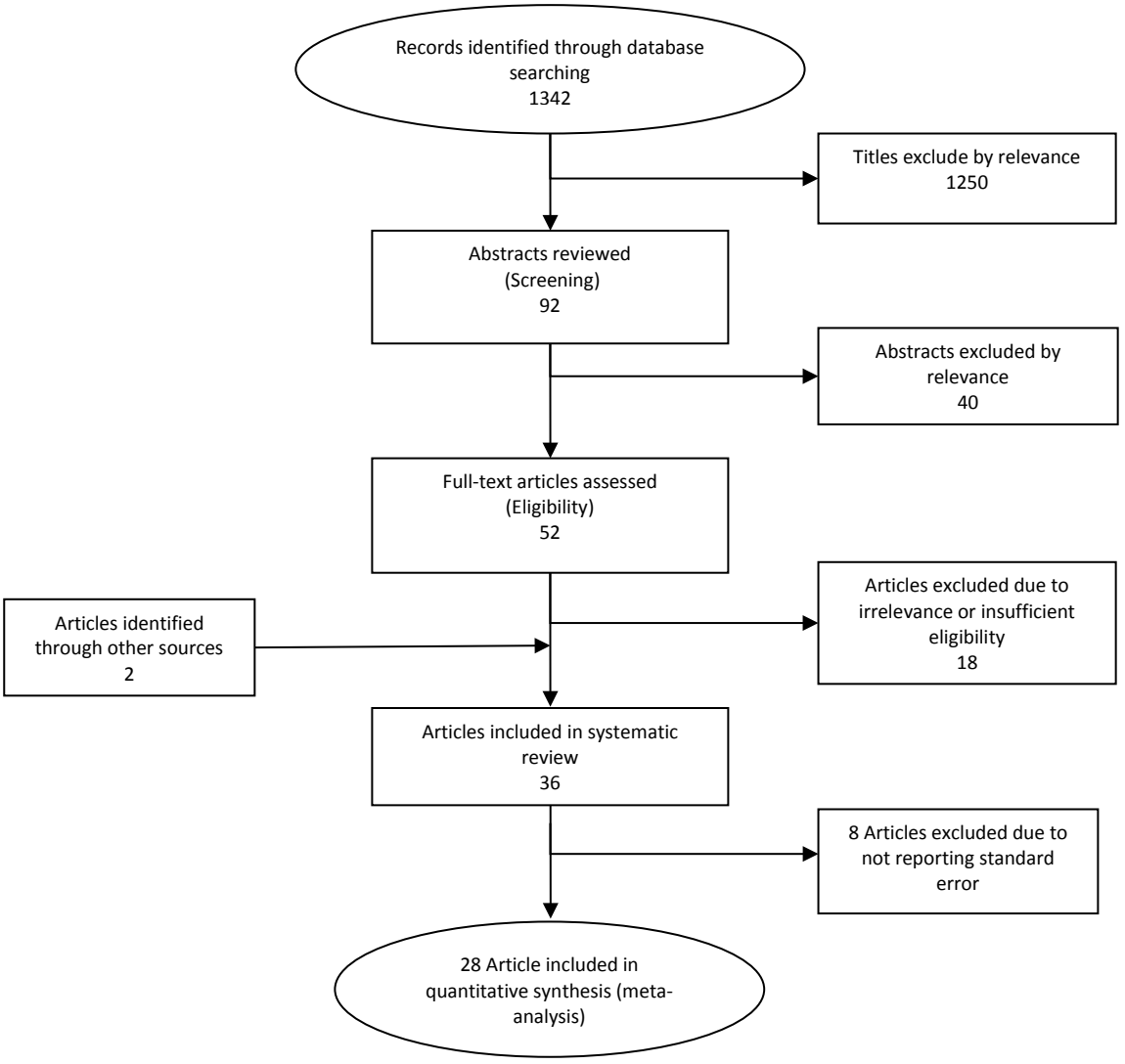

Figure 1. Search diagram and selection process.

Reliability of triage scales as ordinal data is measured by Kappa value (K) [3] which is the degree of agreement among clinicians or clinicians and a gold standard.

Some studies report un-weighted kappa value while others assess reliability by weighted kappa (linear weighted or quadratic weighted). Weighted kappa gives importance to seriousness of disagreement in ratings and un-weighted kappa does not treat so; therefore, weighted kappa is better for ordinal scales such as triage [4].

The K ranges from 0 to 1 where 0 means there is no agreement except what is by chance and 1 means complete agreement. As described by Altman, Kappa agreement is defined as very good (0.81–1.00), good (0.61–0.80), moderate (0.41–0.60), fair (0.21–0.40), or poor (< 0.20) [5].

Articles reviewed independently and duplicated articles were discarded. For data extraction from included studies the extraction table was used. The extracted data include author, year, country, the name of triage scale that its reliability is assessed, type of the triage scale (e.g. 5 level, 4 level, …), type of reliability that is assessed (inter-rater or intra-rater), subjects (patients or case scenarios) and their amount, triageurs' profession (nurse or physician) and their amount, and the kappa values. Figure 1 summarizes the search and selection process. Results from studies that met inclusion criteria were extracted and synthesized. After synthesizing, the overall kappa score of scales was calculated. Another comparison was performed on scales based on type of triage scales.

Results

Detailed results of included studies are presented in appendix 1. They assess reliability of CTAS [6-11], ESI [12-19], Adaptive process triage (ADAPT) [20], South African Triage Scale (SATS) [21, 22], MTS [19, 23, 24], ATS [25, 26], Italian four-level triage system (I-4L) [27, 28], Triage Emergency Method (TEM) [28], French Emergency Nurses Classification in Hospital scale (FRENCH) [29], Hong Kong Accident and Emergency Triage Guidelines (HKAETG) [30], Rapid Emergency Triage and Treatment System –Hospital Unit West (RETTS-HEV) [31], and some local scales with no name. All included studies had observational designs. Table1 shows inter-rater reliability of included triage scales.

Table 1. Inter-rater reliability of triage scales.

Triage scale	Type	Number of Ratings*	Agreement value	Triage scale	Type	Number of Ratings	Agreement value
CTAS	5-level	7614	K= 0.46 LWK= 0.71	CTAS ATS	5-level 5-level	1000 2492	QWK= 0.80 K= 0.42
CTAS	5-level	500	QWK= 0.44	ATS	5-level	2492	K= 0.56
CTAS	5-level	205	QWK= 0.80	ESI	5-level	40	LWK= 0.76
CTAS	5-level	205	QWK= 0.82	ESI	5-level	200	LWK= 0.78
CTAS	5-level	205	QWK=0.76	ESI	5-level	250	LWK= 0.898
CTAS	5-level	205	QWK= 0.73	ESI	5-level	2100	LWK= 0.87
CTAS	5-level	3589	K= 0.263 QWK= 0.531	FRENCH	5-level	1800	K= 0.69 LWK= 0.77
Local scale (USA)	3-level	610	LWK= 0.53	HKAETG	5-level	200	QWK= 0.699
				HKAETG	5-level	200	QWK= 0.717
ESI	5-level	606	LWK= 0.68	HKAETG	5-level	200	QWK= 0.766
Local scale (Canada)	5-level	60	K= 0.662	Local scale (Taipei)	4-level	4324	K= 0.32
ESI	5-level	1240	LWK= 0.80	CTAS	5-level	188	QWK= 0.50
ESI	5-level	209	LWK= 0.73	MTS	5-level	600	K= 0.76
Local scale (Sweden)	5-level	855	K= 0.529	RETTS-HEV	5-level	292	K= 0.60
ADAPT	5-level	570	K= 0.472	ESI	5-level	500	K= 0.46 QWK= 0.73
SATS	5-level	78	QWK= 0.92				
SATS	5-level	72	QWK= 0.94	I-4L	4-level	1230	LWK= 0.73
SATS	5-level	150	QWK= 0.57	I-4L	4-level	1230	LWK= 0.76
SATS	5-level	100	QWK= 0.65	I-4L	4-level	1476	LWK= 0.79
MTS	5-level	1027	K= 0.61	MTS	5-level	100	K= 0.48 QWK= 0.62
ATS	5-level	9954	K= 0.412				
ESI	5-level	702	LWK= 0.8	SATS	5-level	500	QWK= 0.76
ESI	5-level	718	QWK= 0.89	SATS	5-level	500	K= 0.68
I-4L	4-level	945	LWK= 0.73	SATS	5-level	500	LWK=0.72
TEM	4-level	945	LWK= 0.79	SATS	5-level	1000	K= 0.51
CTAS	5-level	500	QWK= 0.84	SATS	5-level	1000	LWK=0.58
CTAS	5-level	500	QWK= 0.83	SATS	5-level	1000	QWK= 0.66

*** Number of ratings= $N_{observer}$ * N_{case}**
K: Kappa; **LWK:** Linear Weighted Kappa; **QWK:** Quadratic Weighted Kappa

Some of these studies (8 studies) were excluded from meta-analysis due to lack of standard error for kappa measures.

Results of Cochran's Q tests showed a low p-value (p-value <.001) which means a significant heterogeneous results among different studies. So, a random-effects model was used in the meta-analysis in order to take into account heterogeneity of the studies.

Figure 2 displays funnel graph for kappa of all triage scales, where standard errors are arrayed on the Y-axis and kappa estimates on the X-axis. The vertical line indicates the random-effect mean and diagonal lines are the 95% confidence interval.

In the absence of publication bias, a funnel plot should be symmetric about the mean effect size. The graph indicates positive bias because there is greater dispersion of estimates to the right of the mean, and a clear indication of "missing" studies to the lower left of the mean.

A method for correcting the publication bias is the trim-and-fill which uses an iterative algorithm to add missing values until

observations are symmetric about a recomputed mean effect size. Using trim-and-fill, the random-effects mean for kappa is .426 (.018).

Figure 3 displays funnel graph for linear weighted kappa, which seems to be symmetric about the mean effect size so there is no publication bias for linear weighted kappa. There was similar result for quadratic weighted kappa (Figure 4).

Table 2 and 3 summarize the results of meta-analysis for each triage scale and for each type of scale respectively (five-level, four-level and so on) based on a random effect model.

As it is shown in Table 2, the ESI and the CTAS have been studied more than other scales. According to our findings, five-level triage acuity scales are used much more than the others in EDs around the world (Table 3). Table 3 also indicates the fact that five-level scales are more reliable in triaging patients in emergency department than others.

Using un-weighted kappa only the FRENCH, the MTS and a local scale from Canada [32] had good inter-rater reliability (K>0.6). On linear weighted kappa the ESI, TEM, FRENCH, I-4L, CTAS, and SATS all were good in inter-rater reliability (K>0.6). Only the local scale from USA was moderate. And the other scales' inter-rater reliability was not assessed by linear weighted kappa. Using quadratic weighted kappa as inter-rater reliability measure introduced ESI as the only triage scale that has very good reliability (K>0.8). The SATS, HKAETG, CTAS, and the MTS were good in quadratic weighted kappa (K>0.6). Other scales have not been assessed by this type of kappa.

Table 4 shows the results of studies on intra-rater reliability of the scales. There are a few studies on this subject and intra-rater reliability of most of the triage scales have not been measured to date with kappa.

With the few studies in hand, five-level scales are slightly better in intra-rater agreement than the four-level ones (Table 5).

Table 2. Inter-rater reliability of each triage scale.

Triage scale	Type	Number of studies	Frequency of measurement	Number of Ratings*	Combined agreement value (CI 95%)
CTAS	5-level	6	2	11203	K=.373 (.18, .56)
			1	7614	Lwk=.71 (.60,.82)
			10	7097	Qwk=.72 (.65, .78)
ESI	5-level	8	1	500	K=.46 (.37, .55)
			8	5347	Lwk=.79 (.75, .84)
			2	1218	Qwk=.81 (.64, .97)
ADAPT	5-level	1	1	570	K=.47 (.45, .50)
SATS	5-level	2	2	1500	K=.59 (.43, .76)
			2	1500	Lwk =.65 (.55, .74)
			6	1900	Qwk=.78 (.68, .88)
MTS	5-level	3	3	1727	K=.61 (.48, .75)
			1	100	Qwk=.62 (.41, .83)
ATS	5-level	2	3	14938	K=.46 (.39, .52)
I-4L	4-level	2	4	4881	Lwk=.75 (.67, .83)
TEM	4-level	1	1	945	Lwk=.79 (.61, .97)
FRENCH	5-level	1	1	1800	K=.69 (.61, .77)
			1	1800	Lwk=.77 (.64, .89)
HKAETG	5-level	1	3	600	Qwk=.73 (.58, .87)
Local scale (USA)	3-level	1	1	610	Lwk=.53 (.40, .66)
Local scale (Canada)	5-level	1	1	60	K=.66 (.33, .99)
Local scale (Sweden)	5-level	1	1	855	K=.53 (.51, .54)
Local scale (Taipei)	4-level	1	1	4324	K=.32 (.27, .36)
RETTS-HEV	5-level	1	1	292	K=.60(.48, .72)

*** Number of ratings= $N_{observer}$ * N_{case}**
K: Kappa; **LWK:** Linear Weighted Kappa; **QWK:** Quadratic Weighted Kappa. Random-effects model was applied to obtain combined agreement values.

Table 3. Inter-rater reliability of triage scales based on scale type.

Triage scale type	Number of triage scales	Number of studies	Frequency of measurement	Number of Ratings	Combined agreement value (CI 95%)
5-level	11	27	16	33445	K=.53 (.48, .57)
			12	16261	Lwk=.76 (.72, .81)
			23	10915	Qwk=.74 (.69, .79)
4-level	3	4	1	4324	K=.32 (.28, .36)
			5	5826	Lwk=.76 (.67, .84)
3-level	1	1	1	610	Lwk=.53 (.37, .69)

Table 4. Intra-rater reliability of triage scales.

Triage scale	type	Number of studies	Frequency of measurement	Number of Ratings	Agreement value
MTS	5-level	1	1	550	K=.84 (.73, .95) Qwk=.9 (.83, .97)
Local scale (Canada)	5-level	1	1	60	K=.76 (.65, .87)
ESI	5-level	1	1	450	K=.65 (.59, .71) Qwk=.85 (.8, .9)
I-4L	4-level	1	1	945	Qwk=.82 (.67, .97)
TEM	4-level	1	1	945	QwK=.78 (.62, .94)

Table 5. Intra-rater reliability of triage scales by scale type.

Triage scale type	Number of triage scales	Number of studies	Frequency of measurement	Number of Ratings	weighted Average of agreement value
5-level	3	3	3	1060	K= .74 (.62 , .86)
			2	1000	Qwk=.87 (.82 , .92)
4-level	2	2	2	1890	Qwk=.80 (.69 , .91)

Discussion

Findings from this review showed widespread use of five-level triage scales rather than others (11 out of 15 scales). This indicates a global tendency among emergency care givers to use five-level scales. Comparing five-level triages with the others proves them as being more reliable.

Reliability of emergency department triage is of high importance since it prioritizes patients and determines treatment area in the ED. Lack of reliable ratings in triage may result in mis-triage, time loss, inappropriate care [6], wasted resources and unnecessary costs.

Although scientific evidence from this review suggests the FRENCH, MTS, TEM, I-4L, CTAS, SATS, HKAETG as having good inter-rater reliability and the ESI triage system as having very good, other factors should also be considered when adopting a triage system. We should keep this fact in mind that most of these scales are developed in high-income countries and may have problems when implementing in developing countries or other countries with cultural differences. Yet some studies assessed reliability of these scales in settings other than the origin country [6, 17, 18, 23] and additional studies in this field are needed to be conducted to clarify the subject.

Lack of a single, consensus criterion to measure reliability of triage scales and use of different types of kappa (un-weighted kappa, linear weighted kappa, and the quadratic weighted kappa) makes it difficult to simply compare triage scales. As a result, some of these scales have greater reliability using un-weighted kappa, others show better when assessed by weighted kappa. This may be due to sensitivity of triage scales in dividing patients into distinct levels. Also there may be differences in circumstances and/or methodology used in different studies which had found different agreement values for the same scale. As an example, using case scenarios or real patients to assess reliability may influence the acuity ratings [33] and in some cases triage nurses may get further information when triaging real patients rather than scenarios [6].

On the intra-rater reliability, there are a few studies; therefore, it shows a need for further studies to be conducted in this field. In fact, intra-rater agreement of most of triage scales has not been measured to date.

Reliability is, with no doubt, one of the essential elements in assessing triage scales, but it is not the only criterion. Other factors such as the validity of the scale in predicting patient's outcomes should also be considered.

With regard to the fact that the use of information technology is growing in healthcare system and with respect to spread of computerized triage in EDs [9, 34-41], it is suggested that another review could be performed on validity and reliability of electronic/ computerized triage.

Another point is that some of triage systems [20, 21, 28-30] are newly invented and are not used and studied enough. This could be assumed that if more studies were performed on them, they might show greater reliability.

Conclusion

With the best of our knowledge and according to the findings from this systematic review, five-level triage scales have widespread use around the world with the greatest inter-rater reliability. Thus, priority can be given to them when adopting a triage system. It is suggested to take a similar, scientifically wise methodology in assessing reliability of triage scales. Also there is a need for assessing intra-rater reliability of most of triage scales.

Limitations of the study: In this review only the kappa statistics are assumed as measure of reliability of triage scales, while some studies used Intra-class Correlation Coefficient (ICC) for this purpose. We suggest additional reviews in this area to include these studies. Another Limitation was that only studies that are published in English and Persian languages are included in this review and this may result in language bias. Also, the total number of studies in this area was limited and our meta- analysis and conclusion are based on these few studies especially for intra rater kappa.

Conflict of interests: The authors declare no conflict of interest.

Acknowledgement: The authors thank Central Library of Tabriz University of Medical Sciences for their help on access to the articles.

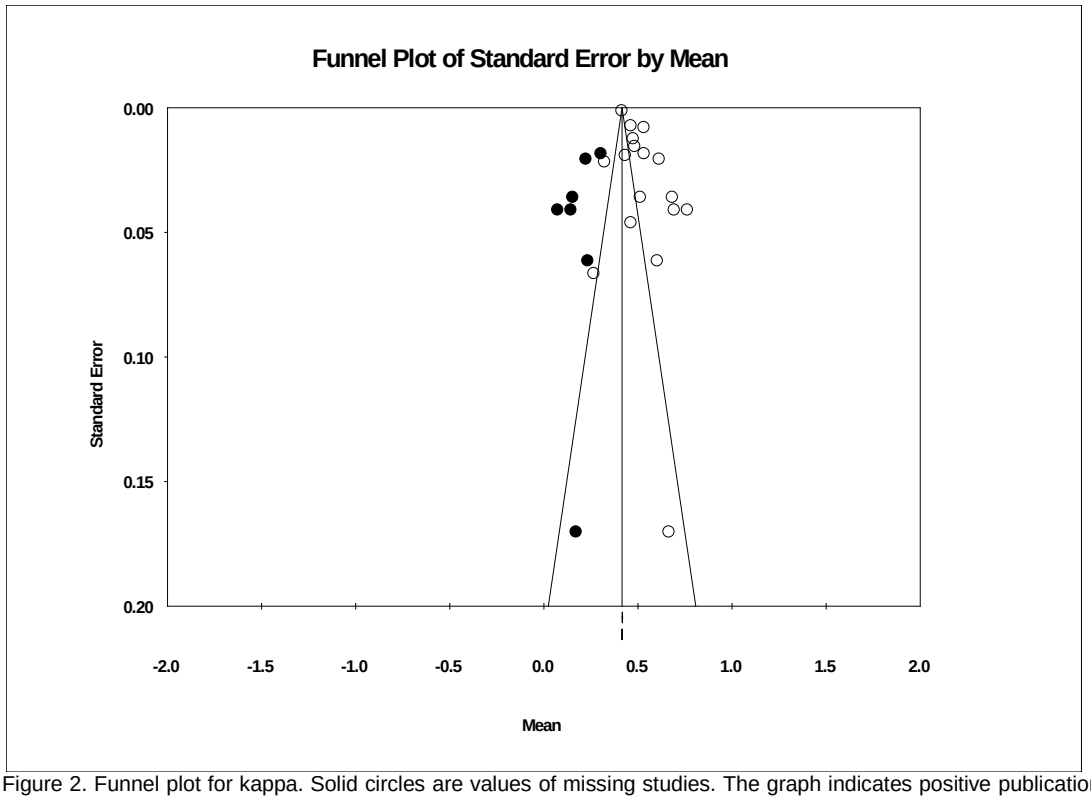

Figure 2. Funnel plot for kappa. Solid circles are values of missing studies. The graph indicates positive publication bias because there is greater dispersion of estimates to the right of the mean, and a clear indication of "missing" studies to the lower left of the mean.

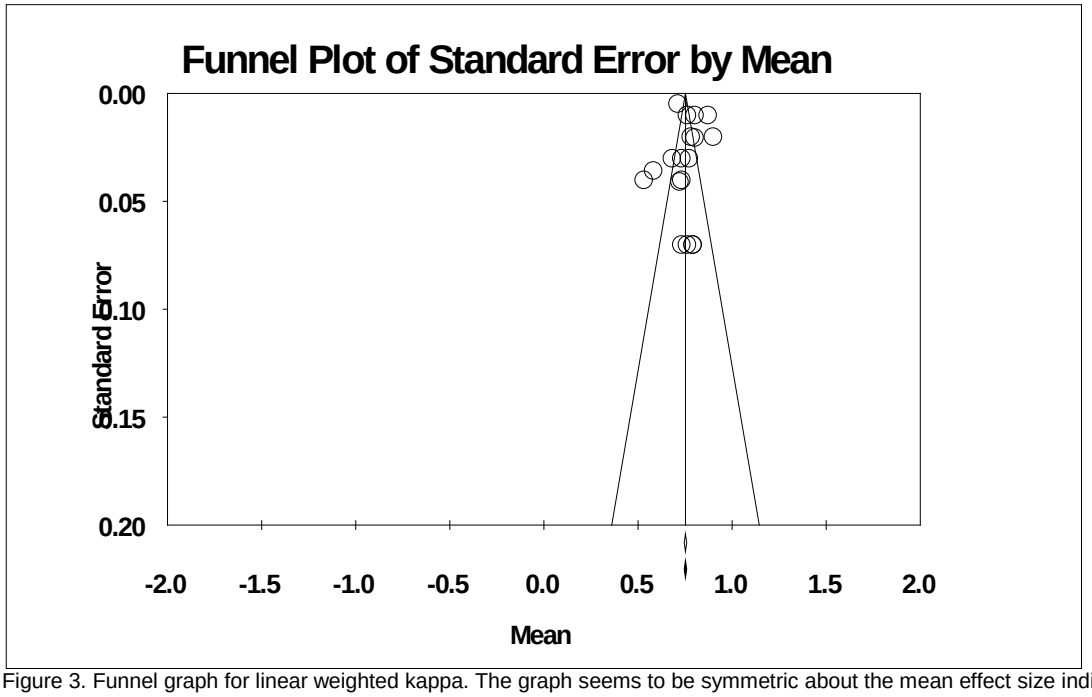

Figure 3. Funnel graph for linear weighted kappa. The graph seems to be symmetric about the mean effect size indicating no publication bias for linear weighted kappa.

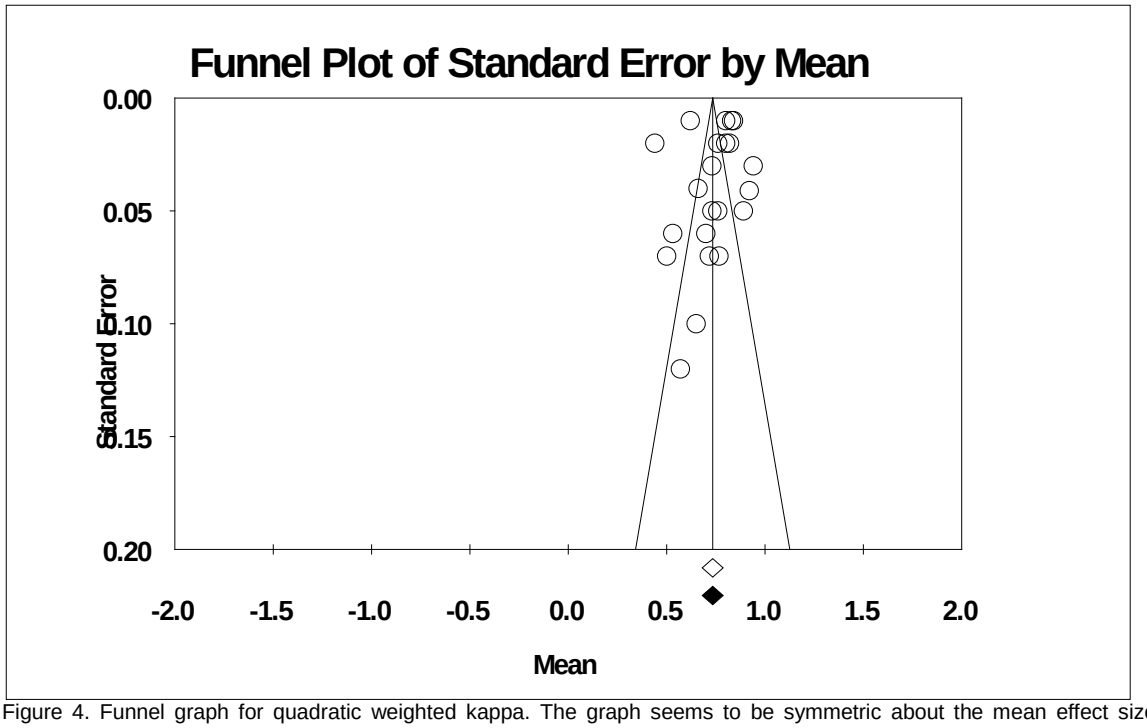

Figure 4. Funnel graph for quadratic weighted kappa. The graph seems to be symmetric about the mean effect size indicating no publication bias for quadratic weighted kappa.

References

1. Durand AC, Gentile S, Gerbeaux P, Alazia M, Kiegel P, Luigi S, et al: Be careful with triage in emergency departments: interobserver agreement on 1,578 patients in France. *BMC Emerg Med* 2011,11:19. doi: 10.1186/1471-227X-11-19.
2. Jobé J, Ghuysen A, Gérard P, Hartstein G, D'Orio V: Reliability and validity of a new French-language triage algorithm: The ELISA scale. *Emerg Med J* 2014. doi:10.1136/emermed-2012-201927.
3. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960, 20(1):37-46.
4. Sim J, Wright CC: The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Phys Ther* 2005;85:257-68.
5. Altman DG: *Practical Statistics for Medical Research*. London: Chapman & Hall; 1999.
6. Göransson K, Ehrenberg A, Marklund B, Ehnfors M: Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19(4):432-8. doi: 10.1111/j.1471-6712.2005.00372.x.
7. Dallaire C, Poitras J, Aubin K, Lavoie A, Moore L. Emergency department triage: do experienced nurses agree on triage scores? J Emerg Med. 2012 Jun;42(6):736-40.
8. Manos D, Petrie DA, Beveridge RC, Walter S, Ducharme J: Inter-observer agreement using the Canadian Emergency Department Triage and Acuity Scale. *CJEM* 2002,4(1):16-22.
9. Dong SL, Bullard MJ, Meurer DP, Colman I, Blitz S, Holroyd BR, et al: Emergency triage: comparing a novel computer triage program with standard triage. *Acad Emerg Med* 2005,12(6):502-7. doi: 10.1197/j.aem.2005.01.005.
10. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S: Reliability of the Canadian emergency department triage and acuity scale: interrater agreement. *Ann Emerg Med* 1999,34(2):155-9.
11. Dallaire C, Poitras J, Aubin K, Lavoie A, Moore L, Audet G: Interrater agreement of Canadian Emergency Department Triage and Acuity Scale scores assigned by base hospital and emergency department nurses. *CJEM* 2010,12(1):45-9.
12. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J: Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs* 2002,28(5):395-400. doi:10.1067/men.2002.127184.
13. Wuerz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and refinement of the emergency severity index. *Acad Emerg Med* 2001,8(2):170-6. doi: 10.1111/j.1553-2712.2001.tb01283.x.
14. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N: Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000,7(3):236-42. doi: 10.1111/j.1553-2712.2000.tb01066.x.
15. Tanabe P, Gimbel R, Yarnold PR, Kyriacou DN, Adams JG: Reliability and validity of scores on The Emergency Severity Index version 3. *Acad Emerg Med* 2004,11(1):59-65. doi: 10.1197/j.aem.2003.06.013.
16. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC: The emergency severity index triage algorithm version 2 is reliable and valid. Acad Emerg Med 2003,10(10):1070-80. doi: 10.1197/S1069-6563(03)00350-6.
17. Grossmann FF, Nickel CH, Christ M, Schneider K, Spirig R, Bingisser R: Transporting clinical tools to new settings: cultural adaptation and validation of

the Emergency Severity Index in German. *Ann Emerg Med* 2011, 57(3):257-64. doi: 10.1016/j.annemergmed.2010.07.021.
18. Kariman H, Joorabian J, Shahrami A, Alimohammadi H, Noori Z, Safari S: Accuracy of emergency severity index of triage in Imam Hossein hospital - Tehran, Iran (2011). *Gorgan Uni Med Sci* 2013;15(1):115-20. [in Persian]
19. Storm-Versloot MN, Ubbink DT, Chin a Choi V, Luitse JS: Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study. Emerg Med J 2009,26(8):556-60. doi:10.1136/emj.2008.059378.
20. Goransson KE, von Rosen A: Interrater agreement: a comparison between two emergency department triage scales. Eur J Emerg Med 2011,18(2):68-72. doi: 10.1097/MEJ.0b013e32833ce4eb.
21. Twomey M, De Sa A, Wallis LA, Myers JE: Inter-rater reliability of the South African Triage Scale: Assessing two different cadres of health care workers in a real time environment. *African Journal for Emergency Medicine* 2011;1(3):113-8. doi:10.1016/j.afjem.2011.08.003.
22. Twomey M, Wallis LA, Thompson ML, Myers JE: The South African Triage Scale (adult version) provides reliable acuity ratings. *International Emergency Nursing* 2012;20(3):142-50. doi:10.1016/j.ienj.2011.08.002.
23. Olofsson P, Gellerstedt M, Carlstrom ED: Manchester Triage in Sweden - interrater reliability and accuracy. Int Emerg Nurs 2009,17(3):143-8. doi: 10.1016/j.ienj.2008.11.008.
24. van der Wulp I, van Baar ME, Schrijvers AJP: Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. Emergency Medicine Journal 2008,25(7):431-4. doi:10.1186/1865-1380-6-1.
25. Gerdtz MF, Collins M, Chu M, Grant A, Tchernomoroff R, Pollard C, et al: Optimizing triage consistency in Australian emergency departments: The Emergency Triage Education Kit. *Emerg Med Australas* 2008;20(3):250-9. doi: 10.1111/j.1742-6723.2008.01089.x.
26. Considine J, LeVasseur SA, Villanueva E: The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Ann Emerg Med* 2004,44(5):516-23. doi:10.1016/j.annemergmed.2004.04.007.
27. Parenti N, Manfredi R, Reggiani MLB, Sangiorgi D, Lenzi T: Reliability and validity of an Italian four-level emergency triage system. *Emerg Med J* 2010,27(7):495-8. doi: 10.1136/emj.2008.070193.
28. Parenti N, Ferrara L, Bacchi Reggiani ML, Sangiorgi D, Lenzi T: Reliability and validity of two four-level emergency triage systems. Eur J Emerg Med 2009,16(3):115-20.
29. Taboulet P, Moreira V, Haas L, Porcher R, Braganca A, Fontaine JP, et al: Triage with the French Emergency Nurses Classification in Hospital scale: reliability and validity. *Eur J Emerg Med* 2009,16(2):61-7. doi:10.1097/MEJ.0b013e328304ae57.
30. Fan MM, Leung LP: Validation of the Hong Kong Accident and Emergency Triage Guidelines. *Hong Kong Med J* 2013,19(3):198-202. doi: 10.12809/hkmj133900.
31. Nissen L, Kirkegaard H, Perez N, Horlyk U, Larsen LP: Inter-rater agreement of the triage system RETTS-HEV. *Eur J Emerg Med* 2014, 21(1):37-41. doi: 10.1097/MEJ.0b013e32836397d9.
32. Fernandes CM, Wuerz R, Clark S, Djurdjev O: How reliable is emergency department triage? *Ann Emerg Med* 1999 ,34(2):141-7.
33. Worster A, Sardo A, Eva K, Fernandes CM, Upadhye S: Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. *J Emerg Nurs* 2007,33(4):319-23. doi:10.1016/j.jen.2006.12.016.
34. Aronsky D, Jones I, Raines B, Hemphill R, Mayberry SR, Luther MA, et al: An integrated computerized triage system in the emergency department. *AMIA Annu Symp Proc* 2008, 2008:16-20.
35. Berman DA, Coleridge ST, McMurry TA. Computerized algorithm-directed triage in the emergency department. *Ann Emerg Med* 1989, 18(2):141-4. doi:0.1016/S0196-0644(89)80102-7.
36. Dong SL, Bullard MJ, Meurer DP, Akhmetshin E, Holroyd BR, Rowe BH: Agreement of a computerized triage tool using written case scenarios. Annals of Emergency Medicine. 2008,51(4):532.
37. Dong SL, Bullard MJ, Meurer DP, Blitz S, Ohinmaa A, Holroyd BR, et al: Reliability of computerized emergency triage. *Acad Emerg Med* 2006, 13(3):269-75. doi: 10.1197/j.aem.2005.10.014.
38. Gravel J, Gouin S, Bailey B, Roy M, Bergeron S, Amre D. Reliability of a computerized version of the Pediatric Canadian Triage and Acuity Scale. *Acad Emerg Med* 2007,14(10):864-9. doi: 10.1197/j.aem.2007.06.018.
39. Ng CJ, Yen ZS, Tsai JC, Chen LC, Lin SJ, Sang YY, et al: Validation of the Taiwan triage and acuity scale: a new computerised five-level triage system. *Emerg Med J* 2011,28(12):1026-31. doi: 10.1136/emj.2010.094185.
40. Rajkumar GN, Small DR, Conn IG: Computerised triage in a prostate assessment clinic. *Prostate Cancer and Prostatic Diseases* 2004,7(2):118-21. doi:10.1038/sj.pcan.4500703.
41. Sung SF, Huang YC, Ong CT, Chen W: Validity of a computerised five-level emergency triage system for patients with acute ischaemic stroke. *Emerg Med J* 2012, 30(6):454-8. doi: 10.1136/emermed-2012-201423.
42. Gerdtz MF, Collins M, Chu M, Grant A, Tchernomoroff R, Pollard C, et al: Optimizing triage consistency in Australian emergency departments: the Emergency Triage Education Kit. *Emerg Med Australas* 2008,20(3):250-9. doi: 10.1111/j.1742-6723.2008.01089.x.
43. Loke SS, Liaw SJ, Tiong LK, Ling TS, Chiang WT: Evaluation of nurse-physician inter-observer agreement on triage categorization in the emergency

department of a Taiwan medical center. *Chang Gung Med J* 2002,25(7):446-52.

Appendix 1:

Extraction table for reliability of triage scales

Row	Author Year Country	Triage scale	Type of Triage Scale	Type of Reliability	Subject and Triageur: Amount, profession	Results: values
1	Goransson et al. 2005[6] Sweden	CTAS	5-level	Inter-rater	18 scenarios 423 RNs	K= 0.46 LWK= 0.71
2	Dallaire et al. 2012 [7] Canada	CTAS	5-level	Inter-rater	100 scenarios 5 RNs	QWK= 0.44
3	Manos et al 2002 [8] Canada	CTAS	5-level	Inter-rater	41 scenarios 5 RNs 5 Drs 5 BLS Paramedics 5 ALS Paramedics	RNs: QWK= 0.80 Dr.: QWK= 0.82 BLS Para.: QWK=0.76 ALS Para.: QWK= 0.73
4	Dong et al. 2005[9] Canada	CTAS	5-level	Inter-rater	97 scenarios 37 RNs	K= 0.263 QWK= 0.531
5	Travers et al. 2002 [12] USA	N.M	3-level	Inter-rater	305 patient records	LWK= 0.53
		ESI	5-level	Inter-rater	303 patient records	LWK= 0.68
6	Fernandes et al. 1999 [32] Canada	N.M	5-level	Inter-rater	5 scenarios 12 RNs	K= 0.662
				Intra-rater	5 scenarios 12 RNs	K= 0.757
7	Wuerz et al. 2001 [13] USA	ESI	5-level	Inter-rater	20 scenarios 62 RNs	LWK= 0.80
					209 patient records	LWK= 0.73
8	Goransson et al. 2011 [20] Sweden	N.M	5-level	Inter-rater	19 scenarios 45 RNs	K= 0.529
		ADAPT	5-level	Inter-rater	19 scenarios 30 RNs	K= 0.472
9	Twomey et al. 2011 [21] South Africa	SATS	5-level	Inter-rater	26 patients 3 ENs	QWK= 0.92
					24 patients 3 FMSs	QWK= 0.94
					26 patients & 3 ENs 24 patients & 3 FMSs	QWK= 0.57
					50 patients 2 teams (each: 1FMS & 1ENA)	QWK= 0.65
10	Olofsson et al. 2009 [23] Sweden	MTS	5-level	Inter-rater	13 scenarios 79 RNs	K= 0.61
11	Gerdtz et al. 2008 [42] Australia	ATS	5-level	Inter-rater	237 scenarios 42 RNs	K= 0.412
12	Wuerz et al. 2000 [14] USA	ESI	5-level	Inter-rater	351 patients RN vs. Drs	LWK= 0.8
13	Tanabe et al. 2004 [15] USA	ESI	5-level	Inter-rater	359 patients RNs	QWK= 0.89
14	Parentia et al. 2009 [28] Italy	I-4L	4-level	Inter-rater	189 scenarios 5 RNs	LWK= 0.73
				Intra-rater	189 scenarios 5 RNs	QWK= 0.82
		TEM	4-level	Inter-rater	189 scenarios 5 RNs	LWK= 0.79
				Intra-rater	189 scenarios 5 RNs	QWK= 0.78
15	Beveridge et al. 1999 [10] Canada	CTAS	5-level	Inter-rater	50 scenarios 10 RNs	QWK= 0.84
					50 scenarios 10 Drs	QWK= 0.83
					50 scenarios 10 RNs vs. 10 Drs	QWK= 0.80
16	Considine et al.	ATS	5-level	Inter-rater	Paper (text) 14 scenarios	K= 0.42

	2004 [26] Australia				Computer (text +photo)	178 RNs 14 scenarios 178 RNs	K= 0.56
17	Eitel et al. 2003 [16] USA	ESI	5-level	Inter-rater	20 scenarios RNs 100 patients RNs		LWK= 0.76 LWK= 0.78
18	Grossmann et al. 2011 [17] Germany	ESI	5-level	Inter-rater	125 patients RNs		LWK= 0.898
19	Taboulet et al. 2009 [29] France	FRENCH	5-level	Inter-rater	300 patient records 6 RNs		K= 0.69 LWK= 0.77
20	Fan et al. 2013 [30] Hong Kong	HKAETG	5-level	Inter-rater	100 patient records 2 RNs 100 patient records 2 RNs 100 patient records RNs vs. expert panel		QWK= 0.699 QWK= 0.717 QWK= 0.766
21	Kariman et al. 2013 [18] Iran	ESI	5-level	Inter-rater	1050 patients 1 RNs vs. 1 Drs		LWK= 0.87
22	Lokeet al. 2002 [43] Taipei	N.M	4-level	Inter-rater	2162 patient records RNs vs. 2 Drs		K= 0.32
23	Dallaireet al. 2010 [11] Canada	CTAS	5-level	Inter-rater	94 patients RNs of two hospitals		QWK= 0.50
24	Nissenet al. 2013 [31] Denmark	RETTS-HEV	5-level	Inter-rater	146 patients RNs		K= 0.60
25	Storm-Verslootet al. 2009 [19] Netherlands	MTS	5-level	Inter-rater Intra-rater	50 scenarios 12 RNs 50 scenarios 11 RNs		K= 0.76 K= 0.84 QWK= 0.90
		ESI	5-level	Inter-rater Intra-rater	50 scenarios 10 RNs 50 scenarios 9 RNs		K= 0.46 QWK= 0.73 K= 0.65 QWK=0.85
26	Parenti et al. 2010 [27] Italy	I-4L	4-level	Inter-rater Inter-rater Inter-rater	246 scenarios 5 RNs 246 scenarios 6 Experts 246 scenarios 5 RNs vs. reference		LWK= 0.73 LWK= 0.79 LWK= 0.76
27	van der Wulp et al. 2008 [24] Netherlands	MTS	5-level	Inter-rater	50 scenarios RNs vs. expert panel		K= 0.48 QWK= 0.62
28	Twomey et al. 2012 [22] South Africa	SATS	5-level	Inter-rater	100 scenarios 5 Drs 100 scenarios 10 ENAs		QWK= 0.76 K=0.68 LWK=0.72 K= 0.51 LWK=0.58 QWK= 0.66

K: Kappa; **LWK:** Linear Weighted Kappa; **QWK:** Quadratic Weighted Kappa; **CTAS:** Canadian Triage and Acuity Scale; **ESI:** Emergency Severity Index; **ADAPT:** Adaptive process triage; **BLS:** Basic Life Support; **ALS:** Advanced Life Support; **N.M:** Not Mentioned; **SATS:** South African Triage Scale; **EN:** Enrolled Nurse; **ENA:** Enrolled Nursing Assistant; **FMS:** Final year Medical Student; **MTS:** Manchester Triage System; **I-4l:** Italian four-level triage system; **TEM:** triage emergency method; **FRENCH:** FRench Emergency Nurses Classification in Hospital scale; **HKAETG:** Hong Kong Accident and Emergency Triage Guidelines; **RETTS-HEV:** Rapid Emergency Triage and Treatment System – Hospital Unit West.