

Reliability of the Emergency Severity Index

Meta-analysis

Amir Mirhaghi, *Abbas Heydari, Reza Mazlom, Farzaneh Hasanzadeh

مُعَوَّلَيَّة مؤشر شدة الطوارئ

التحليل التلوبي

أمير ميرحقي، عباس حيدري، رضا مظلوم، فرزانه حسن زاده

ABSTRACT: **Objectives:** Although triage systems based on the Emergency Severity Index (ESI) have many advantages in terms of simplicity and clarity, previous research has questioned their reliability in practice. Therefore, the aim of this meta-analysis was to determine the reliability of ESI triage scales. **Methods:** This meta-analysis was performed in March 2014. Electronic research databases were searched and articles conforming to the Guidelines for Reporting Reliability and Agreement Studies were selected. Two researchers independently examined selected abstracts. Data were extracted in the following categories: version of scale (latest/older), participants (adult/paediatric), raters (nurse, physician or expert), method of reliability (intra/inter-rater), reliability statistics (weighted/unweighted kappa) and the origin and publication year of the study. The effect size was obtained by the Z-transformation of reliability coefficients. Data were pooled with random-effects models and a meta-regression was performed based on the method of moments estimator. **Results:** A total of 19 studies from six countries were included in the analysis. The pooled coefficient for the ESI triage scales was substantial at 0.791 (95% confidence interval: 0.787–0.795). Agreement was higher with the latest and adult versions of the scale and among expert raters, compared to agreement with older and paediatric versions of the scales and with other groups of raters, respectively. **Conclusion:** ESI triage scales showed an acceptable level of overall reliability. However, ESI scales require more development in order to see full agreement from all rater groups. Further studies concentrating on other aspects of reliability assessment are needed.

Keywords: Triage; Emergency Treatment; Algorithm; Reliability and Validity; Meta-Analysis.

الملخص: الهدف على الرغم من أن لأنظمة الفرز على أساس مؤشر شدة الطوارئ (ESI) العديد من المزايا من حيث البساطة والوضوح، إلا أن الأبحاث السابقة قد شككت في مصداقيتها عند الممارسة. ولهذا كان الهدف من هذا التحليل التلوبي (البعدي) معرفة مغولية مقاييس الفرز. **ط.** الطريقة: تم البحث في قواعد البيانات الالكترونية منذ تاريخ إنشائها إلى مارس 2014، وتم اختيار المواد مطابقة للدلائل الإرشادية لدراسات التبليغ والمغولية والمغولية (GRRAS). وقام باختبار مستقلان بفحص ملخصات مختارة واستخرجها البيانات في الفئات التالية: نسخة مقاييس (الآخر / الأكبر) والمساركون (الكبار / الأطفال)، المقيمون (مرضية، طبيب أو خبير) أو أسلوب المغولية (وسط / بين المقيمين)، والإحصاءات المغولية (الموزونة / غير الموزونة كابا) والمصدر وعام نشر الدراسة. وتم الحصول على حجم التأثير من قبل تحويل معاملات المغولية. وتم تجميع البيانات مع نماذج الآثار العشوائية. وتم إجراء التحوف التالي إحصائياً على أساس طريقة تقدير اللحظات. **النتائج:** تم إدراج ما مجموعه 19 دراسة من ست دول في التحليل. وكان المعامل التبليغى لموازین الفرز ESI كبيراً ويعادل 0.791 (فواصل الموثوقية عند 95% يتراوح بين 0.787–0.795). وكان التوافق أكبر بين الكبار وأحدث إصدارات والكبار الحجم والمقيمين بين الخبراء، مقارنة بالاتفاق مع الإصدارات القديمة والأطفال من المقاييس ومع مجموعات أخرى من المقيمين، على التوالي. **الخلاصة:** أظهرت جداول الفرز ESI وعلى وجه العموم مستوى مقبولاً من المغولية. ومع ذلك، فإن موازین ESI تتطلب المزيد من التطوير حتى يمكن الوصول لاتفاق كامل عند جميع الفئات المقيدة. وهناك حاجة إلى مزيد من الدراسات التي ترتكز على جوانب أخرى من تقييم المغولية.

مفتاح الكلمات: الفرز؛ علاج حالات الطوارئ؛ الخوارزمية: المغولية التحليل التلوبي (البعدي).

ADVANCES IN KNOWLEDGE

- It is important to determine the reliability of triage scales as this reveals the consistency of patient prioritisation within emergency departments.
- The results of this meta-analysis demonstrate that the Emergency Severity Index (ESI) triage scale has substantial agreement and shows acceptable overall reliability when implemented in emergency departments within and outside of the USA.
- However, the ESI was also found to have a tendency to allocate patients to level 2.

APPLICATION TO PATIENT CARE

- A triage scale enables emergency departments to allocate resources to the most critically ill patients. It is therefore important for medical personnel to be aware of the reliability of the scale as inconsistent triage decisions may result in under-triage and put patients' health in danger.

IN EMERGENCY DEPARTMENTS (EDs), PATIENTS are categorised based on their clinical acuity; thus, the more critically ill the patient, the sooner treatment is delivered and care needs are addressed.¹ The Emergency Severity Index (ESI) is a five-level ED triage algorithm designed to stratify patients into groups based on clinical need. The ESI is continuously developed by physicians in the USA and has been adopted by several other developed countries.^{2,3} It has also been endorsed by the American College of Emergency Physicians and the Emergency Nurses Association.⁴

Many studies have investigated the validity and reliability of the ESI triage scale in both adult and paediatric populations.^{2,3,5–10} However, the extent to which the ESI triage scale is used in triage nurses' decision-making outside of the USA is still unclear, especially considering the wide variety of healthcare systems currently in existence around the world.¹¹ With regards to this, Andersson *et al.* addressed contextual influences on the triage decision-making process in rural Sweden.¹² The reliability of triage scales outside of the USA should therefore be assessed for internal consistency, repeatability and inter-rater agreement.^{1,11} While the *kappa* statistic is most commonly used to measure inter-rater agreement, this statistic can be influenced by incidence, bias and levels of scale, potentially generating misleading results.^{13–15} Additionally, weighted *kappa* statistics have been reported to reveal deceptively high reliability coefficients.¹¹ Therefore, a pooled estimate of a reliability coefficient is more practical in the identification of significant differences among reliability methods.

A meta-analysis is a systematic approach for the introduction, evaluation, synthesis and unification of results with reference to a specific research question. It also produces the strongest evidence for intervention and is therefore an appropriate method to gain insight regarding the reliability of triage scales.¹⁶ A review by Christ *et al.* on the reliability of the ESI scale demonstrated *kappa* statistics ranging from 0.46 (moderate) to 0.98 (almost perfect).¹⁷ This considerable variation in *kappa* statistics indicates a discernable gap in the reliability of the triage scale. Thus, the aim of this meta-analysis was to review the reliability of the ESI triage scale in a variety of contexts.

Methods

This study was performed in March 2014. The first phase consisted of a literature search using the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, MEDLINE®/PubMed,

Google Scholar and Cochrane Library databases. Studies published before 1 March 2014 in these databases and found using the following search terms were included: reliability; triage; system; scale; agreement; emergency, and emergency severity index. All studies in English identified by the database search were examined by two researchers to identify eligible articles regarding the reliability of the ESI. Irrelevant or duplicate results were eliminated. Reference lists of acceptable publications were also examined to identify further articles for inclusion in the study.

Articles were chosen for inclusion according to the Guidelines for Reporting Reliability and Agreement Studies.¹⁸ According to these guidelines, studies were only included in the analysis if they reported more than six of the following eight items in sufficient detail: sample size; number of raters; number of subjects; sampling method; rating process; statistical analysis, and reliability coefficient. Disagreements were resolved by consensus with a third researcher. Articles that did not report the type of reliability (either inter-rater reliability, intra-rater reliability or internal consistency) were excluded from the analysis. Researchers also recorded moderator variables relating to participants, raters and the origin and year of publication of the study as well as studies which were conducted based on the latest version (2012) of the ESI triage scale.

In the next phase, further information was retrieved from the articles, including: age group of participants and size of cohort; raters' professions and overall number of raters; instruments used (e.g. live or scenario-based cases); country of origin and year of publication of study; reliability coefficient, and type of reliability. Reliability was determined by inter-rater reliability (weighted or unweighted *kappa* coefficients), intra-rater reliability (reliability statistics including intraclass correlation coefficient, Pearson correlation coefficient or Spearman's rank correlation coefficient) and internal consistency (alpha coefficients) statistics. Authors of research articles were contacted for supplementary information if necessary.

In the meta-regression, each rater sample was considered to be a unit of analysis. If the same sample was reported in two or more articles, it was included only once in the analysis. In contrast, if several samples with different populations were reported in one study, each sample was included as a separate unit of analysis.

Data were pooled for all three types of reliability. Many articles reported a reliability coefficient using the *kappa* statistic; it could be considered an r-type coefficient ranging from -1.00 to +1.00. Standard agreement was categorised as poor ($\kappa = 0.00\text{--}0.20$), fair ($\kappa = 0.21\text{--}0.40$), moderate ($\kappa = 0.41\text{--}0.60$), substantial ($\kappa = 0.61\text{--}0.80$) or almost perfect ($\kappa =$

0.81–1.00).¹³ The *kappa* statistic can be treated as a correlation coefficient in meta-analyses.¹⁹ In order to obtain the correct interpretation, back-transformation (Z to r transformation) of pooled effect sizes to the level of primary coefficients was performed.^{20,21} Fixed- and random-effect models were applied. Data were analysed using Comprehensive Meta-Analysis (CMA) Software, Version 2.2.050 (Biostat Inc., Englewood, New Jersey, USA).

A simple meta-regression analysis was performed according to the method of moments estimator.²² In the meta-regression model, effect size was a dependent variable while studies and subject characteristics were considered independent variables in order to discover potential predictors of reliability coefficients. Z-transformed reliability coefficients were regressed based on variables of country origin and year of publication as well as studies based on the latest version of triage scale versus those of a prior version. Distance was defined as the distance from each study's country of origin to the city of Boston, Massachusetts, in the USA (where the ESI triage scale originated). Meta-regression was performed using a random-effects model due to the presence of significant inter-study variation.²³

This study received ethical approval from the Research & Ethics Committee of Mashhad University of Medical Science in Mashhad, Iran.

Results

A total of 260 primary citations relevant to the reliability of the ESI triage scale were identified during the literature search. However, only 19 unique citations (7.3%) met the inclusion criteria.^{2,3,5–10,24–34} The studies were organised into subgroups according to participants (adult/paediatric); raters (nurses/physicians/experts); method of reliability (intra-/inter-rater); reliability statistics (weighted/unweighted *kappa* statistics), and by country of origin and publication year. The level of agreement among the researchers regarding the final selection of articles for the meta-analysis was almost perfect ($\kappa = 1.0$).

A total of 40,579 cases (both paper-based case scenarios and live triage cases) were included in the initial analysis. Among the 19 studies meeting the required criteria, the reliability of the ESI triage scale had been assessed in six different countries with publication years ranging from 2000–2013 (median year of publication: 2009). In addition, 70% had been conducted using the latest version of the triage scale. The inter-rater reliability method was used in all studies except for one, which used intra-rater reliability. None of the studies in the analysis had used

the alpha coefficient to report internal consistency. The weighted *kappa* coefficient was the most commonly utilised statistic [Table 1].

The overall pooled coefficient for the ESI triage scale was substantial, at 0.791 (95% confidence interval [CI]: 0.752–0.825). Participants' pooled coefficients ranged from substantial (0.732; 95% CI: 0.625–0.812) for nurse-expert agreement to almost perfect (0.900; 95% CI: 0.570–0.980) for expert-expert agreement [Figures 1 and 2]. Agreement regarding adult/paediatric versions of the ESI triage scale was almost perfect for both adult (0.815; 95% CI: 0.753–0.862) and paediatric patients (0.769; 95% CI: 0.747–0.837). Additionally, almost perfect agreement was noted for paper-based scenario assessments (0.824; 95% CI: 0.778–0.861) and substantial agreement was observed for live case assessments (0.694; 95% CI: 0.575–0.784). Inter- and intra-rater reliability agreement was 0.786 (95% CI: 0.745–0.821) and 0.873 (95% CI: 0.801–0.921), respectively. Substantial agreement was found for both weighted (0.796; 95% CI: 0.751–0.834) and unweighted *kappa* statistics (0.770; 95% CI: 0.674–0.841). Agreement for the latest version of the ESI was 0.833 (95% CI: 0.774–0.878), whereas it was 0.808 (95% CI: 0.762–0.846) for previous versions.

Only six studies presented a 5 x 5 contingency table to show the frequency distribution of triage decisions for each ESI level between two raters [Table 2].^{2,3,5,7,9,24} The overall agreement in these studies was 78.55%. Agreement for each ESI level was 1.12%, 23.40%, 19.55%, 18.81% and 15.67% for levels 1 to 5, respectively, while disagreement was 0.25%, 4.07%, 6.10%, 6.90% and 4.12%, respectively. A total of 80% of all disagreements concerned levels 3 to 5 (17.12% out of 21.44%). Only ESI level 2 decisions showed a wide distribution across all levels [Table 2].

Table 3 shows the meta-regression analysis based on the method of moments for moderators (distance from ESI origin, year of publication and ESI version). Studies using the latest version of the ESI scale and which had been published more recently showed significantly higher pooled coefficients. However, higher pooled coefficients were not indicated for studies conducted closest in geographical distance to Boston, USA [Figure 3].

Discussion

The results of this study indicate that the overall reliability of the ESI triage scale is substantial. The ESI showed an acceptable level of reliability which guarantees consistent decisions regarding the allocation of patients to appropriate categories; thereby supporting evidence-based practice in EDs.^{17,35}

Table 1: Studies on the reliability of the Emergency Severity Index triage scale included in the meta-analysis (N = 19)

Author and year of study	Participant	Raters	Instrument	Method	Statistic	CO
Wuerz <i>et al.</i> ⁵ 2000	Adult Adult	EE NP	Scenario Live cases	Inter Inter	Kw	USA
Wuerz <i>et al.</i> ²⁵ 2001	Adult Adult	NN NP	Scenario Scenario	Inter Inter	Kw	USA
Travers <i>et al.</i> ²⁶ 2002	Adult	NE	Scenario	Inter	Kw	USA
Eitel <i>et al.</i> ²⁷ 2003	Adult Adult	NE NN	Scenario Live cases	Inter Inter	Kw	USA
Tanabe <i>et al.</i> ²⁴ 2004	Adult Adult	NE NE	Scenario Scenario	Inter Inter	Kw	USA
Worster <i>et al.</i> ⁶ 2004	Adult	NN	Scenario	Inter	Kw	Canada
Travers <i>et al.</i> ²⁸ 2002	Adult Paediatric	NN NN	Scenario Scenario	Inter Inter	Kw	USA
Choi <i>et al.</i> ²⁹ 2009	Adult Adult	NE NP	Live cases Live cases	Inter Inter	Kw Kw	Korea
Storm-Versloot <i>et al.</i> ³ 2009	Adult Adult Adult Adult	NN NN NN NN	Scenario Scenario Scenario Scenario	Intra Inter Intra Inter	Kuw Kuw Kw Kw	Netherlands
Grossman <i>et al.</i> ² 2011	Adult Adult Adult Adult	EE EE NE NE	Scenario Scenario Scenario Scenario	Inter Inter Inter Inter	r Kw r Kw	Switzerland
Platts-Mills <i>et al.</i> ³⁰ 2010	Adult Adult	NE NP	Live cases Live cases	Inter Inter	Kw	USA
Grossmann <i>et al.</i> ³¹ 2012	Adult	EE	Scenario	Inter	Kw	Switzerland
Baumann <i>et al.</i> ⁷ 2005	Paediatric Paediatric Paediatric Paediatric Paediatric	NE NE NN NP PE	Live cases Scenario Scenario Scenario Scenario	Inter Inter Inter Inter Inter	Kw	USA
Durani <i>et al.</i> ³² 2007	Paediatric Paediatric Paediatric	NN NP PP	Scenario Scenario Scenario	Inter	Kw	USA
Durani <i>et al.</i> ⁸ 2009	Paediatric Paediatric Paediatric Paediatric Paediatric Paediatric	NN NN NP NP PP PP	Scenario Scenario Scenario Scenario Scenario Scenario	Inter	Kuw Kw Kuw Kw Kuw Kw	USA
Travers <i>et al.</i> ⁹ 2009	Paediatric Paediatric	NE NN	Live cases Scenario	Inter	Kw	USA
Green <i>et al.</i> ¹⁰ 2012	Paediatric Paediatric Paediatric Paediatric	NN NN NP NP	Live cases Live cases Live cases Live cases	Inter	ICC Kuw ICC Kuw	USA
Jafari-Rouhi <i>et al.</i> ³³ 2013	Paediatric Paediatric Paediatric	NN NP NP	Scenario Scenario Scenario	Inter	Kuw r Kuw	Iran
Ho <i>et al.</i> ³⁴ 2007	Adult	EE	Live cases	Inter	Kw	USA

CO = country of origin; Inter = inter-rater reliability; Kw = weighted kappa; NP = nurse-physician; NN = nurse-nurse; NE = nurse-expert; Intra = intra-rater reliability; Kuw = unweighted kappa; EE = expert-expert; r = correlation coefficient; PE = physician-expert; PP = physician-physician.

In the ESI, levels 3 to 5 are generally defined by the requirement and availability of resources; because this varies from one setting to another, most disagreements are to do with these levels. This was reflected in the current study. Fortunately, these levels indicate semi- to non-urgent patients which means that misclassifications rarely occur among those who are critically ill. Another interesting observation of the current meta-analysis was the strong tendency towards level 2 categorisation. Although this may prevent the under-triaging of certain patients, it could also create a significant disturbance in the patient flow and interfere with the overall functioning of the ED.

Diverse pooled reliability coefficients were observed regarding participants, patients, raters, reliability methods and statistics among the studies. The results of the meta-analysis found that agreement with the latest and the adult versions of the ESI and among experts was higher than those with previous or paediatric versions and among the other groups of raters. The different results determined by these moderator variables could lead to further studies to explore these variables in more depth. The reliability and consistency of the ESI across EDs in different countries has been documented and supported by scientific evidence.³⁵ This could be due to the fact that the simplicity and objectivity of the ESI algorithm plays a pivotal role in helping clinicians reach optimal agreement.³⁶ In the current study, the analysis of reliability in studies of non-American origin show that the ESI triage scale can be adopted successfully in countries outside of the USA in spite of cultural differences.

Studies using the latest version of the ESI scale and those which were published more recently showed higher agreement. As the ESI triage scale has been

updated several times and its reliability has improved over the years, this indicates that revisions have been effective. Additionally, this emphasises the need for EDs to update their triage systems according to the latest versions of the chosen triage scale.

In general, intra-rater reliability is more satisfactory than inter-rater reliability.³⁷ This was highlighted in the current study, which revealed almost perfect agreement in the former as compared to the substantial agreement yielded by the latter. While intra- and inter-rater reliability are intended to report the degree to which measurements taken by the same and different observers are similar, other methods of examining reliability have remained uncommon in studies regarding triage reliability.³⁸

The current study's analysis demonstrated that the weighted *kappa* coefficient showed substantial agreement. In fact, the weighted *kappa* coefficient reveals higher reliability than the unweighted *kappa* coefficient because it places more emphasis on the larger differences between ratings than on the smaller ones.³⁹ In practical terms, the misallocation of critically-ill patients by even a single ESI level can endanger their clinical outcomes; unweighted *kappa* statistics therefore provide a more realistic estimation of triage scale reliability.¹¹

A number of limitations of this study must be noted. While the ESI showed an acceptable level of reliability, it is important to remember that there is a considerable gap between research and clinical practice even at the best of times.⁴⁰ In addition, almost all of the studies used weighted *kappa* statistics to report reliability coefficients. As weighted *kappa* statistics generally overestimate the reliability of a triage scale, it is necessary to interpret these results with caution.¹¹ Therefore, it is likely that the ESI is in fact only moderately reliable. Furthermore, none of the studies in the analysis reported raw agreement for each

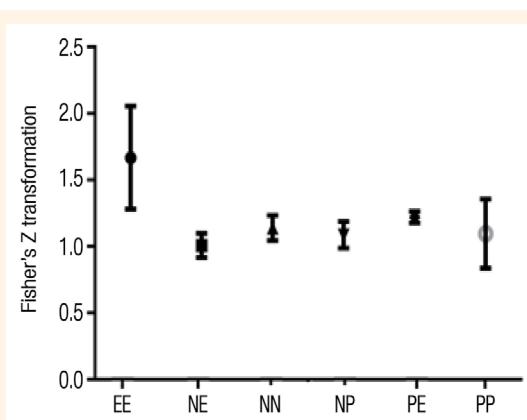


Figure 1: Fisher's Z-transformation showing pooled estimates of participants' reliability (random-effect model) among studies relating to the reliability of the Emergency Severity Index.^{2,3,5–10,24–34}

EE = expert-expert; NE = nurse-expert; NN = nurse-nurse; NP = nurse-physician; PE = physician-expert; PP = physician-physician.

Rater group	Statistics		Correlation 95%			CI
	Correlation	Lower limit	Upper limit	Z value	P value	
Expert-Expert	0.900	0.570	0.980	3.497	0.000	0.00–1.00
Nurse-Expert	0.732	0.625	0.812	9.142	0.000	0.00–1.00
Nurse-Nurse	0.799	0.739	0.846	14.551	0.000	0.00–1.00
Nurse-Physician	0.760	0.582	0.868	5.919	0.000	0.00–1.00
Physician-Expert	0.840	0.813	0.863	28.404	0.000	0.00–1.00
Physician-Physician	0.842	0.479	0.959	3.407	0.001	0.00–1.00
Overall	0.820	0.797	0.840	33.937	0.000	0.00–1.00

Figure 2: Pooled estimates of measures of raters' reliability (random-effect model) using weighted *kappa* statistics among studies relating to the reliability of the Emergency Severity Index.^{2,3,5–10,24–34}

CI = confidence interval.

Table 2: Distribution of triage decision-making relating to each Emergency Severity Index triage category among emergency department raters in six studies^{2,3,5,7,9,24}

Rater 2 ESI category	Rater 1 ESI category n (%)					
	1	2	3	4	5	Total
1	35 (1.12)	9	0	0	0	44
2	7	730 (23.39)	102	26	3	868
3	0	82	610 (19.55)	91	12	795
4	0	18	77	587 (18.81)	85	767
5	0	7	16	134	489 (15.67)	646
Total	42	846	805	838	589	3,120 (100.00)

ESI = Emergency Severity Index.

individual ESI level and only a few studies presented a contingency table for inter-rater agreement. Since this study was limited to overall reliability, some inconsistencies may exist across each ESI level. Finally, research has indicated that raters' experiences could affect the reliability and validity of triage decision-making; these experiences were not reported among any of the studies in this analysis.⁴¹

Conclusion

Overall, the ESI triage scale was shown to display an acceptable level of reliability in this meta-analysis. However, there is a need for further development of the scale in order to reach almost perfect agreement. The reliability of triage scales requires a more comprehensive approach, including a thorough assessment of all aspects of reliability. In light of this, further studies should concentrate on the reliability of triage scales in terms of specific moderator variables, such as the version of the ESI used.

CONFLICT OF INTEREST

Table 3: Meta-regression of Fisher's Z-transformed *kappa* coefficients on predictor variables*

Independent variable	B	SEb	P
Latest ESI version	0.302	0.018	0.00
Distance from ESI origin**	-0.00	0.000	0.53
Publication year	0.015	0.002	0.00

SEb = simultaneous equation bias; ESI = Emergency Severity Index.

*Using studies relating to ESI reliability with weighted *kappa* coefficients.^{2,3,5–10,24–34} **The ESI triage scale originated in Boston, Massachusetts, USA.

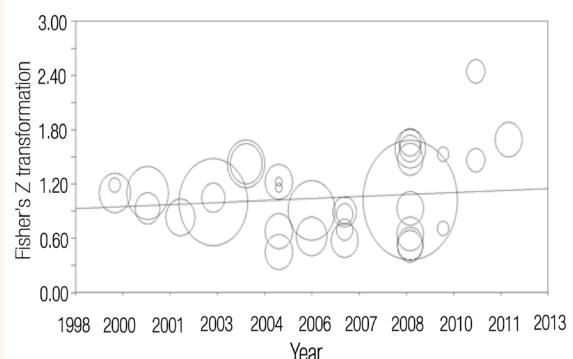


Figure 3: Fisher's Z-transformation of *kappa* coefficients regarding the year of publication among studies relating to the reliability of Emergency Severity Index.^{2,3,5–10,24–34}

The authors declare no conflicts of interest.

References

- Mirhaghi AH, Roudbari M. A survey on knowledge level of the nurses about hospital triage. *Iran J Crit Care Nurs* 2011; 3:167–74.
- 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; 57:257–64. doi: 10.1016/j.annemergmed.2010.07.021.
- 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; 26:556–60. doi: 10.1136/emej.2008.059378.
- Fernandes CM, Tanabe P, Gilboy N, Johnson LA, McNair RS, Rosenau AM, et al. Five-level triage: A report from the ACEP/ENA Five-level Triage Task Force. *J Emerg Nurs* 2005; 31:39–50. doi: 10.1016/j.jen.2004.11.002.
- Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000; 7:236–42. doi: 10.1111/j.1553-2712.2000.tb01066.x.
- Worster A, Gilboy N, Fernandes CM, Eitel D, Eva K, Geisler R, et al. Assessment of inter-observer reliability of two five-level triage and acuity scales: A randomized controlled trial. *CJEM* 2004; 6:240–5.
- Baumann MR, Strout TD. Evaluation of the Emergency Severity Index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med* 2005; 12:219–24. doi: 10.1197/j.aem.2004.09.023.
- Durani Y, Brecher D, Walmsley D, Attia MW, Loiselle JM. The Emergency Severity Index version 4: Reliability in pediatric patients. *Pediatr Emerg Care* 2009; 25:504–7. doi: 10.1097/PEC.0b013e3181b0a0c6.
- Travers DA, Waller AE, Katzenelson J, Agans R. Reliability and validity of the emergency severity index for pediatric triage. *Acad Emerg Med* 2009; 16:843–9. doi: 10.1111/j.1553-2712.2009.00494.x.
- Green NA, Durani Y, Brecher D, DePiero A, Loiselle J, Attia M. Emergency Severity Index version 4: A valid and reliable tool in pediatric emergency department triage. *Pediatr Emerg Care* 2012; 28:753–7. doi: 10.1097/PEC.0b013e3182621813.
- Göransson KE, Ehrenberg A, Marklund B, Ehnfors M. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005; 19:432–8. doi:

- 10.1111/j.1471-6712.2005.00372.x.
12. Andersson AK, Omberg M, Svedlund M. Triage in the emergency department: A qualitative study of the factors which nurses consider when making decisions. *Nurs Crit Care* 2006; 11:136–45. doi: 10.1111/j.1362-1017.2006.00162.x.
 13. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther* 2005; 85:257–68.
 14. 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. doi: 10.1016/j.jclinepi.2010.01.012.
 15. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Fam Med* 2005; 37:360–3.
 16. Petitti D. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis. 1st ed. New York, USA: Oxford University Press, 1994. P. 69.
 17. Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. *Dtsch Arztebl Int* 2010; 107:892–8. doi: 10.3238/arztebl.2010.0892.
 18. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 48:661–71. doi: 10.1016/j.jclinepi.2011.01.016.
 19. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int J Methods Psychiatr Res* 2009; 18:169–84. doi: 10.1002/mpr.289.
 20. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. San Diego, California, USA: Academic Press, 1985. Pp. 76–81.
 21. Rosenthal R. Meta-Analytic Procedures for Social Research. Newbury Park, California, USA: Sage Publications Inc., 1991. Pp. 43–89.
 22. Chen H, Manning AK, Dupuis J. A method of moments estimator for random effect multivariate meta-analysis. *Biometrics* 2012; 68:1278–84. doi: 10.1111/j.1541-0420.2012.01761.x.
 23. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011; 342:d549. doi: 10.1136/bmj.d549.
 24. 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; 11:59–65. doi: 10.1197/j.aem.2003.06.013.
 25. Wuerz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and refinement of the emergency severity index. *Acad Emerg Med* 2001; 8:170–6. doi: 10.1111/j.1553-2712.2001.tb01283.x.
 26. 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; 28:395–400. doi: 10.1067/men.2002.127184.
 27. 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; 10:1070–80. doi: 10.1197/S1069-6563(03)00350-6.
 28. Travers DA, Agans R, Eitel D, Mecham N, Rosenau A, Tanabe P, et al. Reliability evaluation of the Emergency Severity Index version 4. *Acad Emerg Med* 2006; 13:S126. doi: 10.1111/j.1553-2712.2006.tb02227.x.
 29. Choi M, Kim J, Choi H, Lee J, Shin S, Kim D, et al. Reliability of Emergency Severity Index version 4. *Ann Emerg Med* 2009; 54:S95–6. doi: 10.1016/j.annemergmed.2009.06.336.
 30. Platts-Mills TF, Travers D, Biese K, McCall B, Kizer S, LaMantia M, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. *Acad Emerg Med* 2010; 17:238–43. doi: 10.1111/j.1553-2712.2010.00670.x.
 31. Grossmann FF, Zumbrunn T, Frauchiger A, Delport K, Bingisser R, Nickel CH. At risk of undertriage? Testing the performance and accuracy of the emergency severity index in older emergency department patients. *Ann Emerg Med* 2012; 60:317–25. doi: 10.1016/j.annemergmed.2011.12.013.
 32. Durani Y, Brecher D, Walmsley D, Attia M, Loiselle J. The Emergency Severity Index (version 4): Reliability in pediatric patients. *Acad Emerg Med* 2007; 14:S95. doi: 10.1197/j.aem.2007.03.704.
 33. Jafari-Rouhi AH, Sardashti S, Taghizadieh A, Soleimanpour H, Barzegar M. The Emergency Severity Index, version 4, for pediatric triage: A reliability study in Tabriz Children's Hospital, Tabriz, Iran. *Int J Emerg Med* 2013; 6:36. doi: 10.1186/1865-1380-6-36.
 34. Ho A, Tintinalli J, Travers D. Emergency Severity Index for psychiatric triage. *Acad Emerg Med* 2007; 14:S135. doi: 10.1197/j.aem.2007.03.704.
 35. van Veen M, Moll HA. Reliability and validity of triage systems in paediatric emergency care. *Scand J Trauma Resusc Emerg Med* 2009; 17:38. doi: 10.1186/1757-7241-17-38.
 36. Agency for Healthcare Research and Quality, United States Department of Health & Human Services. Emergency Severity Index (ESI): A triage tool for emergency department - 2012 edition of the implementation handbook, version 4. From: www.ahrq.gov/professionals/systems/hospital/esi/esihandbk.pdf Accessed: Apr 2014.
 37. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Phys Ther* 1994; 74:777–88.
 38. Hogan TP, Benjamin A, Brezinski KL. Reliability methods: A note on the frequency of use of various types. *Educ Psychol Meas* 2000; 60:523–31. doi: 10.1177/00131640021970691.
 39. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70:213–20. doi: 10.1037/h0026256.
 40. Le May A, Mulhall A, Alexander C. Bridging the research-practice gap: Exploring the research cultures of practitioners and managers. *J Adv Nurs* 1998; 28:428–37. doi: 10.1046/j.1365-2648.1998.00634.x.
 41. Göransson KE, Ehrenberg A, Marklund B, Ehnfors M. Emergency department triage: Is there a link between nurses' personal characteristics and accuracy in triage decisions? *Accid Emerg Nurs* 2006; 14:83–8. doi: 10.1016/j.aen.2005.12.001.