

The Reliability of the Canadian Triage and Acuity Scale: Meta-analysis

Amir Mirhaghi, Abbas Heydari, Reza Mazlom, Mohsen Ebrahimi¹

Evidence-Based Caring Research Center, Department of Medical-Surgical Nursing, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, ¹Department of Emergency Medicine, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

Abstract

Background: Although the Canadian Triage and Acuity Scale (CTAS) have been developed since two decades ago, the reliability of the CTAS has not been questioned comparing to moderating variable. **Aims:** The study was to provide a meta-analytic review of the reliability of the CTAS in order to reveal to what extent the CTAS is reliable. **Materials and Methods:** Electronic databases were searched to March 2014. Only studies were included that had reported samples size, reliability coefficients, adequate description of the CTAS reliability assessment. The guidelines for reporting reliability and agreement studies (GRRAS) were used. Two reviewers independently examined abstracts and extracted data. The effect size was obtained by the z-transformation of reliability coefficients. Data were pooled with random-effects models and meta-regression was done based on method of moments estimator. **Results:** Fourteen studies were included. Pooled coefficient for the CTAS was substantial 0.672 (CI 95%: 0.599-0.735). Mistriage is less than 50%. Agreement upon the adult version, among nurse-physician and near countries is higher than pediatrics version, other raters and farther countries, respectively. **Conclusion:** The CTAS showed acceptable level of overall reliability in the emergency department but need more development to reach almost perfect agreement.

Keywords: Algorithm, emergency treatment, meta-analysis, reliability and validity, triage

Address for correspondence: Mr. Amir Mirhaghi, Department of Medical-Surgical Nursing, School of Nursing and Midwifery, Chahrrah-e-Doktora, Daneshgah st., Mashhad, Khorasan Razavi, Iran. E-mail: mirhaghia@mums.ac.ir

Introduction

Patients are categorized based on clinical acuity in the emergency departments (EDs) so the more critically-ill patient is, the more immediate treatment and care needs.^[1] The Canadian Triage and Acuity Scale (CTAS) is a five-level emergency department triage algorithm that has been continuously developed in Canada and subjected to several studies.^[2-14] The CTAS, a descriptive 5-point triage scale, has been approved by the Canadian Association of Emergency Physicians, the National Emergency Nurses Affiliation, the Canadian Pediatric Society and

the Society of Rural Physicians of Canada. The CTAS is based on a comprehensive list of patients' complaints is used to ascertain the triage level. Each complaint has been described in details covering high-risk indicators.^[15]

Several studies^[2-14] have investigated the validity and reliability of the CTAS in adult and pediatric populations; but it's still unclear to what extent the CTAS would support consistency in triage nurses' decision making in Canada comparing to other countries, considering the wide variety of health care systems around the world.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Mirhaghi A, Heydari A, Mazlom R, Ebrahimi M. The reliability of the Canadian triage and acuity scale: Meta-analysis. *North Am J Med Sci* 2015;7:299-305.

Access this article online

Quick Response Code:



Website:
www.najms.org

DOI:
10.4103/1947-2714.161243

Besides, some studies^[16,17] have addressed contextual influences on the triage decision making process, therefore it is necessary to discover the impact of these variables on the reliability of triage scale. However some studies reported moderate consistency for CTAS,^[18] it needs to be more explored in terms of participants, statistics, instruments and other influencing criteria as well as mistriage.

The reliability of triage scales should be assessed by internal consistency, repeatability and inter-rater agreement.^[8] However, kappa has been the most commonly used statistics to measure inter-rater agreement, so it is worth-mentioning that kappa statistics could be influenced by incidence, bias and levels of scale lead to misleading results.^[19-21] It is reported that weighted kappa statistics could reveal high and deceiving reliability coefficients.^[8] Therefore computing a pooled estimate of a reliability coefficient could help us identify significant differences among reliability methods.

Meta-analysis is a systematic approach for introduction, evaluation, synthesis and unifying results in relation to studying research questions. It also produces the strongest evidence for intervention.^[22] Therefore; it is an appropriate method to gain comprehensive and deep insight into the reliability of triage scale especially in regard to kappa statistics.

A review on reliability of the CTAS demonstrated that kappa ranges from 0.202 (fair) to 0.84 (almost perfect).^[18-23] The considerable variation in the kappa statistics indicates a real gap in the reliability of triage scale. So in view of the methodological limitations of the triage scale reliability, context-based triage decision making and the necessity of comprehensive insight into scale reliability in the EDs, the aim of this study was to provide a meta-analytic review of the reliability of the CTAS in order to examine to what extent the CTAS is reliable.

Materials and Methods

The University Research ethics committee approved the study. In the first phase of the study, a literature search was conducted through investigating Cinahl, Scopus, Medline, Pubmed, Google Scholar and Cochrane Library databases until the 1st March 2014. Meta-analysis has been performed from Jan to July 2014 and authors started to collect data from March 2014. Searching databases were not limited to time periods. The search terms included "Reliability", "Triage", "System", "Scale", "Agreement", "Emergency" and "Canadian Triage and Acuity Scale" [Appendix A].

Relevant citations in reference lists of final studies were hand-searched to identify additional articles regarding

the reliability of CTAS. Two researchers independently examined the search results in order to recover potentially eligible articles [Figure 1]. Authors of research articles were contacted to retrieve supplementary information if needed.

Irrelevant and duplicated results were eliminated. Irrelevant article has been defined as article which was not related in any manner to the Canadian triage and acuity scale or didn't contain reliability coefficient. Only English language publications were reviewed. Articles have been chosen according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).^[24] According to the guidelines, only those studies that had reported description for sample size, number of raters and subjects, sampling method, rating process, statistical analysis and reliability coefficients were included in the analysis. Each item was graded qualified if described in sufficient detail in the paper. Qualified paper was defined as a paper with qualifying score >6 out of the 8 criteria. Disagreements were resolved by consensus. The articles in which the type of reliability was not reported were excluded from the analyses. Researchers also recorded moderator variables such as participants, raters, origin and publication year of study.

In the next phase, participants (age-group, size), raters (profession, size), instruments (live, scenario), origin and publication year of study, reliability coefficient and method were retrieved. Reliability coefficients were extracted from articles as below:

- Inter-rater reliability: Kappa coefficient (weighted and un-weighted), intraclass correlation coefficient, Pearson correlation coefficient and Spearman rank correlation coefficient.
- Intra-rater reliability: Articles which contained reliability statistics including Pearson correlation

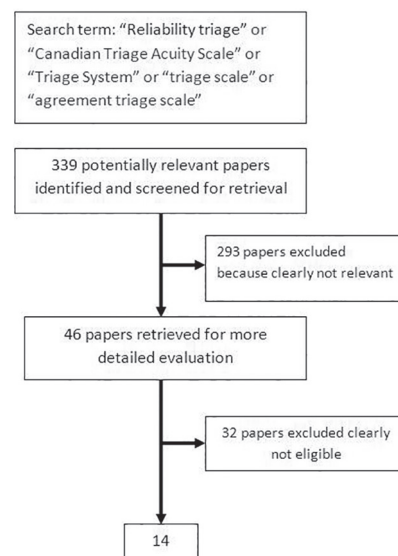


Figure 1: Results of literature search and selection process

coefficient, intraclass correlation coefficient and Spearman rank correlation coefficient were included.

- Internal consistency: Articles that reported alpha coefficients were included in the analyses.

In meta-regression, each sample was considered as a unit of analysis. If the same sample were reported in more than two articles, it was included once. In contrast, if several samples regarding different populations were reported in one study, each sample was separately included as a unit of analysis.

Pooling data was performed for all three types of reliability. The most qualified articles reported reliability coefficient using kappa statistics, so it could be considered as an r type of coefficient ranging from -1.00 to $+1.00$. Standard agreement definition was used as poor ($\kappa = 0.00-0.20$), fair ($\kappa = 0.21-0.40$), moderate ($\kappa = 0.41-0.60$), substantial ($\kappa = 0.61-0.80$), and almost perfect ($\kappa = 0.81-1.00$).^[20] Kappa could be treated as a correlation coefficient in meta-analysis.^[25] 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.^[26,27] Fixed effects and random effects models were applied. Data was analyzed using Comprehensive Meta Analysis software (Version 2.2.050).

Simple meta-regression analysis was performed according to method of moments estimator.^[28] In the meta-regression model, effect size as dependent variable, and studies and subject characteristics as independent variable were considered to discover potential predictors of reliability coefficients. Z -transformed reliability coefficients are regressed on the following variables: Origin and publication year of study. Distance was defined as distance from origin of each study to origin of CTAS (New Brunswick, Canada). Meta-regression was performed using a random effects model because of the presence of significant between-study variation.^[29]

Results

Search strategy introduced 339 primary citations relevant to the reliability of CTAS. Finally, Fourteen unique citations emerged (4.12% out of 339) which met the inclusion criteria [Figure 1]. Subgroups were organized regarding participants (Adult/pediatric), raters (nurse, physician, expert) and method of reliability (intra/inter rater), reliability statistics (weighted/un-weighted kappa) and origin and publication year of study. A. M. has over 10 years of experience in triage and M. E. as an expert, is the head of triage committee in the university. Two clinicians (A. M. and M. E.) and one statistician (R. M.) have reviewed all articles independently. Minor disagreements have been discussed to reach a consensus.

The level of agreement among reviewers through final selection of articles was almost perfect.

A total of 29225 cases were included in analysis [Figure 2]. The reliability of CTAS has been assessed in two different countries. The publication year of studies ranged from 1999-2013 with median 2005. Thirty percent of all studies have been conducted using the latest version of triage scale after 2008. Inter-rater reliability had been used in all studies except for one study using intra-rater reliability. No study in our analysis used alpha coefficient to report internal consistency in reliability analysis. Weighted kappa coefficient was the most common statistics [Table 1]. Overall pooled coefficient for the CTAS was substantial 0.672 (CI 95%: 0.599-0.735).

Participants' pooled coefficients ranged from substantial 0.651 (CI 95%: 0.402-0.811) for nurse-expert agreement, 0.670 (CI 95%: 0.073-0.913) for physician-physician agreement, 0.721 (CI 95%: 0.630-0.793) for nurse-nurse agreement to almost perfect 0.800 (CI 95%: 0.774-0.823) for nurse-physician agreement [Figure 3].

Agreement regarding adult and pediatric version of the CTAS was substantial 0.746 (CI 95%: 0.675-0.804) for adult and 0.598 (CI 95%: 0.375-0.714) for pediatrics.

Agreement regarding paper-based scenario assessment was substantial 0.698 (CI 95%: 0.620-0.762) while it was almost perfect 0.900 (CI 95%: 0.875-0.920) for real live cases assessment.

Agreement regarding inter-rater and intra-rater reliability was substantial 0.708 (CI 95%: 0.629-0.773) and almost perfect 0.800 (CI 95%: 0.773-0.824) respectively.

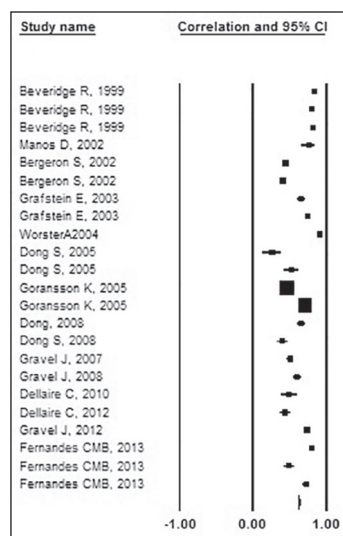


Figure 2: Pooled estimates of participants' reliability coefficients (Random effects model)

Table 1: Studies on reliability of CTAS

Studies	Patient	Participants*	Instrument	Methods**	Statistics***	Country
Beveridge et al., ^[2] 1999	Adult	NN	Scenario	Inter	Kw (0.84)	Canada
	Adult	NP	Scenario	Inter	Kw (0.80)	
	Adult	PP	Scenario	Inter	Kw (0.83)	
Bergeron et al., ^[3] 2002	Pediatrics	NN	Scenario	Inter	Kw (0.45)	Canada
	Pediatrics	PP	Scenario	Inter	Kw (0.41)	
Manos et al., ^[4] 2002	Adult	NN	Scenario	Inter	Kw (0.77)	Canada
Grafstein et al., ^[5] 2003	Adult	NN	Scenario	Inter	Kuw (0.66)	Canada
	Adult	NN	Scenario	Inter	Kw (0.75)	
Worster et al., ^[6] 2004	Adult	NN	Scenario	Inter	Kw (0.91)	Canada
Dong et al., ^[7] 2005	Adult	NE	Scenario	Inter	Kuw (0.26)	Canada
	Adult	NE	Scenario	Inter	Kw (0.53)	
Goransson et al., ^[8] 2005	Adult	NN	Scenario	Inter	Kuw (0.46)	Sweden
	Adult	NN	Scenario	Inter	Kw (0.71)	
Gravel et al., ^[9] 2007	Pediatrics	NN	Scenario	Inter	Kuw (0.51)	Canada
Dong et al., ^[37] 2007	Adult	NN	Scenario	Inter	Kuw (0.66)	Canada
	Adult	NN	Scenario	Inter	Kw (0.40)	
Gravel et al., ^[10] 2008	Pediatrics	NN	Scenario	Inter	Kw (0.61)	Canada
Dellaire et al., ^[11] 2010	Adult	NN	Scenario	Inter	Kw (0.50)	Canada
Dellaire et al., ^[12] 2012	Adult	NN	Scenario	Inter	Kw (0.44)	Canada
Gravel et al., ^[13] 2012	Pediatrics	NE	Scenario	Inter	Kw (0.74)	Canada
Fernandes et al., ^[14] 2013	Adult	NN	Scenario	Intra	Kw (0.80)	Canada
	Adult	NN	Scenario	Inter	Kw (0.50)	
	Adult	NN	Scenario	Inter	Kw (0.73)	

*EE = Expert-expert, NE = Nurse-expert, NN = Nurse-nurse, NP = Nurse-physician, PE = Physician-expert, PP = Physician-physician, **Inter = Inter-rater reliability, Intra = Intra-rater reliability, ***Kw = Weighted kappa, Kuw = Un-weighted kappa, r = Correlation coefficient, ICC = Intraclass coefficient, (Coefficient)

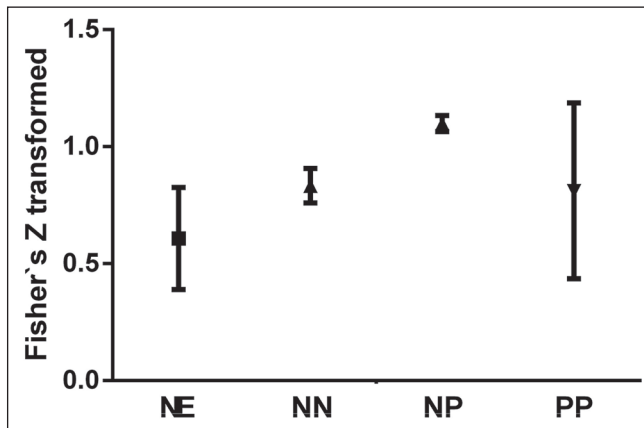


Figure 3: Fisher's Z transformed pooled estimates of participants' reliability (Random effects model) (NE = Nurse-expert, NN = Nurse-nurse, NP = Nurse-physician, PP = Physician-physician)

Agreement relating to weighted kappa was substantial 0.714 (CI 95%: 0.639-0.775) and also moderate 0.475 (CI 95%: 0.127-0.389) for un-weighted kappa. Also agreement regarding most updated version was substantial 0.621 (CI 95%: 0.446-0.751) and 0.742 (CI 95%: 0.655-0.811) for prior version.

Eight studies^[4,7-9,11,30-32] have reported (5 × 5) contingency table to show frequency distribution of triage decisions upon each CTAS level between two raters [Table 2]. Overall agreement was (57.18%). Agreement for each CTAS level was CTAS L-1 (18.14%), CTAS L-2 (7.63%),

CTAS L-3 (7.90%), CTAS L-4 (9.18%), CTAS L-5 (14.33%) and disagreement was (5.80%), (7.89%), (12.84%), (12.54%), (3.76%) respectively. Mistriage decisions were (42.82%), of which overtriage was (25.52%) and undertriage (17.30%) [Table 2].

Meta-regression analysis based on the method of moments for moderators (distance and publication year) was performed. Studies in terms of the distance from the origin of the CTAS in Canada significantly showed lower pooled coefficients, in other terms studies did indicate higher pooled coefficients for the nearest places rather than farther places ($B = -0.00003$; $SE_b = 0.00000$; $P = 0.0000$). Analysis of studies in terms of publication year of study revealed insignificant increase in reliability pooled coefficients, so the reliability of CTAS has not been increased systematically through the years ($B = 0.00005$; $SE_b = 0.00187$; $P = 0.9700$).

Discussion

The overall reliability of the CTAS is substantial in the emergency departments. The CTAS showed acceptable level of reliability to guarantee decisions were made consistently regarding allocating patients to appropriate categories. It supports evidence-based practice in the emergency department.^[18] but generally it's worth-mentioning that there is a considerable gap between research and clinical practice even at the best

of times.^[33] In addition, most studies used weighted kappa statistics to report reliability coefficient [Table 1] and the fact that weighted kappa statistics overestimates the reliability of triage scale^[8] make it necessary to interpret the results with extreme caution. So probably it's important to bear in mind that the CTAS reliability is actually at the moderate level which is congruent with several studies.^[18]

A calculated 42.82% triage decisions were recognized as mistriage. However (25.52%) of them as overtriage decisions are clinically plausible and could extenuate disagreement among raters in favor of patients, an alarming issue is that (13.69%) of triage decisions is related to undertriage in level I and II that it's notable to endanger critically-ill patients life [Table 2].

Post-hoc analysis revealed that level III has been predominant among Dong *et al.*,^[7] Gravel *et al.*,^[9] Dellaire *et al.*,^[11] and Dong *et al.*^[31] on the other hand level I has been prevailing among Manos *et al.*,^[4] Goransson *et al.*,^[8] and Goransson *et al.*^[32] In spite of the fact that (15.21%) of patients have been assigned into level I has been recognized as notable case-mix of patients in the emergency department, it cannot be concluded that CTAS has a tendency toward prioritizing patients into level I because only three studies are congruent with combined result. However ESI has tendency towards categorizing patients as level 2 (23.39% of all), CTAS has appropriately distributed patients among triage levels. In fact, the CTAS appropriately distribute patients into triage categories, so it has not a tendency to allocate patients into any specific level. Therefore it guarantees to prevent influx of patients in specific category. This influx could create significant disturbance in patient flow in the EDs and causes other parts of ED to remain unusable.^[17]

Comparing to other triage scales, mistriage in ESI (10.93%) is notably lower than CTAS and agreement among raters (78.56%) is higher than CTAS. Unlikely, Worster *et al.*, indicated CTAS has higher agreement than ESI.^[6] It can be justified that Canadian hospitals are more familiar with CTAS.

The CTAS show diverse pooled reliability coefficients regarding participants, patients, raters, reliability method and statistics. Results demonstrated agreement upon adult version and among nurse-physician were higher than pediatrics version and the other groups of raters, respectively. This result is congruent with ESI moderators.^[17] All of these moderator variables could lead further studies to explore more exclusively. The CTAS has been documented and supported substantially by scientific evidence in Canada. Only Sweden has reported reliability studies on CTAS except Canada [Table 1]. In this way, metaregression showed there is a

significant difference in terms of distance from origin of CTAS. It shows that CTAS has reached lower reliability coefficients in farther countries [Table 3]. One reason refers to complaint-based nature of CTAS that could be translated changeably in routine practice comparing to the Canada. This result is opposed to ESI triage scale generalizability which has shown the ESI triage scale could be adopted successfully in other countries in spite of cultural diversities.^[17]

The third edition of CTAS has been released^[15] and the reliability of triage scales has not been significantly improved through the years indicating revisions had considerably been effective. Therefore, the CTAS could be enhanced through the years and shown improved reliability [Figure 4].

In general, intra-rater reliability is more satisfactory than inter-rater reliability,^[34] so it revealed almost perfect agreement comparing to substantial agreement for inter-rater reliability. While intra and inter rater reliability are intended to report the degree to which measurements taken by the same and different observers are similar, respectively; other methods of examining reliability has remained uncommon in studies regarding the triage reliability.^[35]

Weighted kappa coefficient showed substantial agreement. In fact, weighted kappa coefficient revealed

Table 2: The Contingency table of triage decision distribution relating to each CTAS category among ED raters^[4,7-9,11,30-32]

ED Raters decisions	ED Raters decisions by CTAS category					Total
	1	2	3	4	5	
1	2698 (18.14)*	715	87	55	6	3561
2	447	1135 (7.63)	384	297	46	2309
3	69	710	1175 (7.90)	852	279	3085
4	5	115	669	1366 (9.18)	1076	3231
5	0	16	69	474	2132 (14.33)	2691
Total	3219	2691	2384	3044	3539	14877 (100.00)

* No (% of total)

Table 3: Meta-regression of Fisher's z-transformed kappa coefficients on predictor variables (Studies with weighted kappa)

Independent variable	B	SEb	P
Distance from origin	-0.00003	0.00000	0.0000
Publication date	0.00005	0.00187	0.9700

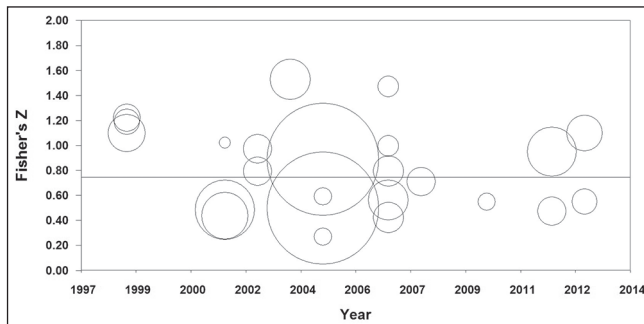


Figure 4: Fisher's Z-transformed kappa coefficients in relation to the Publication year of study

higher reliability than un-weighted kappa coefficient because it put more emphasis on the large differences between ratings than to small differences.^[36] Thus It's worth-mentioning even one category difference in allocating patients into appropriate category could endanger clinical outcomes of critically-ill patients. So un-weighted kappa statistics provides more realistic estimation of triage scales reliability.^[8,38]

A number limitation of this study must be noted. In our analysis, none of these studies have reported raw agreement for each individual CTAS level and only few studies have presented contingency table for inter-rater agreement between raters. Since this study is limited to overall reliability, some inconsistencies may exist across each CTAS level, therefore the results should be interpreted with caution.

Conclusion

Overall, the CTAS triage scale showed acceptable level of reliability in the emergency department, also it appropriately distributes patients into triage categories. Therefore it needs more development to reach almost perfect agreement and decrease disagreement especially undertriage. The reliability of triage scales requires a more comprehensive approach including all aspects of reliability assessment, so it's necessary to conduct further studies concentrating on the reliability of triage scales, especially in different countries.

Financial support and sponsorship

This project has been funded with support from the Mashhad University of Medical Sciences.

Conflicts of interest

There are no conflicts of interest.

References

1. Mirhaghi A, Kooshari H, Esmaeili H, Ebrahimi M. Outcomes for emergency severity index triage implementation in the emergency department. *J Clin Diagn Res* 2015;9:OC04-7.
2. 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;34:155-9.
3. Bergeron S, Gouin S, Bailey B, Patel H. Comparison of triage assessments among pediatric registered nurses and pediatric emergency physicians. *Acad Emerg Med* 2002;9:1397-401.
4. Manos D, Petrie DA, Beveridge RC, Walter S, Ducharme J. Inter-observer agreement using the Canadian Emergency Department Triage and Acuity Scale. *CJEM* 2002;4:16-22.
5. Grafstein E, Innes G, Westman J, Christenson J, Thorne A. Inter-rater reliability of a computerized presenting-complaint-linked triage system in an urban emergency department. *CJEM* 2003;5:323-9.
6. 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.
7. 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;12:502-7.
8. Göransson K, Ehrenberg A, Marklund B, Ehnfors M. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19:432-8.
9. 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;14:864-9.
10. Gravel J, Gouin S, Manzano S, Arsenaault M, Amre D. Interrater agreement between nurses for the Pediatric Canadian Triage and Acuity Scale in a tertiary care center. *Acad Emerg Med* 2008;15:1262-7.
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;12:45-9.
12. Dallaire C, Poitras J, Aubin K, Lavoie A, Moore L. Emergency department triage: Do experienced nurses agree on triage scores?. *J Emerg Med* 2012;42:736-40.
13. Gravel J, Gouin S, Goldman RD, Osmond MH, Fitzpatrick E, Boutis K, *et al.* The Canadian Triage and Acuity Scale for children: A prospective multicenter evaluation. *Ann Emerg Med* 2012;60:71-7.
14. Fernandes CM, McLeod S, Krause J, Shah A, Jewell J, Smith B, *et al.* Reliability of the Canadian Triage and Acuity Scale: Interrater and intrarater agreement from a community and an academic emergency department. *CJEM* 2013;15:227-32.
15. Bullard MJ, Chan T, Brayman C, Warren D, Musgrave E, Unger B. Revisions to the Canadian Emergency Department Triage and Acuity Scale (CTAS) Guidelines. *CJEM* 2014;16:1-5.
16. 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.
17. Mirhaghi A, Heydari A, Mazlom R, Hasanzadeh F. Reliability of the Emergency Severity Index: Meta-analysis. *Sultan Qaboos Univ Med J* 2015;15:e71-7.

18. Farrohknia N, Castrén M, Ehrenberg A, Lind L, Oredsson S, Jonsson H, *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.
19. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68.
20. 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.
21. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Fam Med* 2005;37:360-3.
22. Petitti D. *Meta-analysis, decision analysis, and cost effectiveness analysis*. 1st ed. New York: Oxford University Press; 1994. p. 69.
23. Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. *Dtsch Arztebl Int* 2010;107:892-8.
24. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, *et al.* Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud* 2011;48:661-71.
25. 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.
26. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. 1st ed. Waltham: Academic Press; 1985. p. 76-81.
27. Rosenthal R. *Meta-analytic procedures for social research*. Rev ed. Newbury Park, CA: SAGE Publications; 1991. p. 43-89.
28. Chen H, Manning AK, Dupuis J. A method of moments estimator for random effect multivariate meta-analysis. *Biometrics* 2012;68:1278-84.
29. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549.
30. Bergeron S, Gouin S, Bailey B, Amre DK, Patel H. Agreement among pediatric health care professionals with the pediatric Canadian triage and acuity scale guidelines. *Pediatr Emerg Care* 2004;20:514-8.
31. Dong SL, Bullard MJ, Meurer DP, Blitz S, Ohinmaa A, Holroyd BR, *et al.* Reliability of computerized emergency triage. *Acad Emerg Med* 2006;13:269-75.
32. 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.
33. 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.
34. 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.
35. 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.
36. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.
37. Dong SL, Bullard MJ, Meurer DP, Blitz S, Holroyd BR, Rowe BH. The effect of training on nurse agreement using an electronic triage system. *CJEM* 2007;9:260-6.
38. Ebrahimi M, Heydari A, Mazlom R, Mirhaghi A. Reliability of the Australasian Triage Scale: Meta-analysis. *World J Emerg Med* 2015;6:94-99.

APPENDIX A

("triage"[MeSH Terms] OR "triage"[All Fields]) AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "scale"[All Fields]) AND reliability [All Fields]

("triage"[MeSH Terms] OR "triage"[All Fields]) AND acuity [All Fields] AND reliability [All Fields]

("triage"[MeSH Terms] OR "triage"[All Fields]) AND reliability [All Fields] AND agreement [All Fields]

("triage"[MeSH Terms] OR "triage"[All Fields]) AND reliability [All Fields] AND ("emergencies"[MeSH Terms] OR "emergencies"[All Fields] OR "emergency"[All Fields])

("triage"[MeSH Terms] OR "triage"[All Fields]) AND reliability [All Fields] AND ("emergencies"[MeSH Terms] OR "emergencies"[All Fields] OR "emergency"[All Fields] OR "emergency"[All Fields]) AND Canadian [All Fields] AND ("abstracting and indexing as topic"[MeSH Terms] OR ("abstracting"[All Fields] AND "indexing"[All Fields] AND "topic"[All Fields]) OR "abstracting and indexing as topic"[All Fields] OR "acuity"[All Fields])

("triage"[MeSH Terms] OR "triage"[All Fields]) AND ("emergencies"[MeSH Terms] OR "emergencies"[All Fields] OR "emergency"[All Fields]) AND Canadian [All Fields] AND ("abstracting and indexing as topic"[MeSH Terms] OR ("abstracting"[All Fields] AND "indexing"[All Fields] AND "topic"[All Fields]) OR "abstracting and indexing as topic"[All Fields] OR "acuity"[All Fields])

("triage"[MeSH Terms] OR "triage"[All Fields]) AND ("emergencies"[MeSH Terms] OR "emergencies"[All Fields] OR "emergency"[All Fields]) AND Canadian [All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "scale"[All Fields])