

Review Article

The reliability of the Australasian Triage Scale: a meta-analysis

Mohsen Ebrahimi¹, Abbas Heydari², Reza Mazlom², Amir Mirhaghi³

¹ Department of Emergency Medicine, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

² Evidence-Based Caring Research Center, Department of Medical-Surgical Nursing, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran

³ Department of Nursing, Faculty of Nursing, Neyshabur University of Medical Sciences, Neyshabur, and Evidence-Based Caring Research Center, Department of Medical-Surgical Nursing, School of Nursing & Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran

Corresponding Author: Amir Mirhaghi, Email: Mirhaghia@mums.ac.ir

BACKGROUND: Although the Australasian Triage Scale (ATS) has been developed two decades ago, its reliability has not been defined; therefore, we present a meta-analysis of the reliability of the ATS in order to reveal to what extent the ATS is reliable.

DATA SOURCES: Electronic databases were searched to March 2014. The included studies were those that reported samples size, reliability coefficients, and adequate description of the ATS 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 the method of moment's estimator.

RESULTS: Six studies were included in this study at last. Pooled coefficient for the ATS was substantial 0.428 (95%CI 0.340–0.509). The rate of mis-triage was less than fifty percent. The agreement upon the adult version is higher than the pediatric version.

CONCLUSION: The ATS has shown an acceptable level of overall reliability in the emergency department, but it needs more development to reach an almost perfect agreement.

KEY WORDS: Triage; Emergency treatment; Algorithm; Reliability and validity; Meta-analysis

World J Emerg Med 2015;6(2):94–99
DOI: 10.5847/wjem.j.1920–8642.2015.02.002

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 Australasian Triage Scale (ATS) is a five-level emergency department triage algorithm that has been continuously developed in Australia and subjected to several studies.^[2–7] The ATS, a 5-point triage scale, has been endorsed by the Australasian College for Emergency Medicine and adopted in performance indicators by the Australian Council on

Healthcare Standards. The National Triage Scale (NTS) was implemented in 1993. In the late 1990s, the NTS underwent revisions and was subsequently renamed the Australasian Triage Scale (ATS). The ATS is based on adult physiological predictors (airway, breathing, circulation, and disability).^[8]

Several studies^[2–7] have investigated the validity and reliability of the ATS in adult and pediatric populations; but it's still unclear to what extent the ATS would support consistency in triage nurses' decision making in Australia comparing to other countries, considering the wide

variety of health care systems around the world. Besides, some studies^[9,10] have addressed contextual influences on the triage decision making process, therefore it's necessary to discover the effect of these variables on the reliability of triage scale. However, some studies reported moderate consistency for the ATS,^[11] but it needs to be extensively studied 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.^[12] However, kappa has been the most commonly used statistics to measure inter-rater agreement, and it is worth mentioning that kappa statistics could be influenced by incidence, bias and levels of scale, thus leading to misleading results.^[13-15] It is reported that weighted kappa statistics could reveal high and deceiving reliability coefficients.^[12] 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.^[16] Therefore, it is an appropriate method to gain comprehensive and deep insights into the reliability of triage scale especially in regard to kappa statistics.

A review on reliability of the ATS demonstrated that kappa ranges from 0.25 (fair) to 0.56 (moderate).^[11,17] 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 ATS in order to examine to what extent the ATS is reliable.

METHODS

The study was approved by the Research Ethics Committee of Mashhad University. The databases we searched until March 1, 2014 included Cinahl, Scopus, Medline, Pubmed, Google Scholar and Cochrane Library in the first phase of the study. The search terms included reliability, triage, system, scale, agreement, emergency and Australasian Triage Scale.

Relevant citations in reference lists of final studies were hand-searched to identify additional articles regarding the reliability of the ATS. Three researchers

independently examined the search results in order to recover potentially eligible articles (Figure 1). Authors of the articles were contacted to retrieve supplementary information if needed.

Irrelevant and duplicated results were eliminated. Only English language publications were reviewed. Articles were chosen according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).^[18] 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. According to inclusion criteria, the qualified paper was defined as one with qualifying score more than 6 out of the 8 criteria. Disagreements among the researchers were resolved by consensus. The articles in which the type of reliability was not reported were excluded from the study. The researchers also recorded moderator variables such as participants, raters, origin and publication year of studies.

In the next phase, participants (age-group, size), raters (profession, size), instruments (live, scenario), origin and publication year of studies, reliability coefficient and method were retrieved. The reliability coefficients were extracted from articles including: 1) Inter-rater reliability: kappa coefficient (weighted and un-weighted), intraclass correlation coefficient, Pearson's correlation coefficient and Spearman's rank-order correlation coefficient; 2) Intra-rater reliability: articles which contained reliability statistics including Pearson's correlation coefficient, intraclass correlation coefficient and Spearman's rank-order correlation coefficient were

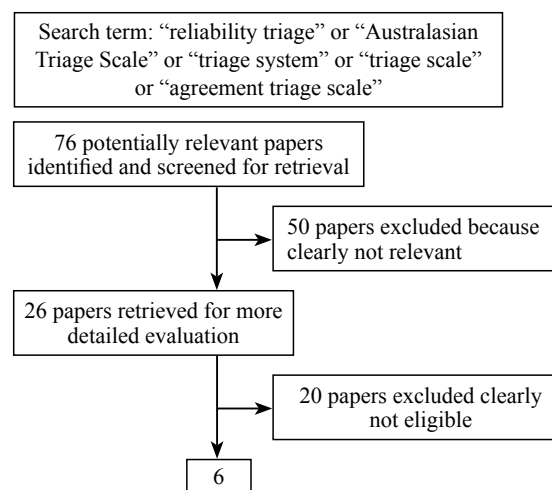


Figure 1. The results of literature search and selection process.

Table 1. Studies on the reliability of the ATS

Studies	Patient	Participants*	Instrument	Methods**	Statistics***	Country
Dilley et al ^[2] 1998	Adult	NN	Scenario	Inter	Kuw (0.25)	Australia
Fernandes CMB et al ^[3] 1999	Adult	NN	Scenario	Intra	Kuw (0.75)	Australia
	Adult	NN	Scenario	Inter	Kuw (0.66)	
Crellin et al ^[4] 2003	Adult	NN	Scenario	Inter	Kuw (0.21)	Australia
Considine et al ^[5] 2004	Adult	NN	Scenario	Inter	Kuw (0.43)	Australia
	Pediatrics	NN	Scenario	Inter	Kuw (0.40)	
Gerdtz et al ^[6] 2007	Adult	NN	Scenario	Inter	Kuw (0.38)	Australia
Gerdtz et al ^[7] 2008	Adult	NN	Scenario	Inter	Kuw (0.41)	Australia

*NN: nurse-nurse; **Inter: inter-rater reliability; Intra: intra-rater reliability; ***Kuw: Un-weighted kappa (coefficient).

included; 3) Internal consistency: articles reporting alpha coefficients were included.

Each sample was considered as a unit of analysis. If the same sample was 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 were analyzed for the 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 ($\kappa=0.81-1.00$).^[19] Kappa could be treated as a correlation coefficient in meta-analysis.^[20] 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.^[21,22] Fixed effects and random effects models were applied. The data were analyzed using comprehensive meta-analysis software (Version 2.2.050).

Simple meta-regression analysis was performed according to the method of moments estimator.^[23] In the meta-regression model, effect size as a dependent variable, and studies and subject characteristics as an independent variable were considered to discover potential predictors of reliability coefficients. Z -transformed reliability coefficients were regressed on the following variables: origin and publication year of studies. Distance was defined as distance from the origin of each study to the origin of the ATS (Melbourne, Australia). Meta-regression was performed using a model of random effects because of the presence of significant between-study variation.^[24]

RESULTS

Literature searching found 76 primary citations relevant to the reliability of the ATS. Finally, 6 (7.89% of

the 76) unique citations, which met the inclusion criteria, were selected (Figure 1). The citations were subgrouped according to participants (adult/pediatric), raters (nurses, physicians, experts) and method of reliability (intra/inter raters), reliability statistics (weighted/un-weighted kappa) and origin and publication year of studies. Two clinicians (AM and ME) and one statistician (RM) reviewed the cited articles independently. Minor disagreements among the reviewers were discussed to reach a consensus. The level of agreement was almost perfect among the reviewers through final selection of the articles.

In the analysis, 4 409 patients were included in the study. The reliability of the ATS was assessed in Australia. The publication year of studies ranged from 1998 to 2007 with a median of 2003. No studies were conducted using the latest version of triage scale. Inter-rater reliability was used in all studies except for one study using intra-rater reliability.^[3] No study in our analysis used alpha coefficient to report internal consistency in reliability analysis. Unweighted kappa coefficient was the only common statistics (Table 1). Overall pooled coefficient for the ATS was moderate 0.428 (95%CI 0.340–0.509). All raters were nurses, so the participants' pooled coefficient was moderate and all studies used paper-based scenario assessment for reporting reliability too.

Agreement on inter-rater and intra-rater reliability

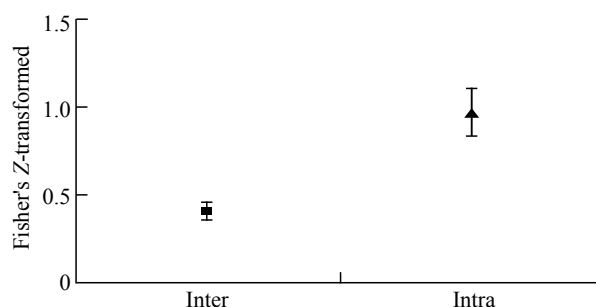


Figure 2. Fisher's Z-transformed pooled estimates of reliability statistics (random effects model) (Inter: Inter-rater reliability; Intra: Intra-rater reliability).

Table 2. The contingency table of triage decision distribution relating to each ATS category among ED raters^[5]

ED raters decisions	ED raters decisions by ATS category					Total
	1	2	3	4	5	
1	357 (7.74)*	61	5	0	1	424
2	275	452 (9.80)	166	6	1	900
3	26	144	887 (19.22)	279	64	1 400
4	2	2	255	890 (19.29)	372	1 521
5	0	0	5	144	220 (4.77)	369
Total	660	659	1 318	1 319	658	4 614 (100.00)

* No (% of total).

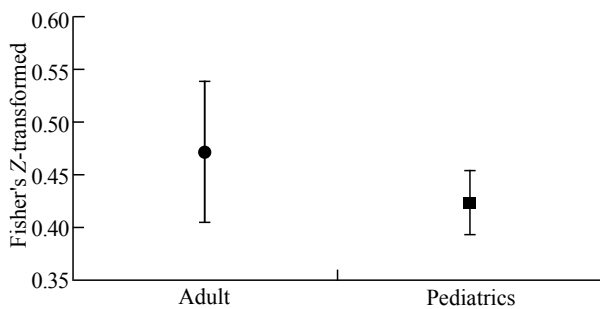


Figure 3. Fisher's Z-transformed pooled estimates of patients' reliability based on age group (Random effects model).

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

Independent variable	B	SE _b	P value
Distance from origin	0.00004	0.00001	0.0000
Publication date	-0.00193	0.00704	0.76219

was fair 0.390 (95%CI 0.307–0.466) and substantial 0.750 (95%CI 0.613–0.843), respectively (Figure 2).

Agreement on adult and pediatric version of the ATS was moderate 0.440 (95%CI 0.329–0.539) for adult and 0.400 (95%CI 0.350–0.448) for pediatrics (Figure 3).

Only one study^[5] reported a contingency table to show frequency distribution of triage decisions upon each ATS level between two raters (Table 2). The rate of overall agreement was 60.81%. The rate of agreement for ATS L-1 was 7.74%, ATS L-2 9.80%, ATS L-3 19.22%, ATS L-4 19.29%, and ATS L-5 4.77%; and the rate of disagreement was 4.10%, 7.10%, 10.23%, 11.49%, and 6.36% respectively. Mistriage decisions accounted for 39.19%, of which overtriage was 20.70% and undertriage 18.49% (Table 2).

Meta-regression analysis based on the method of moments for moderators (distance and publication year) was performed (Table 3). Studies in terms of the distance from the origin of the ATS in Australia significantly showed lower pooled coefficients, in other terms studies did indicate higher pooled coefficients for the nearest places rather than farther places. Analysis of studies in

terms of publication year of study revealed insignificant change in reliability pooled coefficients, thus the reliability of the ATS increased systematically through the years (Table 3).

DISCUSSION

The overall reliability of the ATS is moderate in the emergency departments. The ATS showed a fairly acceptable level of reliability to allocate patients to appropriate categories. However it supports evidence-based practice in the emergency department.^[11] But it is worth mentioning that there is a gap between research and clinical practice even at the best of time.^[25] No study used weighted kappa statistics to report reliability coefficient (Table 1), therefore it is far from weighted kappa bias in reporting reliability coefficients. Weighted kappa statistics overestimates the reliability of triage scale,^[12] thus it is necessary to interpret the results with caution. Therefore it is important to remember that the ATS reliability is actually at the moderate level which is congruent with several studies.^[17]

Approximately 39.19% of triage decisions were recognized as mis-triages. Although it is not highly remarkable, 20.70% were overtrriages and it could extenuate disagreement among raters in favor of patients. In addition, an alarming issue is that 18.49% of triage decisions are related to under-triage in levels I and II which are notable to endanger the life of critically-ill patients (Table 2). Comparing to other triage scales, the rate (10.93%) of mis-triage in ESI is lower than that of the ATS and the rate (78.56%) of agreement among raters is higher than that of the ATS. Only one study compared the reliability of the ATS with ESI. Unlikely, Alpert et al^[26] indicated the ATS has a higher rate of agreement than ESI. It can be justified that the generalizability of result is limited to simulation of triage decisions.

However, ESI has a strong tendency towards categorizing patients as level 2 (23.39% of all), and ATS can appropriately distribute patients in triage levels.

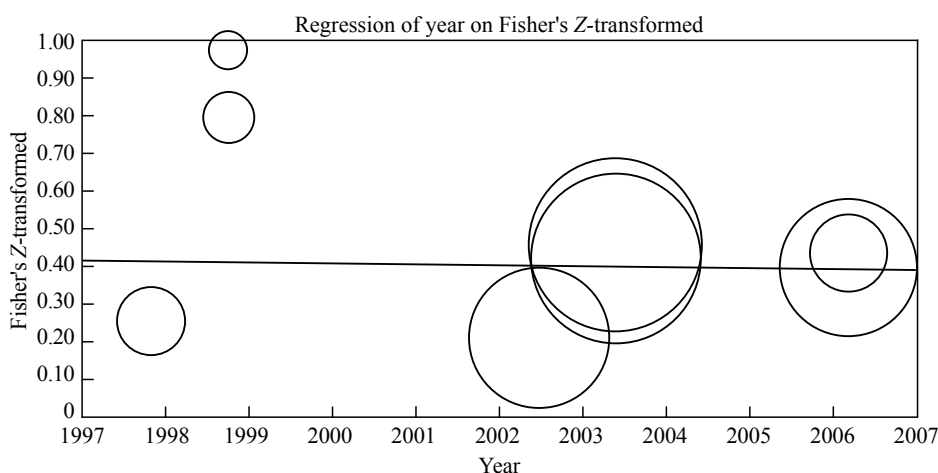


Figure 4. Fisher's Z-transformed kappa coefficients in relation to the publication year of studies.

Therefore, ATS guarantees to prevent influx of patients in specific category. This influx creates significant disturbance in patient flow in the EDs and causes other parts of the ED to remain unusable.^[10]

The ATS shows diverse pooled reliability coefficients regarding participants, patients, raters, reliability method and statistics. The results demonstrated the rate of agreement upon the adult version was higher than the pediatric version. This result is congruent with ESI moderators.^[10] All of these moderator variables could lead further studies to explore more exclusively. The ATS has been documented and supported moderately by scientific evidence in Australia (Table 1). In this way, meta-regression analysis showed that there is a significant difference in distance from origin of the ATS. It shows that the ATS has reached higher reliability coefficients in Australia (Table 3).

The second edition of ATS has been released^[8] and the reliability of triage scale has not been significantly improved through the years. However, Gerdtz et al^[6,7] found that although the improvement has not been significant, marked improvement has been obtained. In fact, the reliability of the ATS increased from a fair reliability coefficient of Dilley et al^[2] in 1998 to moderate reliability coefficient of Gerdtz et al^[7] in 2008, indicating that revision was considerably effective. Therefore, the ATS needs to be enhanced through the years and improved in order to reach almost perfect reliability (Figure 4).

In general, intra-rater reliability is more satisfactory than inter-rater reliability,^[27] so it has revealed substantial agreement comparing to fair agreement for inter-rater reliability. As intra- and inter-rater reliabilities are intended to indicate the similar measurements taken by the same or different observers respectively, other

methods for examining reliability have been uncommon in studies regarding the triage reliability.^[28,29]

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

In conclusion, the ATS triage scale has a fairly acceptable level of reliability in the emergency department, and it appropriately distributes patients into triage categories. Therefore it needs more development to reach almost perfect agreement and decrease disagreement especially under-triage. The reliability of triage scales requires a more comprehensive evaluation including all aspects of reliability assessment, so further studies on the reliability of triage scales are necessary, especially in different countries.

ACKNOWLEDGEMENT

We thank Dr. Ramin Sadeghi for his comments on research methodology.

Funding: None.

Ethical approval: The study was approved by the University Research Ethics Committee.

Conflicts of interest: The authors declare that no competing interest and no personal relationships with other people or organizations that could inappropriately influence their work.

Contributors: Mirhaghi A proposed the study and wrote the first draft. All authors read and approved the final manuscript.

REFERENCES

- 1 Mirhaghi A, Kooshiar H, Esmaceli H, Ebrahimi M. Outcomes for Emergency Severity Index Triage implementation in the Emergency Department. *J Clin Diagn Res* 2015; 9: OC04–OC07.
- 2 Dilley SJ, Standen P. Victorian nurses demonstrate concordance in the application of the National Triage Scale. *Emerg Med* 1998; 10: 12–18.
- 3 Fernandes CM, Wuerz R, Clark S, Djurdjev O. How reliable is emergency department triage?. *Ann Emerg Med* 1999; 34: 141–147.
- 4 Crellin DJ, Johnston L. Poor agreement in application of the Australasian Triage Scale to paediatric emergency department presentations. *Contemp Nurse* 2003; 15: 48–60.
- 5 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: 516–523.
- 6 Gerdtz MF, Bucknall TK. Influence of task properties and subjectivity on consistency of triage: a simulation study. *J Adv Nurs* 2007; 58: 180–190.
- 7 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: 250–259.
- 8 Gerdtz M, Considine J, Sands N, Stewart C, Crellin D, Pollock W, et al. Emergency triage education kit. Department of Health and Ageing, Canberra. 2007; 19.
- 9 Andersson A, 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: 136–145.
- 10 Mirhaghi A, Heydari A, Mazlom R, Hasanzadeh F. Reliability of the Emergency Severity Index Meta-analysis. *Sultan Qaboos University Med J* 2015; 15: 67–73.
- 11 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.
- 12 Göransson K, Ehrenberg A, Marklund B, Ehnfors M. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005; 19: 432–438.
- 13 Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther* 2005; 85: 257–268.
- 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: 1256–1263. Epub 2010/05/01.
- 15 Viera A, Garrett J. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; 37: 360–363.
- 16 Petitti D. Meta-analysis, decision analysis, and cost effectiveness analysis. New York, NY: Oxford University Press; 1994; 69.
- 17 Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. *Dtsch Arztebl Int* 2010; 107: 892–898.
- 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. *Int J Nurs Stud* 2011; 48: 661–671. Epub 2011 Apr 23.
- 19 Julius S, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85: 257–268.
- 20 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: 169–184.
- 21 Hedges LV, Olkin I. Statistical methods for meta-analysis. Academic Press; 1985; 76–81.
- 22 Rosenthal R. Meta-analytic procedures for social research. SAGE Publications; 1991; 43–89.
- 23 Chen H, Manning AK, Dupuis J. A method of moments estimator for random effect multivariate meta-analysis. *Biometrics* 2012; 68: 1278–1284.
- 24 Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011; 342: d549.
- 25 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: 428–437.
- 26 Alpert EA, Lipsky AM, Hertz D, Rieck J, Or J. Simulated evaluation of two triage scales in an emergency department in Israel. *Eur J Emerg Med* 2013; 20: 431–434.
- 27 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: 777–788.
- 28 Hogan TP, Benjamin A, Brezinski KL. Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement* 2000; 60: 523–531.
- 29 Parenti N, Bacchi Reggiani ML, Sangiorgi D, Serventi V, Sarli L. Effect of a triage course on quality of rating triage codes in a group of university nursing students: a before–after observational study. *World J Emerg Med* 2013; 4: 20–25.

Received December 12, 2014
Accepted after revision April 3, 2015