Inter-rater agreement of the triage system RETTS-HEV

Louise Nissen\textsuperscript{a}, Hans Kirkegaard\textsuperscript{c}, Noel Perez\textsuperscript{a}, Ulf Hørlyk\textsuperscript{a} and Louise P. Larsen\textsuperscript{b}

\textbf{Objective} The purpose of this study was to evaluate the inter-rater agreement among nurses using the triage system RETTS-HEV (rapid emergency triage and treatment system – hospital unit west) in a Danish emergency department (ED).

\textbf{Background} The use of triage systems in Denmark has been implemented recently together with structural changes in hospital organization. Testing and evaluation is therefore needed. The RETTS-HEV is a five-scale triage system being used in the ED of Herning, Denmark, since May 2010. The ED is semilarge, with 29,000 annual visits.

\textbf{Materials and methods} Consecutive patients presenting to the ED were assessed by both a duty and a study nurse using RETTS-HEV. Nurses did not receive training before the study. In all, 146 patients were enrolled and a blinded, paired and simultaneous triage was conducted independently to evaluate inter-rater agreement using Fleiss $\kappa$.

\textbf{Results} A total of 155 patients were triaged over a 10-day period and complete data were available for 146 patients. We found the overall agreement to be good [Fleiss $\kappa$ 0.60 (95\% confidence interval 0.48; 0.72)]. The $\kappa$ estimate was higher for the group of patients who required immediate attention [0.83 (95\% confidence interval 0.18; 1.47)].

\textbf{Conclusion} The study found good inter-rater agreement between two independent observers not receiving any new triage training before the study. \textit{European Journal of Emergency Medicine} 21:37–41 \textcopyright 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Keywords: emergency department, inter-rater agreement, reliability, RETTS, triage

Departments of \textsuperscript{*}Emergency, \textsuperscript{b}Occupational Medicine, Regional Hospital of Herning, Herning and \textsuperscript{c}Research Center for Emergency Medicine, University of Aarhus, Aarhus, Denmark

Correspondence to Louise Nissen, MD, Gl. landevej 51A, 1 tv, 7400 Herning, Denmark
Tel: +45 26 13 64 30; fax: +45 78 43 20 61; e-mail: louisenissen84@gmail.com

Received 19 March 2013 Accepted 24 May 2013
RETTS-HEV consists of an algorithm for vital signs and 45 ESS algorithms. In Table 1, the algorithm for vital signs is presented and Fig. 1 shows an example of an ESS algorithm.

There are five categories: (1) red (immediate); (2) orange (can wait 10 min); (3) yellow (can wait 60 min); (4) green (can wait 120 min) and (5) blue (can wait 240 min and can be treated by a trained ED nurse). Triage category 5 patients include only smaller orthopaedic injuries referred to the ED from a general physician; they are classified as category 5 on arrival and do not undergo the full process of triage.

The patient’s vital signs are typed into a computer program that automatically estimates a triage level. Red patients are announced by paramedics who bring patients to the ED and are evaluated and treated immediately by an emergency team. The triage is conducted and typed into the chart when the patient is in a stable condition.

The final triage level can be altered depending on the chief complaint. The nurse can upgrade the level of triage on the basis of the ESS algorithm but can never downgrade a patient from the level determined by vital signs. There are 45 flowcharts describing different ESS algorithms. More than one can be chosen in the case of several complaints.

The triage is a two-step simultaneous process: first, a registration of vital signs in the computer program related to the patient’s electronic chart and second choosing one or more ESS cards. The highest triage level of the two measures equals the final triage level. The triage is normally performed by a single nurse and the process is estimated to last no longer than 10 min [4].

The study was carried out in the ED at the Regional Hospital of Herning, which annually receives around 29 000 visits, of which ~45% (13 000) are admitted. Herning is a provincial town in Denmark and the hospital services a population of ~300 000 people. The ED receives all patients, except cardiac and paediatric medical patients.

Data were collected over a period of 10 weekdays in 7 h shifts between 8 a.m. and 3 p.m. in October 2012. One of 10 study nurses affiliated with this project and one of 20 ED staff nurses simultaneously, but blinded of each other, performed triage of patients in the ED.

The study nurses were also part of the regular ED staff; only they signed in voluntarily to take part in the study. None of the nurses received new RETTS-HEV training before the study.

Patients were selected without preference but limited to the available time of the study nurse who participated in the process of triage evaluating as many patients as possible during that shift. The study nurse moved from one patient to another by selecting the next incoming patient.

Only patients in triage categories 1–4/red–green were included, leaving out category 5/blue/fast-track patients.

The ED staff nurse conducted the interview with the patient. The study nurses’ triage level was based solely on observation of the staff nurses’ interview and observation of vital signs from the monitor. Temperature and respiratory frequency obtained by the staff nurse was shared verbally with the study nurse. Afterwards, both nurses chose the ESS algorithms they found most suitable on the basis of the information provided by the patient during the interview. The nurses chose the ESS algorithm independent of each other.

Both levels of triage were registered in the patient’s electronic chart.

Furthermore, a questionnaire on job experience and RETTS training was handed out to all the participating nurses.
Assessments in the study were collected and analysed using Stata version 11 (StataCorp LP, College Station, Texas, USA). Inter-rater reliability was measured with Fleiss $\kappa$ [7]. Fleiss $\kappa$ was calculated for the two steps of the triage process independently and for the final level of triage. Fleiss $\kappa$ does not assign a different value to agreement according to its magnitude, but states it as either agreement or disagreement [7]. The value 1.0 accounts for perfect agreement, whereas 0 equals chance agreement. Values are interpreted as follows: $\kappa$ value less than 0.40 is considered as poor agreement; 0.40–0.75 as good agreement and greater than 0.75 as excellent agreement [8].

The local Ethics Committee found the study to be exempt from formal ethics review as it was considered a quality assurance study not interfering with or altering standard patient care.

**Results**

The sample size of reliability studies should exceed 50; we included 146 patients [9].

In the 10-day period, 155 patients underwent simultaneous study and staff nurse triage. Nine patients were excluded because the staff nurse did not register them in the electronic chart, leaving 146 double-evaluated patients with 292 processes of triage.
Table 2  Interobserver agreement within each level of triage on the basis of vital signs, emergency symptom and sign algorithm and the final triage level

| Triage | Level  | \( \kappa \) (95% CI) |
|--------|--------|-------------------------|
| Vital signs | \( n = 136 \) |                          |
| Red    | 0.89 (0.61; 1.61) |                          |
| Orange | 0.53 (-0.15; 1.20) |                          |
| Yellow | 0.58 (0.14; 1.03)  |                          |
| Green  | 0.69 (0.04; 1.33)  |                          |
| Overall| 0.65 (0.46; 0.85)  |                          |
| ESS    | \( n = 130 \) |                          |
| Red    | 0.88 (0.45; 1.32)  |                          |
| Orange | 0.63 (0.15; 1.11)  |                          |
| Yellow | 0.49 (-0.01; 1.99) |                          |
| Green  | 0.63 (0.16; 1.09)  |                          |
| Overall| 0.59 (0.46; 0.72)  |                          |
| Final  | \( n = 146 \) |                          |
| Red    | 0.83 (0.18; 1.47)  |                          |
| Orange | 0.66 (0.22; 1.10)  |                          |
| Yellow | 0.51 (0.03; 0.98)  |                          |
| Green  | 0.61 (0.18; 1.05)  |                          |
| Overall| 0.60 (0.48; 0.72)  |                          |

CI, confidence interval.

Twenty staff and 10 study nurses participated in the data-gathering process. The nurses’ had an average working experience of 17.8 years (4–38 years) and an average working experience in the ED of 10.6 years (2 months to 36 years). Twenty of the participating nurses (66%) had received 1 h of formalized training using RETTS-HEV when the system was implemented in 2010 and 10 learned through colleagues.

The mean age of the patients included in the study was 56.5 years (range: 5–95 years) and 50.4% were women. The eight most common chief complaints were abdominal pain (32%), head trauma (7.2%), dyspnoea (6.5%), fever (6.5%), lower extremity injury (4.1%) and dizziness (4.1%).

The interobserver agreement in the triage on the basis of vital signs and ESS, respectively, as well as the final triage score are shown in Table 2.

In the final triage, the nurses agreed on exact triage level in 108 cases (74%) and were within one adjacent triage level in all cases (100%). Among the 38 patients with disagreement, 24 (63%) were between yellow and green, 12 (32%) were between yellow and orange and two (5%) were within red and orange. In all three assessments, the overall \( \kappa \) value and the 95% confidence interval were within the interval for ‘good agreement’.

A clear tendency towards high agreement (0.83) was found in patients who needed immediate treatment (red), although the confidence interval is broad.

On comparing the level of triage on the basis of vital parameters with the final level many patients were upgraded after the selection of an ESS algorithm. The number of patients in the red triage level increased from 10 to 12, patients in orange increased from 11 to 50 and patients in yellow increased from 62 to 142, whereas patients in green decreased from 193 to 88.

In 123 patients, the same ESS was chosen at least once and in seven of the remaining 23 patients, disagreement was because of the selection of different ESS algorithms. In the last 16 cases, the ESS algorithm was only available from one of the nurses.

**Discussion**

Our study showed an overall triage agreement of 0.60 between two ED nurses. Agreement was 0.65 after obtaining vital signs and 0.60 after choosing an ESS algorithm. All results were within the category for good agreement. Agreement was higher within higher levels of triage (red) 0.83.

Other similar studies have found higher values of agreement [10–12] using weighted \( \kappa \) [13,14], whereas we obtained a lower value, as expected, using Fleiss \( \kappa \) [7]. Also, nurses in our study did not receive new triage training before data collection to avoid overestimation of agreement [15].

Our study was designed as recommended in a Swedish review from 2011 with real-life patient cases [16] to ensure that the data were as realistic as possible. Few other studies we reviewed applied this approach to determine reliability [10–12,17], and most of the other studies on reliability were carried out using paper case scenarios [10,15,18–25].

For category 1 (red) patients, in two out of seven cases, there was disagreement between category 1 and 2 (Table 3). One patient with disagreement had two different ESS algorithms selected, leading to disagreement in the final triage category. The other patient had the same vital signs among both nurses but different triage categories on the basis of vital signs. As vital signs were the same among both nurses, we believe that one nurse made a mistake while typing vital signs into the system.

The study nurse did not perform an independent assessment. She obtained her information partly by
overhearing the staff nurse’s interview. The advantage of this method is that the patient only provides one set of information, whereas conducting interviews independently ensures complete blinding but has the disadvantage of potentially obtaining different information from the patient.

Testing of RETTS-HEV in other hospital units is needed to determine the external validity of our results and should be considered for further research. It is important that triage systems are tested in their own environment even though evidence exists from different settings. Patients in Danish EDs may not be comparable with those attending EDs in other countries because of differences in emergency care provided in primary healthcare. For the future, we recommend a more standardized training of nurses to ensure a more uniform conduction of the triage process [26] and further evaluation of RETTS-HEV is needed; currently, validation of the system is ongoing.

Conclusion

This was the first study carried out in Denmark to evaluate inter-rater agreement in a triage system and the first to evaluate agreement for RETTS-HEV by conducting simultaneous and blinded triage of real-life patients.

We have shown that the RETTS-HEV has an overall good inter-rater agreement (κ = 0.60). We found high agreement (κ = 0.83) between nurses evaluating patients who needed immediate treatment.

Acknowledgements

The authors thank Asger Roer Pedersen from the Section of Biostatistics, Department of Public Health, University of Aarhus, for his invaluable help with the statistical analyses. This study was supported by grants from the Region Midtjylland and the Hospital Unit West.

Conflicts of interest

There are no conflicts of interest.

References

1 Lindberg SØ, Lerche la Cour J, Folkestedt L, Hallas P, Brabrand M. The use of triage in Danish emergency departments. Dan Med Bull 2011; 58:A4301–A4303.
2 Skriver C, Lauritzen MM, Forberg JL, Gaardboe-Poulsen OB, Mogensen CB, Hansen CL, Berlac PA. Triage quickens the treatment of the most sick patients. Ugeskr Laeger 2011; 173:2490–2493.
3 Johansen MB, Forberg JL. Nurses’ evaluation of a new formalized triage system in the emergency department – a qualitative study. Dan Med Bull 2011; 58:A4311–A4315.
4 Anderson T, Møller M. Triage-Manual, version 1.8; (January 2012). Available at: http://e dok rm dk e dok e_hoveak nsfSCXIV/ATTACH/RMAP-80VC82/ $FILE/Triage Manual%201.8%20January%202012.doc [Accessed 4 December 2012].
5 Widgren BR, Jourak M. Medical emergency triage and treatment system (METTS): a new protocol in primary triage and secondary priority decision in emergency medicine. J Emerg Med 2011; 40:623–628.
6 Widgren BR, Jourak M, Martinsen A. New accurate triage method. METTS – a yields basis for priority level decisions. Läkartidningen 2008; 105:261–270.
7 Reies JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971; 76:378–382.
8 Armitage P, Berry G. Statistical methods in medical research. Hoboken, NJ: Blackwell Publishing; 1994.
9 Hopkins WG. Measures of reliability in sports medicine and science. Sport Med 2000; 30:1–15.
10 Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity triage algorithm version 2 is reliable and valid. Acad Emerg Med 2004; 11:59–65.
11 Maninges PA, Hume DA, Parker DE, McMurray TA. The Soterion Rapid Triage System: evaluation of inter-rater reliability and validity. J Emerg Med 2006; 30:461–469.
12 Dong SL, Bullard MJ, Meurer DP, Blitz S, Ohimaa A, Holroyd BR, et al. Reliability of computerized emergency triage. Acad Emerg Med 2006; 13:269–275.
13 Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. Scand J Caring Sci 2005; 19:427–431.
14 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.
15 Dalcaire C, Poitras J, Aubin K, Lavoie A, Moore L. Emergency department triage: do experienced nurses agree on triage scores? J Emerg Med 2012; 42:261–270.
16 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; 19:42–55, 1–13.
17 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; 5:323–329.
18 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; 34:155–159.
19 Manos D, Petrie DA, Beveridge RC, Stephen Walter, Ducharme J. Inter-observer agreement using the Canadian Emergency Department Triage and Acuity Scale. CJEM 2002; 4:16–22.
20 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. Emerg Med J 2009; 26:431–434.
21 Grouse AI, Bishop RO, Bannom AM. The Manchester Triage System provides good reliability in an Australian emergency department. Emerg Med J 2009; 26:484–486.
22 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; 57:257–264.
23 Worster A, Sordo A, Eva K, Fernandes CM, Upadhye S. Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. J Emerg Nurs 2007; 33:319–323.
24 Considine J, Urg L, Thomas S. Triage nurses’ decisions using the National Triage Scale for Australian emergency departments. Accid Emerg Nurs 2000; 8:201–209.
25 Tanabe P, Gimbel R, Yamold PR, Kyriacou DN, Adams JG. Reliability and validity of scores on The Emergency Severity Index version 3. Acad Emerg Med 2004; 11:59–65.
26 Harvard Medical Faculty Physicians. Report on the Emergency Care System in Region Midjylland; 2008.