

Interrater Agreement of The Copenhagen Triage Algorithm

Julie Inge-Marie H Borchsenius^{1*}, Rasmus Bo Hasselbalch¹, Morten Lind¹, Lisbet Ravn¹,
Thomas Kalleose², Martin Schultz^{1,4}, Lars Simon Rasmussen^{3,5}, Kasper Iversen^{1,5}

¹ Department of Emergency Medicine, Herlev and Gentofte Hospital, Copenhagen, Denmark

² Clinical Research Centre, Amager and Hvidovre Hospital, Copenhagen Denmark

³ Department of Anaesthesia, Center of Head and Orthopaedics, Rigshospitalet, University of Copenhagen, Denmark

⁴ Department of Internal Medicine, Herlev and Gentofte Hospital, Copenhagen, Denmark

⁵ Institute of Clinical Medicine, University of Copenhagen, Denmark

Correspondence:

Julie Inge-Marie H Borchsenius

Department of Emergency Medicine,

Herlev and Gentofte Hospital

Borgmester Ib Juuls Vej 1

2730 Herlev

E-mail: julieborchsenius@gmail.com

Abstract

Introduction

Systematic triage is performed in the Emergency Department (ED) to assess the urgency of care for each patient. The Copenhagen Triage Algorithm (CTA) is a newly developed, evidence-based triage system, however the interrater agreement remains unknown.

Method

This was a prospective cohort study. The collection of data was conducted in the three sections (Acute/Cardiology, Medicine and Surgery) of the ED of Herlev Hospital. Patients were assessed independently by two different nurses using CTA. The interrater variability of CTA was calculated using Fleiss kappa. The analysis was stratified according to less or more than 2 years of ED experience.

Results

A total of 110 patients were included of which 10 were excluded due to incomplete data. The raters agreed on triage category 80 % of the time corresponding to a kappa value of 0.70 (95% confidence interval 0.57-0.83). Stratified on ED sections, the agreement was 83 % in the Acute/Cardiology section corresponding to a kappa value of 0.73 (0.55-0.91), 79 % in the Medicine section corresponding to a kappa value of 0.64 (0.39-0.89) and 0.56 % in the Surgery section corresponding to a kappa value of 0.56 (0.21-0.90). The experienced raters had an interrater agreement of 0.73 (0.56-0.90), while the less experienced raters had an agreement of 0.76, (0.28-1.24).

Conclusion

A substantial interrater agreement was found for the Copenhagen triage algorithm.

Key words

Interrater agreement, The Copenhagen Triage Algorithm, Triage, clinical assessment

Introduction

Triage is used to prioritize patients in the Emergency Department (ED). The triage tool is based on perceived urgency in an effort to give the best care and lessen the effects of crowding. (1, 2) There is no consensus regarding the optimal method of triage, but most of the applied models are built on the patient's chief complaints and vital signs. (3)

As triage requires time and resources it is problematic that most triage methods are poorly validated. (1, 3) Most of the systems have been developed based on expert opinion rather than data from large prospective cohorts. (1, 3)

An important feature of a triage system is reliability and different nurses should obtain similar triage categories. The reliability of triage has traditionally been measured using interrater agreement, and studies on contemporary triage systems have shown interrater agreement rates with kappa values varying between 0.20 and 0.87. (1, 3)

Recently, The Copenhagen Triage Algorithm (CTA) has been developed with the intent of creating an evidence-based triage model. (4, 5) An example of the chart used for the CTA triage in this study and a brief description of its use are shown in figure 1. CTA classifies patients based on vital signs and a clinical assessment by the ED nurse and has been shown to be a stronger predictor of mortality than a well-known triage system, Adaptive Process Triage (ADAPT), in a large randomized trial. (4, 5) However, the interrater agreement for this model is unknown.

The objective of this study was to examine the interrater agreement of the CTA in the ED.

Methods

Study design and setting

This was a prospective observational cohort study. The collection of data was conducted in the ED at Herlev Hospital on arbitrarily selected dates from June 2018 through March 2019.

This allowed for inclusion throughout the year in order to obtain a representative patient cohort. Herlev Hospital is a 24-hour secondary care unit offering emergency, level-2 trauma, medical, neurological and surgical care. The hospital provides medical care for 425.000 citizens living in the coverage area and has about 150.000 annual admissions to the ED. (6, 7). Patients admitted directly to specific departments in the hospital do not pass through the ED, and this includes paediatric patients, gynaecological and obstetric patients, and trauma patients admitted to tertiary centres of the region.

Patients are admitted to one of three sections (Acute/Cardiology, Medicine, and Surgery,) in the ED according to their primary symptoms upon arrival, and data collection was alternating between sections. The Acute/Cardiology section receives trauma patients, medical and surgical patients triaged as 'Red' (level 1, resuscitation) as well as patients with orthopaedic or cardiac complaints. The Medicine section treats patients within the field of internal medicine, apart from cardiology, gastroenterology, oncology, and haematology. (8) The Surgery section receives patients with gastric or urological complaints. (8)

Selection of participants

All patients above 16 years of age admitted to the ED on the days selected for data collection were included. Patients with minor injuries triaged as blue by both CTA and ADAPT were excluded from the study, since their vital signs were not measured. Patients who were not triaged by two different ED personnel using CTA were also excluded from the study.

On study days patients admitted to the specific section were assessed once using the conventional ADAPT method and twice by different members of the ED triage personnel using CTA. The triage categories of CTA as well as the maximum waiting time to an assessment by a doctor are identical to those of ADAPT. The red category was the most urgent, while the green category was the least urgent, see figure 1. To ensure the clinical presentation and vital signs of the patient did not change between the assessments, the ED

personnel were performing the CTA assessment at the same time blinded to the other's triage. At the beginning of each shift in the study period, the ED personnel received a brief instruction in CTA and how the triage level should be determined, as well as a general presentation of the study. Apart from this, the nurses did not receive any further training in the use of CTA prior to data collection.

Triage staff

The ED personnel performing the triage were nurses, social- and health service assistants (SOSU) and student nurses, in accordance with the usual clinical practice. We classified the raters as either 'experienced' or 'not experienced', where an "Experienced rater" had at least 2 years of work experience in the ED.

Ethics approval

The study was performed without the need for a formal ethics approval in accordance with Danish law. The triage nurses gave consent to participate, as did patients.

Statistics

We hypothesized that the kappa-value of the interrater agreement would be equal to or higher than 0.7. Prior to the study, we performed a power calculation using the "Power4Cates" function in R. With a power of 80 % and a significance level of 0.05, a population of at least 99 patients was required. We assumed that the distribution of patients on the four triage levels from the most to least urgent category would be 5 %, 15 %, 40 %, 40 %, in accordance with previous studies of the CTA. (5)

The interrater agreement was calculated using Fleiss' kappa. We calculated a global kappa and kappa values for each ED section. We also calculated kappa values according to the level of experience of the raters. Kappa levels were interpreted according to Altman's definitions,

shown in Supplementary Figure 1. (9, 10) The data from this study were analyzed using R version 3.6.1.

Results

During the study period 110 patients were included. Of these, 10 had incomplete data and were excluded, thus a total of 100 patients were included in the study (table 1). A team of 51 ED nurses, SOSU's and student nurses performed the assessment of these, 31 (61%) were classified as "experienced raters" (table 2).

The most commonly used category was "green" with 41.5 % of the final triages, the "yellow", "orange" and "red" category were chosen in respectively 37.5 %, 16 % and 5 % of the triages.

The changes made from primary to final triage as a result of the clinical assessment are presented in Figure 2. The final triage level differed from the primary triage in 52 % of the cases. Using the clinical assessment, the triage level was changed to a more urgent category in 40 % of the triages and changed to a less urgent category in 12 %.

The raters agreed on the CTA triage level 80 % of the time. The interrater agreement in all ED teams was $\kappa=0.70$ (95 % Confidence Interval (CI) 0.57-0.83). The highest agreement was found among the raters from the Acute/Cardiology section of the ED, who agreed 83 % of the time, equivalent to a kappa value of $\kappa=0.73$ (95 % CI 0.55-0.91). In the Medicine section the agreement between raters was 79 % and the corresponding kappa was $\kappa=0.64$ (95 % CI 0.39-0.89). Agreement between the raters was lowest in the triages performed in the Surgery section of the ED, who agreed 71 % of the time with a value of $\kappa=0.56$ (95 % CI 0.21-0.90). The rate of agreements and disagreements distributed at the different triage levels are shown

in Table 3. Agreements on changes of the triage level from primary to final triage based on the clinical assessment are shown in table 4.

A sub analysis was performed to investigate whether the raters' experience affected the agreement. The experienced raters performed the triage of 60 cases and had an agreement of $\kappa=0.73$ (95% CI 0.56-0.90), while the less experienced raters performed the triage of 12 cases with an agreement of $\kappa=0.76$ (95% CI 0.28-1.24). In 28 of the cases where an experienced and a less experienced rater triaged the same patient agreements were lower $\kappa=0.54$ (95% CI 0.28-0.81).

Discussion

This is the first study examining the interrater agreement of CTA. We found that the CTA has a good overall agreement. (9)

There is no standardized procedure regarding the validation of triage methods. (2) As a consequence, few models have been investigated sufficiently – or at all – with regard to interrater agreement. (1, 3) This was the case of the currently used Danish modified version of ADAPT, for which the interrater agreement was unknown.

Internationally, the most widespread triage methods are the Australasian Triage Scale (ATS), Canadian Emergency Department and Acuity Scale (CTAS), Manchester Triage Scale (MTS), the Emergency Severity Index (ESI) and the South African Triage Scale (SATS). (3) These methods have all been validated to some extent in terms of interrater agreement, with resulting κ -values in the range of 0.20 to 0.87. (1, 3)

Studies on interrater agreement fall into two categories according to the applied design. Most published studies have been based on paper case scenarios, while relatively few prospective

studies on interrater variability have been performed in an ED setting with live patients. (3) This may not reflect the reality and complexity of the clinical situation owing to the discrepancy between the examined and the real-life scenarios in the ED, hence not directly transferable to clinical practice. A study of the Canadian Triage Acuity Scale comparing the interrater agreement using live patient versus paper case scenarios showed significantly different results between the two methods with a higher interrater agreement ($\kappa=0.9$ vs 0.76) and generally fewer patients assigned to the urgent triage levels in the live patient group (mean triage level 3.4 for live and 3.2 for paper). (11) This conclusion is supported by comparison of different studies concerning widespread triage methods. (12-14) Furthermore, the studies have varying numbers of raters with some only using a few highly experienced research nurses poorly reflecting regular clinical practice (14, 15) These methodological differences make studies of interrater agreement difficult to compare.

Interrater agreement is central to the practical and safe application of a triage method. If the staff is unable to use the method uniformly as intended, the resulting variation will diminish the effect of triage (16) Most of the currently applied triage methods do not include a clinical assessment and are generally structured with the intent of triaging all patients as uniformly as possible using flow-charts and clear cut-offs for each variable. (3, 17) This should, in theory, result in a high interrater agreement, however, in prioritizing interrater agreement these models fail to take advantage of the knowledge and experience of the ED personnel in their assessment of the individual patient. In the CTA model the clinical assessment plays a much more central role which could lead to a larger variation associated with different raters. Actually, we found a substantial level of agreement between the raters overall, when using CTA.

The combination of clinical assessment and objective measures of the patient's current condition is a prominent feature of the CTA and different from other methods of triage. The ESI is to our knowledge the only internationally widespread method, which includes a clinical assessment as a central part of the system. However, in ESI the clinical assessment precedes a series of questions from flowcharts. (18) The interrater agreement on live patient triage using ESI has shown results close to what we show in this study with substantial agreement ($\kappa=0.78$). (15) These similar results were in spite of the difference in experience between the raters in this study and the ESI study in which most of the triage personnel had prior experience using ESI and received a three-hour intensive training course before data collection began. (15) Considering the limited experience with CTA prior to the study, the findings of an overall substantial agreement in our study suggests that CTA is relatively easy to use and will generate consistent results even when the user lacks experience with the method. Higher interrater levels of agreement would be expected if the study was repeated after the CTA had been applied as the standard triage method in the ED for a longer period. Greater levels of agreement may be observed if highly experienced raters use triage algorithms with detailed and standardised structure. Nevertheless, this is not demonstrated by studies on the interrater agreement of CTAS, a triage algorithm based on vital signs and presenting complaint similar to ADAPT, in comparison to our findings. The interrater agreement of the CTAS has been studied in an ED setting, with mostly moderate results (kappa 0.52-0.66), even though most of the studies used extensive training programs before the start of data collection. (3, 14) In general, the level of agreement should be considered in the context of the efficiency of the model on other parameters as well. The method with a slightly lower level of agreement may be preferred, if the model is superior to the compared model on other parameters.

In our study, the interrater agreement varied between the ED sections. The agreements between raters from the Acute/Cardiology and Medicine sections were both within the range of kappa values considered “good” in accordance with the overall agreement, while the agreement in the Surgery team was only “moderate”. This could be because abdominal complaints were more difficult to evaluate. Difficulties concerning the assignment of triage levels to surgical patients have previously been documented for ADAPT, where, critically, only about 15% of patients with gastrointestinal perforation were triaged correctly according to their clinical urgency. (19)

The Copenhagen Triage Algorithm has previously been validated in terms of prediction of 30-day mortality, rate of admission to the intensive care unit, length of stay, waiting time, and rate of readmission within 30 and 90 days. (5) CTA was superior to ADAPT in predicting 30-day mortality and equal to ADAPT regarding the other examined endpoints. (5) The CTA model has emphasis on a clinical assessment of the patient, which has been shown to be superior than currently applied triage models in terms of assigning each patient to the most relevant urgency level. (20, 21) The clinical assessment implies an appraisal of the need for treatment solely based on observation, patient history, and possibly clinical examination of the patient. (20, 22) The clinical presentation may vary between patients, even if their symptoms and vital signs are alike. CTA may help to overcome some of the obstacles associated with the use of triage models based on more structured formulas, including the presenting symptom of the patient. In several triage models, including CTAS, ATS, MTS and ADAPT, the presenting symptom of the patient is central to determining the patient’s level of urgency. (17, 23) The presenting symptom may be a nonspecific determinant of the urgent need for treatment, which is common to a broad spectrum of diseases, frequent in older patients and associated with lower triage priority and increased in-hospital mortality. (24, 25) CTA represents a more flexible method of triage compared to currently implemented

methods such as CTAS, ATS, MTS and ADAPT, since there is a wider possibility of changing the initial triage level in the final step of the algorithm, the clinical assessment to reflect the individual clinical state of the patient. Mistriage is a known pitfall in triage; overtriage may stretch resources, while undertriage may increase morbidity and mortality due to longer waiting time to treatment. (16, 26, 27) In comparison with ADAPT, CTA has been shown to triage at a significantly lower urgency level than ADAPT without it having a negative effect on the examined patient outcomes. (5)

Strengths and limitations

The overall distribution of patients at the four urgency levels in this study is similar to the one found by a larger prospective trial comparing CTA to ADAPT, indicating a representative sample was collected in our study. (5) Although the sample size of 100 seemed sufficient, it is still relatively small compared to the number of annual visits to the ED and the broad spectrum of patients.

The number of raters and their experience range (nurses, nursing students, SOSU assistants) reflect the reality of the triage process according to the working procedures in Danish ED's. This represents a strength regarding the credibility and application of the results to the clinical situation, but the heterogeneity of the group may have resulted in a decreased level of agreement. The large number of raters also introduces an increased statistical uncertainty as reflected in the wide confidence interval.

The ED personnel had limited experience with the use of CTA and received only a brief instruction prior to using the triage system. Because the data was collected alternately in the three sections of the ED and sporadically during the study period, the personnel were not given an opportunity of increasing their experience using the system over time. Their lack of experience with the method may have reduced the level of agreement, and it is possible that

the interrater agreement would increase over time, if the CTA was implemented as the standard triage method in the ED.

Conclusion

The Copenhagen Triage Algorithm is a valid triage system with substantial inter-rater agreement.

List of abbreviations

ED: Emergency Department

CTA: The Copenhagen Triage Algorithm

ADAPT: Adaptive Process Triage

SOSU: Social- and health service assistants

ATS: Australasian Triage Scale

CTAS: Canadian Emergency Department and Acuity Scale

MTS: Manchester Triage Scale

ESI: The Emergency Severity Index

SATS: The South African Triage Scale

Declarations

Ethics approval and consent to participate

The study was performed without the need for a formal ethics approval in accordance with Danish law. The triage nurses gave consent to participate, as did patients.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare, that they have no competing interests.

Funding

Not applicable

Author's contributions

JHJB was the main author of the manuscript and contributed to the collection of and analyzation of data.

RBH was a major contributor in writing the manuscript and analysing the data as well as planning the study.

TK performed the calculations of kappa.

ML and LR contributed to the planning of the data collection.

LSR contributed substantially to the revision of the work.

MS was a major contributor in writing the manuscript as well as planning the study. MS also contributed to the collection of and analyzation of data.

KI had the main idea of the project and contributed to the planning of the study.

All the authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Farrohknia N, Castren 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. *Scandinavian journal of trauma, resuscitation and emergency medicine*. 2011;19:42.
2. Kuriyama A, Urushidani S, Nakayama T. Five-level emergency triage systems: variation in assessment of validity. *Emerg Med J*. 2017;34(11):703-10.
3. Hinson JS, Martinez DA, Cabral S, George K, Whalen M, Hansoti B, et al. Triage Performance in Emergency Medicine: A Systematic Review. *Ann Emerg Med*. 2019;74(1):140–52.
4. Hasselbalch RB, Plesner LL, Pries-Heje M, Ravn L, Lind M, Greibe R, et al. The Copenhagen Triage Algorithm: a randomized controlled trial. *Scandinavian journal of trauma, resuscitation and emergency medicine*. 2016;24(1):123.
5. Hasselbalch RB, Pries-Heje M, Schultz M, Plesner LL, Ravn L, Lind M, et al. The Copenhagen Triage Algorithm is non-inferior to a traditional triage algorithm: A cluster-randomized study. *PLoS One*. 2019;14(2):e0211769.
6. Hospitalets profil. 2020. <https://www.herlevhospital.dk/om-hospitalet/visioner-og-maal/Sider/Hospitalets-profil.aspx>. Accessed 25 Feb 2020.
7. Folkmann T. Nøgletal for Herlev og Gentofte Hospital. <https://www.herlevhospital.dk/om-hospitalet/Organisation/Sider/Noegletal.aspx>. Accessed 25 Feb 2020.
8. Bruun AS. Om Akutmodtagelsen. <https://www.herlevhospital.dk/afdelinger-og-klinikker/akutmodtagelsen/om-afdelingen/Sider/Om-Akutmodtagelsen.aspx>. Accessed 25 Feb 2020.
9. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm*. 2013;9(3):330-8.
10. Altman DG. *Practical statistics for medical research*: Taylor & Francis Ltd; 1990.
11. Worster A, Sardo A, Eva K, Fernandes CM, Upadhye S. Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. *J Emerg Nurs*. 2007;33(4):319-23.
12. Goransson K, Ehrenberg A, Marklund B, Ehnfors M. Accuracy and concordance of nurses in emergency department triage. *Scandinavian journal of caring sciences*. 2005;19(4):432-8.
13. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S. Reliability of the Canadian emergency department triage and acuity scale: interrater agreement. *Annals of emergency medicine*. 1999;34(2):155-9.
14. 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(4):260-6.
15. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity index triage algorithm version 2 is reliable and valid. *Acad Emerg Med*. 2003;10(10):1070-80.
16. Staudenmayer K, Lin F, Mackersie R, Spain D, Hsia R. Variability in California triage from 2005 to 2009: a population-based longitudinal study of severely injured patients. *J Trauma Acute Care Surg*. 2014;76(4):1041-7.
17. Danish Emergency Process Triage. <http://www.deptriage.dk>. Accessed 3rd November 2019.
18. ESI algorithm. <https://www.esitriage.com/esi-algorithm>. Accessed 3rd November.

19. Bohm AM, Tolstrup MB, Gogenur I. Adaptive process triage system cannot identify patients with gastrointestinal perforation. *Dan Med J*. 2017;64(7).
20. Iversen AK S FJ, Schou M, Rasmussen LS, Køber L, Iversen K. Comparison of systematic triage with clinical assessment in prediction of short-term mortality. *Scand J Trauma Resusc Emerg Med*. 2015; 23(Suppl 1): A15.
21. Iversen AK, Kristensen M, Ostervig R, Forberg JL, Schou M, Iversen K. No evidence that formalized triage is superior to informally structured triage. *Ugeskr Laeger*. 2015;177(30).
22. Storm-Versloot MN, Ubbink DT, Kappelhof J, Luitse JS. Comparison of an informally structured triage system, the emergency severity index, and the manchester triage system to distinguish patient priority in the emergency department. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*. 2011;18(8):822-9.
23. M JM. The Canadian Triage and Acuity Scale: A Canadian perspective on emergency department triage. *Emerg Med (Fremantle)*. 2003;15(1):6-10.
24. Kohnlein T. [Chest pain in the emergency department : Differential diagnosis and diagnostic strategy]. *Internist (Berl)*. 2017;58(1):3-7.
25. Djarv T, Castren M, Martenson L, Kurland L. Decreased general condition in the emergency department: high in-hospital mortality and a broad range of discharge diagnoses. *Eur J Emerg Med*. 2015;22(4):241-6.
26. Holst JA, Perman SM, Capp R, Haukoos JS, Ginde AA. Undertriage of Trauma-Related Deaths in U.S. Emergency Departments. *West J Emerg Med*. 2016;17(3):315-23.
27. Barfod C, Lauritzen MM, Danker JK, Soletormos G, Forberg JL, Berlac PA, et al. Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study. *Scandinavian journal of trauma, resuscitation and emergency medicine*. 2012;20:28.

Table 1: Baseline characteristics of patients admitted to the emergency department

Baseline characteristics	N=100
Patients	
Female	53
Age, years (mean, sd)	66 (21)
ED team	
Acute/Cardiology	53
Medicine	33
Surgery	14

Acute/Cardiology, Medicine and Surgery refer to the respective ED sections.

Table 2: Baseline characteristics of raters performing triage in the emergency department

Baseline characteristics N=51	
Raters	
Experienced	31
Less experienced	20
<u>Acute/Cardiology</u>	
Experienced	20
Less experienced	10
<u>Medicine</u>	
Experienced	14
Less experienced	11
<u>Surgery</u>	
Experienced	6
Less experienced	2

The experienced raters were defined by at least 2 years of working in the ED. Some of the raters performed triages in several ED teams, and are therefore included in more than one ED team in the table.

Table 3: Agreements and disagreements on triage level

Number of agreements and disagreements
on the assigned triage level

	Green (Rater B)	Yellow (Rater B)	Orange (Rater B)	Red (Rater B)
Green (Rater A)	37	8	1	0
Yellow (Rater A)		29	9	0
Orange (Rater A)			10	2
Red (Rater A)				4

This table shows how often the raters agreed on the final triage level. The rows represent the final triage level chosen by “Rater A” in each CTA double triage. The columns represent the final triage level chosen by “Rater B”. E.g. the raters agreed on green as the final triage level in 37 out of 100 cases. In 8 out of 100 cases, one rater chose the green triage level in the final triage, while the other rater chose the yellow triage level. The raters agreed on green as the final triage level in 37 % of the cases, yellow in 29 % of the cases, orange in 10 % of the cases and red in 4 % of the cases. In total, the raters agreed on the assigned triages level in 80 % of the cases. The raters disagreed on the assigned triage level in 20 % of the cases.

Table 4: Agreements on changes of the triage level from primary to final triage based on the clinical assessment.

Agreements on changes of the triage level from primary to final triage

based on the clinical assessment.

	Green	Yellow	Orange	Red
2 up				2
1 up		16		3
Unchanged	33	5		4
1 down	4	8		1

The table shows how often the raters agreed on the final triage and how the triage level changed from primary to final triage based on a clinical assessment. The columns (Green, Yellow, Orange and Red) represent the final triage level chosen by each pair of raters. The rows show the changes made from primary to final triage. For example, the raters agreed on uptriage of one triage level with yellow as the final triage 16 out of 100 times. In 42 out of 100 cases, the triage level did not change from primary to final triage. In 38 out of 100 cases, the triage level changed from primary to final triage. The raters agreed on the final triage level 80 % of the time, the disagreements are not shown in this table.

Figures legends

Figure 1 - The Copenhagen Triage Algorithm. The patients are initially classified using a vital sign scoring system, but the assigned category suggested by the score can be altered based on a clinical assessment. CTA allows for a two-class upgrade or a one-class downgrade of the triage category, when the ED nurse assesses that the initial category is not in line with the clinical state of the patient.

Figure 2 – Sankey diagram showing the changes from primary to final triage. The primary triages are displayed on the left side and connected to the final triages on the right side of the diagram. The connections indicate the number of triages with identical primary and final triage.

Supplementary Figure 1 - Altman's kappa strength interpretation criteria

Figure 1: The Copenhagen Triage Algorithm

1 RED Resuscitation	0 min	2 ORANGE Urgent	15 min	3 YELLOW Less urgent	60 min	4 GREEN Not urgent	180 min	5 BLUE 240 min
		Primary triage		Final triage		Oxygen treatment: If the patient receives oxygen treatment after arrival to the ED		
Oxygen treatment	4 points	8-14 points				Vital parameters: Measured prehospitally or upon arrival to the ED		
Systolic BP <100	3 points	4-7 points				Final triage: After the primary triage, the final triage color is decided based on a clinical assessment. It is possible to change the category from one level down to two levels up.		
Heart rate >100	2 points	2-3 points						
Respiratory rate >22	3 points	0 points						
O ₂ saturation <94%	2 points							

Figure 1 The Copenhagen Triage Algorithm. The patients are initially classified using a vital sign scoring system, but the assigned category suggested by the score can be altered based on a clinical assessment. CTA allows for a two-class upgrade or a one-class downgrade of the triage category, when the ED nurse assesses that the initial category is not in line with the clinical state of the patient. Red: The most urgent triage level. The patient should be seen by a doctor immediately. Orange: The urgent triage level. The patient should be seen by a doctor within 15 minutes. Yellow: The less urgent triage level. The patient should be seen by a doctor within 60 minutes. Green: The non-urgent triage level. The patient should be seen by a doctor within 180 minutes.

Figure 2: The changes made from primary to final triage based on a clinical assessment

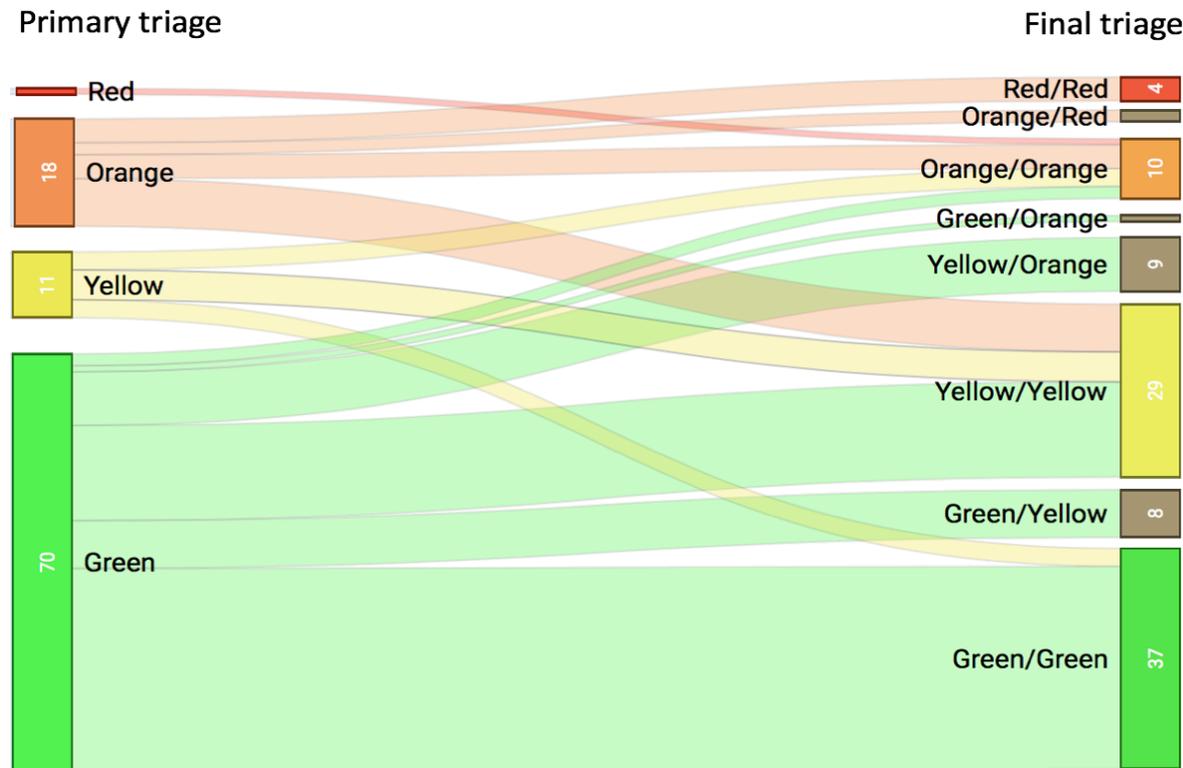


Figure 2 The changes made from primary to final triage based on a clinical assessment. The CTA allows the rater to change the primary triage level (given by the patient’s vital signs) based on a clinical assessment. The diagram illustrates the impact of the clinical assessment on the final triage level. The primary triage is displayed on the left side of the diagram. Green, Yellow, Orange and Red refer to the respective triage levels of the primary triages. On the left, the triage categories chosen by each rater in the final triage are displayed. E.g. Green/yellow: Rater A chose the green triage level in the final triage, while rater B chose the yellow level. The numbers in the bars represent each of the primary triage levels and the combinations of final triage levels represent the number of patients triaged to the respective categories.

Supplementary Figure 1: Altman's kappa interpretation criteria

κ	Strength of agreement
<0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

Supplementary Figure 1: Altman's kappa interpretation criteria