**RESEARCH ARTICLE**

**Open Access**

# FRENCH versus ESI: comparison between two nurse triage emergency scales with referent scenarios

Antoine Aubrion[1,2,3*] , Romain Clanet[1,4], JP Jourdan[5], Christian Creveuil[6], E Roupie[1,7] and Richard Macrez[1,7,8]

## Abstract

**Objectives:**  Acute triage is needed to prioritize care and achieve optimal resource allocation in busy emergency departments. The main objective is to compare the FRench Emergency Nurse Classification in Hospital scale (FRENCH) to the American scale Emergency Severity Index (ESI). Secondary objectives are to compare for each scale the over and under-triage, the triage matching to the gold standard and the inter-individual sorting reproducibility between the nurses.

**Methods:**  This is a prospective observational study conducting among the nursing staffs and nursing students, selected from Caen University College Hospital and Lisieux Hospital Center emergency departments between two months. Each group individually rank 60 referent clinical cases composed by scales designers. An assessment of scale practicality is collected after for each tool. The collected parameters are analyzed by a Cohen kappa concordance test (κ).

**Results:**  With 8151 triage results of gold standard scenarios sorting in two scales by the same nurses, the FRENCH scale seems to give better triage results than the US ESI scale (nurse: FRENCH 60% and ESI 53%, $p = 0.003$ ; nursing students: FRENCH 49% and ESI 42%, $p < 0.001$). In the two groups ESI has also a big tendency to under-sort ($p = 0.01$), particularly for the most severe patients ($p < 0.01$). The interobserver sorting concordance for any experience gives good results for the FRENCH and the ESI without any difference (nurses : FRENCH $K_{PQ}=0.72$ ESI $K_{PQ}=0.78$; $p = 0.32$ ; students $K_{PQ}=0.44$ $K_{PQ}=0.55$; $p = 0.22$).

**Conclusion:**  The ESI and FRENCH scales comparison on 8151 sorting results shows direct validity in favor of FRENCH one and similar interobserver agreement for both scales.

**Keywords:**  Emergency medicine classification, Emergency nursing, Triage, FRENCH scale, ESI

## Introduction

Acute triage is needed to prioritize care and achieve optimal resource allocation in busy emergency departments. Triage is an old process created during the Napoleonic battlefields, then was developed in the civil sector at the beginning of the XXth century [1]. Triage is a key process in emergency departments (ED) organization consisting to decide patient issue by trying to manage patient care and system efficiency [2]. A systematic triage of patients using approved tools is recommended by national and international societies of emergency medicine [3, 4].

Several scales have been developed as decision supports to guide nurses in triage decision. Australasian Triage Scale (ATS, Australia) is based on clinical features in relation to patient presentations [5]. Manchester's triage

*Correspondence: aubrion-a@chu-caen.fr

[3] Department of emergency medicine, Caen-Normandie Hospital (CHU), Caen, France
Full list of author information is available at the end of the article

Aubrion *et al. BMC Emergency Medicine*     (2022) 22:201

Page 2 of 7

system (MTS, United Kingdom) uses series of general and specific determinants to guide decision-making, with patient presentation algorithms [6]. Canadian scale (CTAS, Canada) is based on ATS but includes diagnosis as well [7]. Emergency Severity Index (ESI, USA) published by the Agency for Healthcare Research and Quality in 2005 excludes immediate vital risk and serious disease before considering necessary estimated care [8]. Finally, French Emergency Nurse Classification in Hospital scale (FRENCH, France) is based on 100 determinants (complaints, signs, and vital parameters) [9]. ESI and FRENCH seem to have the best results for appliance and validity. All these scales are based on expert opinion [10]. Using them in ED needs to precisely identify their role and the expected objectives, with a particular attention to their validity and reliability [11]. Under-triage compromises patient's health and over triage consumed medical team and resources [12]. Triage quality and triage-nurse experience are not always linked [13].

There is no gold standard to evaluate these scales. Evaluating a triage scale with real cases is similar to assessing the way a nurse team can sort patients: it can't independently assess the tool itself [9]. Undirect validity criteria like consumption resources prediction or hospitalization are considered as validity standards [14]. Validity represents the accuracy of the triage scale, but is not an evaluation of patient correct triage. Using patients don't allow to compare to gold standard sorting. Before implementation on real patients, a tool should first prove its validity on its own test cases that it proposes. Furthermore direct validity is not influenced by the scale but can only be performed on paper cases.

Our main objective was to compare the French and ESI scales direct validity, by correct response rate for each scale, over and under-triage rates on its example scenarios. Secondary objectives were to compare for each scale the over and under-triage, the triage matching to the gold standard and the inter-individual sorting reproducibility between the nurses.

## Methods

This is a prospective observational study conducting among the nursing staffs and nursing students, selected from Caen University College Hospital and Lisieux Hospital Center emergency departments. For a situation on a real patient, no referral triage score is available. The reliability can only be evaluated on the homogeneity of the answers given by the interobserver agreement. There is a necessity to rely on "gold standard" or "referent" scripted scenarios to allow for comparative assessments during training, as opposed to unpredictable and uncontrolled presentation of actual patients in the ED. The scenarios composed by the designers of each scale provide the

reference response expected by the experts. This answer defines the "gold standard" sort for each paper scenario, and allows to define the correct answers, over and under-sorting.

We designed two groups: a first group with graduated nurses with at least 5 years of experience working daily in the ED, and triage without formalized scale(GN) and a second one composed by second year nursing students (NS) but without prior experience working in ED or patient sorting. An interval of 15 days between the two tests for each nurse, and a reversal of the order of the tests for each half group was organized. Each nurse received triage training for each scale before working on a support including two parts: a scale presentation of five pages, 60 referent clinical cases, composed by the designer's scales with a gold standard triage score. Each nursehad individually rank these 60 same referent clinical cases in 30 min, with the scale handbook available. Then, they assessed the practicality of each scale.

These scenarios were selected from scale handbooks. For the ESI scale, the presentation and the clinical cases have been translated from the ESI Handbook v.4 into French. Two physicians fluent in English and French reviewed the translation. Some vital constants have been converted into European units rounded to the tenth (weight in pounds to kilograms and temperature in Fahrenheit degrees to Celsius). The French version has been adapted nearest as possible of American one.

Statistical analysis were performed with a concordance Cohen's kappa test with linear weighting (κPL) and quadratic weighting (κPQ). The kappa values interpretation results was based on the definitions provided by Altman and Viera. (0-0.2 poor, 0.2–0.4 passable, 0.4–0.6 moderate, 0.6–0.8 good, 0.8-1 very good) [15, 16]. Their comparison was accomplished by a bilateral Student test. The analysis was made at the Unit of Biostatistics and Clinical Research of the Caen University College Hospital with the IBM SPSS and R software programs. Statistical significance was defined as $p < 0.05$.

## Results

In the first group, sixteen graduated nurses (69.6%) accepted to participate to the study, constituting GN group. Among the students, 69 (90.0%) agreed to answer the test, constituting the NS group. 3 students were excluded by a very low response rate ($< 10\%$).

In GN group, based on 1920 cases sorted in a scale (response rate 100%), we obtained for the FRENCH scale 575 correct triage (59.9%) for FRENCH and 508 (52.9%) for ESI ($p = 0.033$; Table 1 A). Over-triage was more commonfor FRENCH (22.5%) compared to ESI (11.5%), $p < 0.01$. Under-triage was less common for FRENCH (17.6%) compared to ESI (35.6%; $p < 0.01$, Table 1 A). In

**Table 1** Triage results with graduated nurses group (GN)

**A- Nurse sorting comparing to referent triage**

| TRIAGE SCALE | | FRENCH | | | | | | ESI | | | | | | p (Overall) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Overall | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Overall | |
| | Scenarios | 5 | 8 | 25 | 16 | 6 | 60 | 12 | 14 | 13 | 10 | 11 | 60 | |
| Agreement | % | 78.8% | 64.8% | 57.0% | 59.0% | 52.1% | **59.9%** | 56.8% | 42.9% | 51.9% | 49.4% | 65.9% | **52.9%** | **0.033** |
| | *n/all* | *63/80* | *83/128* | *228/400* | *151/256* | *50/96* | ***575/960*** | *109/192* | *96/224* | *108/208* | *79/160* | *116/176* | ***508/960*** | |
| Agreement to one level | | 98.8% | 99.2% | 94.0% | 94.5% | 88.5% | **94.7%** | 92.7% | 78.6% | 92.3% | 99.4% | 89.8% | **89.9%** | 0.052 |
| Under triage | | 21.3% | 17.2% | 20.5% | 18.8% | 0.0% | **17.6%** | 43.2% | 50.4% | 37.0% | 43.1% | 0.0% | **35.6%** | < 0.001 |
| Over triage | | 0.0% | 18.0% | 22.5% | 22.3% | 47.9% | **22.5%** | 0.0% | 6.7% | 11.1% | 7.5% | 34.1% | **11.5%** | < 0.001 |
| Under triage > 1 level | | 1.3% | 0.8% | 1.8% | 0.0% | 0.0% | **0.9%** | 7.3% | 21.4% | 7.2% | 0.0% | 0.0% | **8.0%** | 0.0019 |
| Over triage > 1 level | | 0.0% | 0.0% | 4.3% | 5.5% | 11.5% | **4.4%** | 0.0% | 0.0% | 0.5% | 0.6% | 10.2% | **2.1%** | 0.053 |

**B- Nurse triage agreement with the gold standard**

| | Unweighted kappa | | Linear kappa | | Quadratic kappa | |
|---|---|---|---|---|---|---|
| | FRENCH | ESI | FRENCH | ESI | FRENCH | ESI |
| Mean | 0.47 | 0.41 | 0.63 | 0.63 | 0.77 | 0.78 |
| Median | 0.49 | 0.41 | 0.64 | 0.65 | 0.77 | 0.80 |
| 1st quartile | 0.42 | 0.37 | 0.60 | 0.59 | 0.74 | 0.77 |
| 3rd quartile | 0.52 | 0.50 | 0.68 | 0.70 | 0.82 | 0.84 |

**C- Triage agreement between all the nurses**

| | FRENCH | | ESI | | p |
|---|---|---|---|---|---|
| unweighted | 0.33 | (0.26 ; 0.39) | 0.35 | (0.27 ; 0.43) | 0.67 |
| linear | 0.52 | (0.44 ; 0.60) | 0.58 | (0.49 ; 0.67) | 0.34 |
| quadratic | 0.69 | (0.60 ; 0.77) | 0.75 | (0.66 ; 0.83) | 0.34 |

comparison, for level 1, the same users have correctly sorted 63 cases out of 80 (78.8%) with the French, against 109 out of 192 (56.8%) with the ESI ($p < 0.001$). For level 2, 64.8% with the French (83/128), against 42.9% with the ESI (96/224) ; $p < 0.001$ (Table 1 A). Categories 1 and 2 were more under-sorted in ESI (level 1: 43.2%, level 2: 50.4%, $p < 0.001$, Table 1 A) compared to FRENCH scale (level 1: 21.3%, level 2: 17.2%; $p < 0.001$, Table 1 A). For concordance test compared to the gold standard sorting, we obtained an average of $\kappa PL = 0.63$ and $\kappa PQ = 0.77$ for the FRENCH; similar to average of $\kappa PL = 0.63$ and $\kappa PQ = 0.78$ for ESI (Table 1B). For inter-observer triage concordance we obtained $\kappa PL = 0.52$ and $\kappa PQ = 0.69$ for FRENCH and $\kappa PL = 0.58$ and $\kappa PQ = 0.75$ for ESI without any significative difference (p-$\kappa PL = 0.34$; p-$\kappa PQ = 0.34$; Table 1 C).

In the NS group, based on the 6231 cases sorted in a scale (response rate 78% [IC 74–82]), the same users have correctly sorted 1515 cases out of 3101 (48.8%) with the FRENCH scale, against 1327 out of 3130 (42.4%) with the

**Table 2** Triage results with nurses student group (NS)

**A- Student sorting comparing to referent triage**

| TRIAGE SCALE | | FRENCH | | | | | | ESI | | | | | | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Overall | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Overall | |
| | | 5 | 8 | 25 | 16 | 6 | 60 | 12 | 14 | 13 | 10 | 11 | 60 | |
| Agreement | % | 54.9% | 50.6% | 55.1% | 42.6% | 28.9% | **48.8%** | 37.2% | 42.9% | 35.6% | 54.7% | 44.3% | **42.4%** | **<0.001** |
| | *n/overall* | *147/268* | *220/435* | *761/1380* | *288/676* | *99/342* | *1515/3101* | *242/650* | *325/757* | *223/627* | *270/494* | *267/602* | *1327/3130* | |
| Agreement to one level | | 78.0% | 93.6% | 90.9% | 86.5% | 66.8% | **86.6%** | 93.1% | 73.2% | 89.8% | 94.5% | 82.8% | **85.9%** | 0.52 |
| Under triage | | 45.1% | 25.3% | 16.5% | 12.1% | 0.0% | **17.4%** | 62.8% | 48.6% | 38.0% | 21.7% | 0.0% | **35.8%** | <0.001 |
| Over triage | | 0.0% | 24.1% | 28.3% | 45.3% | 71.1% | **33.7%** | 0.0% | 8.5% | 26.5% | 23.7% | 55.7% | **21.8%** | <0.001 |
| Under triage > 1 level | | 22.0% | 6.4% | 2.2% | 0.0% | 0.0% | **3.8%** | 6.9% | 26.8% | 7.7% | 0.0% | 0.0% | **9.5%** | <0.001 |
| Over triage > 1 level | | 0.0% | 0.0% | 6.9% | 13.5% | 33.2% | **9.7%** | 0.0% | 0.0% | 2.6% | 5.5% | 17.2% | **4.7%** | <0.001 |

**B- Student triage agreement with the gold standard**

| | Unweighted kappa | | Linear kappa | | Quadratic kappa | |
|---|---|---|---|---|---|---|
| | FRENCH | ESI | FRENCH | ESI | FRENCH | ESI |
| Mean | 0.31 | 0.28 | 0.45 | 0.49 | 0.58 | 0.66 |
| Median | 0.30 | 0.28 | 0.45 | 0.50 | 0.60 | 0.68 |
| 1st quartile | 0.22 | 0.21 | 0.37 | 0.43 | 0.51 | 0.60 |
| 3rd quartile | 0.38 | 0.33 | 0.51 | 0.56 | 0.66 | 0.73 |

**C- Agreement between all the students**

| | FRENCH | | ESI | | p |
|---|---|---|---|---|---|
| unweighted | 0.18 | (0.13 ; 0.22) | 0.24 | (0.14 ; 0.33) | 0.27 |
| linear | 0.31 | (0.23 ; 0.39) | 0.41 | (0.28 ; 0.53) | 0.18 |
| quadratic | 0.44 | (0.33 ; 0.55) | 0.55 | (0.40 ; 0.71) | 0.22 |

ESI ($p < 0.001$; Table 2 A). Over-triage was more common for FRENCH (33.7%) compared to ESI (21.8%; $p < 0.001$). Under-triage was less common for FRENCH (17.4%) compared to ESI (35.8%; $p < 0.001$). In comparison, for level 1, the same users have correctly sorted 147 cases out of 268 (54.9%) with the French, against 242 out of 650 (37.2%) with the ESI ($p < 0.01$). For level 2, 50.6% with the French (220/435), against 42.9% with the ESI (325/757) ; $p = 0.02$. Categories 1 and 2 were more under-sorted in ESI (level 1: 62.8%, level 2: 48.6%) compared to FRENCH scale (level 1: 45.1%, level 2: 25.3%; $p < 0.01$ and $p < 0.01$. For concordance test compared to the gold standard sorting, we obtained an average of $\kappa PL = 0.45$ and $\kappa PQ = 0.58$ for the FRENCH compared to $\kappa PL = 0.49$ and $\kappa PQ = 0.66$ for ESI (Table 1B). For inter-observer triage concordance we obtained $\kappa PL = 0,31$ and $\kappa PQ = 0.44$ for FRENCH and $KPL = 0,41$ and $\kappa PQ = 0.55$ for ESI without any significant difference for nursing students (p-$\kappa PL = 0.18$ ; p-$\kappa PQ = 0.22$; Table 2C).

Both groups (GN and NS) found significatively more practical FRENCH scale compared to ESI (Table 3).

## Discussion

With French GN or SN, we showed that FRENCH scale gives better triage results than the US ESI scale, whatever nurses experience. For unstable patients, we observed that ESI scale is less performing : more undertriage could potentially lead to adverse outcomes and supports the use of FRENCHThe inter-observer triage concordance for any experience gives no significant difference with 8100 results of dual triage. Moreover, the practicality of the two French user populations are in favor of the FRENCH scale about learning, use facility, and tool security.

We chose to take up the 60-referent cases made by each of the two sorting scales developers; their rating defined a gold standard triage for these fictional patients. The same nurses have sorted the cases of the two scales, after training, but without previous experience of one or the other of the two scales. These clinical situations on paper make it possible to guarantee identical triage conditions: each

nurse is confronted with the same cases, without neither the subjectivity induced towards the patient, the simultaneous sorting with two judges' blind, nor the sorting bias posteriori. This is independently assessed of their triage experience and their service habits [9]. These results are consistent with other studies on these same ESI ($\kappa PL = 0.84$, IC: 0.77–0.91) [17].

For the inter-observer concordance test we obtained correct results with two scales and our data are in accordance to the literature (FRENCH $\kappa PL = 0.77$, $\kappa = 0.64$, ESI v.3 $\kappa PL = 0.89$) [9, 18]. These concordances are lower than in referring articles with a common practice of the evaluated scale for nurses [9]. In order to compare their results, our nurses did not know either of the two scales in common practice. We have not compared the results of nurses with similar experience but with all experience. Both concordances would be artificially increased. On the other hand, in order to test better knowledge, it seems that the referent cases proposed by the developers of each scale are more difficult to sort than the average of real cases. Finally, triage of paper cases usually gives a lower agreement than the identical real cases [19].

This study has several limits. Paper scenarios obtain different results of triage compared to real cases. However it allows a better inter-individual comparability of the triage. However, paper-cases may not be representative of real clinical practice in ED and leave room for imagination. Cases simulated by an actor would not have this limitation. Furthermore, as the clinical scenarios were performed differently for the two scales, the differences observed may be due to differences in the difficulty of the scenarios (level 1–2 scenarios: 13/60 for French and 26/60 for ESI). Using the same scenarios, by consensus of experts on both scales would not have such important limitation. However, the evaluation of the clinical cases by the experts who constructed each scale seemed more robust than a comparative evaluation by independent experts. The same raters participated in each scale test. Despite an interval of 15 days between the two tests for each nurse, and a reversal of the order of the

**Table 3** Users scale practicality evaluation about FRENCH and ESI

| Question | Nurses | | | | Students | | | |
|---|---|---|---|---|---|---|---|---|
| | FRENCH | ESI | IC 95% | p | FRENCH | ESI | IC 95% | p |
| Ease of learning | 8.83 | 6.29 | [1.78 ; 5.22] | $p < 0.01$ | 7.81 | 6.81 | [0.18 ; 0.82] | $p = 0.017$ |
| Ease of use | 8.00 | 5.37 | [2.05 ; 4.95] | $p < 0.01$ | 7.71 | 5.94 | [1.00 ; 2.54] | $p < 0.001$ |
| Speed of implementation | 4.50 | 5.33 | [-1.71;1.38] | $p = 0.793$ | 7.23 | 6.94 | [-0.56 ; 1.14] | $p = 0.498$ |
| Safety of sorting | 7.83 | 4.67 | [3.25 ; 5.42] | $p = 0.039$ | 7.87 | 4.45 | [2.73 ; 4.10] | $p < 0.001$ |
| Sense of use of the scale | 8.66 | 4.20 | [3.25 ; 5.42] | $p < 0.001$ | 7.81 | 5.29 | [1.85 ; 3.18] | $p < 0.001$ |
| Correlation with the clinical feeling of the situation | 7.67 | 6.04 | [-0.07 ; 4.41] | $p = 0.055$ | 7.32 | 4.55 | [2.18 ; 3.37] | $p < 0.001$ |

tests for each half group, contamination between the two scales was indeed possible (Additional file 1: Figure S1). As the difference in choice between two judges is important for the same scale, it seemed important to use the same judges in each scale in order to be able to interpret the results of correct sorting in one or other of the scales [2]. In the case of a randomized trial, the comparison of kappa cannot distinguish the influence of the scales and those of the different judges in the two groups, because of a low intrinsic kappa for the scale.

Experienced nurses worked in the same emergency department but each one had previous experience in other department. The two populations effectives are different but represents a big part of each analyzed group (respectively 69.6% of experience nurses and 90.0% of the nurse students). For both scales, the training received by learners was short. The time offered to answer the questionnaires was limited. The response rate differs between the two groups but is related to the speed of implementation: unlike experienced nurses, most nursing students were unable to full complete the questionnaire. To limit the measurement bias, nurse had to respect the order of the question. A post-training test over several days could provide different results. Subjectivity bias is limited by retaining the same nurses for both scales. Even if the greatest care has been taken for this stage, the translation of the scale and the paper cases of ESI into FRENCH may have lost some nuances. Selection bias is limited by the absence of sampling since all existing scenarios for each scale were included.

Our results seem to be in favor of the FRENCH scale. Indeed each scale was developed in a given context and for a given care organization. Although naive on each scale and in spite of training identical to each one, they may have been influenced by their training,education, and organization French models. The French model defines the nurse as an effector of the medical decision. Whereas the American model leaves a more important place to the decision-making of the nurse, American nurses are more trained and better qualified than in our European system. Thus, for our French nurses, the FRENCH scale may appear more adapted to their practice and more secure. In fact, FRENCH scale leave less freedom to the nurses compared to the ESISo, in a French-style health care system, French nurses and nursing students seem more prone to apply according to the way it has been elaborated, the FRENCH scale than the American ESI scale.

Our study compares two cultures and two ways of thinking. The ESI leaves more freedom for the nurses judgment coupled with an assessment of care needed. FRENCH headed scale approaches an e-sorting scale. A recent paper evaluates an electronic triage system (e-triage) based on machine learning that predicts likelihood of acute outcomes enabling improved patient differentiation [20]. As the FRENCH scale, e-triage is composed of a random model applied to triage data (vital signs, chief complaint) to determine a triage score and seems to improve ESI under-sorting. In both cases the final adjustment of the triage score, depending on the clinical context and the patient's medical history, is based on the evaluator's experience: triage nurse for FRENCH and big data in e-sorting. In view of their similar results with ESI, these two conceptions of final triage fit still need to be compared together in a prospective study.

However in the absence of gold standard scale, any comparison need to be attentive to evaluation criteria. In fact the ESI takes into account the care to be provided and therefore influences the indirect validity. Thus an e-sorting scale will have good concordance results. The direct validity is not influenced by the scale but can only be performed on paper cases. However gold standard scenarios are necessarily developed by the designers of the scale. A lower difficulty of the scenarios can enable them to obtain better scores of adequate sorting (direct validity). Identical scenarios for each scale, sorted by their own experts, would limit this bias.

## Conclusion

Triage is an old process required in any emergency department but without gold standard scale. Different scales are proposed according to countries and cultures. The comparison of two scales requires evaluators without previous experience of one or the other of the two scales and gold standard scenarios for each scale. The ESI and FRENCH comparison on referentpapercase with experience and student French nurses showed direct validity in favor of FRENCH and similar inter-observer agreement for both scales. Triage with these scales appears easily applicable and reproducible and will improve our practices in caring from the emergencies' reception. Further studies are needed, especially to evaluate the effectiveness of guided e-sorting scales.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12873-022-00752-z.

---

**Additional file 1: Figure S1**. Study diagram.

**Additional file 2**.

---

Aubrion *et al. BMC Emergency Medicine*    (2022) 22:201

Page 6 of 7

## Conflict of interest

The authors declare not having any conflicts of interest.

## Authors' contributions

Exp AUBRION A led to A A.. A A., E R, C R, J J.P. led the project, the data collection, the analysis of the results, the drafting and proofreading of the manuscript. C C. carried out the statistical analysis. M R., E R, R R. contributed to the proofreading of the manuscript. The author(s) read approved the final manuscript.

## Availability of data and materials

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

# Declarations

## Ethics approval and consent to participate

This prospective observational study is not subject to the criteria established by article R1121-2 of the public health code, and therefore outside the competence of the committee for the protection of individuals. Indeed, it belongs to non-interventional research without patient participation or access to medical data.Our study proposes the evaluation by professional nurses or apprentices of clinical situations on paper proposed in the learning guide of the Emergency Severity Index (ESI), and those of the French Emergency Nurse Classification in Hospital Scale (FRENCH). These practice cases are the official tests that give the correct triage for each situation. They are considered gold standard cases for testing the validity of the sorting in each scale.
The evaluation of a triage scale with real patients would not allow to compare the 2 scales between them. It would be similar to evaluating how well a team of nurses can triage patients. It cannot independently evaluate the tool itself. As indicated in the bibliography, direct validity can only be achieved by using the paper-based situations proposed in the learning guide for each scale. All participating nurses were informed about the objectives of the survey. Their written consent was obtained before filling in the questionnaire. The clinical situations used were those of the official national guides. No real patients or medical data were used in this study.

## Consent for publication

The clinical situations used were those of the official national guides. No real patients or medical data were used in this study.

## Consent for publication

was obtained for all study participants and all authors of the manuscript. Unfortunately, Pr Eric Roupie died suddenly after the submission phase, and before the exchanges with the reviewers. The corrections have been made with respect to his contributions and the same way of the initial article. We wish to dedicate this work to his memory.

## Competing interests

The authors have no competing interests.

## Author details

[1]Emergency medical service (SAMU 14), Caen University Hospital, Caen, France. [2]Emergency department, Lisieux Hospital, Lisieux, France. [3]Department of emergency medicine, Caen-Normandie Hospital (CHU), Caen, France. [4]Emergency department, Bayeux Hospital, Bayeux, France. [5]Pharmacy department, Public hospital, Vire, France. [6]Department of Biostatistics and Clinical Research, Caen University Hospital, Caen, France. [7]Physiopathology and Imaging of Neurological Disorders, Normandie Univ, UNICAEN, INSERM, UMR-S U1237, Institut Blood and Brain @ CaenNormandie, GIP Cyceron, Boulevard Becquerel, 14074, Caen, France. [8]Normandie Univ, Unicaen, Cermn, 14000 Caen, France.

## References

1. Robertson-Steel I. Evolution of triage systems. Emerg Med J EMJ. 2006;23:154–5.
2. 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. https://doi.org/10.1016/j.annemergmed.2018.09.022.
3. Le triage en structure des urgences. Recommandations formalisées d'experts, société française de médecine d'Urgence http://www.sfmu.org/fr/vie-professionnelle/outils-professionnels/referentiels-sfmu; 2013. Accessed 25 Apr 2021.
4. Gilboy N, Tanabe T, Travers D, Rosenau AM. Emergency severity index (ESI): a triage tool for emergency department care, version 4. Implementation handbook 2020 edition. AHRQ Publication. Rockville, MD. Agency for healthcare research and quality. 2020. https://www.ena.org/docs/default-source/education-document-library/triage/esi-implementation-handbook-2020.pdf?sfvrsn=fdc327df_4.
5. Australasian college for emergency medicine. guidelines on the implementation of the ATS in emergency departments. https://www.acem.org.au/getattachment/d19d5ad3-e1f4-4e4f-bf83-7e09cae27d76/G24-Implementation-of-the-Australasian-Triage-Scal.aspx; 2013. Accessed 14 Nov 2021.
6. Santos AP, Freitas P, Martins HMG. Manchester Triage System version II and resource utilisation in the emergency department. Emerg Med J EMJ. 2014;31:148–52.
7. Canadian association of emergency physicians (CAEP), national emergency nurses Affiliation of Canada (NENA). Association des médecins d'urgence du Québec (AMUQ). implementation guidelines for the Canadian emergency department triage & acuity scale (CTAS). http://ctas-phctas.ca/wp-content/uploads/2018/05/ctased16_98.pdf; 1998. Accessed 26 June 2021.
8. Gilboy N, Tanabe P, Travers DA, Rosenau AM, Eitel DR. Emergency Severity Index, Version 4: implementation handbook. AHRQ publication No. 05-0046-2. Rockville: Agency for Healthcare Research and Quality; May 2005.
9. Taboulet P, Moreira V, Haas L, Porcher R, Braganca A, Fontaine J-P, et al. Triage with the french emergency nurses classification in Hospital scale: reliability and validity. Eur J Emerg Med Off J Eur Soc Emerg Med. 2009;16:61–7.
10. Moll HA. Challenges in the validation of triage systems at emergency departments. J Clin Epidemiol. 2010;63:384–8.
11. Zachariasse JM, Nieboer D, Oostenbrink R, Moll HA, Steyerberg EW. Multiple performance measures are needed to evaluate triage systems in the emergency department. J Clin Epidemiol. 2018;94:27–34.
12. Fernandes CMB, Tanabe P, Gilboy N, Johnson LA, McNair RS, Rosenau AM, et al. Five-level triage: a report from the ACEP/ENA five-level triage Task Force. J Emerg Nurs. 2005;31:39–50.
13. Martin A, Davidson CL, Panik A, Buckenmyer C, Delpais P, Ortiz M. An examination of ESI triage scoring accuracy in relationship to ED nursing attitudes and experience. J Emerg Nurs. 2014;40:461–8.

14. FitzGerald G, Jelinek GA, Scott D, Gerdtz MF. Emergency department triage revisited. Emerg Med J. 2010;27:86–92.
15. Altman D. Practical statistics for Medical Research. London: Chapman & Hall; 1991.
16. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37:360–3.
17. Bergs J, Verelst S, Gillet J-B, Vandijck D. Evaluating implementation of the emergency severity index in a belgian hospital. J Emerg Nurs. 2014;40:592–7.
18. Maningas PA, Hime DA, Parker DE, McMurry TA. The soterion rapid triage system: evaluation of inter-rater reliability and validity. J Emerg Med. 2006;30:461–9.
19. Worster A, Sardo A, Eva K, Fernandes CMB, Upadhye S. Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. J Emerg Nurs. 2007;33:319–23.
20. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann Emerg Med. 2018;71:565–74.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.