

Reproducibility and Clinical Impact of the Manchester Triage System: Insights from a Multicenter Vignette Study

Authors: Arian Zaboli RN, MSN¹; Francesco Brigo MD¹; Gabriele Magnarelli RN, MSN²; Hugh Gorick PhD Student³; Tiziano Garbin RN, MSN⁴; Patrick Clauser RN⁴; Serena Sibilio RN, MSN⁵; Gloria Brigiari MD⁶; Magdalena Massar MSc¹; Michael Mian MD^{1, 7}; Norbert Pfeifer MD²; Gianni Turcato MD^{8, 9}.

Affiliations:

- 1) Innovation, Research and Teaching Service (SABES-ASDAA), Teaching Hospital of the Paracelsus Medical Private University (PMU), Bolzano, Italy.
- 2) Department of Emergency Medicine, Hospital of Merano-Meran (SABES-ASDAA), Merano-Meran, Italy; Lehrkrankenhaus der Paracelsus Medizinischen Privatuniversität, Salzburg, Austria.
- 3) School of Health Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, England.
- 4) Department of Emergency Medicine, Hospital of Bolzano-Bozen (SABES-ASDAA), Bolzano-Bozen, Italy; Lehrkrankenhaus der Paracelsus Medizinischen Privatuniversität, Salzburg, Austria.
- 5) Universitat Basel Department Public Health, Institute of Nursing Science, Basel, BS, Switzerland
- 6) Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, Padova, Italy.
- 7) College of Health Care-Professions Claudiana, Bozen, Italy.
- 8) Department of Internal Medicine, Intermediate Care Unit, Hospital Alto Vicentino (AULSS-7), Santorso, Italy.
- 9) UniCamillus-Saint Camillus International University of Health Science, Rome, Italy

Corresponding author:

Arian Zaboli RN, MSN.

Adress: via A. Volta, 13°; Bolzano; Italy

E-mail: zaboliarian@gmail.com

Contribution statement (CRediT):

Arian Zaboli: is responsible for the overall content [as guarantor].

Arian Zaboli: Conceptualization, Methodology, Investigation, Formal analysis, **Francesco Brigo:** Supervision, Writing-Original draft preparation; **Gabriele Magnarelli:** Investigation, Data Curation; **Hugh Gorick:** Conceptualization Investigation, Writing-Original draft preparation; **Tiziano Garbin:** Investigation; **Patrick Clauser:** Investigation; **Serena Sibilio:** Conceptualization, Writing-Original draft preparation; **Magdalena Massar:** Investigation, Data Curation; **Gloria Brigiari:** Formal analysis; **Michael Mian:** Supervision; **Norbert Pfeifer:** Supervision, Resources; **Gianni Turcato:** Conceptualization, Methodology, Formal analysis, Writing-Original draft preparation.

Authors' disclosure statements: Not applicable.

Competing Interest statement: No competing interest.

Funding statement: Not applicable.

Keywords: Triage; Manchester Triage System; Standard; Agreement; Reference; Application; Simulation.

Reproducibility and Clinical Impact of the Manchester Triage System: Insights from a Multicenter Vignette Study

Abstract

Background: While several studies have evaluated the performance of the Manchester Triage System (MTS), none have specifically examined its accurate application by triage nurses and its association with clinical outcomes. This study focuses on the agreement between nurse-assigned MTS codes and those assigned by an expert group, as well as their ability to predict clinical outcomes.

Methods: This multicenter simulation study was conducted from January 1 to March 31, 2024 across four Emergency Departments (EDs) in the region of South Tyrol (Italy) employing MTS in routine clinical practice. Two emergency physicians developed 30 vignettes derived from real clinical cases to encompass a diverse range of triage scenarios and priority codes. An expert MTS group, composed of three experienced nurses, assigned MTS priority codes following the guidelines outlined in the official MTS textbook. Subsequently, the vignettes were presented to triage nurses, who independently assigned MTS codes. Error rate, agreement between nurse-assigned and expert MTS group codes, and the predictive ability for secondary clinical outcomes (mortality within 72 hours, hospitalization, life-saving intervention, severe condition in the ED, and time-dependent pathology) were compared between the MTS priority assigned by the expert MTS group codes and nurse-assigned triage codes.

Results: Eighty-nine nurses from four EDs participated. The triage code assignment error rate was 28.6% (660/2310). The overall agreement between the triage and expert nurses yielded a Cohen's kappa of 0.590 (95% CI: 0.583-0.598). Expert MTS group applications performed better compared to nurse-assigned codes in predicting clinical outcomes. The mean error rate per nurse was 30% (9/30). Nurses with more years of ED activity and triage expertise had error rates higher than the mean.

Conclusion: The application of MTS using case vignettes was suboptimal in our setting, with more senior nurses having higher error rates. Correct application of the MTS had better prediction of clinical outcomes. It is important to conduct future studies to understand how to best support nursing clinical decision-making in triage.

Key messages

1) What is already known on this topic:

The Manchester Triage System (MTS) stands as one of the most widely adopted standardized triage systems, designed to minimize individual subjectivity in code assignment and ensure equitable access to care. Prior studies have shown that the MTS has only moderate reliability, suggesting variability in its application. However, the underlying factors contributing to these discrepancies, such as nurse characteristics or the quality of their training, remain largely unexplored.

2) What this study adds

This study compared the MTS assignments made by triage nurses to those made by an expert MTS group for 30 clinical vignettes based on real patient cases. There was a 28.1% error rate by the nurses compared to the expert MTS group, with more experienced nurses having a higher error rate than less experienced nurses.

The expert group assignments had better prediction for the actual outcomes in these cases.

3) How this study might affect research, practice or policy

Using proxy outcomes, MTS does appear to have good validity, however it does not appear to be reliably used by nurses in our setting. It is necessary to consider support strategies and tools for nursing clinical decision-making.

Introduction:

Triage serves to assess the priority level of patients upon arrival at the ED, distinguishing those who can wait from those needing immediate medical attention [1, 2]. This distinction is crucial, given the growing ED patient volumes and resource constraints [2, 3, 4]. Standardized triage systems have evolved over decades to ensure consistent risk assessment across diverse healthcare settings.

Among these, the Manchester Triage System (MTS) stands out as extensively studied and widely implemented, utilizing flowcharts for 53 potential presenting symptoms [5, 6]. Despite its broad adoption and substantial research into its functionality, questions persist regarding its reliability in achieving consistent application by triage nurses across different settings [1, 5, 6].

A prior investigation into MTS reliability revealed notable discrepancies among nurses in assigning triage codes to identical patients [7]. Surprisingly, there is limited research into the reasons for these errors or strategies to enhance standardization of assignments. Understanding the factors contributing to these discrepancies is crucial, as errors in code assignment can lead to under-triage, potentially delaying care for critically ill patients, or over-triage, which may unnecessarily burden ED resources and compromise overall efficiency.

Furthermore, existing studies focus on the ability of the MTS assigned to predict clinical outcomes, without evaluating the correct application of the triage system by nurses. As a result, it remains unclear whether discrepancies between triage assignment and outcomes are due to subjective interpretation by the nurses or a lack of understanding of the MTS system itself [8, 9, 10]. In fact, studies on other triage systems demonstrate that between 20% and 30% of assigned triage codes are incorrect according to the triage system, indicating a suboptimal application of triage systems [9, 11].

This study employed vignettes derived from real clinical cases to evaluate the appropriate application of the MTS system by triage nurses and identify nurse characteristics most associated with application errors. We also examined how well triage assignments were associated with clinical outcomes.

Methods

Study design and setting

This was a prospective, vignette-base study conducted between January 1, 2024, and March 31, 2024 across four hospitals in South Tyrol, Italy: Bolzano Hospital, which had 91,000 ED accesses in 2023, Merano Hospital which had 65,000 ED accesses in 2023, Brunico Hospital which had 41,000 ED accesses in 2023 Bressanone Hospital which had 37,000 ED accesses in 2023. The EDs of these hospitals have utilized the MTS as the triage system since 2014. All nurses performing triage must complete a two-day course on conducting triage assessments and using the MTS system which is standardized across all hospitals in the region and uses the same instructors, ensuring uniform training for all nurses. Additionally, nurses must have at least one year of work experience in the ED before they are eligible to perform triage.

Vignette Development

Two emergency physicians with triage expertise and who are responsible for the implementation triage and the structure of triage training in the healthcare organization selected 30 real clinical cases from visits by patients 18 or older between January 1, 2023, to June 31, 2023, using the electronic database of the ED at Merano Hospital. The cases were selected to encompass a broad range of presenting complaints and triage codes, typical of daily practice. From the cases, the physicians wrote vignettes that provided the information available at triage: patient age, gender, mode of arrival to the ED, a brief description of the patient and their presenting symptoms, and vital signs (Supplementary File 1) [7].

Subsequently, the vignettes were anonymized and presented to three expert MTS nurses. These nurses each had over 5 years of triage experience and are involved in translating the MTS manual into Italian; they are responsible for monitoring mistriage incidents and training new triage nurses in the region. The expert group were asked to assign triage codes to each vignette, based on the

correct application of MTS as per the instructions in the MTS textbook [12]. Each nurse assigned a code separately and then agreement was determined using Cohen's kappa, which was 0.956 (95% CI: 0.901 – 1.000). There was only one clinical case where one triage expert assigned a different code than that provided by the other two experts; in this case, the disagreement was resolved through discussion and consultation with the MTS textbook. Thus, for each case there was a single “correct” triage code assigned by the expert group which served as the reference standard for this study. The vignettes were then translated into two languages (Italian and German) by a professional translator. Subsequently, they were reviewed and approved by three native speakers.

Participants

The study was introduced during departmental meetings by local study investigators, who emphasized that participation was voluntary and anonymous and that non-participation would not result in any consequences. The triage nurses received an email sent by department's nursing coordinator, inviting them to participate in the study. The email specified that participation was voluntary and highlighted the study's relevance as an effort to better understand the nurses' skills. To confirm their participation, the nurses had to independently contact the local study investigator to arrange data collection according to their work and personal schedules.

Once enrolled, nurses provided the following information: 1) Age; 2) Gender; 3) Number of years working as a nurse; 4) Number of years working in the ED; 5) Number of years working in triage; 6) Years since the last triage course attended; 7) Degree of education; 8) Personal interest in triage (rated on a scale from 0 to 10, where 0 indicated no interest and 10 maximum interest); 9)

Confidence in assigning triage codes (rated on a scale from 0 to 10, where 0 indicated maximal insecurity and 10 the highest confidence level); 10) ED of employment.

Nurses were also asked for information regarding their opinion of the triage system with the following questions: 1) Performance of the current triage system (rated on a scale from 0 to 10, where 0 indicated the worst performance and 10 perfect performance); 2) Accuracy of the triage

system in prioritizing patients (under-triage, accurate triage classification, or over-triage); 3)

Preference for changing the current triage system (options included: no change, Emergency Severity Index [ESI], South African Triage Scale [SATS], Australasian Triage Scale [ATS], Canadian Triage and Acuity Scale [CATS], or an unvalidated Italian triage system called Gruppo Formazione Triage [GFT]);

The nurses were then asked to read each of the 30 vignettes and assign a code using MTS and not to incorporate their subjectivity into the assessment. Additionally, nurses were not permitted to use digital tools or textbooks, and each nurse completed the form individually. The nurses completed the questionnaire under the supervision of one study investigator to ensure that no additional aids were used.

Outcome

The primary outcome of this study was the frequency of errors in MTS code assignment, using the triage code decided by the expert nursing group as the reference standard.

Secondary outcomes were actual clinical outcomes of the patients as determined by the physicians who had created the vignettes. The MTS score assigned to the patient by the expert group and the nurses were assessed for their ability to predict these outcomes.

Death within 72 hours: Obtained through the registry office of the province of Bolzano, which is responsible for birth and death data for the entire region of South Tyrol.

Life-saving intervention (LSI): This outcome encompasses life-saving measures required by the patient in the ED, as defined by Platts-Mills et al. [13]. These included:

- a) Airway and breathing support, specifically intubation or emergency noninvasive positive pressure ventilation;
- b) Electrical therapy, including defibrillation, emergency cardioversion, or external pacing;
- c) Procedures, such as chest needle decompression, pericardiocentesis, or open thoracotomy;

d) Hemodynamic support, including significant intravenous fluid resuscitation in the context of hypotension, blood administration, or control of major bleeding;

e) Emergency medications, including naloxone, dextrose, atropine, adenosine, epinephrine, or vasopressors.

Hospitalization: Indicates admission of the patient to a regular ward.

Severe condition in the ED: Assessed by the two physicians who had selected the cases to develop the vignettes.

Time-dependent condition: Assessed by the two physicians who had selected the cases to develop the vignettes.

Statistical analysis

Analyses of agreement were conducted for the nurses overall and according to their hospitals.

Concordance in code assignment was assessed using a contingency table comparing the code assigned by expert MTS group versus MTS code assigned by triage nurses. Although the triage code variable is categorical, in this case it was treated as a continuous variable to facilitate representation and assess the mean deviation from the code indicated by the expert MTS group.

Agreement was evaluated using Cohen's kappa, with 95% confidence interval (95% CI). The inter-rater agreement interpreted as follows: 'poor' if Cohen's kappa was <0.00 , 'slight' if between 0.00 and 0.20, 'fair' if between 0.21 and 0.40, 'moderate' if between 0.41 and 0.60, 'substantial' if between 0.61 and 0.80, and 'almost perfect' if between 0.81 and 1.00 [7].

To ensure the forms were not completed randomly, we calculated the expected agreement and compared it to the observed agreement. Incorporating the expected agreement into Cohen's kappa analysis provides a more rigorous interpretation of inter-rater agreement, enabling us to distinguish genuine consistency from chance. This approach enhances both the validity and reliability of the study's findings.

To assess the different distributions of the codes, a bar plot was created comparing the frequency of the codes given by expert MTS group to codes given by the nurses at each hospital and the group as a whole.

For secondary outcomes, we compared the predictive value of correctly applied triage codes to misclassifications for the patient outcomes. We applied logistic regression models for each secondary outcome (e.g., mortality, hospitalization). Each outcome was coded as a binary variable (presence or absence of the event), and the assigned MTS code served as the independent variable. Using these models, we generated predicted probabilities for each outcome, which were then used to construct Receiver Operating Characteristic (ROC) curves. The ROC curves measure the model's ability to distinguish between patients who experienced the outcome and those who did not. The Area Under the ROC Curve (AUROC) quantifies this discriminatory power, with values approaching 1 indicating better discrimination. This analytical approach aligns with methodologies used in previous studies assessing triage systems in relation to patient outcomes [14, 15]. All AUROCs were reported with their respective 95% CI.

Finally, univariate comparisons were conducted to investigate the association between nurse characteristics and error rate. For the group as a whole, the mean number of errors among the 30 vignettes was determined by summing the errors each nurse made across all vignettes and calculating the mean error rate for the entire group. Nurses were then categorized into two groups: those who made more errors than the cohort mean and those who made fewer errors. The correlation between the number of errors and the nurses' age, years of work in the ward, and years of work in triage was indicated using Spearman or Pearson correlation, depending on the distribution of the variable.

Continuous variables were reported as mean and standard deviation (SD) or median and interquartile range (IQR), depending on the distribution. For univariate comparisons of continuous variables with a normal distribution, Student's t-test was applied. For continuous variables with a non-normal distribution, Mann-Whitney or Kruskal-Wallis tests were used, depending on the

number of groups being compared. Categorical variables were reported as absolute numbers and percentages and compared using Fisher's exact test or chi-square test, depending on the expected cell counts.

All analyses were considered statistically significant with a p-value <0.05. Analyses were conducted using the statistical software STATA 16.0 and R 4.3.3.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Ethical consideration

The study was conducted in accordance with the local ethics committee (Comitato etico per la sperimentazione clinica, Azienda Sanitaria dell'Alto Adige, Bolzano, Italia, protocol number Prot.0153572-BZ) and adhered to the Declaration of Helsinki regarding Ethical Principles for Medical Research Involving Human Subjects. In Italy, written consent is not required for studies involving professionals like nurses, provided that participation is voluntary and does not involve sensitive personal data. In this study, participation was entirely voluntary, and nurses implicitly consented by agreeing to take part after receiving an invitation. They were fully informed about the study's objectives, procedures, and their right to withdraw at any time.

Results

Among the four study centers, 77 triage nurses completed the assessment of all 30 vignettes. In Bolzano, 96.6% (29/30) of nurses participated, in Merano 83.3% (20/24), in Brunico 95.0% (19/20), and in Bressanone 90% (9/10). No participants were excluded from the study due to incomplete responses, as all recruited nurses successfully completed the assessment of all 30 vignettes.

Comparing the MTS code assigned by nurses to the code assigned by expert MTS group, the overall error rate was 28.6% (660/2310). (Table 1) Error rates varied from 23.5% (141/600) in Bolzano to 38.2% (218/570) at Brunico. (Table 2) The expected agreement between the code assigned by the expert MTS group and the triage nurses was 30.4%, while the actual agreement obtained was 71.5%, showing that the choice of triage code was not random. Cohen's kappa of 0.59 (0.57-0.60) and $p < 0.001$, demonstrates moderate agreement between the codes assigned by the expert MTS group and the triage nurses. Cohen's kappa by hospital ranged from 0.47 for Brunico to 0.66 for Merano.

Agreement per individual code is shown in Supplemental file 2. The greatest error rate was for the red code, where only 48% of nurses assigned this code when the expert group did. 77.4% of nurses assigned an orange code appropriately, and correct assignments were above 70% for the yellow and green codes (Table 1 and Supplementary File 2). Comparison by hospital is provided in Supplementary File 3.

Looking at the error rate by vignette (Figure 1), there are vignettes where the mean triage-assigned code always overlaps with the one of expert MTS group, while there are others where hospitals have, on mean, attributed a different priority code (Figure 1). For the red code, 2.6% (4/154) of the vignettes received a yellow code (under-triage) by the nurses. 18 vignettes that were classified as green by expert MTS group were classified as red by the MTS nurse code (over-triage).

Secondary outcomes

The discriminatory ability of the MTS codes was evaluated using the AUROC, and in all measured outcomes, the codes assigned by the expert group outperformed those assigned by nurses. (Table 3) Specifically, the expert-assigned codes achieved AUROC values ranging from 0.81 to 0.95, while the nurse-assigned codes had AUROC values ranging from 0.75 to 0.88.

Nurse characteristics and errors

The mean error rate per questionnaire was 30.0% (9/30). Nurses who had greater clinical experience in the ED and more experience in triage had more errors than the mean (Table 4). As the frequency of errors followed a normal distribution, we were able to use Pearson correlation analysis to investigate the relationship of errors and nurse characteristics. The analysis revealed significant positive correlations between frequency of errors rate and nurse age ($r = 0.430$, $p < 0.001$), years of experience in the ED ($r = 0.349$, $p = 0.002$), and years of triage experience ($r = 0.303$, $p = 0.007$). These moderate correlations indicate that older nurses and those with greater experience tended to make more errors when completing the triage questionnaires. When comparing nurses who made more errors to those who made fewer errors (relative to the mean), both groups demonstrated similar abilities in predicting secondary outcomes (Supplementary File 4).

Finally, regarding the nurses' perspectives on the MTS, nurses who made fewer errors were more likely to report that the system tends to overestimate patient severity (Table 4). Nurses with higher error rates were more likely to express satisfaction with the current system and indicate no preference for changes.

Discussion

This study found that, in our setting, the MTS is incorrectly implemented by nurses when applying it to vignettes from real cases. Agreement with the priority assigned by expert nurses using the MTS manual fell within the moderate and substantial range. Notably, there were specific nursing characteristics that were associated with poorer agreement, which included older age, greater nursing experience and greater triage experience. We also demonstrated that triage codes assigned by the expert group were more accurate at predicting clinical outcomes than those assigned by nurses the assignments differed. This underscores the importance of strict adherence to MTS

guidelines to minimize variability. However, further investigation is needed to determine whether deviations arise from forgetting the standards or are influenced by other factors.

Despite significant attention over the past two decades towards developing validated and standardized triage systems, there is limited literature on their accuracy of application. A previous study in two EDs assessed the reliability of MTS by presenting 50 vignettes to 45 nurses, yielding an unweighted Cohen's kappa of 0.48 (95% CI: 0.45-0.50) [7]. While our study had similar findings, our work went further to examine specific factors contributing to these errors, demonstrating the importance of understanding how these factors influence decision-making in triage.

The adoption of validated triage systems such as MTS, ESI, SATS, etc., is supported by evidence of their efficacy and objectivity, contrasting with earlier systems heavily reliant on clinical expertise alone [1, 16, 17]. In a systematic review by Zachariasse et al., which evaluated 33 triage systems, the top performers were those most rigorously validated and studied [16]. Nonetheless, no systematic review has assessed the operator's impact on triage code assignment and its broader implications for the triage process.

Our study reveals a paradoxical finding: triage nurses with greater clinical and triage experience are more likely to make errors when assigning priority codes to vignettes. These findings align with the results from a single-center study, which found an association between greater years of work in emergency department and errors in code assignment, with an odds ratio of 1.039 (95%CI: 1.000-1.080) [9].

This finding suggests that more experienced nurses may rely heavily on their subjective impressions developed over years of practice, potentially deviating from MTS textbook indications. To address this, implementing periodic refresher training and conducting structured re-evaluations of triage protocols may help ensure that even highly experienced nurses remain aligned with standardized practices. This approach would reduce variability and minimize potential errors in code assignment.

Considering that triage represents a pivotal moment in the ED and significantly influences the entire patient pathway, it is worth questioning the notion of treating the triage course as a one-time event without subsequent retraining or updates. Analogous to cardiopulmonary resuscitation (CPR), which adheres to standardized protocols even in high-stress scenarios, triage procedures should also be subject to regular review and reinforcement [18, 19]. Incorporating regular, structured retraining sessions for triage could help mitigate errors resulting from over-reliance on clinical intuition, particularly among experienced nurses. Recent evidence underscores the benefits of ongoing supervision and feedback in enhancing triage accuracy [9]. For instance, a study showed that daily auditing with fixed feedback significantly reduced errors in MTS application from 23.3% to 9.7% ($p < 0.001$) [9].

An alternative for improving triage performance could be the use of artificial intelligence (AI) systems to support nurses' clinical decision-making. A recent study by Levin et al. demonstrated that machine learning, a branch of artificial intelligence, improves predictive ability in triage code assignment, potentially serving as a valuable support tool in triage decision-making [20]. However, the effectiveness of AI tools in triage significantly depends on the type of AI employed. For instance, machine learning approaches have shown promise in improving predictive abilities in triage. In contrast, recent evaluations of Large Language Models (LLMs) have revealed limitations in clinical triage predictions [21, 22]. This discrepancy underscores the importance of selecting appropriate AI methodologies, with machine learning being better suited to address the complexities of triage scenarios compared to more general-purpose LLMs [21, 22]

Routinely used patient outcomes to evaluate triage systems have significant weaknesses, especially considering that patients can be severely ill without experiencing outcomes such as death or hospitalization [17, 23]. Despite these limitations, using multiple outcomes improves the understanding of triage system validity; moreover, subjective outcomes defined by physicians have been shown to assist in identifying acute and severely ill patients [17]. For this reason, our study incorporated both objective and subjective outcomes to provide a multidimensional assessment of

the MTS. This comprehensive approach allowed us to compare the performance of the expert MTS to the MTS acuity assigned by nurses, revealing that the MTS priority, assigned using textbook standards, was a better predictor of the clinical outcomes. This finding highlights the need for standardized approaches to ensure patient safety and equitable care.

Indeed, standardized triage tools are indispensable for the efficient operation of EDs and for guaranteeing equitable patient access to care [2, 12]. However, our findings also highlight a potential disparity: patients accessing different EDs using the same triage system may receive different triage codes. This further underscores the need for improved oversight and enhanced training programs for triage nurses to ensure greater standardization and reliability in code assignment.

This study has several limitations. First, although the vignettes used were derived from real clinical cases, they cannot fully replicate the complexity and nuances of real-life ED scenarios, and thus triage assignments may not represent what the nurses would do in practice with a patient in front of them. While vignettes are widely used to assess the standardization of triage systems, the lack of real patient interaction may affect the external validity of our results [7]. Second, restricting access to the MTS algorithms during vignette assessments may have affected the accuracy of the assigned triage codes. This approach was intentional to mimic real-world conditions, where triage nurses typically rely on their training and experience without accessing the MTS guidelines. However, if allowed to consult the algorithms the frequency of errors in this study might have been lower. Third, the secondary outcomes employed in this study, while providing valuable insights into system performance, are not perfect surrogates for real-world patient acuity and may not fully capture all dimensions of patient severity. However, we mitigated this limitation by combining multiple outcomes to provide a comprehensive evaluation of the MTS and included critical actions which may be the closest approximation of a measure of acuity. Finally, although the response rate among hospitals was high, variability in participation due to the voluntary involvement of triage nurses may have introduced selection bias. Future studies should aim to validate these findings in

real clinical settings and examine the impact of variability in triage codes on actual patient outcomes.

Conclusions

This study has shown that the MTS is applied inconsistently across EDs, and between nurses, with more experienced nurses having higher error rates. Ensuring that standardized systems like the MTS are applied consistently is essential for promoting equitable and safe patient care regardless of the nurse triaging or the ED they attend.

References

- 1) Hinson JS, Martinez DA, Cabral S, et al. Triage Performance in Emergency Medicine: A Systematic Review. *Ann Emerg Med.* 2019;74(1):140-152.
doi:10.1016/j.annemergmed.2018.09.022
- 2) Battisti D, Camporesi S. A proposal for formal fairness requirements in triage emergency departments: publicity, accessibility, relevance, standardisability and accountability. *J Med Ethics.* Published online August 24, 2023. doi:10.1136/jme-2023-109188
- 3) Tam HL, Chung SF, Lou CK. A review of triage accuracy and future direction. *BMC Emerg Med.* 2018;18(1):58. Published 2018 Dec 20. doi:10.1186/s12873-018-0215-0
- 4) Zaboli A, Brigo F, Sibilio S, et al. The impact of COVID-19 pandemic on the urgency of patients admitted to the emergency department. *Int Emerg Nurs.* 2022;65:101229.
doi:10.1016/j.ienj.2022.101229
- 5) Zachariasse JM, Seiger N, Rood PP, et al. Validity of the Manchester Triage System in emergency care: A prospective observational study. *PLoS One.* 2017;12(2):e0170811. Published 2017 Feb 2. doi:10.1371/journal.pone.0170811
- 6) Zaboli A, Sibilio S, Cipriano A, et al. Italian validation of the Manchester Triage System towards short-term mortality: a prospective observational study. *Emergency Care Journal*, 19(3) 2023.
<https://doi.org/10.4081/ecj.2023.11443>

- 7) van der Wulp I, van Baar ME, Schrijvers AJ. 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*. 2008;25(7):431-434. doi:10.1136/emj.2007.055228
- 8) Ausserhofer D, Zaboli A, Pfeifer N, et al. Errors in nurse-led triage: An observational study. *Int J Nurs Stud*. 2021;113:103788. doi:10.1016/j.ijnurstu.2020.103788
- 9) Zaboli A, Sibilio S, Magnarelli G, et al. Daily triage audit can improve nurses' triage stratification: A pre-post study. *J Adv Nurs*. 2023;79(2):605-615. doi:10.1111/jan.15521
- 10) Nishi FA, Polak C, Cruz DALMD. Sensitivity and specificity of the Manchester Triage System in risk prioritization of patients with acute myocardial infarction who present with chest pain. *Eur J Cardiovasc Nurs*. 2018;17(7):660-666. doi:10.1177/1474515118777402
- 11) Sax DR, Warton EM, Mark DG, et al. Evaluation of the Emergency Severity Index in US Emergency Departments for the Rate of Mistriage [published correction appears in *JAMA Netw Open*. 2024 Jun 3;7(6):e2423536. doi: 10.1001/jamanetworkopen.2024.23536]. *JAMA Netw Open*. 2023;6(3):e233404. Published 2023 Mar 1. doi:10.1001/jamanetworkopen.2023.3404
- 12) Mackaway-Jones K, Marsden J, Windle J. *Emergency Triage: Manchester triage group, Third edition*. John Wiley & Sons, Ltd, West Sussex. 2014.
- 13) Platts-Mills TF, Travers D, Biese K, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. *Acad Emerg Med*. 2010;17(3):238-243. doi:10.1111/j.1553-2712.2010.00670.x

- 14) Brouns SHA, Mignot-Evers L, Derkx F, Lambooij SL, Dieleman JP, Haak HR. Performance of the Manchester triage system in older emergency department patients: a retrospective cohort study. BMC Emerg Med. 2019;19(1):3. Published 2019 Jan 7. doi:10.1186/s12873-018-0217-y
- 15) Brutschin V, Kogej M, Schacher S, Berger M, Gräff I. The presentational flow chart "unwell adult" of the Manchester Triage System-Curse or blessing?. PLoS One. 2021;16(6):e0252730. Published 2021 Jun 3. doi:10.1371/journal.pone.0252730
- 16) Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. Performance of triage systems in emergency care: a systematic review and meta-analysis. BMJ Open. 2019;9(5):e026471. Published 2019 May 28. doi:10.1136/bmjopen-2018-026471
- 17) Kuriyama A, Urushidani S, Nakayama T. Five-level emergency triage systems: variation in assessment of validity. Emerg Med J. 2017;34(11):703-710. doi:10.1136/emered-2016-206295
- 18) Anderson R, Sebaldt A, Lin Y, Cheng A. Optimal training frequency for acquisition and retention of high-quality CPR skills: A randomized trial. Resuscitation. 2019;135:153-161. doi:10.1016/j.resuscitation.2018.10.033
- 19) Chien CY, Fang SY, Tsai LH, et al. Traditional versus blended CPR training program: A randomized controlled non-inferiority study. Sci Rep. 2020;10(1):10032. Published 2020 Jun 22. doi:10.1038/s41598-020-67193-1
- 20) Levin S, Toerper M, Hamrock E, et al. 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(5):565-574.e2.

doi:10.1016/j.annemergmed.2017.08.005

21) Masanneck L, Schmidt L, Seifert A, et al. Triage Performance Across Large Language Models, ChatGPT, and Untrained Doctors in Emergency Medicine: Comparative Study. *J Med Internet Res*. 2024;26:e53297. Published 2024 Jun 14. doi:10.2196/53297

22) Zaboli A, Brigo F, Sibilio S, Mian M, Turcato G. Human intelligence versus Chat-GPT: who performs better in correctly classifying patients in triage?. *Am J Emerg Med*. 2024;79:44-47.

doi:10.1016/j.ajem.2024.02.008

23) Challen K. How good is triage, and what is it good for?. *Emerg Med J*. 2017;34(11):702.

doi:10.1136/emered-2017-206973

Overall						
	MTS code according to the triage nurse					
MTS code according to expert MTS group	Red	Orange	Yellow	Green	Blue	Total
Red	74 (48.0)	76 (49.4)	4 (2.6)	0 (0.0)	0 (0.0)	154 (100)
Orange	<u>17 (2.8)</u>	477 (77.4)	98 (15.9)	24 (3.9)	0 (0.0)	616 (100)
Yellow	<u>1 (0.2)</u>	<u>99 (18.4)</u>	381 (70.7)	57 (10.6)	1 (0.2)	539 (100)
Green	<u>18 (1.8)</u>	<u>78 (7.8)</u>	<u>176 (17.6)</u>	719 (71.8)	10 (1.0)	1001 (100)
Blue	<u>0 (0.0)</u>	<u>0 (0.0)</u>	<u>0 (0.0)</u>	<u>0 (0.0)</u>	0 (0.0)	0 (100)

Table 1: Contingency table comparing the code that should have been assigned by the MTS

(vertical priority code) to that subjectively assigned by the nurses (horizontal priority code).

Vignettes with complete agreement between nurses and MTS are reported in gray, under-triage by nurses are reported in bold, and over-triage by nurses is reported with underlining. Numbers in each square are a count of subjects and those in brackets are percentages.

Hospital	Agreement	Expected agreement	Cohen's kappa	95% CI
MTS Bolzano	74.3%	31.0%	0.63	0.57-0.65
MTS Merano	76.6%	31.4%	0.66	0.63-0.68
MTS Brunico	61.7%	28.3%	0.47	0.42-0.50
MTS Bressanone	71.5%	30.5%	0.59	0.56-0.66

Table 2: Agreement between the MTS code assigned by the triage nurse and that expected by MTS.

The expected agreement was reported to assess the non-randomness of the collected data.

Subsequently, Cohen's kappa and its respective 95% confidence interval were reported.

	Expert MTS group	Overall	MTS Bolzano	MTS Merano	MTS Brunico	MTS Bressanone
72 hours mortality	0.91 (0.89-0.92)	0.80 (0.78-0.82)	0.79 (0.77 - 0.81)	0.83 (0.80 – 0.87)	0.75 (.69-0.81)	0.80 (0.74-0.85)
Hospitalization	0.83 (0.80-0.84)	0.81 (0.79-0.83)	0.78 (0.75-0.81)	0.79 (0.76-0.83)	0.84 (0.81-0.87)	0.79 (0.74-0.85)
LSI	0.94 (0.93-0.95)	0.86 (0.84-0.87)	0.84 (0.82-0.86)	0.88 (0.86-0.91)	0.86 (0.82-0.89)	0.88 (0.84-0.92)
Severe condition in the ED	0.81 (0.78-0.82)	0.80 (0.78-0.82)	0.77 (0.74-0.80)	0.81 (0.78-0.84)	0.85 (0.83-0.88)	0.81 (0.76-0.86)
Time-dependent pathology	0.95 (0.94-0.95)	0.86 (0.85-0.88)	0.85 (0.82-0.87)	0.88 (0.86-0.91)	0.85 (0.81-0.88)	0.86 (0.82-0.91)

Table 3: Area Under the Receiver Operating Characteristic for all secondary study outcomes,

where the predictive ability of MTS is reported to reflect the predictive ability of the code assigned by nurses, both overall and for each hospital.

Variables	Less than 9 errors per nurse	Equal or more than 9 errors per nurse	p-value
Number of nurses	47	30	
Sex, n (%)			0.452
Female	34 (72.3)	21 (70.0)	
Male	13 (27.7)	9 (30.0)	
Nurse's age, years, mean (SD)	35.4 (9.4)	41.5 (11.9)	0.013
Years of nursing experience, mean (SD)	10.7 (8.7)	20.0 (14.0)	0.006
Years of experience in the ED, mean (SD)	7.5 (6.2)	12.6 (11.9)	0.011
Years of experience in triage, mean (SD)	5.8 (5.2)	9.4 (7.2)	0.013
Years since the last triage training, mean (SD)	6.2 (2.3)	6.5 (2.2)	0.611
Degree of education, n (%)			0.260
BSC	43 (91.5)	30 (100.0)	
Master's degree or advanced training	1 (2.1)	0 (0.0)	
MSC	3 (6.4)	0 (0.0)	
Personal interest in triage, median (IQR)	9 (8-10)	9 (8-10)	0.708
Confidence in assigning triage codes, median (IQR)	8 (7-9)	8 (7-8)	0.721
How well does the current triage system work, median (IQR)	7 (3-8)	8 (7-9)	0.095
Accuracy of the triage system, n (%)			0.275
Underestimates the risk (under triage)	2 (4.4)	1 (3.3)	
Accurate triage classification	20 (44.4)	19 (63.3)	
Overestimates the risk (over triage)	23 (51.1)	10 (33.3)	
Would you change your current triage system, n (%)			0.038
No	28 (59.6)	27 (90.0)	

CTAS	1 (2.1)	0 (0.0)	
ESI	5 (10.6)	1 (3.3)	
GFT	13 (27.7)	2 (6.7)	
SATS	0 (0.0)	0 (0.0)	
ATS	0 (0.0)	0 (0.0)	

Table 4: Characteristics of the nurses who participated in the study, divided according to those with an error rate above and below the mean.

Figure 1: The chart represents the 30 vignettes in a circular manner. The outer circle, numbered from 1 to 30, represents the vignette analyzed on that axis of the chart, while the concentric circles representing the severity code according to MTS range from 1 to 5, where 1 represents the red code and 5 the blue code. The expected code by MTS is shown in dashed black, while each colored line represents the mean code assigned by the nurses from individual hospitals. A vignette where the assigned code closely matches the expected MTS code will have the apex of each colored line at the same point (e.g., vignette 28), whereas a vignette with a greater distance between the apex predicted by expected MTS and the colored line corresponds to greater mean variability in code assignment by the nurses at the center (e.g., vignette 15).