The accuracy of the Manchester Triage System in an emergency service

Acurácia do Sistema de Triagem de Manchester em um serviço de emergência

Precisión del Sistema de Triage Manchester en un servicio de emergencia

ABSTRACT

Objective: To verify the accuracy of the Manchester Triage System (MTS) and the outcomes of adult patients in an emergency hospital service.

Method: Cross-sectional study, conducted through an inspection of records of risk classification of adult patients treated in the emergency service of a hospital.

Results: The patients (n = 400) were classified according to priority levels, in red (0.8%), orange (58.2%), and yellow (41.0%). The accuracy levels between auditors and nurses were substantial for the flowchart (K = 0.75), and moderate for discriminating factors (k = 0.46) and priority levels (k = 0.42). The accuracy of the MTS was 68.8% with regard to priority levels. Regarding outcomes, 60% of patients were discharged, 37% were transferred to other units, and 3% died.

Conclusion: The MTS proved to be a good predictor of the assessed outcomes, showing that 65.9% of Low Urgency patients progress to discharges, and 3.8% of High Urgency patients progress to death. The accuracy of the MTS was moderate, which suggests the need to implement inspections in emergency services.

Keywords: Triage. Emergency service, hospital. Data accuracy. Sensitivity and specificity. Clinical audit.

RESUMO

Objetivo: Verificar a acurácia do Sistema de Triagem de Manchester (STM) e os desfechos dos pacientes adultos em um serviço de emergência hospitalar.

Método: Estudo transversal, realizado por meio de um processo de auditoria dos registros de triagem com classificação de risco de pacientes adultos atendidos em um serviço de emergência hospitalar.

Resultados: Os pacientes (n = 400) foram classificados de acordo com o nível de prioridade em Vermelho (0,8%), Laranja (58,2%) e Amarelo (41,0%). A confiabilidade entre os auditores e enfermeiros foi substancial para fluxograma (K = 0,75), moderada para discriminadores (k = 0,46) e nível de prioridade (k = 0,42). A acurácia do STM foi de 68,8% no nível de prioridade. Em relação aos desfechos avaliados, 65,9% dos pacientes de Baixa Urgência evoluíram para alta, e 3,8% dos pacientes de Alta Urgência evoluíram para óbito.

Conclusão: A acurácia do STM foi moderada. O STM se mostrou um bom preditor dos desfechos avaliados, evidenciando que a maioria dos pacientes de Baixa Urgência evoluem para alta, e 3,8% dos pacientes de Alta Urgência evoluem para óbito.


RESUMEN

Objetivo: Verificar la precisión del Sistema de Triage Manchester (STM) y los resultados de pacientes adultos en un servicio de emergencia en un hospital.

Método: estudio transversal, realizado a través de un proceso de auditoría de registros de detección con clasificación de riesgo de pacientes adultos atendidos en un servicio hospitalario de emergencia.

Resultados: Los pacientes (n = 400) se clasificaron según el nivel de prioridad en rojo (0,8%), naranja (58,2%) y amarillo (41,0%). Niveles de flujo de trabajo entre auditores y enfermeras (K = 0,75), moderados para discriminadores (k = 0,46) y nivel de prioridad (k = 0,42). La precisión del STM fue del 68,8% en el nivel de prioridad. En cuanto a los resultados, el 60% de los pacientes fueron dados de alta, el 37% fueron transferidos a otras unidades, el 3% falleció.

Conclusión: El STM demostró ser un buen predictor de los resultados evaluados, mostrando que el 65,9% de los pacientes de baja urgencia progresan al alta y el 3,8% de los pacientes de alta urgencia progresan a la muerte. La precisión del STM fue moderada, lo que sugiere la necesidad de implementar auditorías en los servicios de emergencia.

INTRODUCTION

The triage in the emergency system seeks to provide a dynamic evaluation of the users. It contributes for an organized and structured attention using a systematized protocol, promotes the satisfaction of the user and of the health team, and optimizes the waiting time and the resources used[1–3].

The Manchester Triage System (MTS) is a protocol based on identifying the complaint of the patient, based on which the nurse chooses the most appropriate flowchart to conduct the reasoning during the process. These are guided by discriminating factors that are distributed in priority levels[4–6]. The discriminating factors present themselves in the form of structured questions (signs and symptoms) which start with the highest priority cases and go down towards the lowest ones[4–6].

According to the MTS, the patient may be classified in five different levels of priority, which are identified by number, name, color, and target-time until the medical evaluation starts. These are: priority 1, Emergency (red - immediate care); priority 2 very urgent (orange - care should start in at most 10 minutes); priority 3, urgent (yellow - care should start in at most 1 hour); priority 4, little urgency (green - care should start in up to 2 hours); priority 5, not urgent (blue - care should start in up to 4 hours)[6].

The Portuguese Triage Group also implemented the color White in the Manchester Protocol, to identify and monitor users who are there due to reasons related to the administration, such as coming back for reevaluations after discharge, exams, elective procedures, unscheduled therapies, among others[6].

The adoption of a triage system with risk classification, such as the MTS, requires monitoring through a systematic and continuous analysis, to determine whether the methodology being used and its results are satisfactory with regards to the previously established criteria, so that the expected performance can be guaranteed, promoting the quality of the actions carried out in the health services[3,6].

The quality of a triage system can be assured through tests of the reliability and validity of the process. Reliability assessments show the consistency of the method and can be done by verifying the replicability of the results. The validity or accuracy of the system can be measured by assessing its sensitivity and specificity. Evaluating the sensitivity means identifying its correctness for patients who actually belong to the High Urgency categories, while the specificity represents the correctness of the system in identifying patients who do not belong in the Higher Urgency categories[7–8].

A triage process with an inadequate risk classification can lead to the over or undertriage of the patient. The overtriage is when the patient is attributed a category that indicates higher urgency than their real needs. On the other hand, the undertriage is the attribution of a category that is inferior to the actual urgency. Undertriage can lead to serious harm, due to delays on the treatment. Similarly, overtriage leads the institution to waste resources[9].

A correct risk classification is paramount to guarantee the quality of assistance and the management of resources. However, it depends on the continuous training and on the professional experience of the professional who carries it out. In this setting, inspections to monitor the consistency and performance of working processes and whether they need to be improved, as well as to elaborate institutional indexes, is essential[5,6].

The guiding question of this research is: how accurate are the risk classifications carried out in a hospital emergency service? Is there any association between the priority level and the outcomes of the care offered? Therefore, the objective of this research was assessing the accuracy of the MTS and the outcomes of adult patients in an emergency hospital service.

METHOD

Cross-sectional study with a quantitative analysis of the triage records with risk classification. The electronic records used were from patients cared for in the emergency service of a public hospital located in the South of Brazil, which offers attention under the Brazilian Single Health System (SUS). The triage of patients is carried out by nurses, through the MTS. This study was approved by the Research Ethics Committee of the institution, under legal opinion 1.554.135.

The population of this study included 65,160 records of clinical adult patients who underwent a triage with risk classification in 2015. The sample calculation was carried out through the statistical program WINPEP, version 11.43. The expected accuracy was that of 80%, with an accepted margin of error of 4% and a confidence level of 95%, which led to a minimum sample of 385 records. The sample was increased to 420 records, so the same number of records could be analyzed from each month of the year.

The sample selection was randomized, using secondary data from queries. Data was provided from the information technology sector in the form of Excel® spreadsheets originating from the information database of the hospital.

Data collection

All electronic records that contained the risk classification of adult patients who were referred to a clinical medicine...
specialty were included, except those marked with the color white. This includes those who, after the triage, were attributed the colors Green or Blue (and were thus referred to primary or secondary health care services). The risk classifications carried out by the supervisory nurse, who works in the emergency service of the institution.

The variables collected were the number of the record, the date of attention, age, sex, complaint, vital signs, flowchart, discriminating factor, level of priority.

To evaluate the outcomes of the attention considering the emergencies, death during the emergency and internal transfers were considered to be the primary outcomes of serious emergencies. The variables considered as secondary outcomes of the emergencies were the time of permanence of the patient in the emergency sector, and the total hospitalization time. Data collection took place from July to August 2016.

Inspection

The inspection of the process was carried out through the electronic records, in which the nurse and the physician responsible for the inspection, both certified by the Grupo Brasileiro de Classificação de Risco (GBCR - the Brazilian Group of Risk Classification), separately evaluated each classification carried out by the nurses, by carefully applying the MTS to the records.

The inspection, in this study, evaluated the selection of the Flowchart, of the Discriminating Factor, and of the Priority Level by the nurse. It also evaluated the records of vital signs, capillary glucose, and the Glasgow Coma Scale (GCS).

Each inspector, individually, evaluated and expressed agreement or disagreement with the risk classification carried out by the nurse. When they disagreed, they indicated which Flowchart, Discriminating Factor, or Priority Level would be the most appropriate for the complaint of the patient. The records of vital signs and other parameters were also considered, to detect undertriage or overtriage situations.

After each inspector evaluated the entire sample, they reached a consensus and a degree of accordance was analyzed with regards to the variables: Flowchart, Discriminating Factor, and Priority Level. Later, the degree of accordance between the answers of the inspectors and the nurses responsible for classification were also verified.

The agreement between the inspectors was considered to be the golden standard and was compared to the classification of the nurses responsible for the classification; the disagreements were considered to be undertriages or overtriages.

To measure the predictive value of the MTS, an analysis was carried out of the primary and secondary outcomes, as well as of the exposure of the patients to the categories dichotomized in “High Urgency” or “Low Urgency”. The primary outcomes were: discharge from the emergency, death in emergency, and internal transfers. Secondary outcomes were: time of permanence of the patient in the emergency, and time total hospitalization time.

Data analysis

Data collected were organized in Excel® spreadsheets and transferred to the SPSS® software, version 18.0, where they were submitted to descriptive and analytical statistics. To analyze the normality of the sample, the Shapiro-Wilk test was used. Continuous variables were presented with a median and an interquartile range (25 and 75 percentiles) or with means and standard deviation, depending on the normality of the data. The association between qualitative variables was verified by Fisher’s exact test, with a significance of α = 0.05.

The analysis of the variance was used to compare the age between the priority categories.

The Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.

To analyze the accuracy, that is, the validity, a descriptive analysis was carried out between the categories. Relative and absolute frequencies were established for the proportion of patients whose triage was carried out correctly, or in which cases there was undertriage or overtriage. These measures were carried out using different measures of diagnostic performance: sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), and odds ratio (OR).

To calculate the measures of diagnostic performance, the MTS was dichotomized, meaning that the patients classified in the categories Emergency/Red and Very Urgent/Orange were grouped under a single category, named “High Urgency,” while those classified in the category “Urgent/Yellow” were classified as “Low Urgency”. The diagnostic performance measures were calculated through the website VassarStates, considering a 95% confidence interval.

For the initial analysis, the median and the interquartile range were calculated, as well as the p-value (found through the Mann-Whitney test). To measure the strength of the agreement, the Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.

To analyze the accuracy, that is, the validity, a descriptive analysis was carried out between the categories. Relative and absolute frequencies were established for the proportion of patients whose triage was carried out correctly, or in which cases there was undertriage or overtriage. These measures were carried out using different measures of diagnostic performance: sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), and odds ratio (OR).

To calculate the measures of diagnostic performance, the MTS was dichotomized, meaning that the patients classified in the categories Emergency/Red and Very Urgent/Orange were grouped under a single category, named “High Urgency,” while those classified in the category “Urgent/Yellow” were classified as “Low Urgency”. The diagnostic performance measures were calculated through the website VassarStates, considering a 95% confidence interval.

For the initial analysis, the median and the interquartile range were calculated, as well as the p-value (found through the Mann-Whitney test). To measure the strength of the agreement, the Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.

To analyze the accuracy, that is, the validity, a descriptive analysis was carried out between the categories. Relative and absolute frequencies were established for the proportion of patients whose triage was carried out correctly, or in which cases there was undertriage or overtriage. These measures were carried out using different measures of diagnostic performance: sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), and odds ratio (OR).

To calculate the measures of diagnostic performance, the MTS was dichotomized, meaning that the patients classified in the categories Emergency/Red and Very Urgent/Orange were grouped under a single category, named “High Urgency,” while those classified in the category “Urgent/Yellow” were classified as “Low Urgency”. The diagnostic performance measures were calculated through the website VassarStates, considering a 95% confidence interval.

For the initial analysis, the median and the interquartile range were calculated, as well as the p-value (found through the Mann-Whitney test). To measure the strength of the agreement, the Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.

To analyze the accuracy, that is, the validity, a descriptive analysis was carried out between the categories. Relative and absolute frequencies were established for the proportion of patients whose triage was carried out correctly, or in which cases there was undertriage or overtriage. These measures were carried out using different measures of diagnostic performance: sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), and odds ratio (OR).

To calculate the measures of diagnostic performance, the MTS was dichotomized, meaning that the patients classified in the categories Emergency/Red and Very Urgent/Orange were grouped under a single category, named “High Urgency,” while those classified in the category “Urgent/Yellow” were classified as “Low Urgency”. The diagnostic performance measures were calculated through the website VassarStates, considering a 95% confidence interval.

For the initial analysis, the median and the interquartile range were calculated, as well as the p-value (found through the Mann-Whitney test). To measure the strength of the agreement, the Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.

To analyze the accuracy, that is, the validity, a descriptive analysis was carried out between the categories. Relative and absolute frequencies were established for the proportion of patients whose triage was carried out correctly, or in which cases there was undertriage or overtriage. These measures were carried out using different measures of diagnostic performance: sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), and odds ratio (OR).

To calculate the measures of diagnostic performance, the MTS was dichotomized, meaning that the patients classified in the categories Emergency/Red and Very Urgent/Orange were grouped under a single category, named “High Urgency,” while those classified in the category “Urgent/Yellow” were classified as “Low Urgency”. The diagnostic performance measures were calculated through the website VassarStates, considering a 95% confidence interval.

For the initial analysis, the median and the interquartile range were calculated, as well as the p-value (found through the Mann-Whitney test). To measure the strength of the agreement, the Kappa coefficient test was used to measure the degree of agreement. The results were considered to be in agreement when the Kappa was lower than zero. The agreement was found to be minimal when results were between 0 and 0.20; reasonable they were between 0.21 and 0.40; moderate when they were from 0.41 to 0.50; substantial when varying from 0.61 to 0.80; and almost perfect when between 0.81 and 1.00.
association between exposure and outcome, the categories “High Urgency” and “Low Urgency” were dichotomized. Mortality was considered as the primary outcome to analyze the severity of each case. Regarding the secondary outcomes, a cutoff point was established for the length of stay in the emergency (24 hours or more) and total length of hospitalization (3 days or more). The analysis was carried out using a univariate logistic regression, which made it possible to calculate the OR, the 95% confidence interval, and Wald’s Chi-squared P-values < 0.05 were found to be significant.

**RESULTS**

The initial sample included 420 records, from which 20 were excluded: in 18 cases there was evasion from the patient during the attention, and 2 of them referred to workers who suffered workplace accidents out of the working shift, in the Occupation Medicine Services.

Among 400 records of adult patients classified and cared for in the Emergency Service, 52.2% of users were female, and their mean age was 55.4±18.7 years. Patients were classified according to the MTS as Emergency/Red (0.8%), Very Urgent/Orange (58.2%), and Urgent/Yellow (41.0%).

Table 1 shows the characteristics of the sample, according to sex and age, distributed according to priority levels. It should be noted that there were no significant differences between the categories in these variables.

The comparison between the consensus of the inspectors (golden standard) showed that nurses had a 73.5% level of correctness with regards to the Flowchart, of 47.8% with regards to the Discriminating Factor, and of 68.8% with regards to the Priority Level. The combined level of correctness between these three variables was 39.5%.

The degree of agreement of the auditors (golden standard) with the nurses responsible for the classification was substantial for the Flowchart (κ=0.75), moderate for the Discriminating Factor (K=0.46) and for the Priority Level (68.8%).

The correctness between the consensus of the inspectors (golden standard) and the nurses responsible for classifying the level of priority showed an accuracy of 68.8% of the risk classifications carried out by the nurses in the service; 25.5% of the risk classifications were considered to be undertriages, while 5.7% were seen as overtriages.

102 of the undertriages were found to be the result of the absence of records indicating vital signs or other parameters (n=85) which were required by the Flowcharts and Discriminating Factors. Among them, capillary glucose stands out, which was not verified or registered in 50 cases, despite being required; the same can be said for 33 cases in which the Glasgow Coma Scale was not registered, meaning that the Discriminating Factor “Sudden Change in Consciousness” was not included in the Flowcharts that required this evaluation. In other two cases related to the pain, there were also no records to rule out the discriminating factor “Intense pain”.

Table 1 – Characterization of the sample of adult patients cared for in a hospital emergency service and classified through the MTS. Porto Alegre, RS, Brazil, 2016

<table>
<thead>
<tr>
<th>Variable</th>
<th>Red (n=3 (0.8%))</th>
<th>Orange (n=233 (58.2%))</th>
<th>Yellow (n=164 (41.0%))</th>
<th>Total (n=400 (100%))</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1 (33.3)</td>
<td>117 (50.2)</td>
<td>91 (55.5)</td>
<td>209 (52.2)</td>
<td>0.480*</td>
</tr>
<tr>
<td>Male</td>
<td>2 (66.7)</td>
<td>116 (49.8)</td>
<td>73 (44.5)</td>
<td>191 (47.8)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(years)†</td>
<td>59.3±21.2</td>
<td>56.1±18.0</td>
<td>54.3±19.6</td>
<td>55.4±18.7</td>
<td>0.609**</td>
</tr>
</tbody>
</table>

Source: Research data, 2016.

*Fisher’s Exact test; **Anova; †Mean±standard deviation.
Table 2 presents the accuracy of the MTS according to the dichotomization of the categories in High Urgency and Low Urgency.

Regarding the primary outcomes, discharges (60%) were the most frequent, followed by transfers (37%) and death (3.0%). Among secondary outcomes, the median of the length of time in the emergency service was 30.5 hours, and the median of hospitalization length was 3 days. There was a significant difference for these outcomes in the categories (Table 3).

High Urgency patients are 1.69 times more likely to stay more than 24 hours in the emergency services and 2.1 times more likely to remain hospitalized for more than 3 days, when compared to Low Urgency patients (Table 4).

Table 2 - Accuracy of the Manchester Triage System in adult patients cared for in a hospital emergency service. Porto Alegre, RS, Brazil, 2016

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>High Urgency x Low Urgency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated value (IC 95%)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>93% (89-96)</td>
</tr>
<tr>
<td>Specificity</td>
<td>57% (49-64)</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>76% (70-80)</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>84% (76-90)</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>16.87 (9.43-30.20)</td>
</tr>
</tbody>
</table>

Source: Research data, 2016.
CI 95%: 95% confidence interval.

Table 3 - Outcomes of adult patients in a hospital emergency service. Porto Alegre, RS, Brazil, 2016

<table>
<thead>
<tr>
<th>Variable</th>
<th>High Urgency</th>
<th>Low Urgency</th>
<th>Total</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Outcomes n(%)</strong></td>
<td></td>
<td></td>
<td>400</td>
<td>0.006*</td>
</tr>
<tr>
<td>High</td>
<td>132 (55.9)</td>
<td>108 (65.9)</td>
<td>240 (60.0)</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>9 (3.8)</td>
<td>3 (1.8)</td>
<td>12 (3.0)</td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td>95 (40.3)</td>
<td>53 (32.3)</td>
<td>148 (37.0)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>236 (100%)</td>
<td>164 (100%)</td>
<td>400 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

**Secondary Outcomes†**

<table>
<thead>
<tr>
<th>Variable</th>
<th>High Urgency</th>
<th>Low Urgency</th>
<th>Total</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of stay in the Emergency Service (hours)</td>
<td>39(14-101)</td>
<td>23(8.3-70)</td>
<td>30.5(11-87.3)</td>
<td>0.005**</td>
</tr>
<tr>
<td>Total hospitalization time (days)</td>
<td>4(1-12)</td>
<td>1(1-8)</td>
<td>3(1-11)</td>
<td>&lt;0.001**</td>
</tr>
</tbody>
</table>

Source: Research data, 2016.
*Fisher’s Exact test; **Mann-Whitney U test; †Median (interquartile range 25-75).
This study evaluated the accuracy of the MTS in the risk classification of an emergency hospital service whose profile is predominantly targeted at attending SUS clinical emergencies.

Users from all ages (in this study, from 17 to 102 years of age) seek attention in hospital emergency services, which could be related to the highest capacity this service offers in reaching positive results in this service, since they offer attention at any time, with no need for scheduling consultations or exams, which is not true for other levels of health care. In this study, the mean age of users was 55.4 years of age, with a discrete predominance of women (52.2%). This is the opposite of the findings of other national studies, which also evaluated the risk classification of emergencies, finding a predominance of men.

Regarding the priority of care, most cases were classified as Very Urgent/Orange (58.2%), a result that is not in agreement to other studies, in which the categories Little Urgency/Green and Urgent/Yellow predominated.

The inspection showed that the nurses responsible for the classification had a correctness index of 73.5% regarding the Flowchart, of 47.8% regarding Discriminating Factors, and of 68.8% with regard to the Priority Levels. The combined correctness was low (39.5%). A study carried out involving 361 nurses from 21 Brazilian states, evaluated 28 clinical cases, indicating a mean of 20.6 correct answers in the Flowchart, 16.4 in Discriminating Factors, and 21.7 in the Priority Levels. These findings show that different discriminating factors lead to the same level of priority, regardless of the flowchart being used. As a result, they increase the reliability of this system of triage, guaranteeing that the protocol is a safe measure to determine the priority level considering different flowcharts and discriminating factors.

In this study, the reliability level between inspectors was nearly perfect (K=0.92) regarding the Priority Level. On the other hand, the reliability between the consensus of the auditors (golden standard) and the nurses responsible for the classification was moderate (K=0.42) for the Priority Level. The same study mentioned above, which sought to analyze the reliability of the MTS in determining the priority level of the patients, presented a substantial agreement (K=0.69) in the Priority Level variable, regarding external reliability.

In Germany, on the other hand, a country where the MTS version is different from the English one in terms of Flowcharts and discriminating factors, although the five levels of stratification are the same, a study was carried out aiming to investigate the validity and reliability of the MTS. The results showed an almost perfect (k=0.954) reliability between the inspectors (triage nurses and specialists) for the priority levels of the MTS. A systematic and meta-analytical revision that sought to determine the extension of the reliability of the MTS, analyzing inter- and intra-evaluator reliability in seven studies, found a substantial global agreement reliability (K=0.751; CI 95%: 0.677-0.810). In turn, in the analysis of the sub-groups, the agreement between the triage nurses was substantial (K=0.768; CI 95%: 0.694-0.826), while the agreement between specialist nurses was almost perfect (K=0.863; CI 95%: 0.762-0.982).

Considering this setting, it can be stated that the results found in this study reinforce the need for regular and systematic inspections of the risk classification processes through internal inspections. It also reiterates the need for the permanent education of nurses, so that the decision-making process can be standardized, and inconsistencies can be detected and corrected early.

The 68.8% accuracy in this study represented correctness in the priority level of the risk classifications carried out by the nurses in the service. However, there were cases of

Table 4 - Odds ratio for the outcomes length of stay in the emergency service, hospitalization, and mortality. Porto Alegre, RS, Brazil, 2016

<table>
<thead>
<tr>
<th>Outcome</th>
<th>High Urgency x Low Urgency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RO</td>
</tr>
<tr>
<td>Length of stay in the emergency service &gt; 24 h</td>
<td>1.69</td>
</tr>
<tr>
<td>Total hospitalization time &gt; 3 days</td>
<td>2.15</td>
</tr>
<tr>
<td>Mortality</td>
<td>2.13</td>
</tr>
</tbody>
</table>

Source: Research Data, 2016.
OR: Odds Ratio; CI 95%: 95% confidence interval; *Wald Chi-squared.
undertriage (25.5%) and overtriage (5.7%). The main reason for the high levels of undertriage was the absence of the record of certain vital signs or required parameters, especially the ECG and the capillary glucose test. An European multi-center test\(^{(17)}\), whose objective was determining the validity of the MTS in three emergency departments, found overtriage levels in adult patients (from 26.9% to 44.0%) that were lower than the undertriage levels (from 6.2% to 14.1%). The correctness found for adult patients varied from 49.7% to 61.6%. The variability found can be associated with the different settings in which the MTS is being used, as well as with the experience of the nurse who is carrying out the triages. Both situations generate preoccupation, since they interfere in the flow of attention of the patient.

From the point of view of patient safety, the situations of undertriage can lead to delays in the attention and treatment, which favor the clinical deterioration of the patient, since they increase morbidity and mortality and increase the time of permanence of the patient in the health service. On the other hand, from the point of view of management, overtriage situations lead to the waste of human and material resources that could be invested in other patients who really need them.

The level of accuracy found in this study for the correct diagnosis of the MTS level showed a high sensitivity for the detection of High Urgency patients, while the specificity for the detection of Low Urgency patients was moderate. The odds ratio for High Urgency patients to be recognized is of 16.87 times higher than that of Low Urgency ones. As a result, this result of moderate specificity directly impact in the assistance, and can contribute for the overcrowding of the service, since it shows that Low Urgency patients are less likely to be recognized and classified. A prospective cohort study, carried out in Switzerland\(^{(18)}\), evaluated the performance of the MTS in adult patients, finding a sensitivity that was inversely proportional to the priority levels. The specificity, on the other hand, was directly proportional to them. It is believed that the profile of attention and the health network available can influence this data.

Regarding the primary outcomes, the discharge was the most common, mostly for Low Urgency patients, followed by transfers and death, mostly for High Urgency patients. A Brazilian study\(^{(19)}\) which analyzed the predictive value of the MTS with regards to the outcomes in adult patients, found the discharge as its main outcome (83.5%), for which there was a predominance of the Yellow category (40.3%), followed by death (12.3%), an outcome that mostly took place for those classified in the Orange category (4.8%). Another study carried out in an Emergency Service in the South of Brazil\(^{(19)}\) also showed that most patients progressed into a discharge (88.4%), which was followed by hospitalizations (11.4%) and deaths (0.2%). The Red (7.8%) and Orange (1.6%) categories were most frequent in the latter cases.

Regarding the length of stay, it was found that High Urgency patients had a higher median than Low Urgency ones. These results are in accordance to a previous study\(^{(10)}\), in which the Red category had a median of 8 days and the Orange and Yellow ones had a median of 5 days. These findings reiterate the capacity of the MTS of predicting which patients are low gravity cases, since most of these progress into favorable outcomes, such as discharge, lower length of stay, and lower total time of hospitalization.

As expected, High Urgency patients are more likely (OR 1.69) to remain more than 24 hours in the emergency services and 2.1 times more likely to remain hospitalized for more than 3 days, when compared to Low Urgency patients. A national study\(^{(10)}\) showed that high-priority patients are 2.5 times more likely to die and 1.5 times more likely to be hospitalized for more than 5 days, when compared to low-priority patients. Similar data was found in another Brazilian study\(^{(19)}\), which attempted to identify the association of sociodemographic and clinical variables and priority levels with the outcomes of adult patients, using an institutional protocol. This study showed that patients classified in the high priority group present a 9.41 times higher change of death and a 6.05 times higher chance of being hospitalized, when compared patients classified as low priority. The profile of patients seeking emergency services, the management of beds in the institutions, and the availability of trained human resources are factors that, together, can influence the outcomes evaluated.

The limitation of this study is related to the evaluation of the data that was objectively recorded in the electronic records. It is also possible that some of the overtriage found results from the fact that its analysis depends on more subjective discriminating factors, such as pain intensity, which may lead the professional responsible for the triage to overvalue the patient. This cannot be objectively evaluated by a retrospective inspection of records.

**CONCLUSION**

This study found that there was a 68.8% accuracy in the risk classification of patients through the use of the MTS in the adult emergency services of a public hospital. The MTS showed itself to be a good predictor of the outcomes evaluated, showing that 65.9% of patients in Low Urgency are later discharged, while 3.8% of the High Urgency ones die.

The inspection of the MTS triage process showed that there is a moderate agreement between the consensus of
Corresponding author: 

Jaqueline Pereira da Costa
E-mail: jpcosta@hcpa.edu.br

## REFERENCES


