BMJ Open

Real-time prediction of patient disposition and the impact of reporter confidence on mid-level triage accuracies: an observational study in Israel


ABSTRACT

Aim The emergency department (ED) is the first port-of-call for most patients receiving hospital care and as such acts as a gatekeeper to the wards, directing patient flow through the hospital. ED overcrowding is a well-researched field and negatively affects patient outcome, staff well-being and hospital reputation. An accurate, real-time model capable of predicting ED overcrowding has obvious merit in a world becoming increasingly computational, although the complicated dynamics of the department have hindered international efforts to design such a model. Triage nurses’ assessments have been shown to be accurate predictors of patient disposition and could, therefore, be useful input for overcrowding and patient flow models.

Methods In this study, we assess the prediction capabilities of triage nurses in a level 1 urban hospital in central Israel. ED settings included both acute and ambulatory wings. Nurses were asked to predict admission or discharge for each patient over a 3-month period as well as exact admission destination. Prediction confidence was used as an optimisation variable.

Result Triage nurses accurately predicted whether the patient would be admitted or discharged in 77% of patients in the acute wing, rising to 88% when their prediction certainty was high. Accuracies were higher still for patients in the ambulatory wing. In particular, negative predictive values for admission were highly accurate at 90%, irrespective of area or certainty levels.

Conclusion Nurses prediction of disposition should be considered for input for real-time ED models.

INTRODUCTION

Overcrowding in the emergency department (ED) has become such a common phenomenon that it is become a routine working environment in many hospitals. The strain on staff and hospital resources has an impact on the ability to provide adequate medical services and directly correlates with the quality of patient care and overall hospital experience. Multiple studies have demonstrated that ED overcrowding has a negative effect on many outcomes, including patient mortality and waiting times, door to needle time in patients suspected of having acute myocardial infarct (door to needle time is the elapsed time between the arrival of a patient with acute myocardial infarction (MI) to the hospital and the start of coronary arteries catheterisation). It is generally accepted that sub 90 min provides optimal outcomes, pain management and delays in antibiotic administration. Additionally, overcrowding is a major contributing factor in staff burnout. Overcrowding is, therefore, a frequent topic of internal auditing and research

Strengths and limitations of this study

- This study was conducted on a large cohort of patients, very few of whom were excluded from analysis, thus strengthening the reliability of the results.
- To our knowledge, this is the first study of its kind conducted in Israel, and the fact that the data support that of previous studies from other regions is reassuring.
- The study was limited to data collected from one emergency department (ED) in one institution in Israel and did not take into account the nurses’ experience or educational background, limiting both its internal and external validity.
- We are unable to draw conclusions on prediction accuracy related to specific diseases or presentation (ie, chest pain/acute coronary syndrome) as this was beyond the scope of our study.
- We believe that the results of the study indicate that predictions could be effectively used as part of a more holistic real-time, machine learning ED analysis tool as an accurate, cost-efficient and quick input metric.
In Israel, a national survey conducted in 2018 revealed that EDs on average operated at 104% capacity, with an average length of stay of 3.0 hours. The Tel Aviv Sourasky Medical Center (TASMC) ED, the location of this study, is a particularly busy inner-city hospital, with a patient length of stay of 3 hours on average and even higher for those requiring admission (51% staying over 5 hours). Improvement in real-time analysis and computational models of ED overcrowding are expected to facilitate better provision of medical treatment and allocation of resources, thus improving patient outcome in the ED as well as in the admitting hospital departments. There are many tools designed for retrospective analysis of ED disposition prediction and overcrowding. Several studies have shown that tools combining objective metrics with triage nurses’ disposition predictions are able to produce good patient admission prediction as early as at time of triage. In recent years, there have been attempts to construct real-time overcrowding models, often using triage scores and bed availability as inputs. Examples include The National Emergency Department Overcrowding Scale, the Emergency Department Work Index and the Risk Management, Economic Sustainability and Actuarial Science Development in Indonesia. No study has, as of yet, compared the efficacy of these tools.

In TASMC’s ED, nurses triage patients using the Canadian Triage and Acuity Scale (CTAS), which is a model combining subjective metrics such as presenting complaint and severity of pain with objective metrics such as vital signs, evidence of bleeding, presence of rash, etc. CTAS levels range from 1 to 5 and indicate the urgency in which patients require medical attention. A score of 1 indicates patient who require immediate attention in the resuscitation bay, whereas 5 indicates non-urgent cases with the lowest priority. In the USA, for comparison, the Emergency Severity Index (ESI) triage method is the most commonly used.

Many studies have demonstrated that triage nurses are able to predict patient disposition with a high degree of accuracy, based on their experience and the limited information available to them at the time of triage. For example, Danette et al published a study, in which triage nurses were able to predict admission with 71.5% sensitivity and discharge with 88.0% specificity. The negative predictive value (NPV) for discharge was also particularly high at 90%. Predictions were most accurate for young patients and for patients with a low (level 1) or high (levels 4–5) ESI score. Another study looking at overall disposition predictions demonstrated similar results (sensitivity 73.6% and specificity 84.5%). Importantly, when nurses were asked to assign a level of confidence to their predictions, a high degree of certainty correlated with improved accuracy of disposition prediction (sensitivity 83.6%, specificity 93.1%, NPV 95%). However, a similar study from the UK was unable to demonstrate high accuracy of triage disposition predictions (sensitivity and specificity 68% and 85%, respectively). The accuracy of nurse triage in Israel has never been assessed in.

In addition to triage nurse predictions, several studies have explored the possibility of using objective metrics to predict patient disposition. A 2009 retrospective study examined 1100 patient cases in 6 medical centres, excluding trauma, psychiatric and obstetrics and gynecology patients. That study used a variant automatic prediction model available during triage: age over 60, chest pain, shortness of breath, dizziness, weakness or syncope, history of cancer, history of diabetes. Each variant was ascribed a weight with a total combined score of 0–14. When the total score was above 4 (34% of cases), the likelihood of admission was 77%, and when the score was above 5 (29% of cases), the likelihood rose to 80%.

Another study attempted to build a prediction model based on data that are routinely collected during triage. This retrospective study included approximately 300 000 ED case files. Of these cases, 60% were used to train the model and 40% were used to validate it. The data used as input for training included demographic characteristics (age, sex, ethnicity), recent (<3 months) hospital admissions or ED visits, method of arrival, patient acuity category and the presence of chronic illness (eg, diabetes, hypertension, dyslipidaemia). The variables that were found to be significant for hospitalisation prediction were age, method of arrival and patient acuity category.

The concept of combining triage predictions and admission prediction models was explored by Cameron et al. In their research, they compared the prediction ability of triage nurses to that of a simple clinical tool, the Glasgow Admission Prediction Score (GAPS) is a score based on age (a point is given for each decade) triage urgency level (20 points for level 1 and 5 points for level 3); 10 points are given if the patient was referred by a doctor to the ED; 5 points are given if the patient was brought in by ambulance or was admitted in the last 12 months. The model also gives a point for each point received by the National Early Warning Score,a score based on vital signs (NEWS score). This tool was found to be efficient in predicting admission. Their research demonstrated that in most cases, GAPS was superior at predicting patient admission outcome over triage nurses (accuracy of 0.810 vs 0.759). The exception was in cases where nurses were very with their prediction, supporting previous findings. The authors proposed a combination of both triage and admission prediction models. By allowing nurses to overrule GAPS when they were certain of their prediction, overall accuracy was improved to 0.892. It is important to note that GAPS is not an objective tool as it takes into account the triage level as determined by the triage nurse. Riodan et al acknowledged that this in their 2017 publication examining patients with ESI level 3. They experimented with several variables including age, pulse, systolic blood pressure and pain in an attempt to build a regression model capable of predicting patient discharge.
As with many areas of medicine, there is growing interest in the field of artificial intelligence, in particular, machine learning, to predict patient admission outcome at the triage level. One such study found that a trained algorithm outperformed classical methods, especially when predicting outcomes for patients with moderate scores (ie, CTAS level 3). This is a field that is expected to develop rapidly in the coming years. Another interesting study by Tahayori et al analysed the use of natural language processing (NLP) to predict the disposition of patients. The algorithm developed was applied to ED triage notes with a relatively high level of accuracy. Such tools are only as robust as the algorithm developed and the data that were input and used to train them, so, at present, it is necessary to continue to develop human approaches to data analysis.

**METHODS**

This is a single centre, observational, retrospective study to determine the accuracy of nurse predictions of patient disposition and destination. Data were gathered between the period of 1 April 2019 and 30 June 2019 in TASMC ED, a tertiary hospital in central Israel, for all adult patients.

### Table 1 Disposition prediction accuracy by wing

<table>
<thead>
<tr>
<th>Acute wing</th>
<th>Ambulatory wing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases predicted to be admitted</td>
<td>Number of cases predicted to be admitted</td>
</tr>
<tr>
<td>Actual number of admitted cases</td>
<td>Actual number of admitted cases</td>
</tr>
<tr>
<td>Accuracy of triage predictions %</td>
<td>Accuracy of triage predictions %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Acute wing</th>
<th>Ambulatory wing</th>
<th>Ambulatory wing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery</td>
<td>687</td>
<td>40</td>
<td>7.3</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>3345</td>
<td>54.8</td>
<td>32.6</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>15</td>
<td>6.7</td>
<td>36.4</td>
</tr>
<tr>
<td>Cardiology</td>
<td>295</td>
<td>41</td>
<td>25</td>
</tr>
<tr>
<td>Orthopaedics</td>
<td>337</td>
<td>51.3</td>
<td>59.7</td>
</tr>
<tr>
<td>Oncology</td>
<td>9</td>
<td>22.2</td>
<td>100</td>
</tr>
<tr>
<td>ENT</td>
<td>97</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Dermatology</td>
<td>79</td>
<td>31.6</td>
<td>48.6</td>
</tr>
<tr>
<td>Neurology</td>
<td>337</td>
<td>41.8</td>
<td>35.6</td>
</tr>
<tr>
<td>Urology</td>
<td>119</td>
<td>30.3</td>
<td>38.5</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>189</td>
<td>32.3</td>
<td>52.4</td>
</tr>
</tbody>
</table>

Figure 1  Triage predictions according to wing.
All the nurses who took part in this study were graduates of the Emergency Medicine Nursing Course. No data were collected on the nurses themselves. The medical team was blinded to the nurses’ triage predictions to avoid bias.

The participating nurses were asked to fill out a questionnaire that was embedded in the ED’s patient managing software. The nurses were aware of the study and completed the questionnaire in a short period of time with no interference with their work. For each patient,
the nurse provided disposition predictions (admission or discharge), exact admission destination prediction (where relevant) and level of certainty in the prediction (high, medium and low).

Patient demographic data were gathered (patients ID number, sex, age) as well as time of arrival and discharge from the ED (home vs admission), triage placement in the ambulatory wing or acute ED wing, triage level (1–5) according to CTAS, vitals (blood pressure, heart rate, oxygen saturation, respiratory rate, temperature) and pain level (according to Numeric pain Assessment Scale). Textual data regarding the reason of ED visit (ie, presenting complaint) were also included. Selection criteria included any patient visiting the ED and seen by the triage team in said period of time, excluding patients seen by the paediatric team.

Data were processed in order to calculate the sensitivity, specificity, positive predictive value (PPV), NPV and accuracy of nurses’ prediction as well as the influence of various patient characteristics on these parameters.

This manuscript was prepared in accordance with the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) statement for improved reporting of outcomes from observational studies.

RESULTS
Between April and June 2019, data were gathered from 33,685 ED visits, of which 11,143 were referred to the ambulatory wing (33%) and 22,542 to the acute department (67%). The average patient age was 51 years old. The men to women ratio was approximately 52:48. A total of 6566 cases (20%) had incomplete triage prediction forms and were excluded from the results. A total of 27,119 questionnaires were included in the analysis—19,146 (71%) acute and 7973 (29%) ambulatory. No statistically significant difference regarding disposition was found between the group that had complete triage prediction forms and the group that was excluded.

In the ambulatory wing of the ED, discharge was predicted in 7307 cases (92%), of which 6950 cases were actually discharged. For this group, the accuracy of nurse predictions was high with 95% accuracy rate. Nurses predicted hospital admission in 666 cases, of which only 312 were actually admitted. Here, the nurses’ predictive accuracy was much lower at 47%. Combined accuracy was 91%. PPV and NPV were 95% and 46%, respectively. For the purpose of this calculation, admission was defined as a positive test result and discharge negative. Sensitivity and specificity were 47% and 95%, respectively (figure 1).

In the acute wing of the ED, discharge was predicted in 13,145 cases, of which 10,816 were actually discharged (overall number of discharges 12,867), with a prediction accuracy of 82%. Hospital admission was predicted in 6001 cases, of which 3950 were actually admitted (overall number of admissions—6279), with a lower accuracy of 66%. Combined accuracy was 77%. PPV and NPV were 84% and 62%, respectively. Sensitivity was 63% and specificity was 84% (figure 1).

Nurses did not demonstrate a high level of accuracy in predicting the receiving admission department in the hospital at the time of triage for both acute and ambulatory wing settings. The exception to this was for admissions to the oncology department; however, this was a very small cohort (table 1).

No significant difference was found in prediction accuracy between male and female patients in either wing. There was also no significant difference in the prediction accuracy for patients with normal vital signs (pulse, BP, oxygen saturation, temperature) compared with patients with abnormal vitals, remaining approximately 90% in the ambulatory wing and 76% in the acute wing. The exception to this was predictions in patients with abnormal temperatures in the ambulatory wing, which reduced prediction accuracy to 72%.

CTAS triage level had a significant influence on prediction accuracy (figure 2). “As expected, with mid-CTAS levels (specifically level 3), predictions were less accurate. In the ambulatory wings, there was only one case of CTAS level 1 and less than 1% of cases were CTAS level 2. In comparison, 50% of cases were CTAS level 4.

In the acute wing, about 1% of patients were CTAS level 1. Most patients were CTAS levels 3 and 4 (50% and 38%, respectively). In this department, predictions in cases with a CTAS level 3 were particularly inaccurate.

The impact of the time of nurses’ working shift on the accuracy of prediction was also evaluated. Nurse shifts in the ED were divided into the morning (07:00–15:00), evening (15:00–23:00) and night (23:00–07:00). During the data collection period for this study, the ambulatory wing closed at 23:00; therefore, only morning and evening shifts were analysed there. In the acute wing, average prediction accuracy was 85% during the night shift, significantly better than the evening (78%) and morning (71%) shifts. The total number of cases recorded in this study was similar for the morning and evening shifts, however, for the night shift; the number of cases was 50% smaller. There was no significant difference in the proportion of cases recorded as CTAS levels 1 and 2 between shifts, although a larger proportion of CTAS level 5 cases was seen during night shifts. For these patients, prediction accuracy was high and contributed to the overall higher accuracy level.

The degree of reporter certainty when making a prediction had a significant impact on accuracy (figure 3). In the ambulatory wings, when a nurse stated that the prediction was made with high certainty, the accuracy of the prediction was over 96%. Most predictions in this wing were stated to be highly certain or moderately certain (5235 and 2541 accordingly), and only a minority were given with low certainty (458, approximately 5.5%).

In the acute wing, a similar increase was observed for predictions reported as having a high degree of certainty—88% accurate, compared with 77% for the
wing as a whole. In this wing, prediction uncertainty was considerably higher (2114, 11%), and the accuracy of these predictions was just 60% (compared with 70% in the ambulatory wing).

Importantly, CTAS level 3 cases with a high degree of reporter confidence were highly accurate (93% for ambulatory wing and 85% for acute wing), significantly greater than CTAS level 3 accuracies as a whole. It is important to point out that the likelihood of a high certainty prediction for triage level 3 cases is lower than average (figure 3, table 2A,B).

**DISCUSSION**

The results of this study support the results of previous studies, namely that trained triage nurses can accurately predict patient disposition during the triage process. At the extremes of CTAS/triage scores (1 and 5), these predictions were more accurate, as is to be expected. Additionally, reporter confidence is also positively correlated to prediction accuracy, potentially highlighting a particularly useful as well as easy metric to measure. We anticipate that the model we presented can be supplemented by machine learning and NLP, such as that presented in Tahayori et al to assist in early identification of patients who require hospitalisation and provide early notice to admitting hospital departments.

After a discussion with nurses who participated in the study, the structure of the questionnaire itself may be the cause of the inaccuracy in predicted admission destination. However, patients are not always admitted to the most suitable ward due to factors outside the control of the ED, such as bed availability. The subject of destination prediction and the varying limiting factors will be further evaluated in future studies.

Regarding the difference in the prediction accuracy between different shifts, it seems that the higher accuracy in the acute wing during night shifts may be in part due to a greater percentage of CTAS level 5 triage patients in that wing during this shift, as ambulatory patients are also seen there at night. As level 5 cases were predicted with a greater degree of accuracy, this may explain the results.

Careful consideration was given to the analysis of CTAS level 3 patients in this study. These patients represent a substantial percentage of presentations to the ED. In general, reporters struggled to accurately predict disposition for this group. It was demonstrated, however, that when the triage nurse was confident in their prediction for this group, the accuracy was also high. This metric may, therefore, allow for accurate predictions for subset of level 3 patients.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Certainty level</th>
<th>% Rate</th>
<th>Discharge</th>
<th>Hospitalisation</th>
<th>Total</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Triage level 3, ambulatory wing</td>
<td>Very certain</td>
<td>29%</td>
<td>15</td>
<td>39</td>
<td>54</td>
<td>72%</td>
</tr>
<tr>
<td></td>
<td>Somewhat certain</td>
<td>52%</td>
<td>73</td>
<td>25</td>
<td>98</td>
<td>26%</td>
</tr>
<tr>
<td></td>
<td>Not certain</td>
<td>19%</td>
<td>32</td>
<td>4</td>
<td>36</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>100%</td>
<td>120</td>
<td>68</td>
<td>188</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td>Very certain</td>
<td>55%</td>
<td>806</td>
<td>27</td>
<td>833</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td>Somewhat certain</td>
<td>38%</td>
<td>524</td>
<td>57</td>
<td>581</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>Not certain</td>
<td>7%</td>
<td>101</td>
<td>13</td>
<td>114</td>
<td>89%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>100%</td>
<td>1431</td>
<td>97</td>
<td>1528</td>
<td>94%</td>
</tr>
<tr>
<td></td>
<td>Grand total</td>
<td></td>
<td>1551</td>
<td>165</td>
<td>1716</td>
<td>87%</td>
</tr>
<tr>
<td>(B) Triage level 3, acute wing</td>
<td>Very certain</td>
<td>34%</td>
<td>1747</td>
<td>255</td>
<td>2002</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>Somewhat certain</td>
<td>53%</td>
<td>2422</td>
<td>691</td>
<td>3113</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>Not certain</td>
<td>12%</td>
<td>498</td>
<td>223</td>
<td>721</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>100%</td>
<td>4667</td>
<td>1169</td>
<td>5836</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>Very certain</td>
<td>31%</td>
<td>210</td>
<td>910</td>
<td>1120</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>Somewhat certain</td>
<td>56%</td>
<td>834</td>
<td>1154</td>
<td>1988</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td>Not certain</td>
<td>13%</td>
<td>248</td>
<td>205</td>
<td>453</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>100%</td>
<td>1292</td>
<td>2269</td>
<td>3561</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>Grand total</td>
<td></td>
<td>5959</td>
<td>3438</td>
<td>9397</td>
<td>74%</td>
</tr>
</tbody>
</table>
An additional study, ongoing at the time of writing, will evaluate the ability of triage predictions to improve the accuracy of a machine learning algorithm designed to predict overcrowding and patient disposition, especially in areas that demonstrated poor accuracy (ie, CTAS level 3).

This research demonstrated that it is possible to predict future discharge with a high degree of certainty for over 60% of ED patients even as early as initial triage. This group includes all ambulatory wing patients, patients at either extreme end of triage severity levels (1 and 5) and any patient for whom the triage nurse is certain of their prediction.

LIMITATIONS
The major disadvantage of the use of triage predictions as part of an overcrowding analysis tool is the added workload for nursing staff. It is our opinion that additional evidence of the effectiveness of this method is required before recommendations are made.

It is evident from the data concerning disposition predictions that they are, in general, not accurate enough in their raw form to greatly influence the management of the ED. However, it is our belief that such data can be used as a part of a real-time ED overcrowding analysis tool, capable of assisting bed managers and improving patient flow as well as allowing for better allocation of resources.

CONCLUSION
Triage nurses are able to accurately predict disposition with a high degree of accuracy, particularly for patients with on either extreme end of the CTAS score. With the introduction of prediction confidence as a metric, accuracy increased for all predictions, including those made for patients with middle-range CTAS scores. However, predictions for patient destination once admitted were not accurate. We believe that implementing these metrics into a machine learning overcrowding tool may improve overall performance and assist in maximising flow through the ED, thus decreasing length of stay.

Contributors DT was involved with conceptualisation, visualisation, drafting the manuscript, reviewing and editing the manuscript; and is guarantor; NS was involved with conceptualisation, investigation, methodology, supervision, drafting the manuscript and reviewing and editing the manuscript; JM was involved with reviewing and editing the manuscript; NNG was involved with data curation; DEF was involved with reviewing and editing the manuscript; SA was involved with project administration and software; AC was involved with data curation; DEF was involved with reviewing and editing the manuscript; NNG was involved with data curation; YM was involved with conceptualisation, visualisation, drafting the manuscript and reviewing and editing the manuscript; AS was involved with project administration and visualisation; MKS was involved with formal analysis; GP was involved with data curation and project administration. DT, NS, JM, NNG, YM, DEF, SA, AC, MKS and GP agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

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.

Patient consent for publication Not applicable.

Ethics approval Ethical approval was sought and approved by a Helsinki committee, reference 2223-19. The Helsinki committee waived the requirement for written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data used in this study was collected from the emergency department of Tel Aviv Sourasky Medical Center. Data summaries are available in the study itself. Deidentified raw data may be available upon request from the Noa Shopen (2nd author) - email: noashopen@gmail.com Reuse will be permitted in the event of collaborative efforts with appropriate accreditation where applicable.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD Aya Cohen http://orcid.org/0000-0003-1996-3571

REFERENCES

