Initial validation of the International Crowding Measure in Emergency Departments (ICMED) to measure emergency department crowding

Adrian Boyle, James Coleman, Yasmin Sultan, Vijayasankar Dhakshinamoorthy, Jacqueline O’Keeffe, Pramin Raut, Kathleen Beniuk

ABSTRACT

Introduction Emergency department (ED) crowding is recognised as a major public health problem. While there is agreement that ED crowding harms patients, there is less agreement about the best way to measure ED crowding. We have previously derived an eight-point measure of ED crowding by a formal consensus process, the International Crowding Measure in Emergency Departments (ICMED). We aimed to test the feasibility of collecting this measure in real time and to partially validate this measure.

Methods We conducted a cross-sectional study in four EDs in England. We conducted independent observations of the measure and compared these with senior clinician’s perceptions of crowding and safety.

Results We obtained 84 measurements spread evenly across the four EDs. The measure was feasible to collect in real time except for the ‘Left Before Being Seen’ variable. Increasing numbers of violations of the measure were associated with increasing clinician concerns. The area under the receiver operating characteristic curve was 0.80 (95% CI 0.72 to 0.90) for predicting crowding and 0.74 (95% CI 0.60 to 0.89) for predicting danger. The optimal number of violations for predicting crowding was three, with a sensitivity of 91.2% (95% CI 85.1 to 97.2) and a specificity of 100.0% (92.9–100). The measure predicted clinician concerns better than individual variables such as occupancy.

Discussion The ICMED can easily be collected in multiple EDs with different information technology systems. The ICMED seems to predict clinician’s concerns about crowding and safety well, but future work is required to validate this before it can be advocated for widespread use.

INTRODUCTION

Emergency department (ED) crowding is increasingly recognised as a global public health problem. Crowding impacts on both patients and staff. Delays to diagnosis and treatment are well described. Privacy and dignity are compromised and mortality is greater in patients who are admitted through crowded EDs. Crowding decreases staff and privacy and dignity are compromised and mortality is greater in patients who are admitted through crowded EDs. Crowding decreases staff and patient satisfaction and increases burnout. Recruitment and retention of ED staff is harmed, training suffers. While there is widespread acceptance that ED crowding is important, there is little consensus on the best way to measure this. Hwang and Concato stress the need to have some conceptual discipline because ‘separating factors that cause overcrowding from the phenomenon itself can assist in the development of interventions that ease overcrowding’. Measuring crowding is important because measurement allows the development and evaluation of interventions. Many studies use the need for ambulance diversion as a proxy measure for ED crowding, though this practice is not widespread and depends on factors other than ED crowding. Occupancy is also used as a measure, as it is easy to collect. The proportion of patients leaving being seen is also easy to measure, but does not adequately measure the problem. The National Emergency Department Overcrowding Score (NEDOCS) measure has been developed and is used inconsistently. NEDOCS measure correlates well with the proportion of patients leaving before being seen. The Emergency Department Work Index (EDWIN) is limited by relying on a five-point triage scale, which is less relevant to many EDs which use a three-point scale.

Our research group has previously developed, by a formal consensus process involving 40 emergency physicians across the world, a measure of ED crowding. This is an eight-point measure, called the International Crowding Measure in Emergency Departments (ICMED), shown in table 1. Most measures of crowding have not been validated. Validation of a crowding measure needs to demonstrate several forms of validity. Demonstrating face validity is important to ensure uptake among clinicians. The aims of this study were to assess the feasibility of recording the components of this scale, partially validate the ICMED, determine a cut-off level at which we would identify an ED as crowded or not and finally, obtain data to plan a sample size for a definitive validation study.

METHODS

We performed a series of cross-sectional observations across four regional hospitals in the East of England; these serve mixed urban and rural areas in autumn 2012. One was a very large teaching hospital, one was a large district general hospital and the remaining two were small district general hospitals. One of us, JC, recorded the status of the ED using seven of the eight measures in real time every hour. The ‘left before treatment’ variable cannot easily be recorded in real time, but is best reported historically. At the same time, the consultant in charge of the ED was asked to record his/her opinion of how crowded and dangerous the ED was on a 10 cm Visual Analogue Scale ranging from ‘0’ (not at all crowded) to 10 (extremely crowded).
The clinician was blinded to the scores recording violations of the crowding scales, but fully aware of the pressure that their department was under. We purposively sampled mainly through the afternoon and the evening and a mix of weekdays and weekends, as when historical data indicated the demands for service were greatest and showed the greatest changes.

We were advised by our Local Research and Ethics Committee that formal ethical approval was unnecessary as this was a service evaluation and no patient contact was proposed. We analysed the data as a diagnostic test, with the component variables and increasing numbers of violations of the ICMED being correlated, Spearman’s rho=0.60 and p<0.001, though danger tended to attract a lower score. The consultants graded their departments as more than five on the crowding Visual Analogue Scale 40% of the time and as more than five on the danger Visual Analogue Scale 15% of the time. Figure 1 shows that as the number of violations increased, the mean perceptions of crowding and danger increased. Table 2 shows the mean perceptions of concern.

Table 3 shows the diagnostic performance of the ICMED against clinician’s perception of crowding. Numbers of violations of the ICMED tended to perform better than individual items. Three violations of the ICMED seemed to have the optimal performance in predicting clinician’s concerns about crowding.

Table 4 shows the diagnostic performance of the ICMED against clinician’s perception of danger. The ICMED performed less well at predicting clinician’s concerns about danger, in part because there were fewer times where an ED was felt to be dangerous.

The ICMED had an area under the ROC curve (AuROC) of 0.80 (95% CI 0.72 to 0.90) for predicting crowding and 0.74 (95% CI 0.60 to 0.89) for danger, respectively. The measure performed less well at predicting clinician’s perceptions of danger, in part because there were not very many data points that indicated a high perception of danger.

![Mean clinician perception of crowding](image-url)
DISCUSSION

We have evaluated our derived measure of ED crowding. We have demonstrated that the IC MED can easily be collected in four diverse EDs that all have different information systems and that seven measures can be collected in real time. We have also demonstrated that increasing numbers of violations are associated with increased clinician perceptions of crowding and danger.

We have also demonstrated that the combination of variables predicts a clinician’s concerns better than individual variables. This is important as many studies use occupancy as their sole measure of crowding. Our results suggest that the optimal number of violations that predict clinician’s perceptions of crowding is three. The ‘Left Before Being Seen’ variable was not feasible to collect in real time, but we suggest that this has value in operational performance management as it is easily collected and has high face validity.

The IC MED predicted clinician's perception of crowding well, with an AuROC of 0.80. There are several other measures of ED crowding reported in the literature. This AuROC is comparable with the highest reported values of NEDOCS (0.92) and EDWIN (0.84), the best studied measures of ED crowding.

When we used the three violations as a predictor of ED crowding, IC MED exhibited higher sensitivity (91.2) and specificity (100) than other measures of ED crowding. There are also concerns whether the NEDOCS measure works well outside the USA and whether it is too complex.

LIMITATIONS

There are some important limitations to our work. We took a convenience sample. Our sampling frame contained very few times when an ED was under extreme pressure. We did not sample between 22:00 and 08:00. It is possible that our results would differ overnight, as the causes of adverse incidents may be very different after midnight. Our anecdotal experience is that crowding mainly occurs in the afternoon and evening. Our gold standard of clinician perceptions of crowding and danger is not validated, but there is no widely accepted, properly validated measure. It has been suggested that clinician perception of crowding alone could be used, however the validity and reliability of this is uncertain, we think the IC MED would perform better as it addresses the components of crowding and is more objective. Some North American studies have evaluated crowding.

### Table 3 Performance of the IC MED against clinician perception of crowding

<table>
<thead>
<tr>
<th>Violations</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive predictive value (95% CI)</th>
<th>Negative predictive value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One violation</td>
<td>92.3 (86.6 to 98.0)</td>
<td>23.4 (16.1 to 34.7)</td>
<td>18.5 (10.2 to 26.8)</td>
<td>100.0 (82.4 to 100)</td>
</tr>
<tr>
<td>Two violations</td>
<td>92.3 (86.6 to 98.0)</td>
<td>40.9 (30.3 to 51.4)</td>
<td>22.2 (13.3 to 31.1)</td>
<td>96.7 (92.8 to 100)</td>
</tr>
<tr>
<td>Three violations</td>
<td>61.5 (51.1 to 71.9)</td>
<td>67.6 (57.6 to 77.6)</td>
<td>25.8 (16.5 to 35.2)</td>
<td>90.6 (84.3 to 96.8)</td>
</tr>
<tr>
<td>Four violations</td>
<td>61.5 (51.1 to 71.9)</td>
<td>87.3 (80.2 to 94.4)</td>
<td>47.1 (36.4 to 57.7)</td>
<td>92.5 (86.9 to 98.2)</td>
</tr>
<tr>
<td>Five violations</td>
<td>53.9 (43.2 to 64.5)</td>
<td>97.2 (93.6 to 100.0)</td>
<td>77.8 (68.9 to 86.7)</td>
<td>92.0 (86.2 to 97.8)</td>
</tr>
<tr>
<td>Six violations</td>
<td>53.9 (43.2 to 64.5)</td>
<td>98.6 (91.1 to 100.0)</td>
<td>87.5 (80.4 to 94.6)</td>
<td>92.1 (86.3 to 97.9)</td>
</tr>
<tr>
<td>Seven violations</td>
<td>23.1 (14.1 to 32.1)</td>
<td>100.0 (94.9 to 100)</td>
<td>100.0 (29.2 to 100.0)</td>
<td>87.7 (80.6 to 94.7)</td>
</tr>
</tbody>
</table>

IC MED, International Crowding Measure in Emergency Departments.
FUTURE WORK
Our results should be regarded as exploratory and should not be used in routine ED operational management yet. A larger validation study with ‘harder’ endpoints, such as inpatient mortality, length of hospital stay and clinical incidents, is required before this measure can be adopted for routine health management. Our measure should also be compared with other widely used measures of ED crowding, the most commonly used measures such as the NEDOCS and the EDWIN. Our measure should also be compared with patient satisfaction scores.

CONCLUSIONS
The ICMED is feasible to collect in UK EDs. A combination of violations, probably three, predicts clinician concerns better than individual violations. Future work is required to validate this before this can be used.

Contributors AB and JC conceived the study. JC collected the data. AB, JC and PR had conducted the study internationally.

REFERENCES