Quantifying emergency department crowding: comparison between two scores

Jérôme Jobé, Anne-Françoise Donneau, Béatrice Scholtes & Alexandre Ghuysen

To cite this article: Jérôme Jobé, Anne-Françoise Donneau, Béatrice Scholtes & Alexandre Ghuysen (2017): Quantifying emergency department crowding: comparison between two scores, Acta Clinica Belgica, DOI: 10.1080/17843286.2017.1410605

To link to this article: https://doi.org/10.1080/17843286.2017.1410605

Published online: 05 Dec 2017.
Quantifying emergency department crowding: comparison between two scores

Jérôme Jobé, Anne-Françoise Donneau, Béatrice Scholtes and Alexandre Ghuysen

ABSTRACT

Introduction: Emergency department (ED) crowding is a major international concern with a negative impact on both patient care and providers. Currently, there is no consensus regarding measure of crowding. Therefore, emergency physicians have to choose between numerous scoring systems, from simple to more complex. The aim of the present study was to compare the complex National Emergency Department Overcrowding Scale (NEDOCS) with the simple ED Occupancy rate (OR) determination. We further evaluated the correlation between these scores and a qualitative assessment of crowding.

Method: This study was conducted in two academic and one regional hospital in Liege Province, in Belgium; each accounting for an ED census of over 40,000 patient visits per year. Crowding measures were sampled four times a day, over a two-week period, in January 2016.

Results: ED staff considered overcrowding as a major concern in the three ED. Median OR ranged from 68 to 100, while NEDOCS ranged from 64.5 to 76.3. A significant correlation was found in each ED between the OR and the NEDOCS (Pearson $r = 0.973, 0.974$ and $0.972$), as well as between the OR, the NEDOCS and the subjective evaluation by the ED staff ($p = 0.001$).

Conclusion: Crowding evaluation in ED requires validated scores. Our study in three different hospitals demonstrates that simple OR appeared as accurate as more sophisticated NEDOCS. Furthermore, this measure is perfectly correlated with the feeling of ED staff.

1. Introduction

Emergency department (ED) crowding appears when the need for emergency services exceeds the resources. ED crowding has become an international concern with negative impact on patients and providers [1,2]. The impact of adaptative strategies and potential solutions has been difficult to measure due to the absence of an optimal definition or measurement and lack of consensus for a universal metric of crowding [3].

Historically, crowding was estimated by the frequency of ambulance diversion or an evaluation of the number of patients leaving without being seen (LWBS). However, such measures were not very reliable because criteria triggering ambulance diversion may vary between ED.

Consequently, numerous quantitative measures have been developed [4–11] but, in practice, their use has been limited by the requirement to gather large amounts of data [3].

Quantifying hospital crowding is measured by the hospital occupancy rate, the ratio between the total number of inpatients in the hospital and the total number of hospital beds during a given period. McCarthy et al. proposed a similar measure of ED occupancy rate (OR). The simple OR is the ratio of the total number of patients in ED to the total number of ED treatment beds [12].

Interestingly, previous work evaluating crowding measures revealed that the complex quantitative measures shared high discriminatory power for predicting current ambulance diversion status, but none performed better than the simple OR. In the absence of a criterion-standard measurement of ED crowding, scales have also been evaluated by their ability to match ED clinicians' perception of crowding. In that perspective, most crowding scales lack scalability and do not perform as designed. However, in that perspective NEDOCS revealed the highest predictive power for instances of perceived ED crowding in comparison with READI, EDWIN and EDCS [13].

Such multidimensional scales work usually well in EDs where they were derived but commonly fail in the difficult challenge of capturing ED-specific variation across and within multiples EDs [3].

Moreover, global crowding measure may mask much of the variation in crowding that occurs within a 24-h period. In that perspective, a simple and parsimonious...
score may be most adapted for real-time ED crowding monitoring.

These results reinforce the need for further evaluation of OR in various settings, with various crowding instruments and criteria, to determine if this simple score could ultimately measure ED crowding as reliably and validly as more sophisticated scores.

Therefore, the aim of this study is to conduct a comparative evaluation of OR and NEDOCS to measure crowding. More specifically, we intended to determine their respective validity in comparison with the perception emergency providers in three different hospital facilities.

2. Methods

2.1. Study design and setting

This was a prospective study designed to assess the relative interest of NEDOCS and OR to evaluate ED crowding in different types of Belgian hospitals. The ethics committee of the University Hospital of Liege approved the study. This study was conducted in two facilities (CHUST and CHUNDB) of an academic hospital, the University Hospital of Liege and in a regional Hospital, the Centre Hospitalier Peltzer-La Tourelle (CHPLT) in Verviers. At the time of the study, the number of beds with hospital agreement was 426 in CHPLT, 263 in CHUNDB and 622 in CHUST. ED census in CHUST and CHUNDB is around 40,000 ED patients per year, while CHPLT receives annually 35,000 patients. CHUST is a tertiary care ED located in the suburban area of Liege, while CHUNDB and CHPLT are located in more urban zones and are secondary care hospitals.

The study was conducted in two phases. During the first, a qualitative retrospective assessment of overcrowding was made through a specific form completed by staff members investigating their perception of ED crowding related to the past 3 months.

The second period was dedicated to a prospective sampling of data required to fulfil OR and NEDOCS in real time. In addition, a survey of the attending nurse on their actual perception of ED crowding level was made at corresponding sampling instances.

2.2. Data collection

2.2.1. Retrospective assessment, site sampling form

Staff members of the three hospitals received a detailed form at the beginning of January 2016, in order to evaluate their feelings concerning overcrowding in their ED over the previous six months. Completion was on a voluntary basis and staff were blinded to OR and NEDOCS score. Investigators developed the form on the basis of previous work, using multiple choices questions. Results were gathered at the end of February 2016.

2.2.2. Prospective assessment

Data required to calculate the scores were collected between 27 February and 11 March 2016. Measures were made simultaneously in the three EDs, at 5–11 am–5 and 11 pm during two weeks for feasibility reason. Most of the data were collected though the department’s electronic patient tracking system and investigators created a checklist to help the staff to collect the required data.

At corresponding sampling instances, the attending nurse in charge was surveyed to assess the perceived level of ED crowding, based on a simplified four-scale system: not busy at all (one point), steadily and easily keeping up (two points), busy (three points), overcrowded (four points).

2.2.2.1. NEDOCS model

NEDOCS was calculated using the following formula:

\[
\text{NEDOCS} = \left( \frac{P_{\text{bed}}}{B_t} \right) \times 85.8 + \left( \frac{P_{\text{admit}}}{B_s} \right) \times 600 + \frac{W_{\text{time}} \times 5.64}{A_{\text{time}} \times 0.93} + R_n \times 13.4 - 20
\]

\( P_{\text{bed}} \) the number of patients in licensed beds and overflow location, such as hallway beds or chairs, \( B_t \) the number of licensed treatment beds, \( P_{\text{admit}} \) the number of admitted patients, \( B_s \) the number of hospital beds, \( W_{\text{time}} \) the waiting time for the last patient put into bed, \( A_{\text{time}} \) the longest time since registration among boarding patients and \( R_n \) the number of respirators in use in the ED or the number of trauma beds. These items entered into the algorithm yielded a score between 1 and 200, with less than 100 considered not overcrowded and more than or equal to 100 considered overcrowded.

2.2.2.2. ED occupancy rate model

ED OR is the ratio of the total number of patients in ED to the total number of ED treatment beds:

\[
\text{Occupancy rate(OR)} = 100 \times \frac{P_{\text{bed}}}{B_t}
\]

\( P_{\text{bed}} \) is the number of patients in licensed bed and overflow locations and \( B_t \) the number of licensed treatment beds.

OR over 100% identifies situations associated with exceeded ED capacity.

2.3. Statistics

Results were expressed as means and standard deviations for normally distributed data and as medians and inter-quartile range for skewed distributed data. Qualitative variables were summarized using counts and percentages. Quantitative variables were compared between the three EDs by an analysis of variance (ANOVA) followed by the application of a multiple comparison test or non-parametric Kruskal–Wallis. Chi-squared or Fisher exact tests were performed to compare qualitative variables between different EDs. Correlation coefficients (classical Pearson or non-parametric Spearman) were calculated to measure the association between two quantitative variables.
The agreement between the crowding classification derived under NEDOCS and OR approaches was evaluated in each ED using the Kappa Cohen coefficient and corresponding interval. All results were considered to be significant at the 5% critical level ($p < 0.05$). Data analysis was carried out using SAS (Version 9.4 for Windows) statistical package.

3. Results

3.1. ED staff data

During the study period, the number of ED licensed treatment beds in CHPLT was 15, 25 in CHU NDB and 28 in CHUST. Thirty-nine nurses and 15 physicians in CHUST, 43 nurses and 17 physicians in CHUNDB, 32 nurses and 12 physicians in CHPLT composed staffs.

3.2. Retrospective ED crowding perception

Seventy-one percentage of nurses and 63.6% of physicians completed the form. Mean age was 37 (35–40) years and 62.6% of nurses were women. Most members of staffs perceived crowding as a major, and sometimes severe, problem. This perception was correlated to seniority only in CHUST (CHPLT: $r = 0.18$ ($p = 0.33$); CHUNDB: $r = 0.48$ ($p = 0.001$) and CHUST: $r = 0.31$ ($p = 0.073$)).

Crowding occurrence was reported weekly in CHUNDB and CHPLT, and every day in CHUST. ED staff from three institutions reported patients waiting in the ward for in-hospital bed admission daily. LWBS were reported once a week at least in CHUST and CHPLT and several times a week in CHUNDB. Full waiting rooms were described every day in CHUST and CHPLT, and several days a week in CHUNDB. This situation was correlated to LWBS patients in CHUST. Saturated short stay units were reported daily in CHUST and several times a week in CHUNDB and CHPLT. Patients staying for more than 24 h in these units were also reported daily in CHUST and several times a week in CHPLT and once a week in CHUNDB (Figure 1).

3.3. Prospective crowding data

During the study period, 147 sampling instances were completed: 53 in CHPLT, 52 in CHUNDB and 41 in CHUST.

3.3.1. Perceived ED crowding

Global nurses perception of crowding ranged from $2.11 \pm 0.81$ in CHPLT, to $2.44 \pm 0.85$ in CHUNDB and $2.58 \pm 0.98$ in CHUST ($p = 0.03$). Nurses reported overcrowding in 3.7% of sample times in CHPLT, 11.5% in CHUNDB and 19.5% in CHUST. ED was reported as busy in 15% of samples times in CHPLT, 32.6% in CHUNDB and 33.3% in CHUST.

3.3.2. NEDOCS

Median NEDOCS (First and third quartile) during the study period was 64.5 (32.0–91.3) in CHPLT, 53.4 (18.3–95.0) in CHUNDB and 76.3 (45.2–99.3) in CHUST ($p = 0.062$). Indeed, mean number of patients in licensed beds and overflow location was 12 ± 7 in CHUPLT, 18 ± 9 in CHUNDB and 28 ± 11 in CHUST ($p < 0.0001$).

Analysis of parameters used to complete the score revealed that the mean waiting time for the last patient put into a bed was 11.7 (3.00–45.0) minutes in CHPLT, 15.0 (1.80–42.6) minutes in CHUNDB, 19.8 (4.80–37.2) minutes in CHUST ($p = 0.45$). The longest time since registration among boarding patients was 159.0 (60.0–240.0) minutes in CHPLT, 147.3 (81.6–312.3) minutes in CHUNDB and 390.0 (250.2–480.0) minutes in CHUST ($p < 0.0001$). The mean number of respirators in use in ED was 2 ± 2 in CHPLT, 4 ± 4 in CHUNDB and 3 ± 3 in CHUST ($p = 0.038$).

Figure 1. Retrospective ED crowding perception.
Calculation of the score identified varying distribution of the values, allowing crowding level categorization (Table 1).

NEDOCS was above 100 in 17.0% of measurements in CHPLT, 21.2% in CHUNDB and 22.0% in CHUST ($p = 0.81$). Time varying analysis of NEDOCS indicates similar profile in the three EDs. In each, there was a significant ($p < 0.05$) difference between diurnal measures (11am and 5 pm) and night (11 pm and 5am).

Indeed, median diurnal NEDOCS was 84.5 (70.3–102.8) in CHPLT, 100.1 (64.4–118.0) in CHUNDB and 99.5 (79.9–119.3) in CHUST ($p = 0.28$). NEDOCS was above 100 in 28.6% of measurements in CHPLT, 45.8% in CHUNDB and 45.0% in CHUST ($p = 0.37$).

Diurnal NEDOCS reported crowding in 28.6% of measurements in CHPLT, 45.8% in CHUNDB and 45.0% in CHUST (Table 2).

### 3.3. OR
Median OR (First and third quartile) during the study period was 73.3 (53.5–113.3) in CHPLT, 68.0 (40.0–94.0) in CHUNDB and 100.0 (64.3–132.1) in CHUST ($p = 0.0032$).

OR was $> 100$% in 30.2% of measurements in CHPLT, 17.3% in CHUNDB and 48.8% in CHUST ($p = 0.005$). Median diurnal OR was 103.3 (86.7–120.0) in CHPLT, 96.0 (74.0–124.0) in CHUNDB and 132.1 (105.4–144.6) in CHUST ($p = 0.010$). Diurnal OR was $< 100$% in 50% of measurements in CHPLT, 37.5% in CHUNDB and 80% in CHUST ($p = 0.016$). Time varying analysis of OR indicates similar profile in the three EDs (Figure 2).

### 3.3.4. Correlation between NEDOCS and OR
Highly significant ($p < 0.0001$) positive correlation was observed between the two scores in different locations

#### Table 1. Overall distribution of the NEDOCS score according to categorization.

<table>
<thead>
<tr>
<th>NEDOCS category</th>
<th>CHPLT</th>
<th>CHUNDB</th>
<th>CHUST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not busy (0–20)</td>
<td>10 (18.9)</td>
<td>14 (26.9)</td>
<td>4 (9.76)</td>
</tr>
<tr>
<td>Busy (20–60)</td>
<td>16 (30.2)</td>
<td>17 (32.7)</td>
<td>8 (19.3)</td>
</tr>
<tr>
<td>Severely busy (60–100)</td>
<td>18 (33.9)</td>
<td>10 (19.2)</td>
<td>20 (48.8)</td>
</tr>
<tr>
<td>Overcrowded (100–140)</td>
<td>6 (11.3)</td>
<td>7 (13.5)</td>
<td>7 (17.1)</td>
</tr>
<tr>
<td>Severely overcrowded (140–180)</td>
<td>2 (3.77)</td>
<td>4 (7.7)</td>
<td>1 (2.44)</td>
</tr>
<tr>
<td>Dangerously overcrowded (180–200)</td>
<td>1 (1.89)</td>
<td>0 (0.00)</td>
<td>1 (2.44)</td>
</tr>
</tbody>
</table>

#### Table 2. Diurnal and night distribution of the NEDOCS score in each ED according to categorization.

<table>
<thead>
<tr>
<th>NEDOCS category</th>
<th>CHPLT ($n = 28$)</th>
<th>CHUNDB ($n = 24$)</th>
<th>CHUST ($n = 20$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not busy (0–20)</td>
<td>0 (0.00)</td>
<td>10 (40.0)</td>
<td>0 (0.00)</td>
</tr>
<tr>
<td>Busy (20–60)</td>
<td>4 (14.8)</td>
<td>12 (48.0)</td>
<td>0 (0.00)</td>
</tr>
<tr>
<td>Severely busy (60–100)</td>
<td>16 (57.1)</td>
<td>2 (8.0)</td>
<td>9 (37.5)</td>
</tr>
<tr>
<td>Overcrowded (100–140)</td>
<td>5 (17.9)</td>
<td>1 (4.0)</td>
<td>0 (0.00)</td>
</tr>
<tr>
<td>Severely overcrowded (140–180)</td>
<td>2 (7.14)</td>
<td>0 (0.00)</td>
<td>4 (16.7)</td>
</tr>
<tr>
<td>Dangerously overcrowded (180–200)</td>
<td>1 (3.57)</td>
<td>0 (0.00)</td>
<td>0 (0.00)</td>
</tr>
<tr>
<td>$P$-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**3.4. Crowding scores and ED crowding perception**

#### 3.4.1. Retrospectively
Logistical regression analysis indicated a significant correlation between the level of NEDOCS and the perception of crowding by staff members in CHPLT ($p = 0.001$, coefficient $B = 0.804$), CHUNDB ($p = 0.001$; $B = 0.567$) and CHUST ($p = 0.001$; $B = 0.709$). A significant correlation was also found for OR in CHPLT ($p = 0.001$; $B = 0.751$), CHUNDB ($p = 0.003$; $B = 0.595$) and CHUST ($p = 0.001$; $B = 0.600$).

#### 3.4.2. During sample instances
A highly significant ($p < 0.0001$) positive correlation was observed between NEDOCS score and actual nurse perceptions about crowding in different locations (CHPLT: $r_s = 0.80$, CHUNDB, $r_s = 0.57$ and CHUST, $r_s = 0.70$). The same observation holds for OR (CHPLT: $r_s = 0.75$, CHUNDB, $r_s = 0.59$ and CHUST, $r_s = 0.70$).

### 4. Discussion

Results from this study indicate a significant correlation between the two scores, independently of the facility studied. We also found similar performance of these scores in terms of staff perception of ED crowding.

In this work, we found a significant correlation between NEDOCS and OR during the study period, in every location. The median NEDOCS score originally reported in the validation study stood at 93 (72–112) [5], while another team reported higher values of 124 ± 44 (Mean ± SD) [14]. In our three facilities, the median NEDOCS noted during the study period appeared lower, ranging from 64.5 (CHPLT) to 76.3 (CHUST), a condition with lower performance levels.

Although clinician perception of ED crowding has been the least commonly used type of crowding measure, correlation between this perception and other crowding measures have been widely used in validation studies. When describing the specific situation of their ED during the past 3 months, staff members of the three facilities perceived crowding as a major or severe problem.
We found a significant correlation, across all facilities, between the level of crowding perceived retrospectively or in real time by ED staff and NEDOCS score levels measured at that moment.

Our results with the NEDOCS score are consistent with previous studies, notably those originally reported, demonstrating a high predictive power of perceived ED crowding with NEDOCS [5]. However, generalization of this performance across diverse ED settings has sometimes revealed some weakness and discrepancies. Indeed, an Australian team found no significant relationship between NEDOCS and subjective scores due to unreplicable patient flow or staff profiles [15]. Interestingly, in our study, the fit between the estimation of OR and ED crowding staff perception shared the same performance as NEDOCS. Of note, nurses also perceived the highest LWBS patients occurrence in the facility with the highest OR and NEDOCS levels. However, we have to recognize that it was not a measure but was rather an ‘impression’ of the staff.

The most commonly described input measure reported previously is numerical counts or percentage of patients as they arrived to ED. The interest of facilities involved in the present work, was to share approximatively the same numerical ED census, but to have very different settings. Such constraints might account for qualitative discrepancies in terms of severity of illness notably, with a tertiary university hospital involved in complex and multi-morbid admissions and secondary urban and suburban facilities dealing with more ambulatory and fast track cases. We postulated that these differences would account for some variation in terms of crowding. Indeed, sub-analysis of NEDOCS data supports this idea, with a longest time since registration among boarding patients are significantly longer in CHUST.

Concerning output measures, the longest mean waiting time for the last patients to be placed in a hospital bed and the longest stay in ED short stay unit pointed again to the CHUST as the facility, the most impacted by crowding.

Assessing crowding in ED remains very challenging. Indeed, apart from the global perception of staff members, necessarily subjective and reflecting the global and usual situation of the department, day-to-day, minute by minute situation and its permanent, apparently chaotic, fluctuation makes its approach difficult.

Time analysis of scores indicated similar profile in the three EDs and confirmed that crowding was mainly a diurnal phenomenon, sometimes extending to the evening, but rarely at night (Figure 2). However, though ED occupancy or demand may appear chaotic they are accurately forecastable. Indeed, time series analysis has been shown to provide useful, readily available tools for predicting emergency department workload that can be used to plan staff roster and resource. For that purpose, however, real-time data acquisition requires simple and easy to obtain measures such as OR.

This study did not intend to explore specifically the causes or consequences of OR in EDs studied. Indeed, we limited the evaluation of consequences of crowding to data potentially revealing crowding issues and it would have been interesting to explore this aspect more deeply. For technical reasons, data sampling occurred in

Figure 2. OR week profile for the three EDs.
a limited period of time and further studies should be conducted to confirm our findings.

ED crowding has been reported to be like the elephant standing in the room, it is very difficult to describe how heavy he is, how bad he smells and just when the floor might give [17]. Moving one step closer to the ever-elusive operational definition of ED crowding, a call for standardization of overcrowding measures has recently emerged. To fulfil practical requirements and ED specifications, the best candidate should be simple, valid, reliable and sensitive to changes through times. We believe that OR is an easily available, simple and consistent measure that offer a real advantage in terms of generalizable estimation of ED crowding than more complex multidimensional measures. According to Table 3 making a cross-classification of both score in each ED, we could even advance that OR is more sensitive to crowding than NEDOCS. So, OR may not only challenge difficulties of specific variations across EDs from large, academic, inner city ED to more modest, rural or locals ED from district hospitals but also provide simple data for forecasting models.

5. Conclusion
Crowding evaluation in ED requires validated, easy-to-use and generalizable scores. In that perspective, our study in three hospitals facilities with various setting, simple OR appeared as accurate as more sophisticated NEDOCS, in quantifying overcrowding objectively. Furthermore, this measure is perfectly correlated with the perception of ED staff.

Disclosure statement
No potential conflict of interest was reported by the authors.

References