CLINICAL PRACTICE

Development and Validation of a New Index to Measure Emergency Department Crowding

Steven L. Bernstein, MD, Vinu Verghese, MD, Winifred Leung, BS, Anne T. Lunney, RN, Ivelisse Perez, RN, BS

Abstract

Objectives: To develop a quantitative measure of emergency department (ED) crowding and busyness. **Methods:** A five-week study in spring 2002 in an urban teaching ED compared a new index (the Emergency Department Work Index [EDWIN]) with attending physician and nurse ratings of crowding. EDWIN is defined as $\sum n_i t_i / N_a(B_T - B_A)$, where n_i = number of patients in the ED in triage category i, t_i = triage category, N_a = number of attending physicians on duty, B_T = number of treatment bays, and B_A = number of admitted patients in the ED. The triage system used is the Emergency Severity Index (ESI), which was modified by reversing the ranking of triage categories; that is, an ESI score of 1 represented the least acute patient and 5 the sickest. EDWIN was calculated every two hours in a convenience sample of 60 eight-hour shifts. With each measurement, the charge attending physician and nurse estimated how busy/crowded the ED was, using a Likert scale. Nurse and physician assessments were averaged and compared with EDWIN scores. Data were analyzed with SPSS 10.0 (SPSS Inc., Chicago, IL). **Results:** A total of 2,647 patients aged 18 years and older were assessed at 225 time points over 35 consecutive days. Nurses and physicians showed good interrater agreement of crowding assessment (weighted κ 0.61, 95% confidence interval = 0.53 to 0.69). Median EDWIN scores and interquartile ranges (IQRs) when the ED was rated as not busy, average, and very busy were 1.07 (IQR = 0.80 to 1.55), 1.55 (IQR = 1.16 to 1.93), and 1.83 (IQR = 1.42 to 2.45) ($p < 0.001$). The ED was on diversion for 17 time blocks (6.5% of all blocks), with a median EDWIN of 2.77 (IQR = 1.83 to 3.63), compared with an EDWIN of 1.45 (IQR = 1.05 to 2.00) when not on diversion ($p < 0.001$). EDWIN scores correlated weakly with various process-of-care measures chosen as secondary end points. **Conclusions:** EDWIN correlated well with staff assessment of ED crowding and diversion. The index can be programmed into tracking software for use as a ''dashboard'' to alert staff when the ED is approaching crisis. If validated across other sites, EDWIN may provide a tool to compare crowding levels among different EDs. **Key words:** emergency department; overcrowding; quality of care. ACADEMIC EMERGENCY MEDICINE 2003; 10:938–942.

Emergency department (ED) crowding is thought to be associated with adverse outcomes, error, and decreased patient satisfaction. Although emergency physicians have an intuitive sense of when an ED is becoming crowded, there is no universally accepted quantitative index of ED crowding, and ED crowding remains difficult to define.[1,2]

The goal of this study was to develop a simple, quantitative, reproducible, valid measure of ED busyness, which could be integrated into clinical information systems and used as a real-time tool. This index would have the potential to alert personnel when the ED is approaching ''crisis,'' allowing a stepped-up response to increasing volume or acuity or both; assist in triage to various sections of the ED; and be used for administrative, quality improvement, educational, and research activities. It also would allow EDs of different sizes, capacities, and staffing patterns to compare levels of activity.

A secondary goal of this study was to examine whether adverse events are more common during periods of peak ED crowding. There are scant data exploring the relationship between ED crowding and quality of care; in one study, from a single ED in Spain, 72-hour returns and ED mortality were more common during busier periods.[3] Two surveys of 323 ED directors in New York, Florida, Texas, and California cited 13 poor clinical outcomes from crowding.[4,5]

METHODS

Study Design. This was a prospective, observational study to measure agreement between nurse and physician perceptions of ED crowding and a new instrument, the Emergency Department Work Index (EDWIN). The requirement to elicit informed consent

From the Department of Emergency Medicine, Newark Beth Israel Medical Center, St. Barnabas Health Care System, Newark, NJ (SLB, VV, WL, ATL, IP); Mount Sinai School of Medicine, New York, NY (SLB); and Clinical Epidemiology Unit, Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine, New Haven, CT (SLB).

Received December 30, 2002; revisions received February 10, 2003, and March 5, 2003; accepted March 11, 2003.

Presented in part at the American College of Emergency Physicians Research Forum, October 2002, Seattle, WA.

Address for correspondence and reprints: Steven L. Bernstein, MD, Department of Emergency Medicine, Montefiore Medical Center, Albert Einstein College of Medicine, 111 East 210th Street, Bronx, NY 10467. Fax 718-798-0730; e-mail: sbernste@montefiore.org.

doi:10.1197/S1069-6563(03)00311-7

from staff and patients was waived by the hospital's institutional review board.

Study Setting and Population. The study was conducted from March 28, 2002, to May 2, 2002, at an urban teaching hospital, Level 2 trauma center, with an annual volume of 78,000 visits. The ED is divided into three components: general adult, pediatrics, and fast-track. The study was based in the general adult ED.

Construction of the Index. The EDWIN is defined as follows:

$$\text{EDWIN} = \sum n_i t_i / N_a (B_T - B_A)$$

where n_i = the number of patients present in the ED in triage category *i*; t_i = the triage category (ordinal scale 1–5, 5 being most acute); N_a = the number of attending physicians on duty at a given time; B_T = the total number of beds, or treatment bays, available in the ED; and B_A = the number of admitted patients (holds) in the ED.

When a patient is admitted, he or she is counted as a ''hold'' and is removed from the numerator, which is a sum of the triage categories of all active patients in the ED. The number of treatment bays (B_T) is derived from the ED's original blueprint. It omits the various beds placed in hallways and corners by ED personnel.

The triage system used was the Emergency Severity Index (ESI), a five-level instrument that has high interobserver agreement and is associated with resource use and hospitalization rates.[6,7] ESI was chosen because of its reproducibility and simple, algorithmic quality. The ESI was modified slightly by reversing the ordinal ranking of triage categories; that is, 1 was the least acute patient and 5 the sickest. This was done to maintain the arithmetic sensibility of the index: As the numerator increases in value, the busyness of the ED increases. The units of EDWIN may be represented as ''patient triage units per attending physician per available bed.''

Study Protocol. The nurse and attending physician in charge were approached independently, by a trained study observer (medical student, nurse, or emergency medicine resident), every two hours in the ED. Each was shown the following question and asked to answer it using the five-point Likert scale: ''How busy would you say the ED is right now? Please take into account your workload, the workload of all other doctors and nurses, the numbers of patients in the ED and waiting room, and numbers of holds (admitted patients waiting for beds):

1 not busy at all, not crowded
2 steady, easily keeping up
3 average: working hard, but keeping up
4 more crowded and busy than desirable
5 extremely busy, very crowded.''

Nurse and physician were blinded to each other's answers.

Main Outcome Measures. Responses were averaged and reduced to three groups (not busy, average, and busy) by merging the 1s and 2s into one group and the 4s and 5s into another. Values that fell between two groups were assigned to the busier group. The combined nurse/physician measurement of ED crowding was compared with the EDWIN score calculated for that time point.

Secondary Outcome Measures. Individual EDWIN scores also were calculated for all patients in the study, by averaging the scores for the consecutive time points during which each patient was in the ED. If a patient was in the ED between noon and 6 PM, EDWIN scores for that interval were averaged and assigned to that patient as his or her overall score. Several standard process-of-care measures were examined to determine if EDWIN scores for those patients were higher than for others. These measures included patients who left the ED without treatment, patients who left against medical advice (AMA), patients who returned within 72 hours of the initial visit (by examining EDWIN scores for the initial visit), cases referred to the department of quality improvement, and radiology overreads. Because of their relative paucity, the 72-hour returns, quality improvement, and radiology overread cases were grouped into a composite variable. The walkout and AMA patients also were grouped into a composite variable.

Data Analysis. The ED's electronic patient tracking system (Emergency Department Information Manager, Medamerica, Inc., Livingston, NJ) was used to obtain data concerning individual patients, including basic demographic and clinical information and length of stay in the ED. Performance improvement cases were obtained from the hospital's Center for Performance Improvement. Radiograph overreads were obtained from the department of radiology, and 72-hour returns were obtained from the ED's electronic medical record.

Data were entered into SPSS 10.0 (SPSS Inc., Chicago, IL). All stochastic comparisons were performed using nonparametric statistics. Univariate analyses are presented as medians with interquartile ranges (IQRs) except where noted. EDWIN scores for low-activity, medium-activity, and high-activity times (as assessed by nurse and physician) were compared using the Kruskal-Wallis test. The Mann-Whitney-Wilcoxon test was used for two-group comparisons. The weighted κ test statistic was used to assess nurse–physician agreement for crowding and physician–physician

agreement for clinically pertinent radiograph overreads. All comparisons are two-tailed and use a p-value of < 0.05 to denote statistical significance.

RESULTS

A total of 4,816 patients were treated in the ED during the study period, of whom 2,647 (55.0%) had EDWIN scores calculated. During the study period, 225 (53.6%) measurements of EDWIN were made, out of 420 potential readings. Readings were taken between midnight and 8 AM 53 times (23.5%), between 8 AM and 4 PM 103 times (45.8%), and between 4 PM and midnight 69 times (30.7%). Weekends were proportionately represented. Demographic and clinical features of the study patients are shown in Table 1.

Summary Results for the Emergency Department. The median EDWIN score for the entire study period was 1.50 (IQR = 1.07 to 2.11). Scores tended to fall between midnight and 6 AM, because of decreasing patient volume and rise rapidly between 8 AM and noon, when new patient registrations outpaced attending coverage (data not shown).

EDWIN scores were compared with the mean of the nurse and physician Likert scores for busyness/crowding, as shown in Figure 1. There was excellent correlation between EDWIN scores and nurse/physician assessment of crowding (Kruskal-Wallis chi-square test for difference in median EDWIN scores = 42.9, $p < 0.001$). Nurses and physicians had very good-excellent agreement with rating of ED busyness and crowding (weighted κ 0.61, 95% confidence interval = 0.53 to 0.69).[8]

TABLE 1. Characteristics of Study Patients Seen during the Study Period

	EDWIN Scores Available	EDWIN Scores Not Available	p-value
No. patients	2,647	2,169	
Female	1,686	1,356	0.696
Age (yr, median, IQR)	38 (27 to 52)	35 (26 to 48)	< 0.001
Insurance			0.19
Medicaid	247	163	
Medicare	282	226	
Self-pay	879	739	
Other	938	805	
Unknown	301	236	
Disposition			< 0.001
Discharge	1,771	1,578	
Admit	586	335	
Walkouts, AMA	158	144	
Transfer to psychiatric ED*	112	81	
Other (transfer, died in ED)	20	31	

*Patients with primarily psychiatric complaints initially are screened in the general emergency department (ED), then transferred to the psychiatric ED for further evaluation.
IQR = interquartile range; AMA = against medical advice.

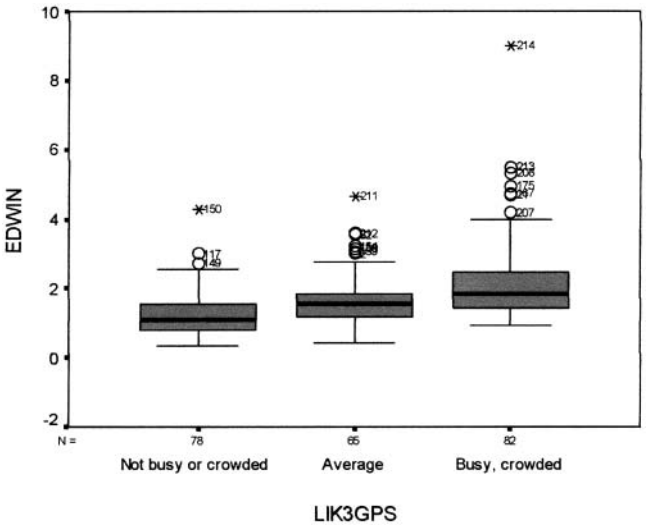

Figure 1. Box-and-whiskers plot of pilot EDWIN data ordinalized into three zones. *LIK3GPS* refers to the five-point Likert scale, reduced to three groups. The left-side box plot shows EDWIN values during a nurse/attending physician–reported assessment of busyness as 1 or 2 (on a five-point Likert scale, with 5 maximally busy), the center box plot shows scores during a reported assessment of busyness as 3, and the right-side box plot shows EDWIN scores during a reported assessment of busyness as 4 or 5. Median values with interquartile ranges: not busy or crowded, 1.07 (0.80 to 1.55), average, 1.55 (1.16 to 1.93), and busy/crowded, 1.83 (1.42 to 2.45). In the not busy group, there were 78 measurements; in the average group, 65; and in the busy/crowded group, 82. Kruskal-Wallis chi-square test for difference in median EDWIN scores = 42.872, $p < 0.001$.

A strong association was found between EDWIN scores and diversion. Of the 225 time points in the study, the ED was on diversion for 17 (6.5%). The median EDWIN for the time on diversion was 2.77 (IQR = 1.83 to 3.63) compared with a median EDWIN of 1.45 (IQR = 1.05 to 2.00) when not on diversion (Mann-Whitney-Wilcoxon, $p < 0.001$).

Summary Results for Individual Patients. The median EDWIN score for all study patients was 1.67 (IQR = 1.27 to 2.22), which was not statistically different from that of patients who walked out, left AMA, or returned within 72 hours. Because of their small numbers and clinical importance, a composite variable was created consisting of 72-hour returns who were admitted, x-ray overreads, and cases referred for quality improvement. These 34 patients had a median EDWIN of 1.90 (IQR = 1.45 to 2.33) compared with 1.66 (IQR = 1.25 to 2.22) for all other patients (Mann-Whitney-Wilcoxon, $p = 0.10$).

Lastly, patient data were analyzed by assigning each patient to the active, busy, or crowded groups, depending on EDWIN score. Using the median scores in the figure, ED activity was divided into three groups: not busy, average, and busy, with EDWIN ranges of 0 to 1.31, 1.32 to 69, and 1.70 to highest score. When analyzed this way, the 67 patients who returned in 72 hours or had x-ray overreads or quality improvement issues had EDWIN scores significantly

higher than other patients (Wilcoxon-Mann-Whitney, $p = 0.03$). There were no statistically significant differences between the patients in the composite variable and the other study patients with respect to age, sex, insurance status, or ED length of stay.

DISCUSSION

Analysis of ED crowding has been hampered by lack of a uniform definition. Among definitions of crowding in use are periods when the ED is on diversion,[9] daily visit totals exceeding a certain threshold,[3] and all ED beds filled more than eight hours daily.[4] Developing a quantitative measure of ED activity and correlating it with adverse outcomes achieves several goals. First, it allows EDs of different sizes and capacities to compare their workloads. Second, it broadens the concept of ED crowding into the wider context of overall levels of activity in the department. This may give a more appropriate lens with which to analyze relationships among crowding, activity levels, and adverse outcomes. Third, EDWIN allows emergency physicians to link the issue of ED crowding to the quality of care and patient safety movements.[10–12]

EDWIN was found to be strongly associated with nurse and attending physician assessment of ED crowding and the presence of diversion. It was weakly associated with a composite variable of three quality-of-care measures: 72-hour returns, x-ray overreads, and quality improvement cases.

The variables used in calculating EDWIN were chosen for their ease of enumeration, clinical sensibility, and universality. The numerator of EDWIN, the summary triage score, has been described previously as one measure of real-time ED activity.[13] To maintain conceptual and calculational simplicity, other variables were omitted that may affect the degree of busyness of an ED. These variables include the following:

1. **Nursing staffing:** Nurse staffing levels may fluctuate considerably from day to day, even within a shift, as nurses may be assigned different tasks in various parts of the ED (e.g., triage, fast-track, adult, or pediatric treatment areas). Omission of nurse staffing was done for practical reasons, not to deny the crucial role of nursing in patient flow.
2. **Resident staffing:** Because most EDs do not employ residents, and because the contribution of residents to patient throughput is unclear, resident presence in the ED was omitted.
3. **Emergency department size:** EDs with more square footage per patient may provide greater throughput than those with less, but there are no data to support this.
4. **Ancillary services:** The availability and timeliness of ancillary services, such as radiology, clinical pathology, or language interpreters, affect throughput, but quantitative measures of these services do not exist.

EDWIN, if validated at other institutions, has several potential uses. As discussed, it can provide a universal measure of ED activity, to "standardize" the assessment of ED crowding and activity level across departments of varying size and capacity. This would be of use primarily to policy makers and researchers.

EDWIN can be used in real-time by ED staff as a "dashboard" instrument, programmed into the unit's patient tracking software, to allow physicians and nurses to assess ED activity. Specifically, if validated as predictive of adverse clinical or process-of-care outcomes, EDWIN may be used to alert staff when the ED is approaching crisis, to allow for a stepped response to a rising index. The nature of the stepped response may vary, depending on the values of the component terms in EDWIN.

EDWIN scores may be used in part to determine diversion status (for hospitals that allow diversion). Criteria for diversion vary widely across emergency medical services jurisdictions but typically depend on some combination of availability of ED beds and cardiac monitors, hospital inpatient census, and availability of critical care beds. If the various hospitals in an EMS system use EDWIN, diversion status may be given first to the hospital with the highest EDWIN, and so forth.

EDWIN suggests that ED activity may be demarcated into three ordinal zones. Using the scores shown in Figure 1, an active but manageable ED has an EDWIN score less than 1.5, a busy ED has an EDWIN between 1.5 and 2, and a crowded ED has a score greater than 2. These three zones—active, busy and crowded—suggest an "ABC" paradigm to characterize ED activity. A multicenter study is needed to see whether activity at other EDs, of different volumes and staffing patterns, follows a similar demarcation.

LIMITATIONS

This study was performed as a convenience sample of time blocks over a relatively brief interval. Continuous data collection over a longer period might have given a more accurate sampling of ED activity.

We studied ED activity at only one institution. To validate and generalize the index, a multicenter study is needed.

EDWIN does not include a term for nursing staffing. Our tracking software includes fields for physician and nurse; however, nursing data are not entered routinely into the tracking software, making a program to count numbers of nurses on duty unreliable. Decreased levels of nurse staffing on inpatient units have been shown to be associated

with adverse outcomes.[14] Future studies incorporating nursing staffing in EDWIN scores should clarify this issue.

The contribution of residents and medical students to patient flow remains unclear. Future studies can consider staffing by these groups as well.

High EDWIN scores were found to be associated with adverse process-of-care measures. The number of patients in each process-of-care category was small, necessitating use of a composite variable. Several individual measures had p-values approaching 0.05, but the exploratory nature of the analyses in this study makes it important to caution against a type I error.

Although studying associations between EDWIN scores and process outcomes is important, we believe that associating higher scores with adverse clinical outcomes, such as missed myocardial infarction and fracture, would be the more clinically pertinent use of EDWIN.[15]

CONCLUSIONS

The Emergency Department Work Index (EDWIN) is an algebraically simple, clinically sensible measure of ED busyness and crowding. It exhibits face and content validity and, at one institution, was associated with nurse and physician assessment of ED crowding and initiation of diversion and weakly associated with quality-of-care measures. The score may be programmed into patient tracking software for use as a real-time measurement of ED activity. A multicenter study is needed to see whether higher EDWIN scores are associated with adverse measures of quality of care.

The authors thank Jennifer R. Bernstein for assistance with data analysis and Carol L. Barsky, MD, and David Salo, MD, PhD, for careful review of the manuscript.

References

1. McCabe JB. Emergency department overcrowding: a national crisis. Acad Med. 2001; 76:672–4.
2. Derlet RW. Overcrowding in U.S. EDs: a critical condition. Calif J Emerg Med. 2002; 3:24–6.
3. Miro O, Antonio MT, Jimenez S, et al. Decreased health care quality associated with emergency department overcrowding. Eur J Emerg Med. 1999; 6:105–7.
4. Derlet RW, Richards JR. Emergency department overcrowding in Florida, New York, and Texas. South Med Assoc J. 2002; 95:846–9.
5. Richards JR, Navarro ML, Derlet RW. Survey of directors of emergency departments in California on overcrowding. West J Med. 2000; 172:385–8.
6. Wuerz R, ESI Triage Study Group. Emergency Severity Index Triage category is associated with six-month survival. Acad Emerg Med. 2001; 8:61–4.
7. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Acad Emerg Med. 2000; 7:236–42.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–74.
9. Schull MJ, Szalai JP, Schwartz B, Redelmeier DA. Emergency department overcrowding following systematic hospital restructuring: trends at twenty hospitals over ten years. Acad Emerg Med. 2001; 8:1037–43.
10. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press, 2000.
11. Corrigan JM, Donaldson MS, Kohn LT, et al. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
12. Gordon JA, Billings J, Asplin BR, Rhodes KV. Safety net research in emergency medicine: proceedings of the Academic Emergency Medicine Consensus Conference on "The Unraveling Safety Net." Acad Emerg Med. 2001; 8:1024–9.
13. Reeder TJ, Garrison HG. When the safety net is unsafe: real-time assessment of the overcrowded emergency department. Acad Emerg Med. 2001; 8:1070–4.
14. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. N Engl J Med. 2002; 346:1715–22.
15. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a modified-Delphi approach. Acad Emerg Med. 2002; 9:1131–9.