Original Paper

Smartwatch-Measured Clinician Perceptions of Emergency Department Crowding and Their Association With Objective Metrics: Prospective Observational Study

Ji Sim Yoon[1*]; Sookyung Park[2*], MPH; K Jane Muir[3,4], PhD; Pamela B DeGuzman[2], PhD; So Yeon Shin[5]; Hansol Chang[1,6,7], MD, PhD; Taerim Kim[6], MD, PhD; Se Uk Lee[6], MD, PhD; Sungjun Hong[1], PhD; Won Chul Cha[1,6,7], MD, PhD; Junsang Yoo[1], PhD

[1]Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

[2]School of Nursing, University of Virginia, Charlottesville, VA, United States

[3]Center for Health Outcomes and Policy Research, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[4]Departmet of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Nursing, Samsung Medical Center, Seoul, Republic of Korea

[6]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[7]Digital Innovation Center, Samsung Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Junsang Yoo, PhD
Department of Digital Health
SAIHST
Sungkyunkwan University
115, Irwon-ro, Gangnam-gu
Seoul, 06355
Republic of Korea
Phone: 82 2 2148 7703
Email: junnsang@skku.edu

Abstract

Background: Emergency department (ED) crowding compromises patient safety and increases clinical workload. Existing crowding metrics capture structural indicators but not how clinicians subjectively experience crowding in real time. Clinician perception may provide a complementary measure.

Objective: This study aimed to examine the concurrent and temporal associations between clinicians' real-time ED perceived crowding (EDPC), captured via a smartwatch-based ecological momentary assessment (EMA) system, and 3 objective crowding metrics: the ED occupancy ratio (EDOR), the National Emergency Department Overcrowding Score (NEDOCS), and the ED Work Index (EDWIN).

Methods: We conducted a prospective observational study over a continuous 449-hour period (November 25, 2024, 7 AM to December 13, 2024, 11:59 PM) in a single tertiary ED operating under reduced bed capacity and attending-only staffing due to a national physician shortage. We enrolled 28 clinicians (5 attending physicians and 23 charge nurses; mean age 36.3, SD 5.9 years; 21 female and 7 male) by voluntary nonprobability sampling. Clinicians recorded EDPC hourly during active shifts on a continuous 0-200 scale anchored to the NEDOCS crowding categories, from "Not Busy" to "Dangerously Overcrowded." All valid ratings within each hourly period were aggregated into a single median EDPC value. Interrater agreement was assessed with the Gwet weighted agreement coefficient (AC2) at hourly time points with at least 2 ratings. Main analyses used Pearson correlations between hourly median EDPC and median EDOR, NEDOCS, and EDWIN. Linear mixed-effects models additionally accounted for repeated ratings within clinicians and clustering at the same hourly time point. Time-lagged correlations were exploratory. Statistical significance was set at α=.05.

Results: Across 1124 scheduled on-duty assessment opportunities, clinicians completed 1100 individual EDPC ratings (response rate 1100/1124, 97.86%). These ratings yielded 449 hourly aggregated EDPC observations, with a median of 2 (IQR 2-3; range 1-4) clinicians contributing ratings per hour. Interrater reliability was substantial (AC2=0.742, 95% CI 0.712-0.772). Objective

XSL•FO
RenderX

metrics were derived from 1846 ED visits. Hourly median EDPC correlated with median EDOR (r=0.542, 95% CI 0.473-0.604; P<.001) and NEDOCS (r=0.483, 95% CI 0.409-0.551; P<.001), but not EDWIN (r=0.060, 95% CI –0.033 to 0.152; P=.20).

Conclusions: We deployed smartwatch-based EMA to capture high-frequency, in situ clinician perceptions during active shifts, potentially reducing susceptibility to retrospective recall bias while minimizing the spatial constraints of traditional survey methods. Hourly median EDPC was moderately associated with EDOR and NEDOCS in this single tertiary ED during an exceptional staffing crisis. The association with EDWIN was less consistent and became significant only after adjustment for time-of-day patterns. The high response rate (1100/1124, 97.9%) supports the short-term feasibility of repeated wearable-based data collection during routine clinical shifts. These findings support further investigation of EDPC as a complementary operational measure but do not establish that it directly measures cognitive workload or predicts subsequent crowding.

(JMIR Mhealth Uhealth 2026;14:e103317) doi: 10.2196/103317

KEYWORDS

clinician perception; crowding; ecological momentary assessment; EDOR; EDWIN; emergency department; NEDOCS; perceived crowding; real-time assessment; wearable device

Introduction

Problem

Emergency department (ED) crowding compromises the quality of care and increases clinician workload worldwide [1-3]. Crowding occurs when the demand for emergency services exceeds the available capacity of the ED or hospital resources [4], reflecting the combined impact of patient volume, acuity, bed availability, and staffing levels [2,3,5]. An aging population is bringing growing numbers of older patients whose multimorbidities and nonspecific symptoms prolong their length of stay [3,6,7]. Boarding of critically ill patients is a major driver of crowding [8-10], forcing departments to deliver intensive inpatient-level care. Despite this escalating burden, how the clinicians managing it experience crowding in real time remains poorly understood.

Review of Relevant Scholarship

Several metrics quantify different aspects of crowding [11,12]. Recent reviews, however, have emphasized that the performance and interpretability of ED crowding indices vary across hospital types and operational contexts [13], and that no universally accepted gold standard has been established [11-15]. These indices capture structural aspects of crowding, such as bed occupancy, boarding time, and acuity-weighted patient-to-staff ratios [16-18], rather than how clinicians subjectively experience crowding in real time.

Existing indices differ in what they operationalize. The National Emergency Department Overcrowding Score (NEDOCS), the most widely used metric [11,12,15], incorporates multiple operational components, including the number of patients currently in the ED, the number of admitted patients waiting for hospital beds, the total number of hospital beds, the longest waiting time from triage to bed assignment, the longest boarding time, and the number of patients receiving mechanical ventilation [17]. Based on these parameters, NEDOCS categorizes ED status into 6 levels ranging from “Not Busy” (0-20) to “Dangerously Overcrowded” (180-200) [11,12,17]. The ED occupancy ratio (EDOR) also focuses on physical occupancy, comparing active patients with operational beds [16]. Whereas EDOR and NEDOCS are based on patient counts and bed capacity, the Emergency Department Work Index (EDWIN) incorporates patient acuity and staffing levels to reflect the intensity of care required [18].

Clinicians’ subjective perception of crowding has received relatively limited empirical attention [11,14], despite growing interest in how it relates to these objective crowding measures [12,15]. Notably, perceived and objective crowding can diverge: in the original NEDOCS validation study, physicians reported feeling less rushed once the ED became severely overcrowded and all beds were full [17], suggesting that objective crowding may not align with clinicians’ subjective perceptions. This divergence has been repeatedly documented across diverse ED settings [19,20].

In this study, ED perceived crowding (EDPC) refers to clinicians’ real-time subjective perception of ED crowding, distinct from operational crowding, clinical workload, and cognitive workload. These constructs are related but not interchangeable [19], and this study directly measured only clinicians’ perceptions of ED crowding. To operationalize EDPC, we adapted the 6 anchors of the original NEDOCS outcome scale ranging from “Not Busy (0-20)” to “Dangerously Overcrowded (180-200)” as a real-time response scale [17].

Where prior work has captured clinician perception, it has typically relied on convenience samples of discrete time blocks [12,14,17] rather than systematic, repeated measurements across a full shift. In this fast-paced, highly uncertain environment [21], real-time assessment of EDPC may provide complementary operational insights. The role of frontline clinicians is critical to timely care delivery [22,23]; operational metrics alone, however, cannot capture their direct experience of crowding. Recent advances in mobile data collection, particularly ecological momentary assessment (EMA), offer a practical means to capture real-time perception while limiting post hoc rationalization or recall bias [24-26]. However, it remains unclear how clinicians’ real-time perceptions of crowding correspond to established objective crowding measures when both are assessed concurrently at high temporal resolution.

Hypothesis, Aims, and Objectives

To address this knowledge gap, we used a smartwatch-based EMA system to collect real-time EDPC data during clinical shifts. The primary aim of this study was to evaluate the concurrent associations between hourly median EDPC and the

XSL•FO
RenderX

corresponding hourly median values of the established objective crowding metrics: EDOR, NEDOCS, and EDWIN. Secondary analyses examined temporal associations between EDPC and these objective metrics. We also assessed interrater agreement among clinicians as a measurement property of EDPC and described response completion rates to characterize the feasibility of smartwatch-based EMA data collection.

Methods

Study Design

This prospective observational study combined retrospectively extracted electronic health record (EHR) data, used to derive the objective metrics, with prospectively collected clinician-reported crowding measures.

Setting

We recruited clinicians and obtained informed consent from November 12 to November 28, 2024. Data were collected in the ED of Samsung Medical Center, a tertiary referral hospital in Seoul, South Korea, from November 25, 2024, at 7 AM to December 13, 2024, at 11:59 PM, comprising 449 hourly observation time points. The study site has 1766 inpatient beds, including 67 ED beds, and manages approximately 65,000-70,000 ED visits annually. During the study period, ED operations were affected by a national medical workforce shortage following the mass resignation of residents and fellows [27,28]. Consequently, attending physicians delivered all ED care without junior medical staff, and the department operated 40 beds rather than its usual 67.

Participants

Inclusion and Exclusion

Eligible participants were attending physicians and charge nurses providing clinical care in the ED (central or acute care areas) during the study period; residents and fellows were not eligible because they were not involved in ED care at that time.

Sampling Procedures

We identified eligible clinicians from the duty roster, informed them about the study, and recruited them by voluntary nonprobability sampling; those who declined were not approached further, and written informed consent was obtained from all enrolled participants.

Of the 37 eligible clinicians, 28 (75.7%) participated, including 5 (35.7%) of 14 attending physicians and all 23 (100%) charge nurses; the remaining 9 attending physicians did not participate. Accordingly, the enrolled clinicians were not intended to constitute a representative sample of ED staff.

The observation period extended continuously from November 25, 2024, at 7 AM through December 13, 2024, at 11:59 PM, yielding 449 consecutive hourly assessment time points: 17 time points from 7 AM to 11 PM on November 25 and 24 time points per day for the subsequent 18 days. Rather than imposing artificial rater quotas or subsampling responses, we collected all responses from participating clinicians on duty. Based on the department's staffing model, each shift was staffed by at most 2 charge nurses and 1 to 3 attending physicians. Because duty periods could overlap across successive shifts, up to 3 charge nurses could be concurrently on duty during some hourly assessment periods. Because enrollment was voluntary, the number of participating clinicians on duty simultaneously varied across hourly assessment periods. Individual ratings within each hourly interval were aggregated into an hourly median EDPC value.

Participant Characteristics

Clinician demographic characteristics (age, sex, and years of clinical experience) and characteristics of ED patient visits during the study period (age, sex, triage acuity, and disposition) are reported in the Results section (Participant and Patient Characteristics).

Sample Size, Power, and Precision

No a priori sample size calculation was performed. The study size was determined by the prespecified data collection period and the number of eligible clinicians who volunteered. A total of 28 clinicians (5 physicians and 23 nurses) were enrolled, and 1100 individual EDPC ratings were obtained across 449 hourly observation time points. The precision of the estimated associations was assessed using 95% CIs. Because consecutive hourly observations were serially correlated, the 449 hourly time points were not considered statistically independent.

Assessments

Measures and Covariates

The primary outcome was EDPC, representing clinicians' perceived level of ED crowding. A raw EDPC entry was defined as a single clinician-reported rating recorded during an hourly assessment period. We then computed an hourly EDPC value by taking the median of all available entries within each hour. We chose the median value as the main hourly summary to provide a consistent measure of central tendency across hours with varying numbers of raters and to limit the influence of extreme ratings in hours with 3 or more raters. We defined the EDPC categories by anchoring scores to the 6 established NEDOCS severity categories (Not Busy, Busy, Extremely Busy, Overcrowded, Severely Overcrowded, and Dangerously Overcrowded) [17]. As the EDPC was recorded on a 0-200 numeric scale, this categorization followed the same point-based thresholds as the NEDOCS framework. This anchoring allowed direct comparison of clinicians' subjective perceptions with the objective operational thresholds defined by that framework.

The objective metrics EDOR, NEDOCS, and EDWIN [16-18] were calculated at a 1-minute temporal resolution using their respective published formulas implemented directly in R and evaluated as predictors of EDPC. These metrics represent different dimensions of ED crowding. Detailed formulas and variable definitions for all metrics are provided in Multimedia Appendix 1. Physician staffing data for the EDWIN calculation were taken from the duty roster and treated as time-varying. Physician shifts were scheduled as follows: *day* (D), 7:30 AM to 8:30 PM with 2 physicians; *evening* (E), 12 noon to 10 PM with 1 physician; and *night* (N), 8:30 PM to 7:30 AM with 2 physicians. When summarized by hourly assessment periods, under the usual staffing pattern, 2 physicians were on duty from

XSL•FO
RenderX

8 AM to 11:59 AM, 3 from 12 noon to 9:59 PM, and 2 from 10 PM to 7:59 AM the following morning. The only exceptions occurred during the N shifts beginning on November 26 and 27, 2024, when 1 physician rather than 2 was assigned to the N shift. We calculated EDWIN using the actual number of physicians scheduled on duty at each time point according to the duty roster, including these exceptions. We aggregated the objective crowding metrics into hourly summary statistics (mean, median, minimum, and maximum) for direct comparison with the hourly EDPC data.

To characterize the study population, we extracted demographic and clinical variables for each visit, including age, sex, acuity level according to the Korean Triage and Acuity Scale (KTAS), vital signs, and clinical disposition (categorized as discharge, hospital admission, transfer, death, or leaving without being seen). The latter category comprised patients who completed triage but left before entering the ED treatment area or being seen, either returning home or seeking care at another hospital. Because they remained in the ED census while physically present after triage, they contributed to the objective crowding metrics during that period.

Instrumentation

Raw EDPC entries were collected using a smartwatch-based system designed to support EMA [24-26] and to enable real-time assessment of clinicians' perceptions during active clinical shifts. Assessing perceptions in the moment was intended to reduce the recall bias associated with retrospective assessment. The research team developed a custom smartwatch application to support repeated real-time assessment within the clinical workflow. The 0-200 NEDOCS scale and its 6 crowding categories were adopted as reference anchors, giving clinicians an intuitive and standardized framework for rating their perceived crowding.

Four study smartwatches (Galaxy Watch 4 Classic; Samsung Electronics) running the custom application described above were provided for EDPC assessments. Participating clinicians wore the study smartwatches only during their on-duty shifts, and the devices were shared across successive shifts rather than assigned to individual participants. Participants did not use their personal smartwatches.

No formal pilot-testing of the EDPC instrument was conducted before the study.

Quality of Measurements

Before data collection, participants received a brief orientation on using the smartwatch application and the EDPC assessment procedure. We demonstrated the interface, including how to launch the application, enter a rating, and modify one already entered. Participants were informed that the interface displayed the 6 crowding categories based on the NEDOCS framework, with corresponding color-coded score ranges, and were instructed to use these references to rate perceived crowding once each hour.

Psychometrics

The interrater reliability of the raw EDPC entries was assessed using the Gwet weighted agreement coefficient (AC2) with ordinal weighting [29] at 447 hourly time points with at least 2 observed ratings. All available clinician ratings at each time point were included. We selected this approach because the number and composition of raters varied across hourly assessments. AC2 estimates with 95% CIs were calculated using the *irrCAC* package in R. As a complementary analysis of the continuous EDPC ratings, we fitted a crossed random-effects variance-component model with hourly time point and clinician as random intercepts to account for the changing and unbalanced rater composition.

Masking

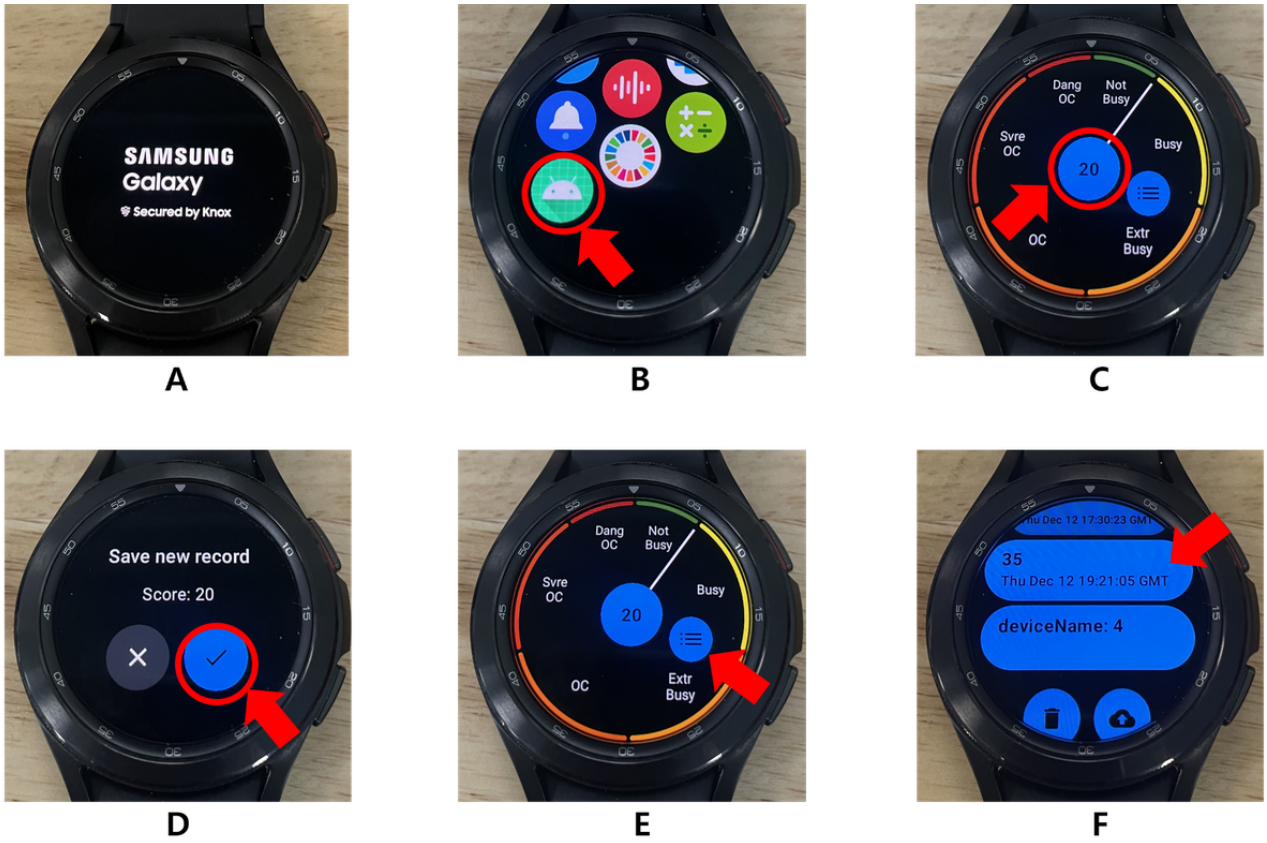
Although the NEDOCS-based categories and score ranges were displayed as reference anchors, participants could not view the actual real-time NEDOCS score while recording EDPC. Participants also had no access to an ED occupancy dashboard or to calculated operational crowding scores, including EDOR and EDWIN.

Data Collection

We prospectively collected EDPC data from enrolled clinicians using the smartwatch-based EMA system. EDPC assessments were conducted hourly. To reduce missed entries, the application delivered a gentle vibration reminder at 30 minutes past each hour. Participants entered their EDPC ratings using the bezel-controlled smartwatch interface shown in Figure 1, with score selection performed using either the rotating bezel or direct touchscreen interaction, followed by touch-based confirmation. Participants could enter one EDPC rating during each hourly assessment period and could correct an erroneous entry within the same period. Each EDPC response was assigned to the clock hour in which it was recorded, regardless of the timing of the reminder (eg, a response recorded at 11:50 AM was assigned to the 11-11:59 AM hourly period). Entries were not mandatory; therefore, clinicians could skip them during urgent clinical situations. If no response was recorded during an hourly assessment period while a clinician was on duty, that assessment opportunity was considered missing. Off-duty periods were excluded from EDPC assessment opportunities. Prompt-level completion rates were calculated as the number of completed EDPC ratings divided by the number of assessment opportunities during clinicians' on-duty periods. These rates were summarized overall, by professional role, and by individual clinician.

XSL•FO
RenderX

Figure 1. Workflow and user interface of the smartwatch-based application for recording raw emergency department perceived crowding (EDPC) entries (Galaxy Watch 4 Classic). (A) Initial smartwatch home screen. (B) Launch of the EDPC application. (C) Main EDPC input interface where clinicians selected their perceived crowding level on a 0-200 scale using either the rotating bezel or direct touchscreen interaction, with the 6 National Emergency Department Overcrowding Score (NEDOCS) crowding category labels, corresponding score ranges, and color coding displayed as visual references. The abbreviated labels displayed on the smartwatch were Extr Busy (extremely busy but not overcrowded), OC (overcrowded), Svre OC (severely overcrowded), and Dang OC (dangerously overcrowded). (D) Touch-based confirmation screen for saving the selected EDPC score. (E) Option to modify or reenter a previously recorded score. (F) Stored EDPC entry with the corresponding time stamp. The interface reproduced the NEDOCS-based category labels, score thresholds, and corresponding color coding as visual reference anchors; the actual calculated NEDOCS score at the time of assessment was not displayed. The use of these NEDOCS-based visual anchors may have introduced anchoring bias in the clinicians' EDPC ratings.



To compare the EDPC with the validated objective metrics, we extracted retrospective patient-level data for all ED visits during the study period from Samsung Medical Center's Clinical Data Warehouse (DARWIN-C). To ensure the accuracy of crowding assessment, we excluded visits that did not contribute to operational crowding: (1) cases immediately transferred to independent care units (eg, pediatric zones or delivery rooms), (2) visits that did not use ED clinical resources (eg, dead-on-arrival cases or nonclinical visits), and (3) erroneous records identified during data validation. ED and hospital bed capacities used in the calculations were obtained from institutional operational records.

Data Analysis

Data Diagnostics

Missing EDPC ratings were assessed at the individual clinician-rating level. The Little missing completely at random (MCAR) test evaluated whether missingness was completely at random.

As a sensitivity analysis, multiple imputation (MI) was performed for missing EDPC ratings using predictive mean matching to generate 50 imputed datasets. The imputation model included clinician ID, hourly median EDOR, NEDOCS, and EDWIN, sine and cosine terms for time of day, and study day. Although clinician ID and temporal covariates were included as predictors, within-clinician and serial correlation structures were not explicitly modeled. Hourly median EDPC values were recalculated for each imputed dataset, and Pearson correlation coefficients were pooled using Rubin's rules after Fisher *z* transformation.

Analytic Strategy

Clinician characteristics and key study variables are summarized using descriptive statistics. Categorical variables are presented as frequencies and percentages, and continuous variables as means (SD) or medians (IQR), as appropriate.

We performed exploratory descriptive analyses of hourly EDPC and the objective metrics to examine their distribution and diurnal patterns over the study period. We then assessed the association between hourly EDPC and each objective metric at contemporaneous time points.

The main analyses examined contemporaneous Pearson correlations between hourly median EDPC and the hourly median values of EDOR, NEDOCS, and EDWIN; no single

XSL•FO
RenderX

correlation was prespecified as the primary comparison. Exploratory analyses additionally examined Pearson correlations between hourly median EDPC and the minimum, maximum, and mean hourly values of each objective metric. Correlation strength was interpreted according to Cohen conventional benchmarks as weak (<0.3), moderate (0.3-0.5), or strong (≥0.5) [30].

In addition, we analyzed individual clinician-level EDPC ratings with linear mixed-effects models to account for the multilevel structure of the data. In the overall analysis, clinician and hourly time point were included as random intercepts to account for repeated ratings within clinicians and clustering of ratings obtained at the same time point. Profession-stratified analyses were performed separately for nurses and physicians, with clinician included as a random intercept to allow for repeated ratings within individuals.

As a post hoc sensitivity analysis, we calculated Spearman rank correlations using the same hourly summary statistics as in the Pearson analyses, to assess the robustness of the main results to departures from linearity and distributional assumptions. For the median-to-median associations, Spearman rank correlations and scatterplots were also examined overall and separately for physicians and nurses, with least absolute deviations (LAD) regression lines used to visually characterize these associations.

To account for temporal dependence, additional sensitivity analyses used generalized least squares (GLS) regression with first-order autoregressive—AR(1)—errors; sine and cosine terms for time of day simultaneously adjusted for common 24-hour diurnal patterns. Residual autocorrelation was assessed using the Ljung-Box test. In addition, we obtained percentile-based 95% CIs for the original Pearson correlation coefficients from a calendar-day block bootstrap with 10,000 resamples, resampling entire study days with replacement so that within-day temporal dependence was preserved.

As an exploratory analysis, we performed time-lagged correlation analyses by shifting each objective metric relative to hourly EDPC within a lag window of ±2 hours. A positive lag indicated that the objective metric preceded EDPC (eg, lag +1 h compared EDPC at time t with the objective metric measured at time t–1 h), whereas a negative lag indicated that the objective metric followed EDPC (eg, lag –1 h compared EDPC at time t with the objective metric measured at time t+1 h). Lag 0 represented contemporaneous measurements. This window was selected to align with the hourly sampling frequency and to capture short-term associations while avoiding more distant relationships that are less likely to be meaningful. No adjustment for multiple comparisons was applied.

The significance level for all statistical tests was set at α=.05. Analyses were performed using R (version 4.3.2; R Foundation for Statistical Computing) [31].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional review board (IRB) of Samsung Medical Center. Prospective collection of EDPC data from clinicians was approved under IRB number 2024-02-102, and written informed consent was obtained from all participating physicians and nurses before enrollment. Retrospective extraction of operational and clinical data from the hospital's Clinical Data Warehouse was approved under IRB number 2025-02-012, with a waiver of informed consent for the use of deidentified records. Clinician smartwatch responses and clinical data were deidentified using unique study identifiers and stored on encrypted, password-protected institutional servers to protect privacy and confidentiality. Participating clinicians were compensated KRW 3000 (US $2.10) per hour of participation. Compensation was calculated based on each clinician's total working hours during the study period. The mean participation time was 40.5 hours per clinician, corresponding to a mean compensation of approximately KRW 121,393 (US $84.96). The total compensation for all 28 participants was KRW 3,399,000 (US $2379). No identifying information about clinicians or patients is presented in the manuscript figures, user interface images, or supplementary materials.

Results

Participant Flow

Among 37 eligible clinicians, 28 (75.7%) participated in the study, including 5 (35.7%) of 14 eligible attending physicians and all 23 (100%) eligible charge nurses. Across the 449 continuous hourly observation intervals (November 25, 2024, 7 AM to December 13, 2024, 11:59 PM), clinicians completed 1100 of 1124 scheduled on-duty EDPC assessments, yielding an overall response rate of 97.86% and a missed-assessment rate of 2.14%. The completion rate was 99.45% (897/902) among nurses and 91.44% (203/222) among physicians (Table 1). The 24 missed assessments were concentrated in 7 of the 28 participants (range 1-15 missed prompts; Table S1 in Multimedia Appendix 1). Little test rejected the assumption that these data were MCAR (χ^2_3=12.3; *P*=.006). In the MI sensitivity analysis, the pooled correlations were similar to the observed-data estimates for EDOR (*r*=0.538, 95% CI 0.468-0.602; *P*<.001), NEDOCS (*r*=0.480, 95% CI 0.405-0.549; *P*<.001), and EDWIN (*r*=0.063, 95% CI –0.030 to 0.155; *P*=.19; Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Medical staff characteristics and emergency department perceived crowding (EDPC) assessment compliance (n=28)[a].

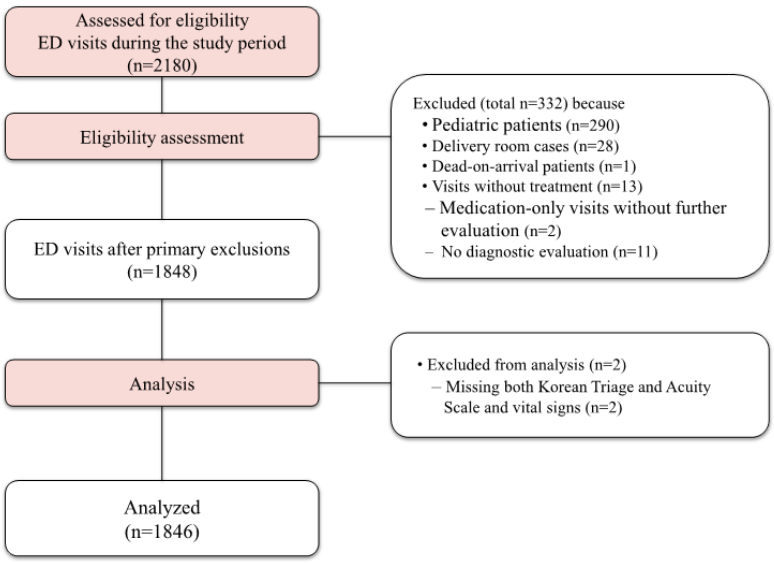
	Physician (n=5)	Nurse (n=23)	Total (n=28)
Demographics			
Age (years), mean (SD)	42.0 (6.7)	35.1 (5.0)	36.3 (5.9)
Sex, n (%)			
Female	3 (60.0)	18 (78.3)	21 (75.0)
Male	2 (40.0)	5 (21.7)	7 (25.0)
Total clinical experience (years), mean (SD)	16.2 (7.0)	10.8 (4.4)	11.8 (5.3)
EDPC assessment compliance			
Scheduled on-duty opportunities, n	222	902	1124
Completed EDPC ratings, n (%)	203 (91.44)	897 (99.45)	1100 (97.86)
Missed assessments, n (%)	19 (8.56)	5 (0.55)	24 (2.14)

[a]Clinical experience represents total years of practice rather than emergency department–specific experience. Participating physicians were attending physicians providing clinical care in the emergency department, and participating nurses were charge nurses assigned to the central or acute care areas.

Across all 449 scheduled hours, the median number of clinicians contributing ratings per hour was 2 (IQR 2-3; range 1-4). Specifically, 2 (0.4%) hours had 1 rater, 255 (56.8%) hours had 2 raters, 180 (40.1%) hours had 3 raters, and 12 hours (2.7%) had 4 raters. Hourly EDPC values were calculated as the median of all available clinician ratings within each time point, resulting in 449 aggregated EDPC observations. All 449 hourly time points had complete corresponding objective metric data and were retained for the association analyses. The 447 hourly time points with ratings from at least 2 clinicians were included in the interrater reliability analysis. The remaining 2 time points, each with only 1 clinician rating, were not eligible for that analysis but were retained for the association analyses.

During the study period, 1848 adult ED visits met the eligibility criteria. Two visits were excluded because both KTAS and vital sign data were missing, resulting in a final analytic sample of 1846 ED visits (Figure 2). “Visits without treatment” referred to visits in which patients did not undergo standard ED evaluation and consequently were not assigned a KTAS level. These included medication-only visits without further evaluation (n=2) and visits with no diagnostic evaluation (n=11). This group was distinct from patients categorized as “Left without being seen,” who completed triage but left before entering the ED treatment area or receiving treatment. Because visits without treatment were neither triaged nor included in the ED census, they were excluded from the study population and did not contribute to the objective crowding metrics.

Figure 2. Flow diagram of emergency department (ED) visits included in the study. This prospective observational study was conducted at a tertiary ED in Seoul, South Korea, from November 25 to December 13, 2024. The diagram shows ED visits assessed for eligibility, exclusions and their reasons, and the final analytic sample.



XSL•FO
RenderX

Recruitment

In total, 28 clinicians participated in this study, including 5 (17.9%) physicians and 23 (82.1%) nurses (Table 1).

Statistics and Data Analysis

Participant and Patient Characteristics

The mean age of the participating clinicians was 36.3 (SD 5.9) years, with a mean clinical experience of 11.8 (SD 5.3) years. Clinical experience varied by professional role, with an average of 16.2 (SD 7.0) years for physicians and 10.8 (SD 4.4) years for nurses. Overall, 75% (21/28) of the participants were women.

The mean patient age was 57.7 (SD 20.4) years, and 52.1% (961/1846) of visits involved male patients. KTAS level 3 was the most common acuity category (1070/1846, 58%), followed by levels 4 (397/1846, 21.5%) and 2 (311/1846, 16.8%). Most patients were alert upon arrival (1785/1846, 96.7%). Regarding disposition, 31.9% (589/1846) of visits resulted in discharge, 25.2% (466/1846) in hospital admission, and 38.8% (717/1846) in leaving without being seen. Patient characteristics and outcomes varied substantially across KTAS levels. KTAS 1-2 visits involved older and more clinically unstable patients, with high admission rates (43.5%-47.6%) and a mortality rate of 39.1% (9/23) among KTAS 1 visits. In contrast, KTAS 4-5 visits involved younger patients and substantially higher rates of leaving without being seen (71.3%-91.1%; Table S3 in Multimedia Appendix 1).

Reliability of EDPC Measurements

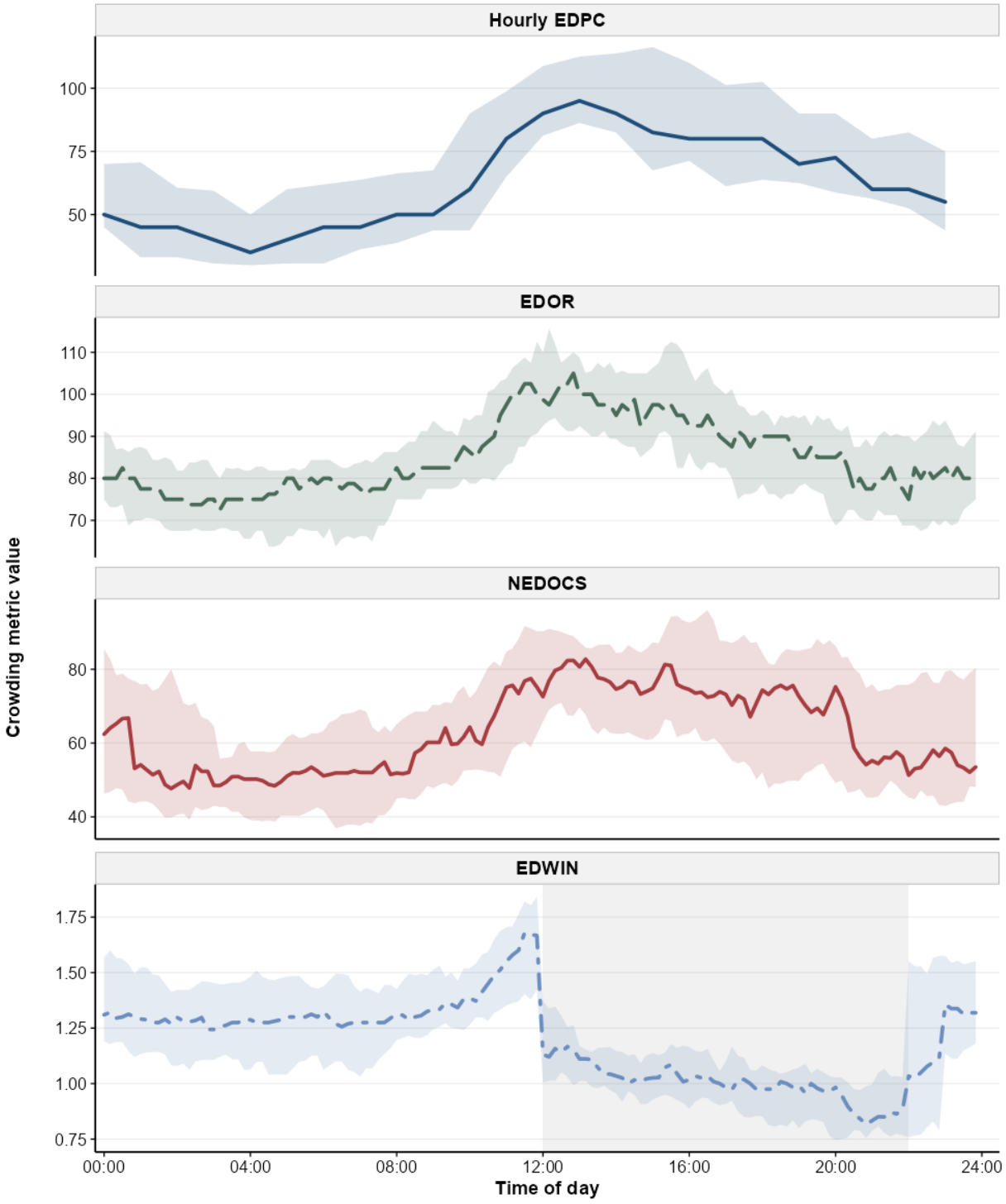
Overall and profession-specific EDPC assessment completion rates are presented in Table 1, and individual clinician-level assessment completion data are provided in Table S1 in Multimedia Appendix 1. EDPC ratings from at least 2 clinicians were available at 447 (99.6%) of the 449 hourly time points. Interrater reliability analysis demonstrated substantial agreement for raw EDPC scores (AC2=0.742, 95% CI 0.712-0.772) and almost perfect agreement after conversion to NEDOCS categories (AC2=0.811, 95% CI 0.790-0.833). In the complementary variance-component analysis of continuous EDPC ratings, 25.2% of the total variance was attributable to hourly time points, 23.5% to between-clinician differences, and 51.3% to residual variation.

Diurnal Patterns of Hourly EDPC and Objective ED Crowding Metrics

The distributions of hourly median EDPC and the hourly median values of EDOR, NEDOCS, and EDWIN are summarized in Table S4 in Multimedia Appendix 1. Both the perceived and objective crowding measures showed clear diurnal variation (Figure 3). The hourly EDPC across all raters increased after 7 AM, peaked at 1 PM, and gradually declined thereafter. Temporal patterns varied by professional role: nurse-reported hourly EDPC peaked at 1 PM, whereas physician-reported hourly EDPC remained elevated between 2 and 4 PM (Figure S1 in Multimedia Appendix 1). These role-specific patterns are presented descriptively and should be interpreted cautiously given the small number of participating physicians.

XSL•FO
RenderX

Figure 3. Diurnal patterns of hourly emergency department perceived crowding (EDPC) and objective emergency department crowding metrics. Data were collected at a tertiary ED in Seoul, South Korea, from November 25 to December 13, 2024. Lines represent median values calculated from observations pooled within each 10-minute time-of-day bin across the study period. Shaded bands represent the IQR (25th-75th percentile) of day-specific median values within each corresponding bin, indicating day-to-day variability. In the Emergency Department Work Index (EDWIN) panel, the vertical gray band marks the period when 3 physicians were usually on duty (12:00-21:59), and the unbanded areas mark the periods when 2 were (08:00-11:59 and 22:00-07:59 the following day). EDOR: emergency department occupancy rate; NEDOCS: National Emergency Department Overcrowding Score; EDWIN: Emergency Department Work Index.



Objective indices showed overall diurnal patterns similar to those of hourly EDPC, but with distinct features across metrics (Figure 3). NEDOCS showed 2 local peaks at approximately 1 and 3:30 PM. EDOR showed a late-morning peak at approximately 11:30 PM, followed by a generally declining pattern with some variability. EDWIN, in contrast, followed a different profile: a primary peak in the late morning (approximately 11:30 PM), then a rapid drop and a secondary peak in the late evening (approximately 11:20 PM).

XSL•FO
RenderX

Associations Between EDPC and Objective ED Crowding Metrics

Table 2 shows correlations between hourly EDPC and summary statistics of the 3 objective metrics. All Pearson correlation analyses were based on 449 paired hourly observations. EDPC correlated strongly with EDOR (r=0.53-0.56) and moderately to strongly with NEDOCS (r=0.47-0.51), whereas EDWIN showed no significant correlation.

Table 2. Pearson correlation between hourly emergency department perceived crowding and hourly summary statistics of objective metrics[a].

	Pearson *r* (95% CI)	*P* value
EDOR[b]		
Minimum	0.527 (0.457 to 0.591)	<.001
Maximum	0.563 (0.496 to 0.623)	<.001
Mean	0.545 (0.477 to 0.607)	<.001
Median	0.542 (0.473 to 0.604)	<.001
NEDOCS[c]		
Minimum	0.473 (0.398 to 0.542)	<.001
Maximum	0.510 (0.439 to 0.576)	<.001
Mean	0.492 (0.419 to 0.559)	<.001
Median	0.483 (0.409 to 0.551)	<.001
EDWIN[d]		
Minimum	0.014 (–0.079 to 0.106)	.77
Maximum	0.090 (–0.003 to 0.181)	.06
Mean	0.052 (–0.041 to 0.144)	.27
Median	0.060 (–0.033 to 0.152)	.20

[a]Each Pearson correlation analysis was based on 449 paired hourly observations.

[b]EDOR: emergency department occupancy ratio.

[c]NEDOCS: National Emergency Department Overcrowding Score.

[d]EDWIN: Emergency Department Work Index.

Post hoc sensitivity analyses using Spearman rank correlations yielded results consistent with the Pearson analyses. EDPC showed significant positive rank correlations with all summary statistics of EDOR (ρ=0.53-0.57; all P<.001) and NEDOCS (ρ=0.47-0.52; all P<.001), whereas EDWIN showed no significant correlations (ρ=–0.02 to 0.08; P=.09-.63; Table S5 in Multimedia Appendix 1). Scatterplots of hourly median EDPC against the hourly median EDOR, NEDOCS, and EDWIN values, with LAD regression lines and Spearman correlation coefficients, are presented overall and separately for physicians and nurses (Figure S2 in Multimedia Appendix 1).

In clinician-level multilevel analyses using all 1100 individual EDPC ratings and accounting for clustering by both clinician and hourly time point, EDPC was significantly associated with median EDOR (β=.966, 95% CI 0.840-1.091; P<.001) and median NEDOCS (β=.651, 95% CI 0.558-0.745; P<.001), but not with median EDWIN (β=3.865, 95% CI –1.761 to 9.490; P=.18; Table 3). Profession-stratified analyses showed the same pattern of statistical significance in nurses and physicians, with EDOR and NEDOCS significantly associated with EDPC and EDWIN showing no significant association in either professional group. The estimated EDWIN coefficient was positive for nurses (β=4.114) and negative for physicians (β=–.825), although neither association was statistically significant (Table 3).

Additional temporal sensitivity analyses showed that EDOR and NEDOCS remained significantly associated with hourly EDPC after simultaneously accounting for serial correlation and common time-of-day patterns using GLS AR(1) models (both P<.001). EDWIN was also significantly associated with EDPC after adjustment for both temporal components (P=.004; Table S6-A in Multimedia Appendix 1). Calendar-day block bootstrap analyses of the original Pearson correlations yielded 95% CIs excluding zero for EDOR (r=0.542, 95% CI 0.417-0.661) and NEDOCS (r=0.483, 95% CI 0.389-0.581), but not for EDWIN (r=0.060, 95% CI –0.165 to 0.234; Table S6-B in Multimedia Appendix 1).

In exploratory time-lagged correlation analyses, EDOR consistently showed positive correlations with EDPC across the ±2-hour lag window. NEDOCS followed a similar pattern, whereas EDWIN generally showed weaker associations across the lag conditions. The temporal patterns are presented in Figure S3 in Multimedia Appendix 1, and complete results, including lag direction, number of paired observations, correlation coefficients, 95% CIs, and unadjusted P values, are provided in Table S7 in Multimedia Appendix 1. P values were not adjusted for multiple comparisons because these analyses were exploratory.

XSL•FO
RenderX

Table 3. Multilevel analyses of associations between individual clinician-level emergency department perceived crowding ratings and objective emergency department crowding metrics.

Objective metric	Ratings, n	Clinicians, n	Hourly time points, n	β (95% CI)	SE	*P* value
Overall multilevel analysis[a]						
EDOR[b]	1100	28	449	.966 (0.840 to 1.091)	0.064	<.001
NEDOCS[c]	1100	28	449	.651 (0.558 to 0.745)	0.048	<.001
EDWIN[d]	1100	28	449	3.865 (–1.761 to 9.490)	2.870	.18
Profession-stratified multilevel analysis[e]						
Nurse						
EDOR	897	23	449	.985 (0.853 to 1.116)	0.067	<.001
NEDOCS	897	23	449	.641 (0.545 to 0.738)	0.049	<.001
EDWIN	897	23	449	4.114 (–1.389 to 9.616)	2.807	.14
Physician						
EDOR	203	5	191	.810 (0.620 to 1.000)	0.097	<.001
NEDOCS	203	5	191	.693 (0.549 to 0.838)	0.074	<.001
EDWIN	203	5	191	–.825 (–8.251 to 6.602)	3.789	.83

[a]Linear mixed-effects models were fitted using individual clinician-level emergency department perceived crowding ratings as the outcome and the median value of each objective crowding metric at the corresponding hourly time point as the predictor. Overall models included random intercepts for both clinician and hourly time point to account for repeated ratings within clinicians and clustering within the same time point. β represents the unstandardized regression coefficient.

[b]EDOR: emergency department occupancy ratio.

[c]NEDOCS: National Emergency Department Overcrowding Score.

[d]EDWIN: Emergency Department Work Index.

[e]Profession-stratified linear mixed-effects models included clinician as a random intercept to account for repeated ratings within clinicians. A separate hourly time point random intercept was not retained in the stratified models because within-hour replication was sparse in the physician subgroup. β represents the unstandardized regression coefficient.

Discussion

Support of Original Hypotheses (Primary Findings)

Clinician-perceived crowding, measured as the hourly EDPC, was associated with established objective crowding measures, although the magnitude and consistency of these associations differed across metrics. EDPC showed consistent associations with EDOR and NEDOCS, whereas its association with EDWIN depended on whether common time-of-day patterns were explicitly adjusted for. These findings suggest that clinician-perceived crowding may provide complementary information to objective crowding metrics. The high response rate supports the feasibility of repeated real-time EDPC assessment during routine clinical shifts.

Similarity of Results

Alignment of Hourly EDPC With Objective Crowding Measures

Hourly EDPC correlated moderately to strongly with occupancy-based measures, such as EDOR [15,16] and NEDOCS [17]. Clouston et al [12] demonstrated that ED occupancy, despite its simplicity, had a predictive accuracy similar to that of complex crowding scores such as NEDOCS and was useful for real-time tracking. Furthermore, recent before-and-after evidence indicates that transition units can reduce boarding and NEDOCS-defined crowding, illustrating that occupancy- and boarding-based measures may be particularly responsive to changes in hospital access block [32]. Our findings are consistent with this observation. Occupancy-based metrics may align well with EDPC because clinicians managing a growing number of patients directly experience the physical saturation of treatment spaces; however, this study cannot establish the specific mechanism underlying this association.

Hourly EDPC showed significant positive correlations with EDOR and NEDOCS across all 4 hourly summary statistics (minimum, maximum, mean, and median). Descriptively, correlations of EDPC with the maximum values of EDOR and NEDOCS were slightly higher than those with the corresponding mean or median values. However, these differences were small, and the dependent correlation coefficients were not formally compared; therefore, this observation should not be interpreted as evidence that clinician-perceived crowding is more strongly associated with peak values. Prior work has suggested that traditional crowding metrics may not fully capture the dynamic, rapidly fluctuating nature of ED operations [33], and that the relationship between objective crowding and clinician perception may be nonlinear [19]. These observations motivate further study of whether within-hour fluctuations or peak crowding conditions disproportionately shape clinician perceptions. Although we examined both median and peak crowding metrics,

XSL•FO
RenderX

this study cannot determine which measure best represents clinicians' subjective experience.

Comparison With Previous Real-Time Perception Studies

This study builds on the work of Boverio et al [19], who demonstrated the potential of real-time subjective assessments using EMA. Boverio et al [19] used smartphone text messages twice daily, a method limited by low measurement frequency and by the need to use a smartphone during clinical workflows, whereas our study used a smartwatch-based application with hourly prompts. Subtle wrist-vibration reminders prompted clinicians to complete hourly EDPC assessments through the application during routine clinical shifts. This approach was intended to reduce interruption and recall bias relative to retrospective, single–time point assessments, although these potential benefits were not directly measured and require further evaluation.

Patient-centered research further highlights discrepancies between objective measures and subjective perceptions of ED crowding. Wang et al [34] examined patient-perceived crowding at the end of ED encounters and observed variable agreement with objective indices. However, their study relied on a single–time point measurement, whereas our study captures clinicians' perceptions repeatedly throughout their clinical shifts, allowing for a more dynamic assessment.

Schippers et al [35] developed an AI-driven predictive model using routinely collected operational parameters to predict ED crowding, with ambulance diversion status as a binary outcome. In contrast, our study examines perceived crowding across varying levels of crowding through repeated, real-time assessments, rather than focusing on the threshold of ambulance diversion. Our findings demonstrate associations between objective crowding metrics and clinician-reported EDPC, suggesting that these metrics reflect certain aspects of clinicians' experiences of crowding but may not fully capture their subjective perceptions. Future qualitative research could explore discrepancies between operational indicators and clinicians' experiences in greater depth, providing insights into the contextual factors that shape perceived crowding beyond what quantitative measures can capture.

Interpretation

Interrater Agreement and Variability

Weighted AC2 indicated substantial ordinal agreement among clinicians' EDPC ratings (AC2=0.742). However, the complementary variance-component analysis showed that 25.2% of the total variance was attributable to hourly time points, 23.5% to between-clinician differences, and 51.3% to residual variation. These findings reflect different aspects of agreement and variability. Ordinally weighted AC2 allows partial agreement between nearby scores, whereas the variance-component analysis quantifies different sources of variability in the continuous ratings. Taken together, these findings suggest partial concordance in perceived crowding rather than interchangeability of individual clinician ratings.

Temporal Robustness and Interpretation of Associations

The associations of EDPC with EDOR and NEDOCS remained consistent across analyses accounting for serial correlation, common time-of-day patterns, and within-day temporal dependence, supporting their robustness to the temporal structure of the data. In contrast, the results for EDWIN were less consistent across the temporal sensitivity analyses. Although the original Pearson correlation and calendar-day block bootstrap analysis did not show a significant association between EDPC and EDWIN, a significant association emerged in models that explicitly adjusted for time-of-day patterns. These findings indicate that the EDPC-EDWIN association was sensitive to the adjustment for diurnal variation and should therefore be interpreted cautiously.

Occupancy-based metrics showed more consistent associations with hourly EDPC than EDWIN did. Notably, the diurnal profile of EDWIN diverged in shape from that of EDPC, EDOR, and NEDOCS: whereas the latter 3 measures peaked in the late morning to early afternoon, EDWIN exhibited an additional secondary peak in the late evening (approximately 11:20 PM) that was not mirrored by a corresponding rise in EDPC. This divergence in diurnal shape, rather than the shared presence of diurnal variation per se, may partly explain the inconsistency of the EDWIN-EDPC association across analytic approaches.

This finding must also be interpreted within the operational context of this study. Although EDWIN has the theoretical advantage of incorporating patient acuity [18], it also incorporates physician staffing, which varied over time during the study period. During a recent disruption in the physician workforce in South Korea [27], only 1 to 3 attending physicians delivered ED care, and the number of operational beds was reduced from 67 to 40. Therefore, the EDWIN findings should be interpreted in the context of this atypical staffing environment. Because EDWIN incorporates physician staffing and bed capacity directly into its formula [18], fluctuations in these structural components may influence the metric independently of clinicians' perceived crowding. This sensitivity analysis suggests that the inconsistent unadjusted association between EDWIN and EDPC in this study may be related to the exceptional staffing context and the staffing-sensitive construction of the metric; however, this study cannot establish that staffing variability was the sole or primary explanation. Considering that EDWIN aligned with staff assessments in its original development study and has been used in other settings [36], our findings should not be interpreted as evidence that EDWIN is not generally useful. Instead, staffing-sensitive metrics such as EDWIN should be applied with caution in smaller EDs and in those that adjust staffing in response to crowding.

In exploratory time-lagged analyses, EDPC showed similar correlations with objective crowding measures at lag 0 and at both +1-hour and –1-hour lags, indicating similar associations when the objective metrics preceded or followed EDPC by 1 hour. Because lagged correlations can arise from autocorrelation and shared diurnal trends rather than a genuine predictive relationship, and because predictive performance was not assessed, this finding should be interpreted as an exploratory

XSL•FO
RenderX

temporal association rather than evidence that EDPC functions as a leading indicator of crowding. Clinicians may detect early cues of worsening crowding, but confirmation of any predictive relationship would require dedicated time-series and prospective prediction analyses.

Role-Specific Differences in Perceived Crowding

Descriptively, nurse- and physician-reported EDPC showed different peak timing across the day (Figure S1 in Multimedia Appendix 1). However, in the profession-stratified multilevel analyses, EDOR and NEDOCS were significantly associated with individual EDPC ratings in both nurses and physicians, whereas EDWIN was not significantly associated with EDPC in either group. Thus, despite the descriptive differences in diurnal patterns, the pattern of associations with the objective crowding metrics was similar across professions. These findings should be interpreted cautiously given the small number of participating physicians (n=5) and require confirmation in larger, more balanced samples. Previous studies have shown that perceived crowding differs significantly across professional roles in the ED, including among nursing roles [19]. Because our study included only charge nurses and attending physicians, future research could incorporate other clinical roles, administrative staff, and patients to examine differences in perceived crowding across stakeholder groups and explore the factors underlying these differences.

Generalizability

The applicability of the assessment approach and the observed associations may vary across settings. Hourly smartwatch-based reporting could be adapted to other EDs, although the short study period limits conclusions about sustained participation and acceptability in routine practice. The magnitude and pattern of associations between EDPC and objective metrics may also vary according to staffing models, operational capacity, and patient populations. In particular, the attending-only staffing model and reduced bed capacity during the exceptional workforce disruption observed in this study may limit generalizability to EDs operating under routine conditions. Multicenter studies with longer observation periods under routine operating conditions are needed to evaluate sustained participation and the consistency of these associations across settings.

Implications

Clinical Importance of Clinician-Perceived Crowding

These findings have important implications for ED care delivery. Previous studies have consistently shown that ED crowding is associated with an increased length of stay, delays in care, reduced care quality, and poor patient safety outcomes [1,3]. Crowding arises when demand for services exceeds a department's available resources [4], and ED clinicians play a critical role in timely patient assessment [22,23], initiation of time-sensitive treatment [37], and recognition of clinical deterioration [38]. Prior literature has also characterized ED clinicians' workload as involving substantial cognitive demands and time pressure that are not fully captured by structural factors such as bed and patient numbers [39].

Objective metrics often assume static patient conditions, whereas the ED environment and patient status can change rapidly [40]. Real-time monitoring tools have been proposed to better capture these fluctuations [19,33]. In this context, hourly EDPC may provide complementary information about clinicians' real-time experience of crowding that is not captured by structural metrics alone; whether this information reflects cognitive load, safety risk, or other unmeasured dimensions of operational burden was not evaluated in this study and remains an open question for future research.

Methodological Contributions and Future Directions

The high overall response rate of 97.9% (1100 of 1124 expected on-duty assessments) supports the feasibility of this approach for frequent EDPC assessments during routine clinical shifts. This finding indicates that clinicians could complete hourly ratings consistently during the study period. However, acceptability, sustainability, and documentation burden were not directly measured relative to other assessment methods and require further evaluation. Documentation burden, the increased time and effort required to record care, is a well-known contributor to professional burnout syndrome [41,42]. Moreover, many health care technologies are abandoned when they introduce excessive complexity or demand substantial changes to established clinical routines [43]. Although the smartwatch interface was designed to minimize this burden through score selection using either the rotating bezel or direct touchscreen interaction, followed by touch-based confirmation, this study did not directly assess whether it achieved this relative to other data collection methods.

Importantly, the implementation did not require each participant to own or be assigned an individual smartwatch. Study devices were shared across successive shifts, so the number required depended on how many participating clinicians were concurrently on duty rather than on total enrollment. This device-sharing approach may reduce hardware requirements and remove the need for clinicians to own compatible devices. Nevertheless, the feasibility, cost, and scalability of this approach in larger or differently staffed EDs were not evaluated in this study and require further investigation.

Despite these methodological advantages, our approach did not entirely overcome the interruptive nature of active EMA. Raw EDPC entries may have been missed more frequently during periods when clinicians were too busy to record them [44]. In addition, as participation was voluntary, it remains uncertain whether this active recording approach is sustainable in routine practice outside the research setting.

Our findings do not imply that subjective perceptions should be used as stand-alone metrics for ED crowding. Relying solely on subjective assessments carries inherent clinical risk. As Garcia-Romero et al [20] have demonstrated, clinicians who are chronically exposed to excessive crowding tend to underestimate objectively measured crowding levels, a pattern consistent with the "normalization of deviance." To address both the interruptive limitations of EMA and individual clinicians' tendency to underestimate crowding, future research should focus on hybrid monitoring models. As a hypothesis-generating direction, future work could combine

XSL•FO
RenderX

active EDPC data with passive smartwatch-based physiological sensing (such as heart rate variability) [45], and objective EHR-based metrics in a multidimensional monitoring approach. Such an approach remains speculative at this stage; whether it could support applications such as early identification of rising operational burden or resource-reallocation alerts would require dedicated development and prospective validation before any clinical implementation could be considered.

Beyond technical feasibility, future research should use effectiveness-implementation hybrid designs to evaluate the acceptability, adoption, and long-term sustainability of smartwatch-based EDPC assessments. These studies could also investigate strategies to minimize reporting burden and protect frontline staff well-being. Furthermore, research should explore how real-time EDPC assessments could inform actionable thresholds for resource allocation and operational decision-making.

Limitations

This study had several limitations. First, it was conducted in a single tertiary hospital over a relatively short period, limiting its generalizability. Although EDPC and objective crowding metrics were measured at high temporal resolution, the limited study duration did not allow us to assess seasonal variation in ED crowding or in clinician perceptions.

Second, the EDPC instrument used NEDOCS-derived anchors, categories, and score ranges, and has not been independently validated against other subjective or objective crowding measures. Although participants could not view the actual real-time NEDOCS score while recording EDPC, the shared framework may have influenced clinicians' subjective ratings through anchoring or criterion contamination and contributed to the observed association between EDPC and NEDOCS. This design-related limitation cannot be resolved through statistical adjustment alone; accordingly, that association should not be interpreted as independent validation of either measure. Future validation studies should use neutral scale anchors that do not reproduce the categories or thresholds of the comparator metric.

Third, the study was conducted during an unprecedented national health care workforce crisis. During this study, attending physicians delivered all ED care, without residents or fellows, and as a crisis response, only 40 of the 67 ED beds were operational. This exceptional staffing and capacity context is not solely a threat to generalizability: because EDWIN incorporates physician staffing and bed capacity directly into its formula, these conditions may have altered the metric's internal behavior, complicating interpretation of its association with EDPC.

Fourth, to minimize disruption of routine clinical care, our study design permitted missed responses. Little MCAR test rejected the assumption that these responses were MCAR, indicating that missingness may have been systematically related to observed or unobserved factors. For example, clinicians may have been unable to record entries during periods of extreme workload or severe crowding. Although MI produced similar correlation estimates, the imputation model did not explicitly account for within-clinician or serial correlation, and missing-data bias remains possible. Additionally, participation in this study was voluntary, and it remains uncertain whether this active recording approach would be sustainable in routine, nonresearch clinical practice.

Fifth, multiple statistical comparisons were performed without adjustment for multiplicity, and no single primary comparison was prespecified. Therefore, the possibility of inflated type I error cannot be excluded, and the findings should be interpreted with caution.

Conclusion

This prospective observational study provides methodological and empirical insights into measuring ED crowding. Using dedicated smartwatch-based EMA, we captured clinicians' perceptions in situ at high temporal frequency during active shifts, potentially reducing susceptibility to retrospective recall bias while minimizing the spatial constraints of traditional survey methods. During an exceptional health care workforce disruption, hourly median EDPC was moderately associated with EDOR and NEDOCS, whereas its association with EDWIN was less consistent across temporal sensitivity analyses. These findings warrant further investigation of EDPC as a complementary measure of ED crowding alongside established objective metrics. However, our findings do not establish that EDPC directly measures cognitive workload or predicts subsequent crowding. The high response rate (1100/1124, 97.9%) supports the feasibility of wearable-based momentary assessments in acute care. Future studies should evaluate the generalizability of these findings across diverse ED settings and further assess the validity of EDPC as a complementary crowding measure using independently anchored scales, appropriate time-series analyses, and multicenter designs.

Acknowledgments

The authors would like to express their sincere gratitude to the emergency department clinicians who generously participated and contributed to this research.

Funding

This work was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number RS-2024-00335925). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

XSL•FO
RenderX

Data Availability

Data supporting the findings of this study are available from the corresponding author upon request. These data are not publicly available due to institutional and ethical restrictions related to patient and clinician confidentiality.

Generative AI Use

During the preparation of this manuscript, the authors used ChatGPT (GPT-5.5/GPT-6; OpenAI) and Gemini (Gemini 1.5 Pro; Google). According to the Generative AI Delegation Taxonomy [46], the following tasks were delegated to generative AI tools under full human supervision: literature search and systematization, data cleaning, proofreading and editing, translation, and reformatting. All outputs produced with these tools were thoroughly reviewed, verified, and critically evaluated by the authors. Responsibility for the final manuscript lies entirely with the authors. These tools are not listed as authors and do not bear responsibility for the final outcomes.

Authors' Contributions

Conceptualization: WCC, JY
Methodology: all authors
Investigation: JSY
Data curation: JSY
Formal analysis: JSY, SP, SH, JY
Validation: KJM, PBD, SYS, HC, TK, SUL, SH
Writing–original draft: JSY, SP, JY
Writing–review & editing: KJM, PBD, SYS, HC, TK, SUL, SH
Supervision: WCC, JY
Project administration: WCC, JY
Funding acquisition: WCC

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary methods, tables, and figures.
[DOCX File , 7019 KB-Multimedia Appendix 1]

References

1. Ahmed AE, Aqeel A, Ghazwani M, Alenzi AN, Alruwaili FM, Alsharari AS, et al. Clinical and operational effects of emergency department crowding: a systematic review. Cureus. 2026;18(1):e100560. [doi: 10.7759/cureus.100560] [Medline: 41625046]
2. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Ann Emerg Med. 2003;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
3. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One. 2018;13(8):e0203316. [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
4. American College of Emergency Physicians (ACEP). Crowding. Policy statement. Ann Emerg Med. 2013;61(6):726-727. [doi: 10.1016/j.annemergmed.2013.03.037] [Medline: 23684339]
5. Kim M, Lee S, Choi M, Kim D, Yoo J, Shin TG, et al. Factors that predict emergency department length of stay in analysis of national data. Clin Exp Emerg Med. 2025;12(1):35-46. [FREE Full text] [doi: 10.15441/ceem.24.309] [Medline: 39414332]
6. Haimovich AD, Berry SD, Landon BE. Prolonged emergency department stays for older US adults. JAMA Intern Med. 2025;185(8):1037-1038. [doi: 10.1001/jamainternmed.2025.2006] [Medline: 40587143]
7. van Dam CS, Peters MJL, Hoogendijk EO, Nanayakkara PWB, Muller M, Trappenburg MC. Older patients with nonspecific complaints at the emergency department are at risk of adverse health outcomes. Eur J Intern Med. 2023;112:86-92. [FREE Full text] [doi: 10.1016/j.ejim.2023.03.018] [Medline: 37002150]
8. Mohr NM, Wessman BT, Bassin B, Elie-Turenne M-C, Ellender T, Emlet LL, et al. Boarding of critically ill patients in the emergency department. J Am Coll Emerg Physicians Open. 2020;1(4):423-431. [FREE Full text] [doi: 10.1002/emp2.12107] [Medline: 33000066]
9. Oskvarek JJ, Zocchi MS, Black BS, Celedon P, Leubitz A, Moghtaderi A, et al. US Acute Care Solutions Research Group. Emergency department volume, severity, and crowding since the onset of the coronavirus disease 2019 pandemic. Ann Emerg Med. 2023;82(6):650-660. [doi: 10.1016/j.annemergmed.2023.07.024] [Medline: 37656108]

XSL•FO
RenderX

10. Jayaprakash N, Pflaum-Carlson J, Gardner-Gray J, Hurst G, Coba V, Kinni H, et al. Critical care delivery solutions in the emergency department: evolving models in caring for ICU boarders. Ann Emerg Med. 2020;76(6):709-716. [doi: 10.1016/j.annemergmed.2020.05.007] [Medline: 32653331]
11. Badr S, Nyce A, Awan T, Cortes D, Mowdawalla C, Rachoin J. Measures of emergency department crowding, a systematic review. How to make sense of a long list. Open Access Emerg Med. 2022;14:5-14. [FREE Full text] [doi: 10.2147/OAEM.S338079] [Medline: 35018125]
12. Clouston R, Atkinson P, Canales DD, Fraser J, Sohi D, Lee S, et al. Emergency department occupancy is useful as a simple real-time measure of crowding. CJEM. 2022;24(1):23-26. [doi: 10.1007/s43678-021-00098-8] [Medline: 33748940]
13. Altun M, Kudu E. Overcrowding in emergency departments: a scoping literature review. Anatolian J Emerg Med. 2025;8:94-100. [doi: 10.54996/anatolianjem.1684632]
14. Jones P, Wells S, Ameratunga S. Towards a best measure of emergency department crowding: lessons from current Australasian practice. Emerg Med Australas. 2018;30(2):214-221. [doi: 10.1111/1742-6723.12868] [Medline: 28941074]
15. Mol S, Gaakeer M, van der Linden MC, Baan-Kooman ECM, Backus BE, de Ridder VA. Crowding, perceived crowding and workload in Dutch emergency departments: should we continue on the same road? Eur J Emerg Med. 2023;30(4):229-230. [doi: 10.1097/MEJ.0000000000001034] [Medline: 37115965]
16. McCarthy ML, Aronsky D, Jones ID, Miner JR, Band RA, Baren JM, et al. The emergency department occupancy rate: a simple measure of emergency department crowding? Ann Emerg Med. 2008;51(1):15-24, 24.e1. [doi: 10.1016/j.annemergmed.2007.09.003] [Medline: 17980458]
17. Weiss SJ, Derlet R, Arndahl J, Ernst AA, Richards J, Fernández-Frackelton M, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). Acad Emerg Med. 2004;11(1):38-50. [FREE Full text] [doi: 10.1197/j.aem.2003.07.017] [Medline: 14709427]
18. Bernstein SL, Verghese V, Leung W, Lunney AT, Perez I. Development and validation of a new index to measure emergency department crowding. Acad Emerg Med. 2003;10(9):938-942. [FREE Full text] [doi: 10.1111/j.1553-2712.2003.tb00647.x] [Medline: 12957975]
19. Boverio R, Signorini F, Rossi C, Bellini RIM, Zanelli C, Ghilardi GI, et al. FENICE network. Measuring emergency department crowding: comparing the NEDOCS and Fenice crowding scores against staff perception. Intern Emerg Med. 2025. [doi: 10.1007/s11739-025-04216-7] [Medline: 41385153]
20. García-Romero M, Rita-Gáfaro CG, Quintero-Manzano J, Bermon A. Escala NEDOCS vs valoración subjetiva, ¿El personal de salud en urgencias es consciente de su sobrecupo? Colomb Med. 2017;48(2):53-57. [doi: 10.25100/cm.v48i2.2131]
21. Muir KJ, McHugh MD, Merchant RM, Lasater KB. Left without being seen: nurse work environment and timely outcomes in New York and Illinois emergency departments. J Emerg Nurs. 2024;50(5):660-669. [doi: 10.1016/j.jen.2023.11.010] [Medline: 38127046]
22. Muir KJ, Golinelli D, Lasater KB, Aiken LH, Merchant R, McHugh MD, et al. Association of emergency department and medical-surgical nurse staffing on patient departures against medical advice. J Emerg Nurs. 2026;S0099-1767(25):00510-0. [FREE Full text] [doi: 10.1016/j.jen.2025.12.009] [Medline: 41575406]
23. Heslin SM, Schwaner R, Viccellio P, King C, Rowe A, Thode H, et al. A physician-nurse team adjacent to triage shortens length of stay and left-without-being-seen rates. Clin Exp Emerg Med. 2025;12(1):26-34. [FREE Full text] [doi: 10.15441/ceem.24.248] [Medline: 39237140]
24. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol. 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
25. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry. 2018;17(2):123-132. [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
26. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud. 2021;22(5):2361-2408. [FREE Full text] [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
27. Yoon JH, Kwon IH, Park HW. The South Korean health-care system in crisis. Lancet. 2024;403(10444):2589. [doi: 10.1016/S0140-6736(24)00766-9] [Medline: 38879246]
28. Park HA, Kim T, Kim HJ, Han S. Epidemiological trends in emergency department visits by age group: a report from the National Emergency Department Information System (NEDIS) of Korea, 2020-2024. Clin Exp Emerg Med. 2025;12(4):405-413. [FREE Full text] [doi: 10.15441/ceem.25.291] [Medline: 41531412]
29. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(Pt 1):29-48. [doi: 10.1348/000711006X126600] [Medline: 18482474]
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ. Lawrence Erlbaum Associates; 1988.
31. R: a language and environment for statistical computing. R Core Team. 2025. URL: https://www.R-project.org [accessed 2026-10-23]
32. Altun M, Kudu E, Yakin F, Korgan MB, Unal E, Demir O, et al. Impact of transition units on emergency department crowding: a before-and-after study. J Emerg Med. 2026;81:68-77. [doi: 10.1016/j.jemermed.2025.12.007] [Medline: 41581296]

XSL•FO
RenderX

33. Strada A, Bravi F, Valpiani G, Bentivegna R, Carradori T. Do health care professionals' perceptions help to measure the degree of overcrowding in the emergency department? A pilot study in an Italian university hospital. BMC Emerg Med. 2019;19(1):47. [FREE Full text] [doi: 10.1186/s12873-019-0259-9] [Medline: 31455226]
34. Wang H, Kline J, Jackson B, Robinson RD, Sullivan M, Holmes M, et al. The role of patient perception of crowding in the determination of real-time patient satisfaction at emergency department. Int J Qual Health Care. 2017;29(5):722-727. [doi: 10.1093/intqhc/mzx097] [Medline: 28992161]
35. Schippers EE, Willemsen SP, Collée LAM, Koenraadt KLM, Schol-Gelok S. Crowding in real-time: an artificial intelligence-based model to manage crowding in the emergency department (CRAMMED Study). BMC Emerg Med. 2026;26(1):232. [FREE Full text] [doi: 10.1186/s12873-026-01647-z] [Medline: 42310553]
36. Chen SM, Stewart G, Sentman E, Helwig S, Rust L, Hoffman J, et al. Validation of the emergency department work index in a pediatric freestanding community emergency department. Pediatr Emerg Care. 2024;40(10):705-710. [doi: 10.1097/PEC.0000000000003247] [Medline: 39051989]
37. Peltan ID, Bledsoe J, Jacobs J, Groat D, Klippel C, Adamson M, et al. Effectiveness and safety of an emergency department code sepsis protocol: a pragmatic clinical trial. Ann Am Thorac Soc. 2024;21(11):1560-1571. [FREE Full text] [doi: 10.1513/AnnalsATS.202403-286OC] [Medline: 38996086]
38. Liu Q, Xie C, Tan J, Xu L, Zhou F, Peng L. Exploring the nurses' experiences in recognising and managing clinical deterioration in emergency patients: a qualitative study. Aust Crit Care. 2024;37(2):309-317. [doi: 10.1016/j.aucc.2023.06.004] [Medline: 37455210]
39. Park S, Yoo J, Lee Y, DeGuzman PB, Kang M, Dykes PC, et al. Quantifying emergency department nursing workload at the task level using NASA-TLX: an exploratory descriptive study. Int Emerg Nurs. 2024;74:101424. [doi: 10.1016/j.ienj.2024.101424] [Medline: 38531213]
40. Martin N, Bergs J, Eerdekens D, Depaire B, Verelst S. Developing an emergency department crowding dashboard: a design science approach. Int Emerg Nurs. 2018;39:68-76. [doi: 10.1016/j.ienj.2017.08.001] [Medline: 28865753]
41. Gesner E, Dykes PC, Zhang L, Gazarian P. Documentation burden in nursing and its role in clinician burnout syndrome. Appl Clin Inform. 2022;13(5):983-990. [FREE Full text] [doi: 10.1055/s-0042-1757157] [Medline: 36261113]
42. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc. 2016;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
43. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11):e367. [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
44. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res. 2019;21(12):e14475. [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
45. Peabody JE, Ryznar R, Ziesmann M, Gillman L. A systematic review of heart rate variability as a measure of stress in medical professionals. Cureus. 2023;15(1):e34345. [FREE Full text] [doi: 10.7759/cureus.34345] [Medline: 36865953]
46. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (generative AI delegation taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res. 2026;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

AC2: Gwet weighted agreement coefficient
AR(1): first-order autoregressive
DARWIN-C: Samsung Medical Center’s Clinical Data Warehouse
ED: emergency department
EDOR: emergency department occupancy ratio
EDPC: emergency department perceived crowding
EDWIN: Emergency Department Work Index
EHR: electronic health record
EMA: ecological momentary assessment
GLS: generalized least squares
IRB: institutional review board
KTAS: Korean Triage and Acuity Scale
LAD: least absolute deviations
MCAR: missing completely at random
MI: multiple imputation
NEDOCS: National Emergency Department Overcrowding Score

XSL•FO
RenderX

Edited by S Brini; submitted 04.Jun.2026; peer-reviewed by F Signorini, E Kudu; comments to author 08.Aug.2026; revised version received 20.Oct.2026; accepted 21.Oct.2026; published 10.Nov.2026

Please cite as:
Yoon JS, Park S, Muir KJ, DeGuzman PB, Shin SY, Chang H, Kim T, Lee SU, Hong S, Cha WC, Yoo J
Smartwatch-Measured Clinician Perceptions of Emergency Department Crowding and Their Association With Objective Metrics: Prospective Observational Study
JMIR Mhealth Uhealth 2026;14:e103317
URL: https://mhealth.jmir.org/2026/1/e103317
doi: 10.2196/103317
PMID:

©Ji Sim Yoon, Sookyung Park, K Jane Muir, Pamela B DeGuzman, So Yeon Shin, Hansol Chang, Taerim Kim, Se Uk Lee, Sungjun Hong, Won Chul Cha, Junsang Yoo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX