8-31-2010

Emergency department crowding and provider workload

Kathryn Coles

Rochester Institute of Technology

College of Liberal Arts

Department of Psychology

EMERGENCY DEPARTMENT CROWDING AND PROVIDER WORKLOAD

A Thesis

By

Kathryn M. Coles

B.A., State University at Buffalo, Psychology, 2003;

M.S., University of Georgia, Exercise Science, 2005

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Applied Experimental and Engineering Psychology

August 31, 2010

We approve the thesis of Kathryn M. Coles:

Date

Faculty Advisor and Chair of the Thesis Committee:	____________________ Esa M. Rantanen, Ph.D. Associate Professor, Dept. of Psychology, RIT	__________
Reader:	____________________ Sandra Schneider, MD Professor, Department of Emergency Medicine, University of Rochester	__________
Reader:	____________________ Matthew Marshall, Ph.D. Associate Professor, Dept. of Industrial Engineering, RIT	__________

Abstract

This observational field study attempted to quantify the objective task load imposed on emergency department (ED) providers, determine the degree of subjective workload they experience, and to correlate these data with ED operational metrics, mainly ED crowding metrics. Participants were a convenience sample of 10 emergency care providers; the 3 female and 7 male participants represented a variety of provider levels (6 physicians, 3 physician assistants, and 1 nurse practitioner). Forty-two hours of data were collected. ED variables were obtained from the hospital's existing information system each hour and included the Emergency Severity Index (ESI), number of people in the waiting room, patient/doctor ratio, patient/nurse ratio, number of patients assigned, number of providers on duty and crowding variables; Emergency Department Work Index (EDWIN) and occupancy level. Providers were shadowed and observed each hour by a researcher who recorded the type of tasks they performed, the number of tasks they performed, the time they spent on each task and the number of times they were interrupted. Subjective workload ratings (NASA-TLX) were obtained from providers at the end of each hour of observation. Correlations were performed to evaluate the relation of observed, subjective and hospital variables. Overall objective task load was quantified using time-on-task data and task difficulty weightings to achieve a single standardized value for overall objective workload (OTLX). OTLX scores were regressed against ED crowding measures of occupancy and EDWIN score. Structured interviews were conducted with each participant following the observation sessions. Results from the study revealed that providers spent 75 percent of their time performing tasks related to communication with staff, direct patient care, and paperwork. The other 25 percent of their time was spent checking test results, admitting patients to the hospital, taking breaks, looking for supplies, checking the electronic whiteboard, and other job-related tasks. ED occupancy was positively correlated to subjective workload and predicted 30 percent of the variance in subjective workload. The EDWIN score, on the other hand, only predicted 9 percent of the variance in subjective workload. This study revealed no correlation between ED crowding and objective task load and ED crowding predicted less than 4 percent of the variance in OTLX scores.

In accordance with “Occam’s razor”, ED occupancy may provide an advantage over more complex compound measures of ED crowding such as the EDWIN score in predicting provider subjective workload and may be more useful in making ED staffing and scheduling decisions. In addition to collected and recorded variables, valuable insights were obtained from ED providers regarding issues of ED crowding, time-pressure and workload. It is apparent from their responses that, in the absence of observable changes in task load, the quantity and status of the “unseen” patient weighs heavily on their minds. Future research should assess the number of patients waiting or the number of patients who have left without being seen (LWBS) not only as a metric of ED crowding but as a predictor of ED provider workload.

Table of Contents

List of Figures

List of Tables

Aknowledgements

I would like to thank my advisor, Esa Rantanen, for presenting me with this opportunity and for his continued guidance, support, and patience. I am especially grateful to Sandra Schneider, Terry Fairbanks, and John Hilmi who familiarized me with the emergency department environment and made it possible for me to conduct research with URMC. I would also like thank Matt Marshall for his careful review of the project and his helpful suggestions. Finally, I want to acknowledge all of the the providers who were willing to let me follow them around for hours while they were treating the sick and injured. I am sincerely grateful for your contribution to this research.

Without the support of my husband and my family, who have gone through the thesis process twice with me, I would not have been able to do this. I'm sure it wasn't easy and I thank you for your support, encouragement and understanding.

Introduction

Emergency department (ED) crowding has become a leading concern among medical professionals. Despite the growing concern and focus on ED crowding, there is a lack of consensus on the terminology used to describe it and the method used to define it (Moskop, Sklar, Geiderman, Schears, & Bookman (2009). According to Moskop et al. (2009), two terms are commonly used to describe the state of having a large number of people relative to the space available; "crowding" and "overcrowding". Although the terms are used interchangeably, overcrowding suggest a more negative circumstance than crowding, which may or may not be valid since both refer to the same concept. For this reason, Moskop et al. (2009) suggest that the term "crowding" be used to refer to this state.

Crowding

Despite the severity of the issue, a common definition of ED crowding is lacking. According to the American College of Emergency Physicians (ACEP; 2008), crowding exists when there is no space left to meet the timely needs of the next patient in need of emergency care. This definition proposes a supply and demand relationship between available space and patient need but does address in the operational characteristics of this state. At what point should an ED be considered crowded? Schneider, Gallery, Schafermeyer, & Zwemer (2003), defined ED crowding, more specifically, in terms of physical crowding and personnel availability. In a random survey of 250 emergency departments across the United States, Schneider et al. (2003) examined the point prevalence of crowding at an index time (Monday, March 12, 2001). Physical crowding was defined in this study as having more patients in the ED than treatment rooms. Personnel shortage was defined as a patient to nurse ratio greater than 4:1 or a physician caring for more than 6 patients. Results from their survey revealed both physical crowding and personnel shortages with an average of 1.1 patients per treatment space, a mean 4.2 patients per registered nurse; and a mean 9.7 patients per physician.

According to the ACEP (2008), causes of ED crowding include unnecessary visits, the use of the ED as a safety net by the poor and uninsured, seasonal variation in illnesses, and boarding of inpatients. The main contributor to ED crowding was boarding of inpatients in the ED (ACEP, 2008), which reduces the department's ability to see and treat new patients (Schneider et al. 2003; Asplin et al., 2003). Emergency department crowding increases the number of patients hospital personnel have to simultaneously care for. This has a direct effect on patients who are boarded in the ER. Hollander and Pines (2007) reported the ratio of nurses to patients to be 1:2 in the intensive care unit (ICU) and 1:10 in the ED. Therefore, patients boarded in the ED do not get the same care as those boarded in the ICU (Hollander & Pines, 2007). The ACEP (2008) reported several negative consequences of ED crowding including increased waiting times, ambulance diversion, increased length of stay, medical errors, sentinel events (unexpected occurrences involving death or serious physical or psychological injury), increased patient mortality, financial losses to hospital and physician, and increased medical negligence claims. These consequences are based on observable data commonly collected by hospitals, yet there are likely other consequences of ED crowding that are not measured. For example, it is likely that crowding contributes to increased physician workload, which may influence overall patient care. This study will explore the relationship between ED crowding and physician workload using cognitive task analysis (CTA) techniques of real-world observation and structured interview.

Measuring Workload in the ED

Although human factors techniques have been widely used in the domains of aviation safety and air traffic control, they have not been heavily utilized by health care industry. Wears and Perry (2002) cited four main contributors to the absence of human factors and ergonomics in health care facilities. These factors include self-blame by medical professionals, lack of human factors resources, decentralization of authority, and persistence of the Guild and Workshop mentality. Wears and Perry (2002) emphasize the need for emergency departments to redirect focus of blame from humans in order to understand how the design of the system as a whole contributes to errors. The application of human factors techniques in emergency department research requires that the ED be viewed as a system. According to Asplin et al. (2003), emergency department crowding can be partitioned into 3 components: input (emergency care, unscheduled urgent care, and safety net care), throughput (patient arrival, triage room placement, diagnoses and treatment, and ED boarding of inpatients), and output (ambulatory care, transfer to other facility, and admittance to hospital). This research will explore provider mental workload in response to system changes, mainly indicated by ED crowding and related ED operational metrics.

Mental Workload

Humans have limited capacity, or resources, to process and respond to information (Navon & Gopher, 1979; Wickens, 1984). Workload is an important factor to consider when studying human behavior as it refers to the portion of resources required to perform a particular task (O'Donnel & Eggemeier, 1986) and can be defined as the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available (Gopher & Donchin, 1986). In essence, workload is defined by the relationship between resource supply and task demand (Wickens & Hollands, 2000).

Mental workload is a multidimensional construct that defies simple definitions and measures (e.g., single metrics). Workload depends on 1) individual factors such as operator capabilities, goals, decision/selection strategies and commitment of mental and physical resources 2) environmental and task demands placed on sensory, cognitive, and psychomotor resources (Wickens, 1984) and 3) performance. Nevertheless, unambiguous quantification of mental workload is critical to human factors engineering efforts in design and evaluation of various systems. For example, consider the design of combat mission aircraft and related equipment. As those aircraft are developed, and the technology advanced, it is critical that additional demands are not placed on operator workload. In order to ensure this does not occur, one must first be able to quantify workload so that it can be examined in response to system changes. Another example comes from the U.S. Army Research Institute of Behavioral of Social Sciences. In response to Special Operation Forces (SOF) modifications to a UH-60 mission aircraft, Bierbaum, Szabo, and Adrich (1989) conducted a detailed analysis of the tasks that must be performed to accomplish the UH-60 combat mission. These tasks include those related to SOF modifications. A key component of this analysis was the estimation of workload associated with the sensory, cognitive, and psychomotor components of each task needed to complete the mission. With findings from the analysis, Bierbaum, Szabo, and Adrich (1989) developed a computer model to predict UH-60 operator workload.

Objective Task Load

It is important to distinguish here between objective task load and subjectively experienced workload. Workload depends on individual factors such as operator capabilities, sensory, cognitive, and motor skills, knowledge base, selection of strategies, and commitment of mental and physical resources, as well as individual task goals, performance, and preconceptions. Task load, unlike workload, is not related to individual operator characteristics or perceptions. Instead, it is defined by the demands placed on the operator, or the performance on a given task/s, and is the same for everyone performing the same task/s under the same conditions. For example, in terms of performance, task load can be objectively measured by the ratio of time required to perform a number of tasks and time available to do so. Performance measures are easy to obtain in controlled laboratory environments where tasks are predetermined by the researcher. In real-world settings, however, measurement of task load is more difficult. This study was conducted in a real-world operating ED where performance measurement was not feasible. Instead, we used cognitive task analysis (CTA) techniques to assess the demands placed on ED providers.

Cognitive task analysis. Cognitive task analysis (CTA) represents a combination of techniques that can be used to determine how work or tasks are performed. Real-world observation is the gold standard of CTA methods (Crandall, Klein, & Hoffman, 2006); however, the combination of several CTA methods provides a clearer picture of the overall action and interactions being observed. Data collected using CTA methods can be classified in terms of time (past, present, future), realism (real world, simulators, or artificial environment), difficulty (routine, challenging, or rare events), and generality (abstract, job/task, or incident/event) (Crandall, et al. 2006). The current study was conducted in a real-world job setting where routine, challenging and rare events occur; objective variables were collected in the present while subjective variables required retrospective analysis.

Concurrent tasks. Physicians are required to perform several tasks during their shift. Some of the primary tasks include conducting patient exams, writing up patient charts, communicating with other physicians, and treating patients. Crowding is likely to make these tasks more difficult as it increases the number of patients a physician has to care for simultaneously and, according to Dismukes, Loukopoulos, and Jobe (2001), concurrent task management is a point of vulnerability that leads to lapses in monitoring and failures to remember to complete deferred actions. Varying levels of crowding and time of day may influence the number of patients assigned to a physician. In a retrospective observational study of workload patterns among ED physician teams, Levin et al. (2007) found that shift changes during peak occupancy periods caused patient load imbalances which led to some residents managing a disproportionate number of patients compared with others.

Interruptions. Interruptions are likely to increase task difficulty. In an observational study of a level-one trauma center, Brixey et al. (2007) identified people, pagers, telephones, and the environment (i.e., missing supplies) as sources of interruptions in the ED. Interruptions are positively correlated with the average number of patients being simultaneously managed (Chisholm, Collision, Nelson, & Cordell, 2000), which in turn is a direct result of crowding. Chisholm et al. (2000) classified emergency physicians as “interrupt-driven” and recorded their interruption rate at 30.9 per 180-minute time period, or about 10 interruptions per hour. In another study conducted in an adult area of an academic ED, Fairbanks, Bisantz, and Sumn (2007) found that physicians were interrupted 6.9 times per hour while bedside nurses were interrupted 0.5 times per hour.

Despite different results in the frequency of interruptions recorded in an ED, it is obvious that interruptions present an opportunity for errors to arise. They require the physician or nurse to reallocate their attention from their current task to another task and, therefore, result in breaks in task. Chisholm et al. (2000) recorded physician breaks in task at 20.7 per 180-minute time period. Concurrent tasks and interruptions increase task difficulty and because workload is closely related to task difficulty (Gopher & Donchin, 1986), crowding is likely to increase physician workload.

Time pressure. The ED personnel work under constant time pressure and hence effective time management and task prioritization are critical to physician performance and patient safety. Concurrent tasks and interruptions likely increase task difficulty as well as time pressure, and therefore also workload. Load on the human-information processing system results directly from the ratio of the time necessary to process the required information to the time available for making a decision (Hendy, Liao, & Milgram, 1997). This ratio is likely to increase as crowding increases.

Subjective Workload

Mental workload may not manifest itself in an observable manner. For this reason, workload is best measured using a variety of techniques rather than a single technique (O’Donnell & Eggemeier, 1986). For example, consider that people are able to cope with increasing task demands by increasing the mental and physical effort they exert. Although demands are increased, objective performance measures may appear stable or unaffected by the change. In these circumstances, the use of an objective method of measurement would result in an inaccurate measure of workload. According to Hart & Staveland (1988), subjective ratings may come closest to tapping mental workload.

Several researchers have attempted to measure subjectively perceived workload rather than performance. One-dimensional scales, such as the Modified Cooper-Harper scale (Wierwille & Casali, 1983) and the Overall Workload (OW) scale (Vidulich & Tsang, 1987) and multi-dimensional scales, such as the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) (Hart & Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT) (Reid, Shingledecker & Eggemeier, 1981) have all been developed as methods to quantify subjectively experienced workload. Hill , et al. (1992), compared the four aforementioned subjective workload scales and found that all were sufficiently acceptable and sensitive tools for measuring variations in workload, but that NASA-TLX and OW were consistently superior in terms of sensitivity and operator acceptance.

The NASA-TLX is perhaps the most widely used and accepted technique for measuring subjective workload. The NASA-TLX is defined by Hart and Staveland (1988) as a "multi-dimensional rating that provides an overall workload score based on a weighted average of ratings on six subscales: mental demands; physical demands; temporal demands; operator performance; effort; and frustration" (p. 3). Operators, workers, or participants provide ratings (1-100) of their perceived subjective workload for a given task/s as it relates to each of the 6 dimensions. In order to calculate an overall score for the NASA-TLX, a weighting procedure is used as a method of individualizing the index to the task and the operator. Fifteen subscale pairings (i.e. mental demand vs. frustration) are presented to each individual. For each pair, the individual is asked to select the subscale (dimension) that contributes most to the workload experienced for the task/s in question. The more times a particular subscale is chosen over another subscale, the higher weight it receives. Weights can range from 0 to 5 depending on how many times a subscale is selected. Individual ratings of perceived workload are then multiplied by these weights, summed and averaged to create an overall score of perceived workload for the task/s performed. The widespread use of TLX can be attributed to the rigorous development and validation procedures employed by Hart & Staveland (1988) and to its ability to obtain more detailed and diagnostic data (Hill et al., 1992) regarding the potential causes, or sources, of workload than other methods. This study will employ the NASA-TLX to examine overall subjective workload as well as individual subjective workload dimensions.

Hospital Statistics and Measures

Hospitals routinely collect ED data and measures related to patients, providers, and general department efficiency. Such measures may include the total number of patients in the ED, the total number of patients in the waiting room, the patient severity level, the number of providers on duty, and the patient doctor or nurse ratios. In most cases, these data are collected electronically in real-time. Patient severity refers to the triage assignment given to each patient using the 5-level Emergency Severity Index where 1 is most urgent and 5 is least urgent (Wuerz, Milne, Eitel, Traver & Gilboy, 2000). Index categories are defined by patient acuity (stability of vital functions, degree of distress), expected resource intensity, and timelessness (expected staff response, time to disposition).

In addition to these data, most hospitals also compute real-time overcrowding metrics which may include some of the basic ED measures. The Emergency Department Work Index (EDWIN), the National Emergency Department Overcrowding Scale (NEDOCS), the Demand Value of Real-time Analysis of Demand Indicators (READI), and Work Score have all been developed to quantify and predict ED overcrowding.

The Emergency Department Work Index (EDWIN) was developed as a simple quantitative measure of ED crowding and busyness to be integrated into real-time clinical information systems (Bernstein, Verghese, Leung, Lunney, & Perez, 2003). The EDWIN takes into account several of the basic ED measures. These measures include the total number of patients in the ED, the severity or acuity of their condition, the number of physicians on duty, the total number of beds available, and the number of admitted patients being held in the ED at any given time. Research has demonstrated the discriminatory and predictive validity for the EDWIN (Weiss, Ernst, and Nick, 2006; Hoot, Zhou, Jones, and Aronsky, 2007)

Weiss, Ernst, and Nick (2006) collected the NEDOCS, EDWIN and an overcrowding measure every two hours for 10 days. The overcrowding measure was based on expert opinion and was measured on a 100-mm visual analogue scale (VAS). Overcrowding was based on the dichotomous overcrowding VAS score (>= 50 = overcrowded, <50, not overcrowded). The ability of each measure to discriminate overcrowding was examined using the area under the ROC curve (AUC) where 1.0 reflects perfect discrimination. The AUC for NEDOCS and EDWIN was 0.83 and 0.80, respectively. They concluded that both scales had high construct validity but that the NEDOCS was slightly preferred over the EDWIN.

In a different study, Hoot, Zhou, Jones, and Aronsky (2007) quantified the potential for monitoring current and near-future emergency department (ED) crowding using the EDWIN, the NEDOCS, the READI, the Work Score and basic ED occupancy level. They calculated these measures at 10-mintue intervals over an 8-week period using ambulance diversion status as an outcome variable for crowding and occupancy as a performance baseline measure. The ability of each measure to discriminate ambulance diversion status was examined using area under the ROC curve (AUC). Predictive power was examined by applying activity monitoring operating characteristic curves to measure the timeliness of early warnings and false alarm rates. The results revealed that the EDWIN, NEDOCS and Work Score monitor current ED crowding with high discriminatory power (AUCs of 0.81, 0.88 and 0.90, respectively). Although occupancy level was intended as a baseline measure, they were interested to find that it, too, had high discriminatory power (AUC = 0.90). In addition, they report that only occupancy level provided more than 1 hour of advanced warning of crowding.

For this study, basic ED metrics related to patients, providers, and general department efficiency as well as crowding metrics (EDWIN and basic occupancy) were collected and examined in relation to objective task variables and subjective workload. EDWIN and occupancy level were chosen as crowding measures based on their demonstrated validity and availability. Occupancy, specifically, was chosen for its simplicity.

Purpose of the Research

The purpose of this research was to objectively quantify the task load imposed on ED providers, determine the subjective workload they experience, and to correlate these data with the ED operational metrics, specifically, ED crowding metrics. Observed variables were recorded using CTA techniques. Providers were shadowed by a researcher who recorded the types of tasks they performed, the number of tasks they performed, the time they spent on each task and the number of times they were interrupted. An Objective Task Load Index (OTLX) was developed to quantify the task load imposed on providers by using task weights and time-on-task data. Subjective variables were obtained from providers and included overall subjective workload ratings and subscale ratings (NASA-TLX). ED variables were obtained from the hospital's existing information system and included basic ED operating variables such as patient severity (ESI), the number of people in the waiting room, patient/doctor ratio, patient/nurse ratio, the number of patients assigned and the number of providers on duty; crowding variables included the EDWIN and occupancy level. Figure 1 details conceptually the expected relation between the main variables of interest (ED crowding, provider objective task load, and provider subjective workload) as well as the component variables measured in an effort to define and quantify them.

Primary hypotheses. We make the following hypotheses

(1) Objective task load will be positively correlated with ED crowding metrics. This hypothesis is based on the assumption that ED crowding increases the concurrent task demands placed on providers.

(2) Provider subjective workload will be positively correlated with objective task load. This hypothesis is based on the assumption that providers' subjective workload is related to the resources demanded by the tasks they are performing. Workload refers to the portion of resources required to perform a particular task (O'Donnel & Eggemeier, 1986). According to Hart & Staveland (1988), subjective ratings may come closest to tapping mental workload.

(3) Provider subjective workload will be positively correlated with ED crowding metrics. This hypothesis is based on the assumption that hypotheses one and two are true.

Secondary hypothesis. ED crowding metrics will be predictive of provider subjective workload ratings. If perceived workload is as strong an indicator of mental workload as Hart & Staveland propose (1988), then it should be sensitive to changes in ED crowding. This hypothesis is based on the assumption that ED crowding will explain a significant portion of the variation in subjective workload ratings. This predictive link is diagrammed in Figure 1.

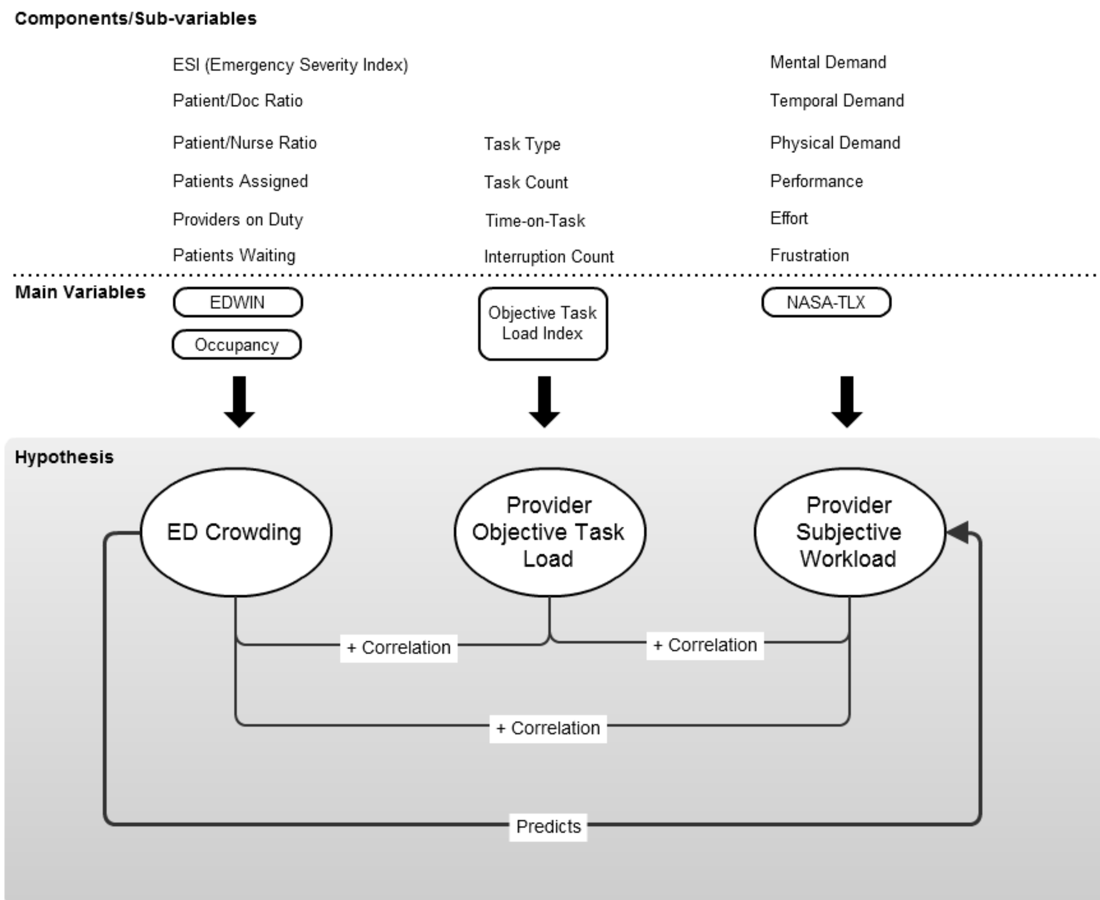

Figure 1. A diagram of the hypothesized relationship between ED crowding, objective task load, and subjective workload. This diagram outlines the expected correlations between the main research variables and the components used to define them. This figure lists collected and/or measured variables used to define ED crowding, objective task load, and subjective workload.

Method

Participants

This was a field study conducted at the University of Rochester Medical Center's (URMC) Highland Hospital Emergency Department between May and August 2009. The study was approved by the URMC's Research Subject's Review Board and the Rochester Institute of Technology's Internal Review Board (approval letters are in Appendix A). Participants were a convenience sample of 10 emergency care providers; the 3 female and 7 male participants represented a variety of provider levels (6 physicians, 3 physician assistants, and 1 nurse practitioner).

Participants were recruited via e-mail and word-of-mouth. The study was presented at a provider staff meeting and recruitment e-mail (Appendix B) was distributed to all ED physicians, physician assistants, and nurse practitioners. Potential participants were asked to respond to the e-mail if they were interested in volunteering for the study. All potential participants were given an informed consent form describing the nature of the study (Appendix C) as well a detailed verbal description of the time and procedures associated with participation in this research. Verbal consent was obtained from each participant. The verbal consent, instead of signed consent, was used in order to protect the participant privacy. Volunteers were included if they were willing to be shadowed and observed for several hours during their work shift and if they were available to be shadowed within the timeframe of the data collection period (May–August, 2009). No monetary incentives were provided for participation in this project; however, data collected from this study may be used to guide decisions on ED staffing, scheduling, and other operations and to improve the overall knowledge about the effects of ED crowding.

Materials

Data were collected in real time from hospital electronic systems, direct observation, and participant feedback. The following materials were used in the data collection process.

Task log. Observational data, including tasks, time-on-task, and notes were recorded with pen on the log (Appendix D) and were subsequently transferred to an electronic spreadsheet (Appendix E).

NASA-TLX subjective workload index. A modified version of the NASA-TLX (Appendix F) was used to assess subjective workload as it related to the work performed during the one-hour observation period. The instructions for participant ratings were condensed and modified to more closely fit the specific circumstances of this study. These instructions are presented in Appendix G.

Structured interview questions. The development of the structured interview questions was based on knowledge from previous workload and time pressure literature (Zakay, 1993) as well as observations from this study. The questions sought to determine how ED crowding affects the provider in terms of task load, subjective workload, and time pressure (Appendix H).

Variables

Emergency department variables. The ED at URMC, like other EDs, routinely monitors and stores a number of statistics related to ED crowding. ED data were retrieved and printed at the end of each observation hour from the emergency department's electronic information system, CareSuite ED PulseCheck (Picis, Wakefield, MA). The following variables were used as indicators of ED crowding.

Occupancy. Occupancy refers to the proportion of licensed ED beds occupied by patients at a given point in time.

EDWIN. The Emergency Department Work Index (EDWIN) is defined as

$$\frac{\sum n_i t_i}{N_a(B_T - B_A)} \tag{1}$$

where n_i = the number of patients in the ED in the triage category *i*, t_i = triage category, N_a = the number of attending physicians on duty. B_T = the number of treatment bays, and B_A = the number of admitted patients in the ED (Bernstein et al., 2003). Triage categories, again, are based on the Emergency Severity Index. The EDWIN score refers to ED crowding at a given point in time. A higher EDWIN score indicates a more crowded ED.

Number of patients waiting. The total number of patients waiting included those waiting at triage, in the waiting room, and at the greeting desk.

Patient/doctor ratio. The patient/doctor ratio refers to the average patient/doctor ratio for the entire ED at a given point in time.

Patient/nurse ratio. The patient/nurse ratio refers to the average patient/doctor ratio for the entire ED at a given point in time.

Total providers. The total number of providers was recorded as the total number of ED providers on duty at a given point in time.

Number of patients assigned. The number of patients assigned was recorded as the number of patients assigned, at a given point in time, to the provider being observed.

Emergency Severity Index (ESI). ESI refers to the triage assignment given to each patient using the 5-level Emergency Severity Index (ESI) where 1 is most urgent and 5 is least urgent (Wuerz, Milne, Eitel, Traver & Gilboy, 2000). For data analysis purposes, the Emergency Severity Index was reversed so that level 5 represented the most urgent cases and level 1 the least urgent cases. ESI was calculated as

$$(n_i t_i) \,/\, n \qquad (2)$$

where n_i = the number of patients in the ED in the triage category i, t_i = triage category, and n = the total number of patients in the ED. The ESI was recorded for a given point in time and reflects overall ED severity as it is not related to any one patient or set of patients. A higher average ESI suggests a more urgent ED status.

Objective task load variables. Altogether four different variables were identified as pertaining to task load and recorded during the shadowing sessions.

Task type. A set of commonly observed ED provider tasks was developed and used to systematically code observed behaviors. The tasks were chosen based on observation, prior research (Levin et al., 2006), pilot testing and expert opinion. The original task set can be viewed in Table 1.

Table 1

Original Task Codes Used to Define and Record Observed Tasks

Task Code	Description
TP	Treating patient - includes procedures and exams (i.e. rectal/pelvic)
VCS	Verbal communication between participant and staff (excludes consults)
VCP	Verbal communication between participant and patient (excludes initial history and exam)
VCF	Verbal communication between participant and patient family member/s
AP	Performing tasks directly related to the admittance of the patient to the hospital
CB	Checking the boards (electronic white board)
CT	Checking test/lab results
HE	Obtaining patient history and performing initial physical exam
SE	Locating/retrieving supplies and equipment - includes charts & stickers
CON	Provider-to-provider verbal exchange of patient information (includes formal and informal consultations)
Break	Performance of any activity or discussion of any topic unrelated to the job
Chart	Written Charting
WO	Writing orders for tests, labs, and meds
OC	Outgoing call
IC	Incoming call
WD	Write discharge instructions
Meet	Any organized meeting with other staff members
Teach	Any activity related to teaching a resident or new staff member
LM	Leaving/sending a message via phone or e-mail
Cultures	Reviewing and initial culture results
RP	Researching patient history
Paperwork	Any paperwork that doesn't fall under other written categories (Chart, WO, WD)

Task count. Each time a task was initiated, a count was recorded for that task. Task count refers to the frequency with which a task was initiated in the 60-minute observation period and is independent of the time spent on that task.

Time-on-task. The time spent on each task was recorded to the nearest minute on the Task Log. This raw time data was summed to yield the total time spent on any given task for that 60-minute observation period.

Interruptions. Interruptions were defined as any event that required the participant to disengage from his or her current task and that was unrelated to the current task. Interruptions commonly led to other tasks and, therefore, interruptions were not logged as an independent task category. Instead, interruptions were logged by count and categorized by type (face-to-face, phone, pager) in the task log. These data were subsequently summed to yield the total number of interruptions (per type) observed during each 60-minute observation period.

Subjective workload ratings. Subjective workload was measured using the NASA-TLX. **R**atings **(1 -100)** were obtained for each of six index subscales: Mental Demand, Temporal Demand, Physical Demand, Frustration, Performance, and Effort. These subscales were weighted and used to calculate a total NASA-TLX score, which was used as the overall measure of subjective workload.

Procedure

First meeting. The purpose of this first session was to explain the details of the research, obtain verbal consent for participation, and prepare for the use of the NASA-TLX. The volunteer met the researcher at Highland Hospital ED where they were informed about the details of the study and were provided with the consent form. After the volunteer read the consent form and verbal consent was obtained, the participant was provided with a background and instructions for using the NASA-TLX (Appendix E). Participants were then provided with the definitions for each of the six NASA-TLX rating subscales (Hart & Staveland, 1988) and were asked to read through them. At this time, participants were encouraged to ask questions, if they had any, about the scales. In order to gather appropriate weightings for each subscale, participants were provided with 15 index cards. Each card listed one pair of subscales (ex, Mental Demand/ Performance) and participants were asked to circle the one that contributed most to the subjective workload they experienced in their job. Following the collection of pair ratings, arrangements were made for scheduling of the observation session/s and final interview. Depending on the provider's schedule, the first observation session either began immediately after this initial meeting or within the next 1 or 2 days.

Observation sessions. During each observation session, the researcher met the provider at Highland Hospital ED and began following the provider while he or she performed normal work activities, including direct patient care. The researcher recorded the tasks being performed by the provider as well as the length of time spent on each task. Interruptions were also recorded. A sample of one hour of observational data is presented in Appendix H. At the end of each hour of observation, the NASA-TLX was administered. Hospital and emergency department data was printed from the electronic information system and retrieved simultaneously or immediately following from the ED PulseCheck system.

Interview sessions. Interview sessions were scheduled after a participant completed the observation session/s. Interviews were conducted one-on-one in the break room or a quiet portion of the ED. During each interview session, the participant was asked to answer structured interview questions regarding the effects of ED crowding on their perceived workload and time pressure (Appendix F) and to clarify any questions the researcher had regarding the observations. Interviews were audio-recorded with participant consent.

Results

A total of forty-two hours were observed. Data presentation is based on one-hour observation sessions. During these sessions, which occurred between 8:00 am and 1:00 am, a total of 901 patients were triaged at the ED. Note that depending on the variable recorded, there are missing data for some hours of observation. After each hour of observation, the researcher had to make a decision regarding data collection. In some circumstances, interruption of the participant to collect subjective data was unwarranted and unsafe. In other circumstances, since computers were shared, operational ED data were not collected because there was no access to a computer at the time of data collection. Statistical analyses were performed using PASW v.18 for Windows (SPSS, Inc., Chicago, IL).

Descriptive Statistics

ED variables. Mean, standard deviation, and range are reported in Table 2 for each ED crowding variable. In general, ED variables were quite variable which means that observational and subjective data can be viewed against a range of ED states. Based on the values for mean occupancy (.59) and mean EDWIN score (0.82), it would appear that the ED, on average, was not near capacity. It is important to note, however, that ED occupancy was calculated based on the number of total licensed beds and does not reflect the number of beds that were staffed. It is possible, and even likely, that the true proportion of occupied licensed and staffed beds reached 1.0. For this reason, it is also important note that the average number of patients waiting was six. Based on the criteria established by Schneider et al (2003), the calculated patient/nurse ratio (6.03) represents a personnel shortage. ESI statistics are presented separately as this is not considered a crowding variable.

Table 2

Descriptive Statistics for Emergency Department Variables

ED Variable	Mean	SD	Minimum	Maximum
Occupancy	.59	.16	.31	.95
EDWIN Score	.82	.30	0.45	1.90
Patients Waiting	6.00	4.74	0.00	18.00
Patient/Doctor Ratio	10.14	4.01	5.33	25.00
Patient/Nurse Ratio	6.03	1.80	3.25	10.00
Total Providers on Duty	11.30	1.94	7.00	15.00
No. Patients Assigned	5.19	1.75	1.00	8.00

Note. N = 37. All data refer to one-hour observation periods.

ESI. The mean ESI was 3.35 (SD = .13, N = 37) and was relative stable across different participants and times of day ranging from 3.13 to 3.36. The small range is likely due, in part, to the small number of patients assigned to levels one (n = 0) and level five (n = 27). In contrast, 300, 532, and 42 patients were assigned to triage levels 2, 3, and 4, respectively.

Observations. Observed tasks were originally assigned to one of twenty-three codes. After reviewing the data, it was apparent that some codes could be combined into more meaningful and useful groupings. Some tasks, such as "Admissions", were unique enough to warrant their own code. New groupings were based on the researcher's observations and best judgment. For example, the original codes of patient history and exam (HE), treatment of patient (TP), verbal communication with patient (VCP) and patient family (VCF) were all combined into the group "Direct Patient Care". The final code groupings are presented in Table 3. All subsequent analyses of tasks will refer to these task groupings.

Table 3

Original Task Codes and Final Task Code Groupings

Task Group	Original Code
Communication With Staff	CON, VCS, IC, LM, OC, Meet, Teach
Direct Patient Care	VCF, VCP, HE, TP
Paperwork	Chart, WO, WD, Cultures, Paperwork
Checking Test Results	CT
Admissions	AP
Check Electronic Whiteboard	CB
Look For Supplies/Equipment	LSE
Break	Break
Other	MR, PR, Transport Paper, Other

Note. Original task codes were consolidated into more meaningful task groups. The final task groupings were used in all subsequent analyses.

Time-on-task. The amount of time spent on individual tasks was analyzed to better understand how providers' time is spent. Mean, standard deviation, and range are presented for the proportion of time/hour spent on each task group in Table 4. The amount of time spent on a given task was quite variable. The three most common activities, communication with staff, direct care, and paperwork accounted for more than 75% of the participants' time.

Table 4

Descriptive Statistics Time Spent (Minutes) On Each Type of Task

Task Group	Mean	SD	Min.	Max.
Communication With Staff	16.07	8.57	1.00	37.00
Direct Care	15.60	8.93	.00	32.00
Paperwork	13.00	7.16	.00	29.00
Checking Test Results	3.48	3.20	.00	13.00
Admissions	3.38	6.22	.00	24.00
Break	2.92	3.95	.00	18.00
Looking for Supplies/Equipment	2.14	2.37	.00	10.00
Check Electronic Whiteboard	1.48	1.85	.00	8.00
Other	1.93	2.74	.00	14.00

Note. N = 42. All data refer to an average one-hour observation periods.

Task count. Task count, or the number of times a task was initiated, was compared to the average time spent on that task. This comparison is similar to the one presented by Levin et al. (2006). Task count and time-on-task data were relatively congruent with the exception of patient communication with staff and direct patient care. Over 25 percent of time was spent on direct patient care while only 16 percent of all initiated tasks involved direct patient care. In other words, although more time was spent on direct patient care, the task itself was not as frequent. Task count and time-on-task are represented, respectively, as a percentage of total task count and time-on-task for each task category in Figure 2.

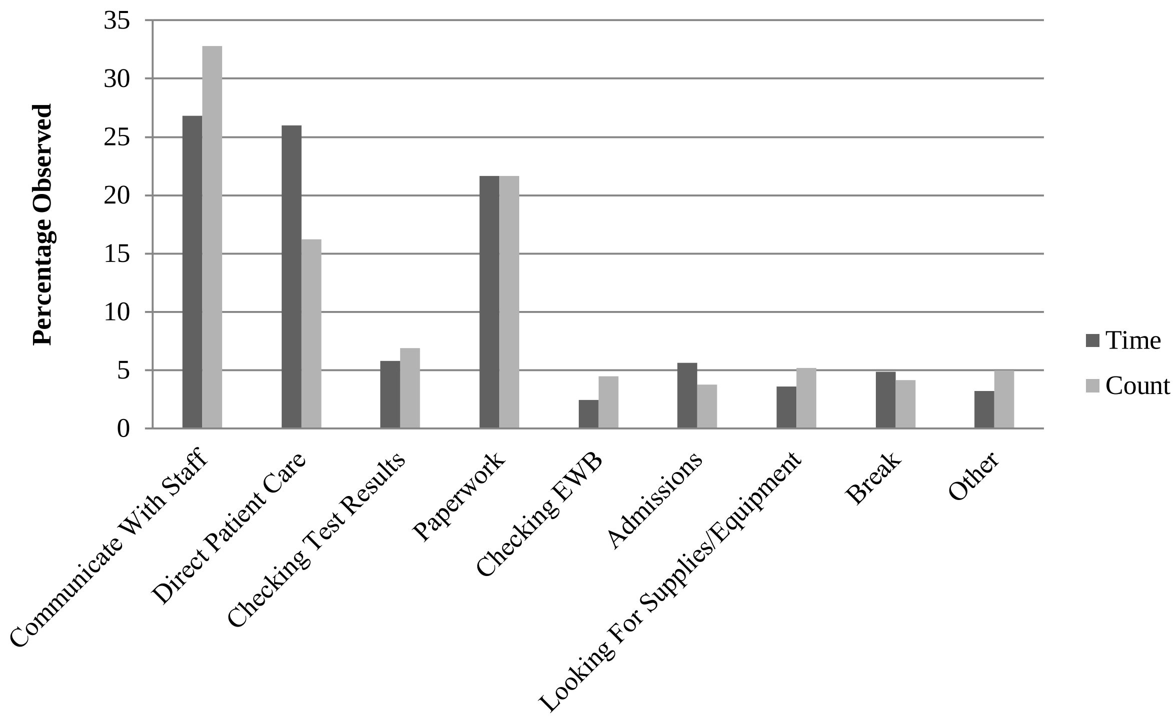

Figure 2. Percentage of observed task count and time-on-task for each task category. This figure compares the time spent on a given task to its frequency (number of times initiated). Mean time-on-task and frequency (count) are presented as percentages for each task.

Interruptions. The number of times a physician was interrupted was analyzed for each observation hour. The number of observed interruptions varied greatly but, in general, there were relatively few (Mean = 3.4, SD = 2.47) per each hour of observation. Face-to-face interruptions were the most commonly observed interruptions. Most often this was a nurse or other health care provider inquiring about a patient's status or paperwork. Mean, standard deviation, and range are presented in Table 5 for each interruption type.

Table 5

Descriptive Statistics for Interruptions per Hour (In Counts)

Interruption Type	Mean	SD	Min.	Max.
Face-to-face	2.69	2.10	.00	11
Phone	.69	.99	.00	4
Pager	.02	.15	.00	1
All	3.40	2.47		

Note. N = 42. Interruptions were recorded and logged by type (face-to-face, phone, or pager). All data refers to an average one-hour observation period.

Subjective workload. Each participant rated the contribution of each NASA-TLX subscale dimension contributed to their subjective job workload. Participants provided the following average ratings (scale is 1 to 5) for Mental Demand (2.83, SD = 1.30), Temporal Demand (3.17, SD = 0.81), Effort (2.75, SD = 1.44), Frustration (3.95 SD, = 1.40), Performance (2.25, SD = 1.18), and Physical Demand (0.08, SD = 0.28). Note that the mean Physical Demand weight was extremely low which reflects its lack of contribution to the participants' perceived workload. Each raw subscale rating (1-100) was multiplied by the weight (1-5) given by the participant for that subscale. The subscales were used to compute the overall NASA-TLX score. The overall NASA-TLX score, which represents overall subjective workload was quite variable and ranged from 18.67 to 75.00 (Mean = 51.84, SD = 14.54, N = 36). Mean, standard deviation, and range are presented in Table 6 for each of the six weighted subscales ratings.

Table 6

Descriptive Statistics for Subjective Workload Subscale Ratings (NASA-TLX)

Subscale	Mean	SD	Min.	Max.
Mental Demand	53.71	20.76	10	95
Temporal Demand	55.05	20.16	17	97
Physical Demand	32.60	24.80	3	92
Frustration	56.26	21.87	14	98
Performance	31.86	15.38	5	75
Effort	59.48	17.20	25	98

Note. N = 42. The maximum possible rating is 100. All data refers to an average one-hour observation period.

Correlations

ED variables and observations. Pearson correlations were performed to assess the relation of ED operational variables and observed variables (time-on-task, task count, and interruptions).

ED variables & time-on-task. An exploratory analysis was conducted to determine the correlation between ED variables and objective time-on-task data (for each task category). All time-on-task data was measured as the proportion of time spent on a task for a given hour. Time-on-task data was not significantly correlated with ED crowding, as measured by EDWIN and occupancy, for any of the task categories. Results, however, did reveal significant correlations between other ED variables and time-on-task data (Table 7).

Table 7

Correlation Matrix: Pearson Correlations of ED Operational Variables and the Proportion of Time Spent on Tasks

	Waiting Room	Providers on duty	P/D ratio	P/N ratio	Patients assigned	ESI
Communicate	-.13	.01	-.13	-.09	.07	-.10
Direct Care	.02	-.07	-.01	.06	-.42**	.11
Check Test	-.08	.06	-.05	-.09	.29	-.07
Paperwork	-.02	-.33**	.04	-.01	.13	-.08
EWB	.19	.17	.21	.27	-.25	-.04
Admissions	.10	.32*	.09	-.05	.41**	.13
Look for S/E	.17	.05	.14	.14	-.24	-.10
Break	.08	-.07	.00	.20	.24	.00
Other	-.02	.15	-.07	-.13	-.18	.04

Note. N = 37. * p < .10, ** p < .05, *** p < .01 (2-tail)

*ED variables & task counts***.** An exploratory analysis was conducted to determine the correlation between ED variables and objective task count data (for each task category). Task count for admission was positively correlated with ED crowding (Occupancy and EDWIN), $r(37) = .361$, $p = .033$. Except for admission, none of the other task counts were significantly correlated to crowding. Results, however, did reveal significant correlations between other ED variables and task count data (Table 8).

Table 8

Correlation Matrix: ED Operational Variables and Task Counts (Frequencies)

	Patients waiting	Providers on duty	P/D ratio	P/N ratio	Patients assigned	ESI
Communicate	-.08	-.05	-.15	-.11	-.02	-.11
Direct Care	-.10	-.04	-.12	-.02	-.29*	.03
Check Test	-.10	-.07	-.09	-.10	.30*	-.17
Paperwork	.00	-.39**	-.00	-.11	.09	-.08
EWB	.33*	.12	.28	.40**	-.33*	-.13
Admissions	.18	.31*	.16	.03	.38**	.03
Look for S/E	.11	-.12	.31*	.13	-.26	-.21
Break	-.10	-.07	-.15	.00	-.03	.17
Other	-.22	.06	-.19	-.24	-.12	-.04

Note. N = 37. * $p < .10$, ** $p < .05$, *** $p < .01$ (2-tail).

ED variables & interruptions. Total interruptions were not significantly correlated to any of the ED variables.

Observations and subjective workload. Pearson correlations were performed to assess the relation of observational data (time-on-task, task count, and interruptions) and subjective workload.

Time-on-task & subjective workload. Subjective workload, as measured by the NASA-TLX, was positively correlated to the proportion of time spent viewing the electronic white board, $r(36) = .35$, $p = .037$. Subjective workload was negatively correlated to the proportion of time spent communicating with staff, $r(36) = .38$, $p = .152$, and taking breaks $r(36) = -.41$, $p = .014$.

Significant correlations were also found between time-on-task and NASA-TLX subscales (Table 9).

Table 9

Correlation Matrix: Proportion of Time Spent Performing Tasks with NASA-TLX Subscale Ratings (Un-Weighted)

	Mental Demand	Temporal Demand	Physical Demand	Effort	Frustration	Performance
Communicate	-.05	-.39**	-.07	-.29*	-.28*	-.16
Direct Care	-.09	.32**	.23	.05	.15	-.00
Check Test	.07	-.31**	-.12	-.06	-.29*	-.28*
Paperwork	-.01	.30*	-.05	.22	.25	.14
EWB	.01	.33**	-.06	.05	.20	.34**
Admissions	.36**	-.23	-.07	.24	.01	.13
Look for S/E	-.12	.15	.41***	-.06	.27*	.08
Break	-.32**	-.34**	-.08	-.37**	-.28*	-.20
Other	.11	.44***	-.31**	.25	.07	.19

Note. All correlations are Pearson correlations. * $p < .10$, ** $p < .05$, *** $p < .01$ (2-tail). N = 42.

Task count & subjective workload. Subjective workload, as measured by the NASA-TLX, was positively correlated to the frequency of viewing the electronic white board, $r(42) = .37$, $p = .026$, and direct patient care, $r(42) = .35$, $p = .037$. Subjective workload was negatively correlated with the frequency of taking a break, $r(42) = -.37$, $p = .025$. Significant correlations were also found between task count (frequency) and NASA-TLX subscales (Table 10).

Table 10

Correlation matrix: Task Frequency (Count) and NASA-TLX Subscale Ratings (Un-Weighted)

	Mental Demand	Temporal Demand	Physical Demand	Effort	Frustration	Performance
Communicate	.03	-.08	.12	-.06	-.02	-.17
Direct Care	.28*	.46***	.45***	.28*	.11	-.16
Check Test	.17	-.16	.11	.06	-.21	-.20
Paperwork	.08	.33**	.02	.28*	.25	.01
EWB	-.05	.33**	.04	.06	.24	.34**
Admissions	.27	-.06	.03	.30*	.15	.06
Look for S/E	-.12	.09	.29*	-.18	.12	.05
Break	-.34**	-.30*	-.05	-.26	-.20	-.16
Other	.13	.40***	-.29*	.20	.02	.12

Note. N = 42. * $p < .10$, ** $p < .05$, *** $p < .01$ (2-tail).

Interruptions & subjective workload. Interruptions were not correlated to overall subjective workload, however, they were positively correlated to ratings on the effort subscale (NASA-TLX), $r(42) = .38$, $p = .013$).

ED variables and subjective workload. Pearson correlations were performed to assess the relation between ED operational variables and subjective workload ratings (NASA-TLX).

ED crowding and overall subjective workload. An exploratory analysis examined the correlation between ED crowding variables (EDWIN and occupancy) and overall subjective workload (NASA-TLX score). As expected, the analysis revealed a significant correlation between the two crowding variables, $r(37) = .572$, $p <. 001$. The analysis also revealed a significant positive correlation between occupancy and subjective workload, $r(32) = .575$, $p < .001$. The correlation between EDWIN and subjective workload was positive and approached significance, $r(33) = .337$, $p = .055$.

ED Crowding & NASA-TLX Subscale Ratings. An analysis was performed to determine how the NASA-TLX subscale ratings contributed to the correlation between subjective workload (NASA-TLX) and ED crowding, and to what extent they contributed. For this analysis, ED crowding was defined by ED occupancy. The results revealed significant positive correlations between ED crowding and 4 of 6 NASA-TLX subscales. Note that the higher the Performance rating was, the greater the perceived failure was. Temporal Demand, $r(37) = .33$, $p = .047$; Performance, r(37) = .40, p = .015; Effort, r(37) = .42, p = .01; and Frustration, r(37) = .40, p = .013 were all positively correlated to ED crowding. The analysis also revealed non-significant positive correlations between ED crowding and Mental Demand and ED crowding and Physical Demand. Figure 3 displays the strength of each NASA-TLX subscale and ED occupancy.

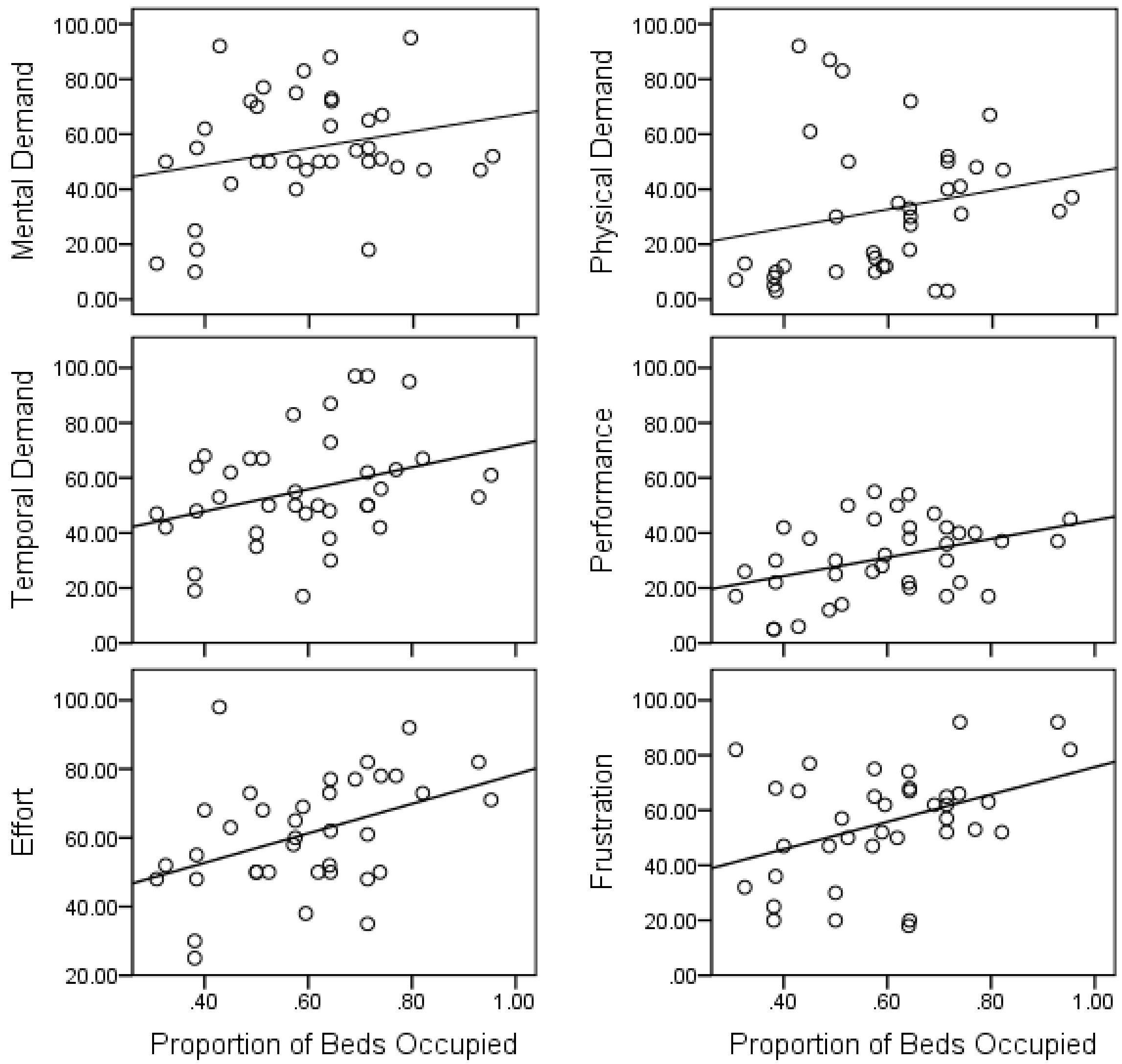

Figure 3. NASA-TLX subscale ratings as a Function of ED occupancy. The scatterplots represent the correlation between ED occupancy and ratings from each of the 6 NASA-TLX subscales.

Other ED variables & overall subjective workload. An analysis was conducted to assess whether or not overall subjective workload was correlated to other ED operational variables (providers on duty, patients assigned, patient/doctor ratio, and patient/nurse ratio, and number of patients in the waiting room). Subjective workload was positively correlated to the number of patients in the waiting room, $r(33) = .381$, $p = .029$. The correlation between subjective workload and providers on duty, $r(33) = .321$, $p = .069$; patient/doctor ratio, $r(33) = .304$, $p = .086$; and patient/nurse ratio $r(33) = .299$, $p = .091$ were all positive and approaching significance. There was no significant correlation between subjective workload and the number of patients assigned.

Other ED variables and NASA-TLX subscale ratings. In addition to the correlation between ED crowding and overall subjective workload, several significant correlations were also revealed between the other ED variables and the NASA-TLX subscale ratings (Table 11).

Table 11

Correlation Matrix: ED Operational Variables and NASA-TLX subscale Ratings (Un-Weighted)

	Mental Demand	Temporal Demand	Physical Demand	Effort	Frustration	Performance
Pt. assigned	.17	-.36**	-.11	.02	.12	.16
Providers	.18	.16	-.05	.26	.15	.34**
Pt. waiting	.05	.13	.24	.13	.41**	.21
Pt./doc ratio	.07	.06	.01	-.05	.33**	.14
Pt./nurse ratio	-.00	.09	.16	.07	.25	.32*
ESI	.08	.15	-.02	.09	-.42**	-.11

Note. N = 37. * p < .10, ** p < .05, ***, p < .01 (2-tail).. Emergency Severity Index (ESI).

Regression Analyses

ED crowding and objective task load index (OTLX). The OTLX was calculated in an effort to define objective workload in terms of a single representative and standardized value. The OTLX was based on proportion of time spent in each task and the weighting of the tasks according to their contribution to overall workload.

Objective task load index (OTLX). A standardized index (values between 0 and 1) for workload was calculated using the following algorithm:

$$\frac{\sum \mathrm{t}_i \mathrm{w}_i / \mathrm{OT}}{\mathrm{W}_{\max}} \tag{3}$$

For each subject and 60 min observation period, the time spent in each task (t_i) was multiplied with a weight associated with that task (w_i), the products added, the sum divided with total observation time (OT), and finally the resulting average workload divided with the maximum possible weight value (W_{max}) in the scheme used, which standardized the OTLX index between 0 and 1. Three different weighting methods were applied 1) weights estimated from the observations in this study, 2) weights modified from Bierbaum, Szabo, and Adrich's (1989) UH-60 crew member workload analysis, and 3) weights based on the coefficients from multiple regression (MR) of overall subjective workload scores (NASA-TLX) against the proportion of time spent on each task.

Method 1. The first task weighting method involved simple estimation on the part of the researcher. Following the completion of the study, the researcher assigned each task a weight from 1 to 5 on an ordinal scale. The weights were based on observation of the tasks being performed in the context the ED environment.

Method 2. Task weights were based on a workload analysis of tasks performed by UH-60 mission aircraft operators. Bierbaum et al. (1989) assessed the mission task demands with respect to Wickens (1984) theory of workload as a multidimensional construct. A key component of this analysis was to estimate the workload associated with the sensory, cognitive, and psychomotor components of each task needed to complete the mission. A described in their research, "the sensory component refers to the complexity of the visual (V), auditory (A), and/or kinesthetic (K) stimuli to which the operator must attend; the cognitive (C) component refers to the level of information processing required from the operator; the psychomotor (P) component refers to the complexity of the operator's behavioral responses", (Bierbaum et al., p. 12). Since the SOF modifications included the use of night vision goggles (visual-aided), the additional sensory component (G) was added.

For each of the aforementioned workload components, subjective judgments were used to derive estimates of the operator's workload associated with each mission task. A pair comparison survey was designed by matching the verbal anchors for each workload component by all other components for a total of 21 pairs. Pairs were presented randomly to subject matter experts. The frequency for which one an anchor was chosen (over another) was recorded and used to compute the rating for each verbal anchor on an approximately equal-interval scale. Verbal descriptors of workload were identified for each task and matched to the appropriate verbal anchor. Numerical estimates of workload were then assigned to represent the level of workload for a particular task component. For example, under the cognitive component, the verbal descriptor "evaluation/judgment (consider several aspects)" is assigned a value of 6.8. With findings from the analysis, they developed a computer model to predict UH-60 operator workload. Workload estimates from this study will be used to analyze tasks for the current study.

For the current analysis, each of the primary tasks was matched to a descriptor/s from the UH-60 task analysis (Bierbaum, 1989) that best represented the task. For example, the task "checking test results" was matched to the following descriptors from that analysis; visually register/detect (detect occurrence of an image), visually inspect/check (discrete inspection/static condition), and visually discriminate (detect visual differences). Workload rating values for those descriptors were reported as 1.0, 4.0, and 3.7, respectively. These values were summed and averaged to calculate a workload value of 2.9 for "check test results". The same procedure was used for the other tasks.

Method 3. A multiple regression (MR) of the overall subjective workload (NASA-TLX) was run against the proportion of time spent on a given task. The resulting Beta coefficients for each task were standardized and used as weights in the calculation of the OTLX.

The results of these weighting structures are outlined in Table 12. Note that the different weighting structures result in very different task rankings. For instance, method one ranked admissions as the most difficult task, whereas methods two and three ranked looking for supplies and equipment and "other" as the most difficult tasks, respectively.

Mean OTLX scores were similar for all three weighting methods. Method one weighting resulted in a mean objective task load index of 0.61 (SD = 0.07) with scores ranging from 0.46 to 0.77. Method two weighting resulted in a mean objective task load index of 0.68 (SD = 0.08) with scores ranging from 0.37 to 0.74. Method three weighting resulted in a mean objective task load index of 0.50 (SD = 0.06) with scores ranging from 0.29 to 0.60.

Table 12

Task Weights Used, By Method, to Calculate Objective Task Load Index (OTLX)

	Weights		
Task Group	Method 1	Method 2	Method 3
Communication With Staff	3.0	3.0	.40
Direct Care	4.0	4.5	.49
Paperwork	2.0	6.5	.73
Checking Test Results	2.0	2.9	.11
Admissions	5.0	4.2	.78
Break	1.0	1.0	.00
Looking for Supplies/Equipment	4.0	7.0	.72
Check Electronic Whiteboard	3.0	4.0	.04
Other	2.5	4.0	.90

Note. Method one task weights (1-5) were estimated from the observations in this study. Method two task weights (1-7) were modified from the UH-60 crew member workload analysis (Bierbaum et al., 1989). Method three task weights (0-1) were based on the coefficients produced from the MR of proportions of times spent in each task against the NASA-TLX scores.

OTLX scores from each weighting method were regressed, separately, against ED crowding. Results revealed that ED crowding (EDWIN and occupancy) did not account for a significant proportion of the variance in objective task load (as measured by OTLX). In fact, ED crowding explained less than 4% of the variance in the scores.

ED crowding and subjective workload. A regression analysis was performed to examine the predictive value of the EDWIN score and the occupancy for determining provider subjective workload, as measured by the NASA-TLX. The EDWIN score did not explain the proportion of variance in subjective workload, $R^2_{adj} = .09$, $F(1, 32) = 4.00$, $p = .055$. ED occupancy, however, explained a significant proportion of variance in subjective workload, $R^2_{adj} = .31$, $F(1, 32) = 15.35$, $p < .001$ (Figure 4).

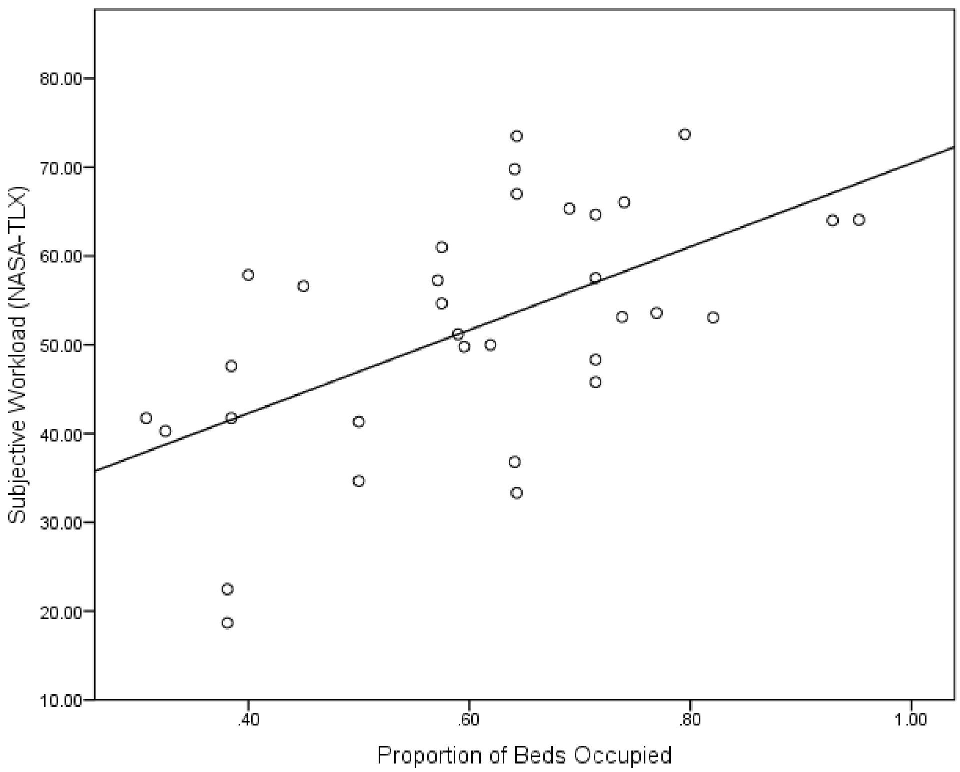

Figure 4. Regression of subjective workload (NASA-TLX) against ED occupancy. $R^2_{adj} = .31$, $p < .001$. N = 33.

Provider Differences

The previous analyses have considered providers as one group. However, it may be important to distinguish between provider levels, especially in terms of subjective workload ratings. For this reason, an analysis was conducted to compare subjective workload ratings by provider level; physician or physician assistant. Since there were only three data points associated with nurse practitioner ratings, these were not included in the analysis. To assess possible differences in overall subjective workload by provider level, an independent samples t-test was conducted on NASA-TLX scores. Mean ratings for physicians and physician assistants (PAs) were 49.03 (SD = 16.68) and 55.47 (SD = 12.87), respectively, and were not significantly different, $t(31) = -1.22$, $p = .271$ (Figure 5).

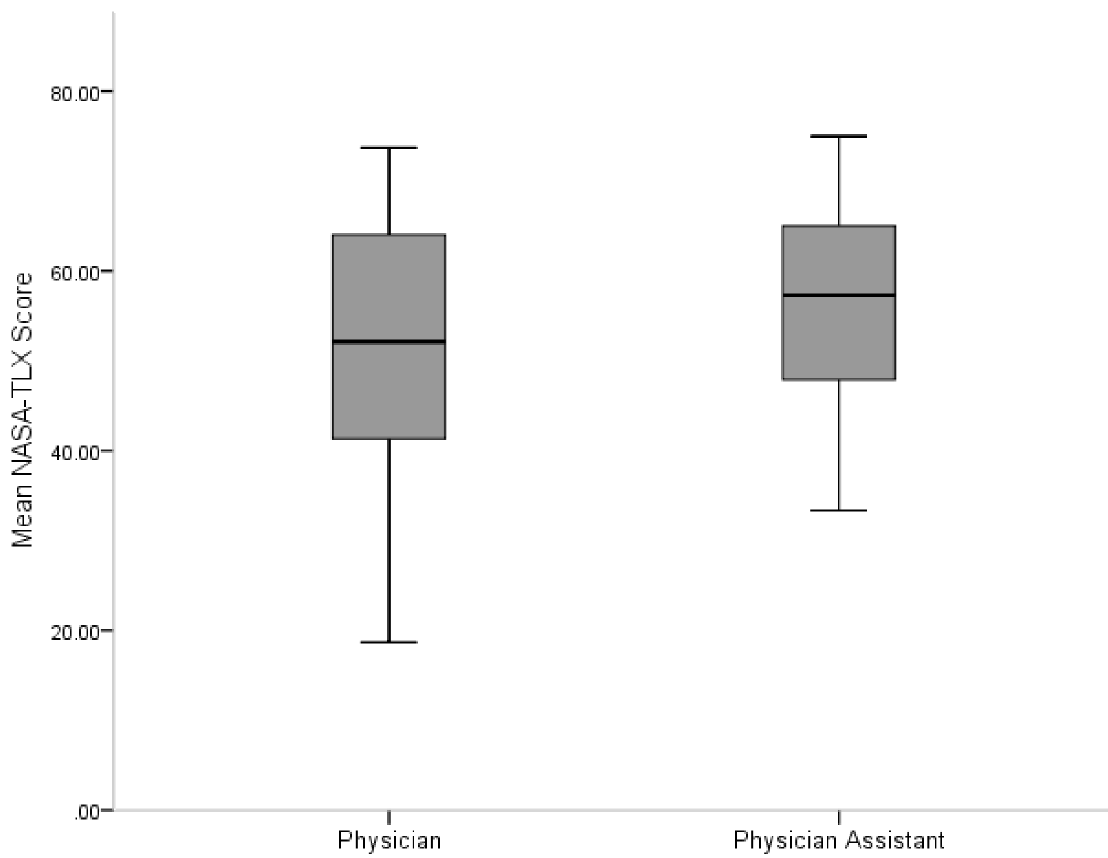

Figure 5. Mean NASA-TLX scores by provider level. N = 37.

A second analysis assessed possible differences in NASA-TLX subscale ratings by physicians and PAs. Independent samples t-tests were conducted on ratings for each of the NASA-TLX subscales. Except for temporal demand ratings, there were no significant differences in ratings by physicians and PAs. Mean ratings of temporal demand by physicians (50.33, SD = 17.46) and PA's (63.87, SD = 23.15) were significantly different, *t*(37) = 2.07, *p* = .049. Figure 6 displays the mean ratings, by provider level, for each NASA-TLX subscale.

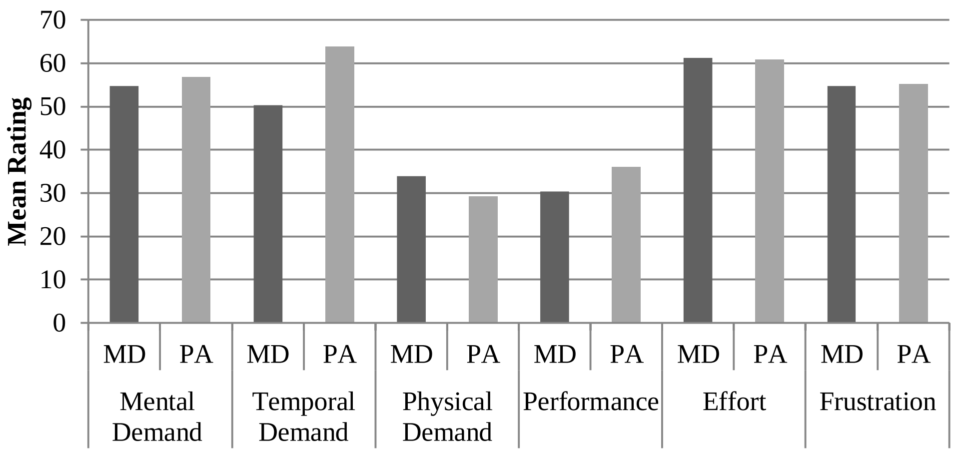

Figure 6. Mean NASA-TLX subscale ratings by provider level. N = 37.

Interviews

Nine of the ten participants completed the interview sessions. The following paragraphs provide a summary of the information gathered from the interview.

Routine and non-routine tasks. The most commonly reported routine tasks included ordering and checking labs/test, getting a patient history, and performing basic procedures such as suturing. The most commonly reported non-routine tasks included differential diagnoses, and more complex or rare procedures such as lumbar puncture and central lines.

Interruptions. Most providers reported that, while many interruptions were necessary and part of the job, other interruptions, such as an incoming call regarding a previously seen patient, were regarded as unnecessary. Eight out of the nine participants felt that interruptions affected their work. They reported that they felt interruptions made it more difficult to return to previously deferred tasks and that it resulted in work that was less efficient.

Time pressure. The providers reported that while some tasks have clear end points (suturing a laceration), other do not (obtaining patient history). When asked what prompts them to switch from one task to another, most providers reported that they most often switched tasks as a natural result of the ending of a previous task or as a result of prioritizing the most urgent task. When asked what proportion/percentage of tasks the providers thought they need to return to in order to complete, the answers ranged from 30 – 100%. The most commonly reported methods for coping with time pressure were list creation (prioritization) and taking a mental break.

Task shedding. In general, the providers reported that they avoided postponing tasks indefinitely or passing them off to other providers. In circumstances when a task absolutely needed to be transferred or postponed indefinitely, providers reported that the decision was most often based on urgency.

Time estimation. When asked whether or not they felt they had an accurate sense of the amount of time it tasks to perform a specific task, 3 reported "yes", 1 reported "no", and 5 reported "sometimes". All providers reported that they felt that their sense of time changes based on task complexity and the number of tasks they are performing. In general, they feel that time seems to pass more quickly as task complexity and load increase.

ED crowding. The definitions reported by the providers for ED crowding varied but almost all referred to the number of unseen patients (e.g. in the waiting room). All providers reported that they felt as though ED crowding affected the demands placed on them while six of the nine reported that it affected the effort they put forth. When asked if they thought that it affected their performance, six reported "yes", one reported "no" and two were unsure. Those who reported "yes" felt as though ED crowding made it more likely they would miss something important.

Discussion

This observational field study attempted to quantify the objective task load imposed on ED providers, determine the degree of subjective workload they experience, and to correlate these data with ED operational metrics; mainly ED crowding metrics. Cognitive task analysis techniques of task coding and time-on-task recording were used a methods to examine ED providers' objective task load as it relates to ED crowding and subjective workload.
Cognitive task analysis revealed that the providers spent 75 percent of their time performing tasks related to communication with staff, direct patient care, and paperwork. The other 25 percent of their time was spent checking test results, admitting patients to the hospital, taking breaks (this included food and bathroom breaks), looking for supplies, checking the electronic whiteboard, and other job-related tasks.

Hollingsworth, Chisholm, Giles, Cordell and Nelson (1998) conducted a time-and-motion study of 39 ED providers (faculty physicians, residents, and nurses). They observed and recorded several provider activities and categorized them as one of three types; direct patient care, indirect patient care and personal activities. Results from this study and the study conducted by Hollingsworth et al. (1998) revealed that the majority of providers' time was spent on tasks related to indirect patient care. For the current analyses, communication with staff, paperwork, checking test results, admissions, checking the electronic whiteboard, and looking for supplies and equipment were considered indirect patient care activities. A comparison of the general findings between these two studies can be seen in Table 13. Note that the current study reports much less time spent on personal activities. Hollingsworth et al. (1998) reasoned that the time spent on personal activities was much higher for emergency nurses and may be a result of needing to "pace" themselves for a 12-hour shift. Except for one nurse practitioner, the current study did not include emergency nurses which may explain why these percentages are so different.

Table 13

A Comparison of Results from This Study and Results from Hollingsworth et al. (1998) of the Percentage of Provider Time Spent on Direct Patient Care, Indirect Patient Care and Personal Activities

	Coles	Hollingsworth (1998)
Indirect Patient Care	69%	47%
Direct Patient Care	26%	32%
Personal Activities	5%	21%

ED Variables and Observed Variables

Hypothesis 1 was not supported. ED crowding was not significantly correlated to objective time-on-task or interruption data. ED crowding (occupancy) was, however, significantly correlated to the frequency of performing admissions related tasks. Not surprisingly, the more patients assigned to providers, the less time the providers spent on direct patient care and the more time they spent on patient admissions. The amount of time providers spent on paperwork was negatively correlated to the number of providers on duty.

Observed Variables and Subjective Workload Variables

Hypothesis 2 was partially supported. Overall subjective workload was positively correlated only to the proportion of time spent viewing the electronic whiteboard and was negatively correlated with the proportion of time spent taking breaks and communicating with staff. This was not necessarily unexpected. Since observations were recorded as proportions, more time spent on one task naturally resulted in less time spent on other tasks.

Since the results are correlational, it is unclear whether providers chose to take breaks and communicate with other staff members when they felt their workload was low or whether spending more time on break or communicating with other staff members resulted in lower perceived workload. In terms of task frequency, overall subjective workload was positively correlated to viewing the electronic white board and direct patient care and negatively correlated with the frequency of taking a break. Overall subjective workload was not correlated to the number of observed interruptions.

ED Variables and Subjective Workload Variables

Hypothesis 3 was supported. Subjective workload was positively correlated with ED crowding metrics. In fact, the links between ED crowding and subjective workload were the strongest of all links revealed in this study. ED occupancy was found to explain over 30 percent of the variation in overall subjective workload. The link between ED crowding and subjective workload was driven by the moderate correlations between ED occupancy and subjective workload assessed by NASA-TLX subscales, Temporal Demand, Performance, Effort, and Frustration. Overall subjective workload increased as the ratio of patients to doctors and patients to nurses increased. It is also important to mention that the moderate correlation between subjective workload and the number of patients in the waiting room as this this was a commonly reported source of worry for the providers.

The positive correlation between the number of providers on duty and subjective workload was unexpected and may be explained by staffing of additional providers during peak times. Another unexpected finding was the lack of a relationship between subjective workload and the number of patients assigned. It is possible that patient characteristics (acuity, status, and behavior) are more closely related to the workload experienced by providers than patient quantity. Patient characteristics, however, were not recorded in this study.

Measurement of Interruptions.

Interruptions are difficult to classify and quantify. In order to classify an interruption as such, one must determine when is an interruption and interruption rather than an inherent element of the task. Making this determination within the ED environment is particularly difficult as some interruptions, such as a return phone call from an admitting physician, are expected. If interruptions are predictable or expected, are they interruptions? Interviews with providers strongly suggest that many interruptions are, in fact, part of the job and often guide them to their next task. This clearly supports the classification of ED physicians by Chisholm et al. (2000) as "interrupt-driven".

The task of quantifying interruptions relies on the interruption having an observable manifestation and not all interruptions are observable. Internal interruptions, or "self" interruptions may be the result of suddenly remember something. On observation, the person being observed may appear to shift quickly from one task to another without any apparent reason. There is no easy way to determine whether or not the person was, in fact, interrupted without asking. It was not feasible in ED environment to capture and quantify these types of self-interruption without adding to the external interruptions. Brixey et al. (2007) resolved this issue by defining self-interruptions as instances when "the subject stopped performing the initial task and performed an interrupting activity without provocation from a source outside the subject" (p. 5). This is a clear operational method for defining a self-interruption but it relies on the assumption that the act of switching a task, before it is complete, is a result of an interruption which may or may not be true.

An interruption was defined, conservatively, in the current study as an event resulting in disengagement of the original task that was unrelated to the original task. The key difference is that an interruption must be unrelated to the task at hand. However, even interruptions that are observable may not reflect the task for which they are interruption (i.e. thought processes). Furthermore, merely counting interruptions provides for only a very limited measure of the severity of an interruption; what is interrupted and the elapsed time before the original task can be resumed may be more detrimental to performance.

Issues with Crowding Metrics

It is important to reiterate that the mean reported occupancy and EDWIN values were relatively low. Based on these values, it would appear that the ED, on average, was not near capacity. It is important to note, however, that ED occupancy was calculated based on the number of total licensed beds and does not reflect the number of beds that were staffed. It is possible, and even likely, that the true proportion of occupied licensed and staffed beds reached 1.0. The study was conducted during a remodeling of the ED which may have rendered some beds theoretically, but not practically, available. Although the occupancy value may be affected by this remodeling effort, relative occupancy should not. In light of this issue, it is important to look at other indicators of crowding. Additional evidence of crowding comes from the average number of patients waiting to be seen (six) and the average patient nurse ratio (6.03) which, according to Schneider et al (2003), represents a personnel shortage.

The relation of occupancy to subjective workload

The linear relation of occupancy to overall subjective workload was made clear in this study however the certainty of this linear relationship comes into question when you consider that extreme crowding may not have been captured. Although the occupancy reached a maximum of 0.95, the limit of this metric can and, in many cases of crowding, does exceed 1.0. A national survey conducted by Schneider et al. (2003) revealed an average of 1.1 patients per treatment space. It is unclear, based on the current results, whether this linear relationship would have persisted given more severe crowding conditions. For instance, one might predict that subjective workload increases in relation to crowding up to a point and after that point increases exponentially.

Quantification of Objective Task Load

ED crowding did not account for a significant proportion of the variance in objective task load (as measured by OTLX). In fact, ED crowding explained less than 4% of the variance in the scores. There are a two possible explanations for these negative findings 1) the method of deriving the OTLX weights was incorrect (2) the raw data used in the calculation of the OTLX scores did not capture the components of workload.

In regard to deriving OTLX weights, it can be argued that method one relied too heavily on the observations made by one person, method two relied too heavily on data weights derived from other domains, and method three relied solely on statistical outcomes from a multiple regression. As for this last method, the number of variables entered exceeded the number recommended based on the number of available data points. A multiple regression of only one independent and dependent variable usually requires 30 observations or data points and an additional 10 observations per each additional independent variable.

The idea that the raw data did not capture the components of workload has two further possibilities. The first possibility is that the times spent on the different tasks were more or less independent from the factors included in the EDWIN score or occupancy and may remain relatively stable in response to crowding. Responses from the interview suggest that the time spent performing a task depends largely on how long that task takes to finish. For the most part, the providers' felt that although they may have endured more time pressure in response to situation of ED crowding, they did not necessarily rush a task and instead spent the necessary time on a task.

The second possibility is that there were little differences between the levels of difficulty of the separately identified tasks, and hence we could not expect to find correlation between the mostly homogenous mixture of tasks and subjective workload stemming from other sources. This latter point may be particularly true of weighting method one for which most weights fell between 2 and 4 and therefore didn't afford a substantial difference from un-weighted tasks.

Limitations of the Research.

This study was limited by its sample size, timing of observation sessions and available data collection methods. The use of a small convenience sample limited the number of cases (hours of data) available for analysis to 42 hours may limit the generalizability of the findings of this study. The small data set increases the likelihood that we may have made a type 2 error and missed smaller effects or links that were present. The timing of the observation sessions, although varied, was not random and did not equally represent all 24 hours for all days of the week. A special point was made, however, to include peak hours of operation in the analyses. As for data collection, time-on-task data was limited by the pen and paper method and was, therefore, recorded to the nearest minute. Considering the nature of the tasks, this type of rounding should not make a significant difference in the outcomes of the study.

Conclusions and Recommendations

Results from the study revealed no correlation between ED crowding variables and observational variables and ED crowding did not predict variations in objective task load, as measured by the OTLX methods. This suggests that either the method of computing OTLX scores needs improvement or that the time providers spend performing different tasks is truly independent of ED crowding. Although there is evidence that the latter may be true, further research should still focus on determining appropriate weighting structures for to use in the computation of OTLX. This could be handled by asking expert and novice providers to make pair-wise task comparisons.

ED occupancy was positively correlated to subjective workload and predicted 30 percent of the variance in subjective workload. The EDWIN score, on the other hand, only predicted 9 percent of the variance in subjective workload. In accordance with “Occam’s razor”, the principal that the simplest explanation is the correct explanation, ED occupancy may provide an advantage over more complex compound measures of ED crowding such as the EDWIN score in predicting provider subjective workload and may be more useful in making ED staffing and scheduling decisions. An effort should be made to distinguish the advantages and disadvantages of using more complex ED crowding metrics.

This study was unique in that the results came from a real-world operational ED and not a laboratory and although this study may not have captured extreme levels of ED crowding, the levels of crowding captured varied enough to allow for correlations with observed and subjective variables. In addition to collected and recorded variables, valuable insights were obtained from ED providers regarding issues of ED crowding, time-pressure and workload. It is apparent from their responses that, in the absence of observable changes in task load, the quantity and status of the "unseen" patient weighs heavily on their minds. Future research should assess the number of patients waiting or the number of patients who have left without being seen (LWBS) not only as a metric of ED crowding but as a predictor of ED provider workload.

References

American College of Emergency Physicians (ACEP) (2008). ACEP task force report on boarding, Emergency department crowding: High-impact solutions. Retrieved, April 20, 2008, from http://www.acep.org/bvtaxonomy.aspx?__taxonomyid=114&path=%5cPractice+Management%5cCrowding&__taxonomyshowall=0.

Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N. & Camargo, C. A. (2003). A conceptual model of emergency department crowding. *Annals of Emergency Medicine, 42*(2), 173 – 180.

Bernstein, S. L., Verghese, V., Leung, W., Lunney, A. & Perez, I. (2003). Development and validation of a new index to measure emergency department crowding. *Academic Emergency Medicine, 10*(9), 938-942.

Bierbaum, C. R., Szabo, S. M. & Aldrich, T. B. (1989). *Task analysis of the UH-60 mission & decision rules for developing a UH-60 workload prediction.* (ADA210763).

Brixey, J. J., Tang, Z., Robinson, D. J., Johnson, C. W, Johnson, T. R., Turley, J. P. et al. (2007). Interruptions in a level one trauma center: A case study. *International Journal of Medical Informatics*, *77*(4), 1-7.

Crandall, B., Klein, G. & Hoffman, R. (2006). *Working minds*: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press.

Chisholm, C. D., Collision, E. K., Nelson, D. R., & Cordell, W. H. (2000). Emergency department workplace interruption: are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine, 7*(11), 1239 – 1243.

Dismukes, R. K., Loukopoulos, L. D. & Jobe, K. K. (2001). The challenge of managing concurrent and deferred tasks. In *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus OH: Ohio State University.

Fairbanks, R. J., Bizantz, A. M. & Sunm, M. (2007). Emergency department communication links and patterns. *Annals of Emergency Medicine, 50*(4), 396 – 406.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman and J. P. Thomas (eds.) *Handbook of perception and human performance* (pp. 1 – 49). New York: Wiley.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland Press.

Hendy, K., Liao, J. & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 39*(1), 30-47.

Hill, S. G, Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L. & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34*(4), 429 – 439.

Hollander, J. E. & Pines, J. M. (2007). The emergency department crowding paradox. The longer you stay, the less care you get (Editorial). *Health Policy and Clinical Practice, 50*(5), 497 – 499.

Hollingsworth, J. C., Chisholm, C. D., Giles, B. K., Cordell, W. H. & Nelson, D.R. (1998). How do physicians and nurses spend their time in the emergency department? *Annals of Emergency Medicine, 31*(1), 87 – 91.

Hoot, N. R., Zhou, C., Jones, I., & Aronsky, D. (2007). Measuring and forecasting emergency department crowding in real time. *Annals of Emergency Medicine, 49*(6), 747 – 755.

Levin, S., France, J., Hemphill, R., Jones, I., Chen, K. Rickard, D., Makowski, R. & Aronsky, D. (2006). Tracking workload in the emergency department. *Human Factors, 48*(3), 526 – 539.

Levin, S., Aronsky, D., Hemphill, R., Han, J., Slagel, J., & France, D. J. (2007). Shifting toward balance: Measuring the distribution of workload among emergency department physician teams. *Annals of Emergency Medicine, 50*(4), 149 – 423.

Mospkop, J. C., Sklar, D. P., Geiderman, J. M., Schears, R. M. & Bookman, K. J. (2009). Emergency department crowding, part 1 – concept, causes, and moral consequences. *Annals of Emergency Medicine, 53*(3), 605 – 611.

Navon, & Gopher, (1979). On the Economy of the Human-Processing System. *Psychological Review, 86*(3), 214-55.

O'Donnell & Eggemeier (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman and J. P. Thomas (eds.) *Handbook of perception and human performance* (pp. 1 – 49). New York: Wiley.

Pines, J. M. & Hollander, J. E. (2008). Emergency department crowding is associated with poor care for patients with severe pain. *Annals of Emergency Medicine, 51*(1), 1-5.

Reid, G. B., Shingledecker, C. A. & Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 522 – 526).

Schneider, S. M., Gallery, M. E., Schafermeyer, R. & Zwemer, F. L. (2003). Emergency department crowding: A point in time. *Annals of Emergency Medicine, 42*(2), 167 – 172.

Vidulich, M. A. & Tsang, P. S. (1985). Absolute magnitude estimation and relative judgment approaches to subjective workload assessment. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 71-75). Santa Monica, CA: Human Factors Society.

Wears, R. L. & Perry, S. J. (2002). Human Factors and Ergonomics in the Emergency Department. *Annals of Emergency Medicine, 40*(2), 206 – 212.

Wickens, Christopher D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds). *Varieties of Attention.* New York: Academic Press.

Wickens, C. D. & Hollands, J. G. (2000). Engineering psychology and human performance. New Jersey: Prentice Hall.

Wierwille, W. W. & Casali, J. G. (1983). A validated rating scale for global mental workload measurement application. In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 129 -133). Santa Monica, CA: Human Factors Society.

Weiss, S. J., Ernst, A. A. & Nick, T. G. (2006). Comparison of the national emergency department crowding scale and the emergency department work index for quantifying emergency department crowding. *Academic Emergency Medicine*, *13*(5), 513 – 518.

Wuerz, R. C., Milne, L. W., Eitel, D. R., Traver, D. & Gilboy, N. G. (2000). Reliability and validity of a new five-level triage instrument. *Academic Emergency Medicine*, *7*(3), 236 – 242.

Zakay, D. (1993). The impact of time perception processes on decision making under time stress. In O. Svenson and J. Maule (eds.), *Time pressure and stress in human judgment and decision making* (pp. 59 - 72). New York: Plenum.

Appendix A

R·I·T

Rochester Institute of Technology

RIT Institutional Review Board for the
Protection of Human Subjects in Research
141 Lomb Memorial Drive
Rochester, New York 14623-5604
Phone: 585-475-2167
Fax: 585-475-4250
Email: jhrpop@rit.edu

Form C
IRB Decision Form

TO: Kathryn Coleu; Esa Rantanen

FROM: RIT Institutional Review Board

DATE: **May 21, 2008**

RE: **Decision of the RIT Institutional Review Board**

Project Title – The Effect of Emergency Department Crowding on Medical Staff's Taskload

The Institutional Review Board (IRB) has taken the following action on your project named above.

☒ Approved, no greater than minimal risk

Now that your project is approved, you may proceed as you described in the Form A. **Note that this approval is only for a maximum of 12 months; you may conduct research on human subjects only between the date of this letter and May 21, 2009.**

You are required to submit to the IRB any:

- **Proposed** modifications and wait for approval before implementing them,
- Unanticipated risks, and
- Actual injury to human subjects.

Return the Form F, at the end of your human research project or 12 months from the above date. If your project will extend more than 12 months, your project must receive continuing review by the IRB.

Heather Foti
Associate Director, Office of Human Subjects Research

Revised 10-18-06

University of Rochester Research Subjects Review Board

Letter of Approval

RSRB: RSRB00025164 **Principal Investigator**: Sandra Schneider

Study Title: Emergency Department Crowding and Physician Workload

Initial Approval: December 27, 2008

Study Approval Expires: December 26, 2009

Risk Level:

- Minimal Risk - Adults only

Review Level: Expedited

Expedited Category(ies):

- 7 - individual or group characteristics or behavior

Additional Remarks:

- Waiver of Documentation of Consent granted

- HIPAA: Does not apply

This approval is contingent upon the investigation being conducted in compliance with the approved study protocol including all requirements and/or determinations of the RSRB. Unless a Waiver of Consent is specified above, consent must be obtained and documented in the manner approved by the RSRB. Please note all remarks and/or attachments. Only consent forms bearing a current 'RSRB Approved' Watermark may be used. Only the most recently approved version of any consent or recruitment document may be used when obtaining consent. **Consent forms/recruitment letters must be printed on department letterhead.**

As the Principal Investigator, you are responsible for the following activities:

- Timely submission of continuing review progress reports. Federal Regulations require that the RSRB conduct continuing review of research. You will receive Progress Report forms from the RSRB
- Requesting any proposed changes in the above research activity. All subject recruitment materials must be approved prior to use. Changes may not be initiated without RSRB approval except when necessary to eliminate apparent immediate hazards to the subject(s) and then a report must be submitted along with the amendment request
- Maintaining all approved study documents in your study file
- Maintaining signed consent forms for at least three years after the research is completed or for a longer term if required by FDA regulations
- Reporting any unexpected serious problems involving risks to subjects or others (including unexpected deaths, hospitalizations or serious injuries) in accordance with the RSRB Adverse Event guidelines
- Submitting a final progress report to the RSRB upon completion of this study

Jeanne Grace, RSRB Chair December 27, 2008

The Department of Health and Human Services has approved a Federalwide Assurance (FWA) with the University of Rochester (FWA9386), which is in effect through September 27, 2010.

Appendix B

Dear Physicians and Physician Assistants,

We are looking for emergency physicians who are interested in volunteering to participate in a study of workload. This research aims at identifying how emergency department crowding affects physician workload.

This study is being conducted by Dr. Sandra Schneider of the University of Rochester's Department of Emergency Medicine in collaboration with Dr. Esa Rantanen and Mrs. Kathryn Coles from the Department of Psychology at the Rochester Institute of Technology.

Those who are interested in participating will be scheduled for an introduction session (lasting approximately 10 minutes) that describes the study and the consent process.

Physicians who choose to participate will be scheduled for additional observation sessions during which they will be shadowed by a researcher during a portion their normal shift (approximately 4 hours at a time). During these sessions, physicians will be asked to provide ratings of their subjectively experienced workload.

Following the observation session, an in-depth interview will be conducted (approximately 30 minutes). During the interview, physicians will be asked to share their thoughts about the tasks they perform and the workload those tasks impose.

If you are interested in participating or have any questions about the study, please respond to this e-mail with your name and contact information.   I will contact you as soon as possible to set up the first session and/or answer any questions you might have.

Sincerely,

Katie Coles

PLEASE REFER TO THE ATTACHED CONSENT FORM FOR MORE INFORMATION

Appendix C

Study Title: EMERGENCY DEPARTMENT CROWDING: PHYSICIAN AND PA WORKLOAD

Principal Investigator: Dr. Sandra Schneider

Introduction:

This consent form describes a research study and what you may expect if you decide to participate. You are encouraged to read this consent form carefully and to ask the person who presents it any further questions you may have before making your decision whether or not to participate. This study is being conducted by Dr. Sandra Schneider of the University of Rochester's Department of Emergency Medicine in collaboration with Dr. Esa Rantanen and Ms. Kathryn Coles from the Department of Psychology at the Rochester Institute of Technology.

You are being asked to participate in this study because you are an expert in emergency department (ED) operations, tasks performed by ED personnel, and the demands placed on them by the tasks and the environment.

Purpose of Study

The purpose of this study is to objectively quantify the task load imposed on medical emergency department (ED) physicians and physician assistants, determine the subjective workload they experience as a result, and to correlate these data with the existing metrics of ED crowding to evaluate the relationship between of ED crowding and subjective workload.

Description of Study Procedures

If you decide to participate in this study a researcher will 'shadow' you during your normal working hours and observe your performance of tasks you encounter during your shift. You will be asked to provide ratings of your subjectively experienced workload during these observation periods and participate in an in-depth interview about the task load imposed on you and the workload you experience.

Number of Subjects

Between 5 and 10 physicians are expected to participate in this study. The eventual number will depend on how many physicians and physician assistants volunteer to participate

Risks of Participation

There are no risks involved with participation in this research.

Benefits of Participation

You will benefit from this research as it aims at identifying how crowding affects your workload as part of ED personnel. This research may guide and support management decisions on ED staffing, scheduling, and other operations that affect ED personnel.

Payments

There will be no payment for your participation in this study.

Sponsor Support

This research is not supported financially by any sponsor.

Confidentiality of Records

While we make every effort to maintain confidentiality, it cannot be absolutely guaranteed. However, no information identifying individual participants will ever be associated with the data collected. All data will be stored and secured only on the investigator's computer. Interviews will occur one-on-one in a private setting. The results of this research study may be presented in meetings or in publications, but no subsequently published results will contain any information that could be associated with individual participants.

Contact Persons

For more information concerning this research, or if you feel that your participation has resulted in any emotional or physical discomfort, please contact: Dr. Sandra Schneider at sandra_schneider@urmc.rochester.edu., Dr. Esa Rantanen at esa.rantanen@rit.edu (585) 475-4412, or Kathryn Coles at kmc1195@rit.edu.

If you have any questions about your rights as a research subject, or any concerns or complaints, you may contact the Human Subjects Protection Specialist at the University of Rochester Research Subjects Review Board, Box 315, 601 Elmwood Avenue, Rochester, NY 14642-8315, Telephone (585) 276-0005, for long-distance you may call toll-free, (877) 449-4441. You may also call this number if you cannot reach the research staff or wish to talk to someone else.

Voluntary Participation

Participation in this study is voluntary. You are free not to participate or to withdraw at any time, for whatever reason, without risk or penalty. In the event that you do withdraw from this study, the information you have already provided will be kept in a confidential manner.

Appendix D

Participant
Date
Hour
Hospital

Task	Time	Notes	Task	Time	Notes

Appendix E

Start Time		Task	End Time	Notes
14:15		ct	14:17	ct scan/chest x-ray
Date		vcs	14:19	RN or tech
21-Jun		cb	14:20	"pick up new patient"
Participant		he	14:25	elderly man fell - is a resident in a nursing home - hard of hearing-has trouble putting in hearing aid
7		se	14:26	get stickers
		wo	14:28	chest x-ray
Interruptions		ct	14:34	test results show anemia,low Na, dehydration, febrile
face	5	wo	14:35	
phone	2	tp	14:38	place bp cuff. Perform rectal exam
page	0	ct	14:39	rectal results
		wo	14:40	
		ap	14:41	find out who the PCP admits to
		chart	14:43	
		ic	14:46	return call from admitting doc
		ap	14:48	talk to other MD about admission
		ap	14:49	page to admit
		chart	14:53	
		pr	14:55	look for old patient records
		chart	14:56	
		ap	14:57	web page to admit
		chart	14:58	
		con	14:59	PA
		chart	15:01	
		con	15:02	MD
		ap	15:05	return page
		con	15:12	MD
		ap	15:13	consult with admitting MD regarding patient condition
		ap	15:15	MD

Appendix F

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand How mentally demanding was the task?

Very Low — Very High

Physical Demand How physically demanding was the task?

Very Low — Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low — Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix G

Modified Instructions for NASA-TLX Ratings

We are interested in assessing the experiences you have during the different job conditions. Right now I am going to describe the technique that will be used to examine your experiences. In the most general sense we are examining the “workload” you experience. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task/job itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you feel.

One way to find out about workload is to ask people to describe feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually. This set of six rating scales was developed for you to use in evaluating your experiences. Please read the descriptions of the scales carefully. If you have any question about any of the scales, please ask me about it. It is extremely important that they be clear to you.

After each hour of observation, you will be given a sheet of rating scales. You will evaluate your experience of your job during that hour by putting an “X” on each of the six scales at the point which matches your experience. Each line has two endpoint descriptors. Note that “own performance” goes from “good” on the left to “bad” on the right. Please consider your responses carefully and consider each scale individually. Your ratings will play an important role in the evaluation being conducted.

Appendix H

Interview Questions

1. **Routine & Non-routine Tasks.** Can you list some routine tasks that you perform on a daily basis that require little effort and/or thought?

    a. What about non-routine tasks that require more thought?

    b. What proportion of your work would you say is comprised of routine tasks versus non-routine tasks?

    c. Do you feel like the time it takes to perform routine tasks is stable under different circumstances or does it vary?

        i. If it varies, what do think causes it to vary?

    d. Do you feel like the time it takes to perform non-routine tasks is stable under different circumstances or does it vary?

        i. If it varies, what do think causes it to vary?

2. **Interruptions.** What do you consider to be an interruption to your work? How would you define it?

    a. Do you think your work is affected by interruptions? ____

        i. If so, how?

3. **Time Pressure.** How do you know how long to spend on a given task?

    a. How do you know if you are spending too much time on a given task?

    b. How do you know when a task is complete?

c. What most often prompts you to switch from working on one task to working on another?
    i. What proportion of tasks must you return to in order to complete?

d. Do you ever feel rushed? ______ If so, How do you cope with this?

4. **Task Shedding.** When you are faced with several tasks that need to be performed, do you ever need to postpone one indefinitely or transfer one to someone else? ______

    a. If so, how do decide which task to postpone or pass on?

5. **Time Estimation.** Do you feel you have an accurate sense of how much time has passed while performing a given activity?

    a. Do you think that your sense of time changes based on the complexity of the activity you are performing? _____ If so, How?

    b. Do you think that your sense of time changes based on the number of activities you are performing? ______ If so, How?

ED Crowding Specific Questions

1. How would you define emergency department crowding (here and in general)?

2. Do you feel that emergency department crowding affects the demands placed on you? _____How so?

    a. How does it affect you mentally?

    b. How does it affect you physically?

    c. How does it affect the time pressure placed on you?

1. Do you feel that emergency department crowding affects the amount of effort you put forth? _____ How so?

a. Is this same for routine/non-routine tasks?

Do you think that crowding affects your performance? _____