



Published in final edited form as:

*IIE Trans Healthc Syst Eng.* 2011 ; 1(2): 131–143. doi:10.1080/19488300.2011.609524.

## MEASURING WORKLOAD OF ICU NURSES WITH A QUESTIONNAIRE SURVEY: THE NASA TASK LOAD INDEX (TLX)

Peter Hoonakker, PhD<sup>1</sup>, Pascale Carayon, PhD<sup>1</sup>, Ayse Gurses, PhD<sup>2</sup>, Roger Brown, PhD<sup>3</sup>, Kerry McGuire, BS<sup>1</sup>, Adjhaporn Khunlertkit<sup>1</sup>, and James M. Walker, M.D., FACP<sup>4</sup>

<sup>1</sup>Center for Quality and Productivity Improvement and Department of Industrial and Systems Engineering, University of Wisconsin-Madison

<sup>2</sup>John Hopkins School of Medicine

<sup>3</sup>School of Nursing, University of Wisconsin-Madison

<sup>4</sup>Geisinger Health System

### Abstract

High workload of nurses in Intensive Care Units (ICUs) has been identified as a major patient safety and worker stress problem. However, relative little attention has been dedicated to the *measurement* of workload in healthcare. The objectives of this study are to describe and examine several methods to measure workload of ICU nurses. We then focus on the measurement of ICU nurses' workload using a subjective rating instrument: the NASA TLX.

We conducted secondary data analysis on data from two, multi-side, cross-sectional questionnaire studies to examine several instruments to measure ICU nurses' workload. The combined database contains the data from 757 ICU nurses in 8 hospitals and 21 ICUs.

Results show that the different methods to measure workload of ICU nurses, such as patient-based and operator-based workload, are only moderately correlated, or not correlated at all. Results show further that among the operator-based instruments, the NASA TLX is the most reliable and valid questionnaire to measure workload and that NASA TLX can be used in a healthcare setting. Managers of hospitals and ICUs can benefit from the results of this research as it provides benchmark data on workload experienced by nurses in a variety of ICUs.

### Keywords

Workload; measurement; intensive care units; nurses

## INTRODUCTION

High workload has been identified as a major concern in health care, in particular in intensive care units (ICUs). Patient care in the ICU is characterized by highly demanding tasks that need to support urgent therapeutic intervention. High workload is one of the most important job stressors reported by ICU nurses (Kiekkas, et al., 2008; Schaufeli & Le Blanc, 1998). High workload can have negative consequences for the nurses and other ICU clinicians, as well as for their patients. Direct effects of high workload on patient care may be related to the lack of time to perform important care tasks, and include complications, poor patient outcomes and increased mortality (Gurses, Carayon, & Wall, 2009; Kiekkas, et al., 2008). Furthermore, lack of time can have a negative impact on nurse-physician communication (Baggs, et al., 1999) and provider-patient communication (Llenore & Ogle, 1999). According to the Joint Commission (Joint Commission, 2007), two-thirds of the root causes of sentinel events (any unanticipated event in a healthcare setting resulting in death

or serious physical or psychological injury to a person or persons, not related to the natural course of the patient's illness) in the period 1995–2005 were communication-related. High workload may indirectly impact patient safety because of its negative impact on communication and decreased job satisfaction, motivation and burnout. High workload has been identified as one of the most important causes of burnout (McManus, Keeling, & Paice, 2004). Burnout, in turn, is associated with decreased well-being and quality of care, and increased costs related to absenteeism and turnover (Embriaco, Papazian, Kentish-Barnes, Pochard, & Azoulay, 2007).

Given the evidence regarding the negative impact of high ICU nursing workload, it is important to examine the dimensions of workload such as physical workload, mental workload (i.e. information overload), time pressure and emotional workload (Alvarado & Carayon, 2007), and the methods for measuring these different dimensions of ICU nurses' workload. In this paper we describe the construct of workload and different approaches to measure workload, compare different instruments to measure ICU nurses' workload, and test if one of the most reliable and valid human factors measures of subjective workload, i.e. the NASA task-load-index (TLX), can be used to measure nursing workload.

## BACKGROUND

Workload is difficult to conceptualize as shown by the following definitions. Workload can be defined as “the relative capacity to respond” (Lysaght, et al., 1989). “Workload is a construct that is used to describe the extent to which an operator has engaged the cognitive and physical resources required for a task performance” (Bucks, Ryan, & Wilson, 1994). “Workload is a multidimensional and complex construct, that is affected by external task demands, environmental, organizational and psychological factors, and perceptive and cognitive abilities” (Weinger, Reddy, & Slagle, 2004). These three definitions show first of all that workload is a difficult to define concept and secondly, that workload consists of several components: (1) there is an operator, using his or her resources to respond to (2) external physical or cognitive demands to (3) perform a certain task. Apart from the external demands, other factors such as environmental, organizational and psychological factors also play a role in this process. Nurses' workload is equally or even more difficult to define. See for example Morris et al. (Morris, MacNeela, Scott, Treacy, & Hyde, 2007) for a literature review on nurses' workload.

To give some indication of the importance of workload in health care, and in particular the ICU environment, we conducted a search in PubMed using the keyword “workload”. Results of this search showed that workload is mentioned in nearly 20,000 articles. The combination of the keywords “workload” AND “nurses” produced 4,247 hits, and the combination of the search terms: “workload”, “nurses” AND “intensive care” OR “critical care” produced 489 hits. However a search using a combination of the following keywords: “intensive care” OR “critical care” AND “nurses” AND “workload measurement” produced only 16 hits. A similar literature search in CINAHL identified 7,543 results for “workload”, 2,577 results for the combination of “workload” AND “nurse”, and 357 results for the combination “workload”, AND “nurse”, AND “critical care” OR “intensive care”. A search in CINAHL using the keywords “workload measurement” led to 19 hits.

To summarize, although workload of nurses has received considerable attention in the literature (see for example Morris et al. (2007) for a review on nurses' workload), issues around the *measurement* of nurses' workload have received considerably less attention. In the literature on ICU nurses' workload, two different approaches to the measurement of workload are proposed: a patient-based approach that uses data on nurse-patient ratio and the acuity of the patient, and an operator-based approach that takes the experience of the

operator (in this case the ICU nurse) into account. Note that these two approaches are not mutually exclusive: they have different purposes. The first approach is used for staffing activities; for organizing the different tasks; and for coping with patient severity and turnover. The second approach is used for research activities; to determine the amount of nursing effort resulting from each activity; and how to cope with it.

### Patient-based measurement of workload

The patient-based approach considers *patient characteristics* such as number of patients per nurse and acuity of the patient (dependency method), and time spent on the patient as percentage of total time (activity method). This approach is driven by increasing effectiveness and efficiency, and the ultimate goal is to deploy nurses more efficiently and thereby reducing nursing costs.

One of the most popular workload measures in healthcare is the Therapeutic Intervention Scoring System (TISS). The TISS is a patient-based workload measure. Cullen et al. (1974) developed TISS in 1974 to measure severity of illness at the patient level and assess the corresponding nursing workload in ICUs. In 1981, the Acute Physiology and Chronic Health Evaluation (APACHE) scores were proposed, using severity of illness for benchmarking case mix and predicting outcomes (Knaus, Zimmerman, Wagner, Draper, & Lawrence, 1981). The introduction of the APACHE scores made TISS less relevant and useful for measuring severity of illness at the patient level; therefore, later versions of the TISS focused on measuring nursing workload. However, the TISS has been criticized (D. R. Miranda, de Rijk, & Schaufeli, 1996) because (1) it is time consuming, (2) the use of the instrument is cumbersome, (3) the items do not adequately reflect indirect patient care activities of nurses such as contact with the family and maintaining supplies, (4) it exclusively scores direct patient care activities and does not reflect several other daily activities of the nursing staff (e.g., meetings, trainee supervision, research) that are equally important to the professionals and to the organization and management of the ICU, and (5) the items are difficult to interpret (Gurses, 2005). In response to the criticisms, a simplified version, the TISS-28, was developed in 1996 (D. R. Miranda, et al., 1996). However, TISS-28 was still found to be fairly time consuming, in particular when used in large epidemiological studies (Rothen, Kung, Ryser, Zurcher, & Regli, 1999).

In response to this criticism, Miranda et al. (2003) developed a new instrument called the Nursing Activities Score (NAS). NAS includes five items (e.g., care of relatives, administrative tasks, monitoring) and 14-sub-items that were not included in TISS-28. The weights of NAS measure time spent on nursing activities at the patient level. As opposed to the weights in TISS-28, which represent the *severity of the patient's condition*, the weights of NAS represent the calculated percentage of *nursing staff's time* (per period of 24 hours) dedicated to the performance of the activities included in the instrument. The sum of the weights of the individual items scored reflects the amount of time that an ICU nurse spends on performing tasks during a particular day. NAS was validated in a study of 99 ICUs in 15 countries: NAS explains 81% of the nursing time, whereas TISS-28 explains only 43%.

Other instruments to measure patient-based workload in ICUs include the Time Oriented Scoring System (TOSS), the Project Research in Nursing (PRN), and the OMEGA system. The OMEGA 1990 version was validated in a multi-centric study (OMEGA versus PRN) and is currently used in European ICUs (Lazard, et al., 1996). It encompasses 47 diagnostic and therapeutic items that are weighted from 1 to 10 points according to the required workload and that are divided into three categories of tasks.

However, research that uses patient load and severity-based measures is limited in its contribution to understanding the impact of workload on the performance and well-being of

health care providers; most of these measures (except NAS) focus on the time spent with the patient. Indirect patient activities and other daily activities that are important to the measurement of nursing workload, such as time spent with the patient's family, are not taken into account (Cohen, et al., 1999; D. R. Miranda, et al., 1996). We need tools for comprehensive workload assessment that can identify sources of excessive workload and provide direction for corrective interventions (Gurses, 2005).

### Operator-based measures of workload

The operator-based approach is a human factors approach that considers characteristics of the *nurse* and interactions between the nurse and the work environment. The goals of this approach are to examine causes of high workload and identify strategies to reduce workload and ultimately improve quality of working life for the nurse and quality and safety of care for the patient. Two operator-based approaches to measuring workload can be distinguished. Both are aimed at measuring the cognitive and physical resources required to perform a task. The psychophysiological or "objective" approach is concerned with the physiological basis of psychological processes, in this case workload. The subjective approach uses feedback from the "operator" to get information on how the operator experiences the different processes, in this case workload.

#### Psychophysiological or "objective" measures of operator-based workload—

There are several psychophysiological measures of workload, such as ocular, cardiac and respiratory responses and brain activity. Although these measures are sometimes referred to as "objective" measures, it is important to note that they are *indirect* measures of workload: they are based on an assumed relationship between cognitive activity and autonomic activity (Tsang & Wilson, 1997). Mental workload has been assessed using ocular measures such as increases in pupil diameter, decreases in blink rate, cardiac responses such as heart rate variability and overall heart rate, respiratory responses such as breathing rate, and brain activity measured with event-related brain potentials (ERP) and electroencephalo-graphic activity (EEG). Most of the physiological measures of workload have been associated with mental workload and/or stress, or with subcomponents of mental workload, mostly in laboratory settings (Kramer, 1991). Most of these studies have been conducted in laboratory settings, and have been rarely used in situations where the subjects have to move around.

**Subjective measures of operator-based workload—**Several subjective rating techniques are available to measure operator-based workload. These rating techniques are called subjective to set them apart from "objective techniques" such as physiological measures (see Lysagh et al. (1989) for a critical evaluation of empirical and analytical approaches to measuring workload). Several arguments can be made for the usefulness of subjective rating techniques (1992). First, according to some researchers, operator ratings are the most *direct* indicators of operator workload (Sheridan, 1980). Workload is measured at "the source", i.e. the nurse, instead of using an indirect method of assessing nurses' workload by focusing on the patient characteristics. That gives the approach more validity. Second, both inter- and intra-*reliabilities* of the instruments used in this approach are higher (Hill, et al., 1992) than in the patient-based approach. Third, operator ratings are among the *least intrusive* of all techniques because they can be administered after the task is completed without disturbing the operator during task performance. Fourth, the subjective techniques are *flexible and portable*; no equipment or special data collection devices are needed. Finally, subjective techniques can be *quick and inexpensive to administer and analyze* (Hill, et al., 1992).

Over the years, several instruments have been developed to measure operator-based subjective workload: the Cooper-Harper Scale, the perceived workload scale, the Subjective

Workload Assessment Technique (SWAT), the Workload Profile (WP), the Rating Scale Mental Effort (RSME) and the NASA-Task Load Index (NASA-TLX).

### Description of the NASA TLX

The NASA task-load index (TLX) was developed more than 20 years ago to measure workload in aviation (Hart, 2006; Hart & Staveland, 1988). The NASA-TLX is one of the most widely used instruments to assess overall subjective workload. In a recent review, Hart estimated that the NASA-TLX has been used in more than 300 studies, mainly in air traffic control, and civilian or military aviation (Hart, 2006). The NASA-TLX is a *multidimensional* instrument that consists of 6 subscales: Mental Demand (MD), Physical Demand (PD) and Temporal Demand (TD), Frustration (FR), Effort (EF), and Performance (PE) (see Appendix A for the NASA TLX). Twenty-step bipolar scales are used to obtain ratings on these dimensions, resulting in a score between 0 and 100. The underlying assumption of the instrument is that the combination of these 6 dimensions is likely to represent “workload” experienced by operators (Hart, 2006). Overall workload (OW) is represented by a combination of the six dimensions. The NASA TLX was developed as part of a large program of laboratory research and the instrument has been demonstrated with a variety of tasks. For example, the NASA TLX has successfully been used in multitask contexts such as in simulated flight tasks, real flight tasks, and air combat.

Originally, a weighting procedure was applied to the raw test scores to develop a composite score tailored to *individual* workload definitions, but many researchers have eliminated the weighting procedure and instead use the raw test scores. The Raw TLX (RTLX) instrument is simpler to apply: the ratings are averaged or added to create an estimate of overall workload (Hart, 2006).

Another modification to the original use of the NASA TLX instrument is to analyze the subscale ratings (MD, PD, TD, FR, EF and PE) in addition to the overall workload (OW) score, or to use the subscale ratings to identify the specific type of workload or performance problem (Hart, 2006). For an extensive description of the NASA TLX in different settings including health care, see Young et al. (Young, Zavelina, & Hooper, 2008).

### Reliability and Validity of NASA TLX

To establish the quality of psychological measures, such as the NASA TLX, we use principles and methods developed in the field of psychometrics. Two key concepts in psychometrics are reliability and validity. *Reliability* refers to the issue of measurement repeatability: when we measure something at two different occasions, the two measures should result in the same outcomes. *Validity* refers to the content of measurement: are we measuring what we think we are measuring? There are several methods to evaluate reliability and validity (Hoonakker, Carayon, & Walker, 2010).

We can evaluate reliability by measuring a concept at two different times (test-retest-reliability), by evaluating the internal consistency (of the answers) of questions that are supposed to measure the same concept, and by comparing with other methods of measurement of equal or higher level, for example standardized (and validated) questionnaires. The test-retest method has been used to examine the reliability of the NASA TLX instrument. Batisse and Bortolussi (1988) reported a test-retest reliability of 0.77 for the NASA TLX.

Three forms of validity can be distinguished (Nunnally, 1978): content validity, construct validity and predictive validity or concurrent validity. While developing the NASA TLX, Hart and Staveland (1988) established *content validity* of the NASA TLX instrument.

*Construct validity* is important when measuring abstract concepts (e.g., satisfaction, workload, stress). The abstract concept (the construct) is operationalized by several questions. Discriminant validity and convergent validity are two methods to measure construct validity. *Discriminant validity* describes the degree to which the operationalization is **not** similar to (diverges from) other operationalizations that it theoretically should not be similar to. *Convergent validity* refers to the degree of agreement between measurements of the same characteristics obtained by different approaches supposed to measure the same characteristics. Results of a laboratory study by Rubio et al. (Rubio, Díaz, Martín, & Puente, 2004) in which three instruments to measure mental workload were compared showed that NASA TLX was highly correlated (correlations between 0.97 and 0.98) to two other instruments (Subjective Workload Assessment Technique (SWAT) and Workload Profile (WP)). Therefore, convergent validity is high for the NASA TLX instrument (Rubio, et al., 2004).

*Predictive and concurrent validity* (also known as criterion validity) refers to the comparison of the measurement under evaluation to another variable that lies outside the domain of the concept being measured. The study by Rubio et al. (2004) compared three subjective workload instruments (NASA TLX, SWAT, WP) and showed that all three instrument had high correlations (between 0.73 and 0.79) with one measure of performance, but that NASA TLX showed a higher correlation than the two other instruments with a second measure of performance. Concurrent validity of the NASA TLX was found to be higher than the concurrent validity of the two other workload instruments.

Finally, *sensitivity* refers to the ability of the instrument to discriminate among different levels, in this case task loading. Results of a comparison between NASA TLX and SWAT showed that the NASA TLX performed better in terms of sensitivity (i.e. the degree to which various measures can distinguish workload levels), especially at low levels of workload (Hill, et al., 1992).

In summary, the literature shows that the NASA TLX is a reliable and valid instrument and is actually more reliable and valid than other subjective workload instruments. Furthermore, NASA TLX is easy to administer, especially when using the raw scores instead of the weighted scores (Hendy, Hamilton, & Landry, 1993; Nygren, 1991). Finally, NASA TLX is one of the few instruments that include a measurement of physical workload. Nurses in ICUs can be exposed to high physical workload; therefore, it is important to use an instrument that captures various dimensions of workload. Most studies to test reliability and validity of NASA TLX were conducted mainly in air traffic control, and civilian or military aviation and few studies have examined the use of NASA-TLX in healthcare. We used data collected in two studies to assess construct validity of NASA TLX in health care and to examine workload of ICU nurses.

## METHODS

### Sample

We combined data collected in two studies to examine workload among ICU nurses (see Table 1). Data for Study 1 were collected in 2004 in a study of performance obstacles and facilitators, workload, quality of working life and quality and safety of care among ICU nurses (Gurses, 2005; Gurses, et al., 2009). Data for Study 2 were collected in a study on the impact of Computerized Provider Order Entry (CPOE) implementation in a repeated cross-sectional design in 2007–2008 (Hoonakker, Carayon, Douglas, et al., 2008; Hoonakker, Carayon, Walker, & Wetterneck, 2008). CPOE is a software used to enter medical orders with the aid of a computer. Data were collected 6 months before CPOE implementation, and 3 months and one year after CPOE implementation.

**Study 1**—Data were collected in a total of 7 hospitals and in 17 ICUs (Adult (3), Burn Unit (1), Cardiac (1), Cardiothoracic (1), Medical/Coronary (1), Medical/Surgery (2), Neonatal (1), Neurosurgery (1), Neurotrauma (1), Pediatric (1), Surgical (2), Medical/Surgical (2) ICUs) in the Midwest of the USA. One participating hospital is academic, and six are non-academic. One hospital is rural, and six are urban. A total of 300 nurses filled out the survey. Eighty-six percent of the nurses in Sample 1 are female. Average age is 36 years. The highest level of education of most of the participating nurses is either an associate degree in nursing (ADN, 29%) or a bachelors of science in nursing (BSN, 69%) degree. Average tenure in the ICU is 6.4 years. On an average, nurses worked 37 hours per week, most of them (80%) in 12-hours shifts, in 8-hours shifts (12%), or in other shifts (8%). Forty-nine percent of nurses worked in night shifts and 51% in day shifts. Response rates in the different ICUs varied from 40% to 100%. The overall response rate was 77%.

**Study 2**—Data were collected during three rounds of data collection in a study to examine the impact of CPOE implementation in a hospital in the Eastern USA ([http://cqpi.engr.wisc.edu/cpoe\\_home](http://cqpi.engr.wisc.edu/cpoe_home)). Most of the respondents in Study 2 are female (88%). Average age is 36 years. Forty percent of the nurses have had some college or technical training; 47% graduated from college; 10% have had some graduate school; 3% have a graduate degree. Most of the ICU nurses (98%) are Caucasian. Average tenure at the hospital is nearly twelve years. Average tenure in the current ICU is 10 years. On average, nurses work 41.5 hours a week, mostly in 12-hour shifts (74%), 8-hour shifts (9%) or a combination of 12- and 8-hour shifts (17%). Nine percent of the nurses work during weekdays; 9% in weekends and 72% in a combination of both. Thirty-six percent of the nurses work day shifts; 4% evening shifts; 33% night shift; and 27% in a combination of shifts. Twenty-nine percent of the nurses work in the Adult ICU (AICU), 27% in the Cardiac ICU (CICU), 14% in the Pediatric ICU (PICU), and 30% in the Neonatal ICU (NICU). One-hundred-seventy-nine nurses in four ICUs filled out the survey questionnaire 6 months before CPOE implementation (Round 1) (response rate (RR)=93%); 118 nurses filled out the survey three months after CPOE implementation (R2, RR=51%); and 160 nurses filled out the survey one year after CPOE implementation (R3, RR=72%). The overall response rate was 72%.

**Combined sample**—The combined sample contains data from 757 nurses, working in 8 different hospitals and 21 different ICUs (see Table 1). Eighty-seven percent of the respondents is female; 44% of the respondents is younger than 35 years; 27% is between 35 and 44 years old; 24% is between 45 and 54; and 5% is older than 55 years. Average age is about 38 years old.

Tenure at their current ICU varies between less than one year to more than 40 years, with an average of nearly 8 years. Most respondents work during 12-hours shifts (76%); 10% work in an 8-hour shift, and 14% work in another shift. Thirty-two percent of the respondents work in a 12-hour day shift; 30% in a 12-hour night shift, and 37% in another shift (4-hour shift, 8-hour shift, evening shift, etc.). The 21 different ICUs were combined into 7 types of ICU: Adult Medical/Surgical ICU (AICU), Burn Unit, Cardiac ICU (CICU), Neurological ICU (Neuro ICU), Neonatal ICU (NICU), Pediatric ICU (PICU) and Surgical ICU (SICU).

## Instrument

The NASA Task Load Index (TLX) consists of six dimensions to assess workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Twenty-step bipolar scales are used to obtain ratings on these dimensions, resulting in a score between 0 and 100. The six scales are combined to create an overall workload scale (0–100). In both studies, the face validity of NASA TLX was tested in a pilot study by asking ICU

nurses to fill out the questionnaires (see for example Hoonakker et al. (2011) for an extensive description of the questionnaire development). Reliability (Cronbach's alpha) of the overall workload scale is 0.72.

## Analysis

Data were analyzed using SPSS© v16.0 and MLwiN v2.1 (Rasbash, Charlton, Browne, Healy, & Cameron, 2009). In order to describe the workload of ICU nurses, we use descriptive statistics: mean score (M), the confidence interval (CI) around the mean, the standard deviation (SD), median (the value at which 50% of the sample will score below, and 50% above, also known as Quartile 2, or Q2), the 25<sup>th</sup> percentile (Q1: the value at which 25% of the sample will score below, and 75% will score above), the 75<sup>th</sup> percentile (Q3: the value at which 75% of the sample will score below, and 25% will score above), and the interquartile range (Q3-Q1). We use box-and-whisker plots to graphically summarize the data using Q1, median and Q3 for the box and  $1\frac{1}{2} * IQ$  for the whiskers. Analysis of variance allowed us to examine differences between different groups, and a two-level multivariate analysis with repeated measures (Goldstein, 2003) examined differences in workload between rounds of data collections at Hospital 8 (Study 2). Results of the two-level multi-variate analysis showed that the samples in the three rounds of data collection in study 2 were independent. We controlled for possible differences in ICUs, and tenure of nurses.

## RESULTS

### Individual characteristics

Data were analyzed to examine the possible relationship between individual characteristics (gender and age) and workload. Results show no statistically significant differences in workload between the four age groups (< 35, 35–44, 45–54, 55 years), with an exception for physical demands. Physical demands *decrease* with age: the older nurses (>55 years) report significantly lower physical workload than younger nurses. In general, *female* nurses experience a higher overall workload ( $t=-2.61$ ,  $p<0.01$ ); mental demands ( $t=-3.14$ ,  $p<0.01$ ), temporal demands ( $t=-2.08$ ,  $p<0.05$ ); effort ( $t=-2.15$ ,  $p<0.05$ ) than male nurses. There are no significant gender differences in physical workload, performance and frustration.

### Job and organizational characteristics

We examined the relationship between the following job and organizational characteristics and workload: hospital, ICU, shift (day vs. night shift) and shift length (8-hour shift vs. 12-hour shift). We also examined the impact of CPOE implementation in hospital 8 with three rounds of data collection (study 2).

**Workload by shift**—Results of a comparison of the effects of day- vs. night-shifts and the length of the shifts (8-hour vs. 12-hour shift) on workload show that nurses report a higher overall workload ( $t=-1.22$ ,  $P<0.05$ ), physical demands ( $t=-2.89$ ,  $p<0.01$ ), and frustration ( $t=-1.87$ ,  $p<0.05$ ) in a 12-hour shift than in an 8-hour shift. Overall workload ( $t=2.44$ ,  $p<0.05$ ), temporal demands ( $t=2.25$ ,  $p<0.05$ ) and effort ( $t=2.48$ ,  $p<0.05$ ) are higher in 12-hour day-shifts than in 12-hour night shifts.

**Workload by hospital**—Figure 1 shows the box-and-whisker plot for overall workload and Table 2 summarizes the descriptive statistics for the 6 dimensions of workload and overall workload in the different hospitals.

There are significant differences between the 8 *hospitals* in overall workload ( $F=16.52$ ,  $p<0.001$ ), mental demands ( $F=17.64$ ,  $p<0.001$ ), physical demands ( $F=3.30$ ,  $p<0.01$ ),



temporal demands ( $F=12.59$ ,  $p<0.001$ ), effort ( $F=11.32$ ,  $P<0.001$ ), performance ( $F=2.35$ ,  $p<0.05$ ) and frustration ( $F=9.62$ ,  $p<0.001$ ). In general, workload is higher in larger hospitals (H1, H5, and H8), with the highest workload in hospital 8.

**Workload in various ICUs**—Figure 2 shows the box-and-whisker plot for overall workload and Table 3 summarizes the descriptive statistics for the 6 dimensions of workload and overall workload in the different ICUs.

Differences in overall workload between the different ICU types are significant ( $F=8.86$ ,  $p<0.001$ ). Results of an ANOVA with Bonferroni correction show that overall workload in adult and cardiac ICUs is significantly higher than in neuro- and neonatal ICUs. Differences in mental demands ( $F=5.26$ ,  $p<0.001$ ), physical demands ( $F=51.879$ ,  $p<0.001$ ) and temporal demands ( $F=4.59$ ,  $P<0.001$ ) are also significant. Mental demands are significantly lower in the neuro ICU than in the other ICUs. Temporal demands are also lower in the neuro ICU than in most other ICUs. Physical demands are significantly lower in the neonatal ICU and to a lesser extent in the pediatric ICU than in other ICUs.

**Effects of CPOE implementation**—Results of a two-level multivariate analysis with repeated measures to examine differences in workload between rounds of data collection (6 months pre-CPOE implementation, 3 months post-CPOE implementation and one year post-CPOE implementation) showed no significant differences in levels of workload, after having controlled for differences in ICUs and tenure of the nurses. At the item-level there was a nearly significant decrease in physical workload between round 1 and 2 ( $\chi^2=3.33$ ,  $df=1$ ,  $p<0.07$ ). However, in the third round, physical workload returned to pre-implementation levels. Finally, results of a comparison of workload levels between the 4 ICUs showed that nurses in the neonatal ICU reported significantly lower workload than nurses in the other ICUs (AICU, CICU and PICU).

### Comparison of different measures of nurses' workload

In Study 1, we used several measures for workload: patient-based measures such as nurse-patient ratio (NPR), the Nursing Activities Score (NAS) and operator-based subjective scales, including the Rating Scale Mental Effort (RSME), the Perceived Workload Scale (PWS) and the NASA TLX. For NAS we used two measures: NAS (the combined NAS depending on the number of patients) and NASMAX: the NAS score for the patient with the highest NAS score. Table 3 shows the correlations between the different measures of workload.

Results show that the different dimensions of NASA TLX are highly inter-related, except for performance (PE). PE is only correlated to FR (frustration). Overall workload (OW) as measured with NASA TLX is highly correlated with the two other subjective rating scales (RSME and PWS,  $r=.77$  and  $r=.81$  respectively). These results confirm the convergent validity of NASA TLX.

Results of the comparison between patient-based measures of workload show that NPR (nurse patient ratio) is moderately correlated ( $r=0.57$ ) with the NAS.

Results of the comparison between patient-based and subjective measures of workload show that NPR (nurse patient ratio) is not significantly related to the subjective measures of workload. The NAS score is moderately correlated with the subjective measures of workload ( $r=0.45$  with overall workload (NASA TLX),  $0.39$  with RSME and  $r=0.42$  with PWS).

In summary, the different workload instruments measure different aspects of workload. For example, interestingly, nurse-patient-ratio is not at all related to subjective measures of workload. The results provide support for discriminant validity of NASA TLX.

## DISCUSSION

In this paper we examined the measurement of workload in healthcare, examined different instruments to measure workload in healthcare, with a focus on the NASA TLX, and used this instrument in large sample of ICU nurses in multiple settings.

Workload is a difficult to define concept and consequently *measuring* workload is equally or even more difficult. Some authors have even referred to workload as “an unquantifiable entity” (Hughes, 1999). There are two broad approaches to measuring nurses’ workload.

The patient-based approach takes into account *patient characteristics* such number of patients per nurse and acuity of the patient (dependency method), and time spent on the patient as percentage of total time (activity method).

The operator-based approach is a human factors approach that considers characteristics of the *nurse* and interactions between the nurse and the work environment. The goal of this approach is to examine causes of high workload and -using a human factors approach- to reduce workload and ultimately improve quality of working life for nurses, and quality and safety of care for patients (Carayon, Alvarado, & Hundt, 2003; Carayon & Gurses, 2005; Gurses, et al., 2009).

Results of this study show that the two approaches to measure workload (patient-based and operator-based) are only moderately (NASA TLX with NAS:  $r=0.45$ ,  $p<0.01$ ) correlated, or not correlated at all (NASA TLX with Nurse-Patient-Ratio (NPR):  $r=0.10$ ,  $p>0.05$ ). This clearly shows the difference between the two approaches: one is developed to organize the different tasks and to cope with patient severity and turnover; the second approach is developed to determine the amount of nursing effort resulting from each activity, and how to cope with it. It also shows how confusing it can be to use the term workload in healthcare.

The literature shows that when comparing the different subjective workload instruments the NASA TLX is one of the most reliable and valid subjective workload instruments available. Applying a previously validated instrument that is developed for a specific setting (NASA TLX was originally developed to measure workload in aviation) to another setting (in this case health care) is never easy. However, results of this study show that it should possible to use NASA TLX in health care. Results of this study show that NASA TLX correlates high with other subjective workload instruments (RSME and PWS; 0.77 and 0.81 respectively) that were not developed for a specific setting.

Results of our study and statistical analysis show support for several forms of validity of the NASA TLX. Results of this study have shown that NASA TLX has *face validity* and different forms of construct validity. For example, results of this study show that NASA TLX has *convergent validity* (it correlates high with PWS ( $r=.81$ ) and RSME ( $r=.77$ ), *discriminant validity* (it correlates very low with nurse-patient-ratio) and is *sensitive* enough to detect differences between different groups (individual characteristics such as age and gender, and job and organizational characteristics such as setting (hospital and ICU), day vs. night shift, and shift length). Results of our study also show that NASA TLX is reliable in health care (Cronbach’s  $\alpha = 0.72$ ). However, several authors have suggested that a test-retest reliability is better measure of reliability for multi-dimensional scales such as NASA-TLX (Battiste & Bortolussi, 1988).

Therefore, in future longitudinal studies on workload in healthcare, test-retest reliability of NASA TLX should be examined. Further, future longitudinal studies should also focus on predictive validity of NASA TLX in healthcare. Finally, in this study we examined reliability and validity of NASA TLX in two studies among *ICU nurses*. Future studies should examine validity and reliability of NASA TLX among other healthcare providers (RN nurses, fellows, attendings, residents, nurse practitioners, nurse-assistants, etc) and in different settings (hospitals, outpatient clinics, etc).

Workload can be very high in health care, especially in intensive care units (ICUs). Results of our study show that -on a scale from 0–100- ICU nurses score an average of 82.8 on mental demands (How much mental activity is required to perform your job (thinking, deciding, calculating, remembering, looking, searching, etc...)), 69.1 on physical demands (How much physical activity is required to perform your job (e.g., pushing, pulling, turning, controlling, activating, etc.)), 70.2 on temporal demands (How much time pressure do you feel due to the rate or pace at which the tasks or task elements occurred?), 76.3 on effort (How hard do you have to work (mentally and physically) to accomplish your level of performance?), 78.4 on performance (How satisfied are you with your performance at your job?), 51.4 on frustration (How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel about your job?), and 71.4 on overall workload.

Subjective measures of workload such as NASA TLX can be used to determine the nurses' workload and can help establish causes of workload. There are several frameworks that can be used to determine possible causes of nurses' workload, for example the work-system model (Carayon, 2009; Carayon, et al., 2006). Results of several studies, using the work-system model, show that different dimensions of workload may be affected by different work system factors. For example, the causes of physical workload are lifting, working in awkward postures, and stooping (Carayon & Alvarado, 2007). Nurses' mental workload is related to the need for ICU nurses to process information, often in a short period of time. For instance, code (crisis) situations require quick decision making and quick processing of a lot of information (Carayon & Alvarado, 2007). Carayon and colleagues (Carayon & Alvarado, 2007; Gurses & Carayon, 2007) provide an extensive description of performance obstacles that can contribute to ICU nurses' workload.

To conclude: results of this study have shown that NASA TLX is a reliable and valid tool to measure workload among ICU nurses. The NASA TLX is easy to administer and allows the researcher to measure different dimensions of workload and overall workload. Results of our research can be used for benchmarking results of future studies on workload of ICU nurses.

## Acknowledgments

This study was supported by grant 1R01 HS015274-01 [PI: P. Carayon] from the Agency for Healthcare Research and Quality (AHRQ), a Health Services Research Dissertation Grant (#1 R36 HS14517) [PI: A. Gursus] from AHRQ, and by grant 1UL1RR025011 from the Clinical & Translational Science Award (CTSA) program of the National Center for Research Resources National Institutes of Health [PI: M. Drezner]. We are indebted to the participating ICU nurses and their supervisors and ICU medical directors for their willingness to cooperate in this research.

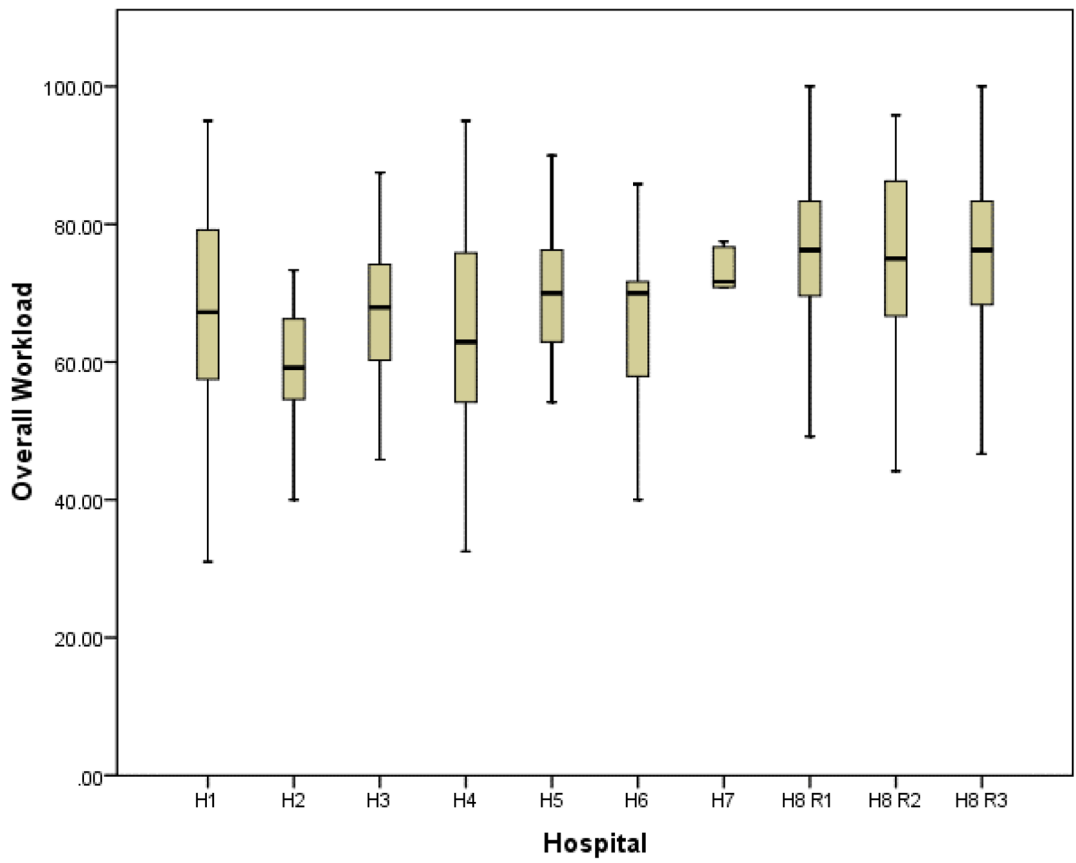
## References

- Alvarado CJ, Carayon P. Workload and patient safety among critical care nurses. *Crit Care Nurs Clin N Am.* 2007; 19:121–129.
- Backs RW, Ryan AM, Wilson GF. Psychophysiological measures of workload during continuous manual performance. *Human Factors.* 1994; 36:514–531. [PubMed: 7989054]

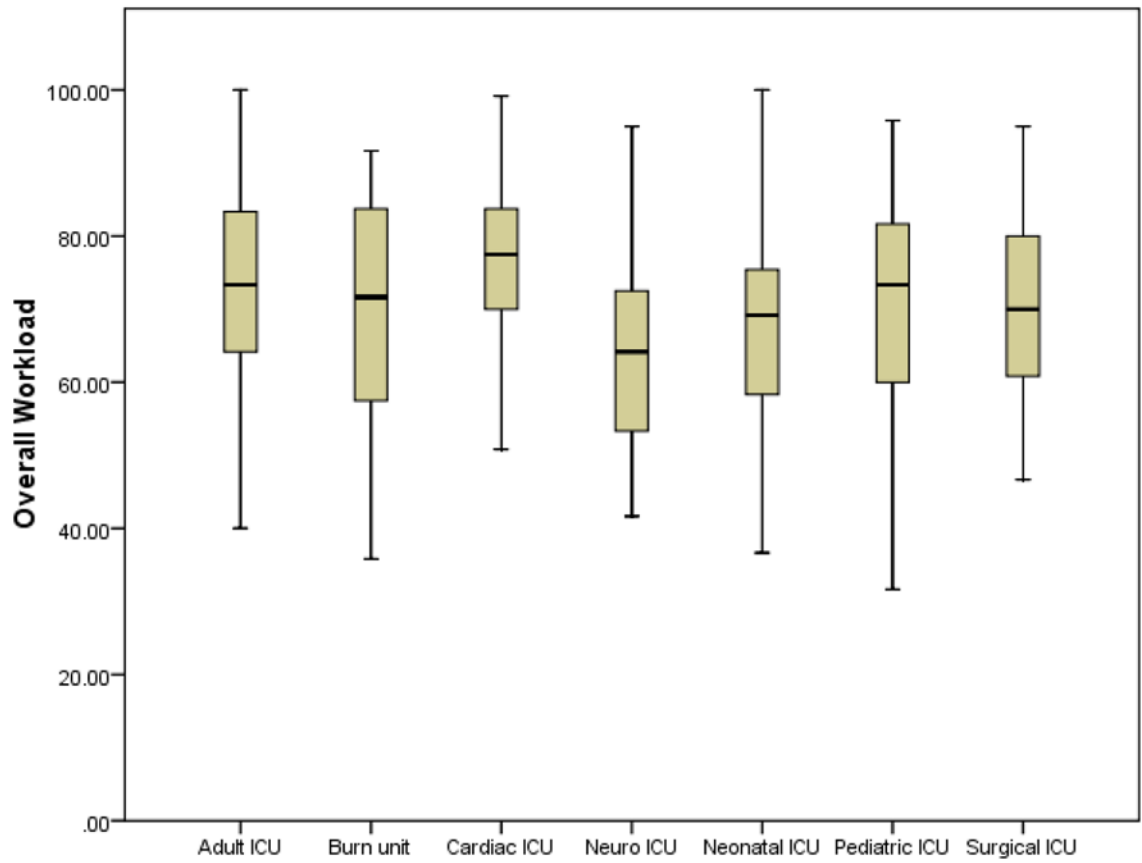
- Baggs JG, Schmitt MH, Mushlin AI, Mitchell PH, Elredge DH, Oakes D, et al. Association between nurse-physician collaboration and patient outcomes in three intensive care units. *Crit Care Med*. 1999; 27(9):1991–1998. [PubMed: 10507630]
- Battiste, V.; Bortolussi, M. Transport pilot workload: A comparison of two subjective techniques. Paper presented at the Human Factors Society Thirty-Second Annual Meeting; 1988.
- Carayon P. The balance theory and the work system model...twenty years later. *International Journal of Human-Computer Interaction*. 2009; 25(5):313–327.
- Carayon, P.; Alvarado, C.; Hundt, AS. Reducing Workload and Increasing Patient Safety Through Work and Workspace Design. Washington, DC: Institute of Medicine; 2003.
- Carayon P, Alvarado CJ. Workload and patient safety among critical care nurses. *Crit Care Nurs Clin N Am*. 2007; 19:121–129.
- Carayon P, Gurses AP. Nursing workload and patient safety in intensive care units: a human factors engineering evaluation of the literature. *Intensive Crit Care Nurs*. 2005; 21(5):284–301. [PubMed: 16182125]
- Carayon P, Hundt AS, Karsh BT, Gürses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: The SEIPS model. *Quality & Safety in Health Care*. 2006; 15(Suppl 1):i50–i58. [PubMed: 17142610]
- Cohen MM, O'Brien-Pallas LL, Copplestone C, Wall R, Porter J, Rose K. Nursing workload associated with adverse events in the postanesthesia care unit. *Anesthesiology*. 1999; 91(6):1882–1990. [PubMed: 10598633]
- Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: A method for quantitative comparison of patient care. *Critical Care Medicine*. 1974; 2(2):57–60. [PubMed: 4832281]
- Embriaco N, Papazian L, Kentish-Barnes N, Pochard F, Azoulay E. Burnout syndrome among critical care healthcare workers. *Current opinion in critical care*. 2007; 13(5):482–488. [PubMed: 17762223]
- Goldstein, H. *Multilevel Statistical Models*. 3. London: Arnold; 2003.
- Gurses, AP. Performance Obstacles and Facilitators, Workload, Quality of Working Life and Quality and Safety of Care among Intensive Care Nurses. University of Wisconsin-Madison; Madison, WI: 2005.
- Gurses AP, Carayon P. Performance obstacles of intensive care nurses. *Nursing Research*. 2007; 56(3): 185–194. [PubMed: 17495574]
- Gurses AP, Carayon P, Wall M. Impact of performance obstacles on intensive care nurses' workload, perceived quality and safety of care, and quality of working life. *Health Services Research*. 2009; 44(2):422–443. [PubMed: 19207589]
- Hart, SG. NASA-Task Load Index (NASA-TLX): 20 years later. Paper presented at the HFES; San Francisco, CA. 2006.
- Hart, SG.; Staveland, LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, PA.; Meshkati, N., editors. *Human Mental Workload*. North-Holland: Elsevier Science Publishers B. V; 1988. p. 139-183.
- Hendy KC, Hamilton KM, Landry LM. Measuring subjective workload: When is one scale better than another? *Human Factors*. 1993; 35(4):579–601.
- Hill SG, Iavecchia HM, Byers JC, Bittner AC, Zaklad AL, Christ RE. Comparison of four subjective workload rating scales. *Human Factors*. 1992; 34(4):429–439.
- Hoonakker, PLT.; Carayon, P.; Douglas, S.; Schulz, K.; Walker, J.; Wetterneck, TB. Communication in Intensive Care Units and the Relation with Quality of Care and Patient Safety. In: Snelwar, LI.; Mascia, FL.; Montedo, UB., editors. *Human Factors in Organizational Design and Management*. Vol. IX. Santa Monica, CA: IEA Press; 2008. p. 715-721.
- Hoonakker PLT, Carayon P, Walker J. Evaluation of CPOE Implementation: Measuring end-user satisfaction with a survey questionnaire. *Applied Clinical Informatics*. 2010; 1(3):268–285. [PubMed: 21423791]
- Hoonakker, PLT.; Carayon, P.; Walker, J.; Wetterneck, TB. Coordination in the ICU and the Relation with Quality of Care and Patient Safety. Paper presented at the Healthcare systems Ergonomics and Patient Safety (HEPS 2008) Conference; Strasbourg, France. 2008 Jun 25–28.

- Hoonakker PLT, Cartmill RS, Carayon P, Walker JM. Development and Psychometric Qualities of the SEIPS Survey to Evaluate CPOE/EHR Implementation in ICUs. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*. 2011; 6(1):51–69.
- Hughes M. Nursing workload: an unquantifiable entity. *J Nurs Manag*. 1999; 7(6):317–322. [PubMed: 10827626]
- Joint Commission. Sentinel Events Statistics, March 31, 2007. 2007. Retrieved August 7, 2007, from <http://www.jointcommission.org/SentinelEvents/Statistics/>
- Kiekkas P, Sakellaropoulos GC, Brokalaki H, Manolis E, Samios A, Skartsani C, et al. Association between nursing workload and mortality of intensive care unit patients. *J Nurs Scholarsh*. 2008; 40:385–390. [PubMed: 19094155]
- Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE- acute physiology and chronic health evaluation: A physiologically-based classification system. *Critical Care Medicine*. 1981; 9(8):591–597. [PubMed: 7261642]
- Kramer, AF. Physiological metrics of mental workload: A review of recent progress. In: Damos, DL., editor. *Multi-task performance*. London, UK: Taylor & Francis; 1991. p. 279-328.
- Lazard T, Metel O, Guidet B, Maury E, Valleron AJ, Offenstadt G. AIDS in a medical intensive care unit. Immediate prognosis and long-term survival. *The Journal of the American Medical Association*. 1996; 276:1240–1245.
- Llenore E, Ogle KR. Nurse-patient communication in the intensive care unit: a review of the literature. *Aust Crit Care*. 1999; 12(4):142–145. [PubMed: 11271028]
- Lysaght, R.J.; Hill, S.G.; Dick, A.O.; Plamondon, B.D.; Linton, P.M.; Wierwille, W.W., et al. *Operator workload: Comprehensive review and evaluation of operator workload methodologies (No ARI Tech Report 851)*. Fort Bliss, TX: U.S. Army Research Institute; 1989.
- McManus IC, Keeling A, Paice E. Stress, burnout and doctors' attitudes to work are determined by personality and learning style: a twelve year longitudinal study of UK medical graduates. *BMC Med*. 2004; 2(29)
- Miranda D, Nap R, de Rijk A, Schaufeli W, Iapichino G, Group TW. Nursing Activities Score. *Critical Care Medicine*. 2003; 31(2):374–382. [PubMed: 12576939]
- Miranda DR, de Rijk A, Schaufeli W. Simplified therapeutic intervention scoring system: The TISS-28 items- results from a multicenter study. *Critical Care Medicine*. 1996; 24:64–73. [PubMed: 8565541]
- Morris R, MacNeela P, Scott A, Treacy P, Hyde A. Reconsidering the conceptualization of nursing workload: literature review. *Journal of Advanced Nursing*. 2007; 57(5):463–471. [PubMed: 17284279]
- Nunnally, J.C. *Psychometric theory*. 2. California: McGraw Hill; 1978.
- Nygren TE. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*. 1991; 31:17–31.
- Rasbash, J.; Charlton, C.; Browne, W.J.; Healy, M.; Cameron, B. *MLwiN Version 2.1*. Bristol, UK: Centre for Multilevel Modelling, University of Bristol; 2009.
- Rothen HU, Kung V, Ryser DH, Zurcher R, Regli B. Validation of “nine equivalents of nursing manpower use score” on an independent data sample. *Intensive Care Medicine*. 1999; 25:606–611. [PubMed: 10416913]
- Rubio S, Díaz E, Martín J, Puente JM. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology: an International Review*. 2004; 53(1):61–86.
- Schaufeli, W.; Le Blanc, P. Personnel. In: Miranda, DR.; Ryan, D.W.; Schaufeli, W.B.; Fidler, V., editors. *Organisation and Management of Intensive Care: A Prospective Study in 12 European Countries*. Springer; 1998. p. 169-205.
- Sheridan T. Mental workload-What is it? Why bother with it? *Human Factors Society Bulletin*. 1980; 23(2):1–2.
- Tsang, P.; Wilson, G.F. Mental workload. In: Salvendy, G., editor. *Handbook of Human Factors and Ergonomics*. Upper Saddle River, NY: Wiley; 1997. p. 417-449.
- Weinger MB, Reddy SB, Slagle JM. Multiple measures of anesthesia workload during teaching and nonteaching cases. *Anesth Analg*. 2004; 98(5):1419–1425. [PubMed: 15105224]

Young G, Zavelina L, Hooper V. Assessment of Workload Using NASA Task Load Index in Perianesthesia Nursing. *Journal of PeriAnesthesia Nursing*. 2008; 23(2):102–110. [PubMed: 18362006]



**Figure 1.**  
Overall workload by hospital



**Figure 2.**  
Overall workload by ICU type



**Table 1**

Sample description

Site	Hospital type	Academic/ non-academic	Location	Staffed beds	Number of participating ICUs	Number of nurses	
Study 1	1	Nongovernment, not-for-profit	Academic	Urban	471	6: Medical/surgical ICU Burn Unit Cardiothoracic ICU Neurosurgery ICU Pediatric ICU Cardiac ICU	55 12 20 23 33 15
	2	Nongovernment, not-for-profit	Non-academic	Urban	178	2: Medical/surgical Neuro-Trauma	12 8
	3	Nongovernment, not-for-profit	Non-academic	Urban	377	1: Adult ICU	16
	4	Nongovernment, not-for-profit	Non-academic	Rural	319	4: Neonatal ICU Pediatric ICU Medical/coronary ICU Surgical ICU	25 10 16 23
	5	Nongovernment, not-for-profit	Non-academic	Urban	295	2: Surgical ICU Medical ICU	4 8
	6	Nongovernment, not-for-profit	Non-academic	Urban	258	1: Adult ICU	15
	7	Nongovernment, not-for-profit	Non-academic	Urban	226	1: Adult ICU	5
	8	Nongovernment, not-for-profit	Non-academic	Rural	424	4:	R1* R2* R3*

Site	Hospital type	Academic/ non-academic	Location	Staffed beds	Number of participating ICUs	Number of nurses
					Adult ICU Cardiac ICU Neonatal ICU Pediatric ICU	51   26   43 48   35   46 55   36   48 25   21   23
Total	8 hospitals	1 academic 7 non-academic	3 rural 5 urban	2548 beds	21 ICUs	179   121   160 757 respondents

\* R1, R2, and R=Round 1, Round 2, and Round 3 of data collection

**Table 2**  
Dimensions of workload and overall workload by hospital (H1–H7), Means, Confidence Interval (CI) and standard deviations

	H1 (N=156)	H2 (N=20)	H3 (N=15)	H4 (N=70)	H5 (N=12)	H6 (N=15)	H7 (N=5)	H8, R1* (N=180)	H8, R2* (N=119)	H8, R3* (N=160)	Total	CI LB*	CI UB*
MD	76.15 (21.59)	60.75 (24.78)	75.67 (19.90)	75.93 (18.30)	86.67 (11.74)	74.0 (17.75)	64.0 (34.17)	87.97 (10.12)	87.10 (12.81)	87.44 (13.09)	82.83 (16.97)	81.62	84.05
PD	69.84 (23.74)	62.75 (21.18)	76.25 (14.55)	56.36 (29.48)	68.33 (22.40)	76.33 (14.70)	75.0 (31.02)	69.75 (24.65)	68.95 (27.93)	72.47 (24.28)	69.14 (25.21)	67.34	70.95
TD	62.98 (26.04)	51.25 (23.33)	60.62 (17.60)	59.43 (24.31)	69.58 (23.01)	62.00 (19.35)	65.00 (33.73)	74.52 (16.64)	77.23 (18.48)	75.78 (16.88)	70.22 (21.52)	68.68	71.76
EF	71.27 (20.73)	56.75 (21.84)	74.06 (14.74)	67.50 (19.92)	76.36 (15.02)	68.00 (17.40)	70.00 (37.58)	80.28 (13.56)	79.54 (16.47)	81.23 (13.98)	76.29 (17.74)	75.02	77.57
PE	75.96 (16.99)	74.00 (18.25)	75.94 (18.91)	80.50 (13.94)	75.42 (20.50)	73.67 (20.31)	64.00 (34.53)	80.53 (12.25)	76.50 (17.67)	80.88 (12.56)	78.42 (15.44)	77.31	79.63
FR	43.77 (26.09)	38.25 (22.90)	36.25 (28.31)	39.93 (24.56)	45.83 (28.83)	38.00 (25.55)	46.00 (32.67)	58.49 (26.27)	58.24 (25.28)	56.00 (27.09)	51.45 (27.14)	49.51	53.40
OW	66.7 (15.4)	60.0 (11.2)	66.3 (11.0)	63.3 (14.2)	70.3 (10.9)	65.3 (12.7)	64.0 (22.9)	75.31 (10.84)	74.60 (12.44)	75.61 (11.48)	71.40 (13.62)	70.43	72.37

\* H8, R1 = Hospital 8, 1<sup>st</sup> round of data collection; CI LB = 95% Confidence Interval, Lower Bound, CI UB = Confidence Interval, Upper Bound

\*\* MD=Mental Demands, PD=Physical demands, TD=Temporal Demands, EF=Effort, PE=Performance, FR=Frustration, OW= Overall Workload

**Table 3**

Workload dimensions and overall workload by ICU, means and confidence interval (CI)

	AICU (N=232)	Burn unit (N=12)	CICU (N=180)	Neuro ICU (N=31)	NICU (N=163)	PICU (110)	SICU (N=25)	Total	CI LB*	CI UB*
MD	82.36	79.17	85.39	67.90	84.29	82.64	80.40	82.83	81.62	84.05
PD	78.19	73.75	80.73	74.35	45.06	64.91	69.20	69.14	67.34	70.95
TD	70.06	67.50	75.03	57.58	71.56	65.27	67.20	70.22	68.68	71.76
EF	78.28	71.67	81.30	64.52	73.83	71.83	74.80	76.29	75.02	77.57
PE	78.23	75.83	79.78	75.32	78.67	78.95	71.60	78.42	77.31	79.53
FR	51.24	43.75	53.10	42.74	51.07	53.09	51.40	51.45	49.51	53.40
OW	73.08	68.61	75.85	63.74	67.42	69.45	69.10	71.40	70.43	72.37

\* CI LB = 95% Confidence Interval, Lower Bound, CI UB = Confidence Interval, Upper Bound

**Table 4**

Comparison and correlations of different workload measures

	MD	PD	TD	EF	PE	FR	OW	RSME	PWS	NPR	NAS	NAS MAX
MD	1											
PD	.38**	1										
TD	.65**	.43**	1									
EF	.72**	.53**	.73**	1								
PE	.02	-.08	-.13*	-.04	1							
FR	.29**	.32**	.46**	.43**	-.36**	1						
OW	.78**	.69**	.83**	.87**	.03	.62**	1					
RSME	.61**	.47**	.68**	.76**	-.04	.44**	.77**	1				
PWS	.65**	.54**	.75**	.71**	-.16**	.51**	.81**	.74**	1			
NPR	-.02	.02	.15**	.11	-.13*	.18**	.10	.05	.10	1		
NAS	.28**	.37**	.41**	.40**	-.08	.31**	.45**	.39**	.42**	.57**	1	
NAS MAX	.37**	.38**	.43**	.44**	-.01	.23**	.48**	.43**	.42**	.13*	.81**	1






\* Correlation is significant at the 0.05 level,

\*\* is significant at the 0.01 level

**Appendix 1**

NASA TLX (can also be a Table)

The following questions deal with the workload that you experience in your job. Please put an ‘X’ on each of the following six scales at the point that matches your overall experience of workload.

	Low	High
1. <b>Mental demand.</b> How much mental activity is required to perform your job (thinking, deciding, calculating, remembering, looking, searching, etc...)?		
2. <b>Physical demand.</b> How much physical activity is required to perform your job (e.g., pushing, pulling, turning, controlling, activating, etc.)?		
3. <b>Temporal demand.</b> How much time pressure do you feel due to the rate or pace at which the tasks or task elements occurred?		
4. <b>Effort.</b> How hard do you have to work (mentally and physically) to accomplish your level of performance?		
5. <b>Performance.</b> How satisfied are you with your performance at your job?		
6. <b>Frustration level.</b> How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel about your job?	