A Real-Time Assessment of Work Stress in Physicians and Nurses

Thomas Rutledge
VA San Diego Healthcare System and University of California

Erin Stucky
Rady Children’s Hospital and UCSD Department of Pediatrics

Adrian Dollarhide and Martha Shively
VA San Diego Healthcare System

Sonia Jain and Tanya Wolfson
University of California

Matthew B. Weinger
Vanderbilt University Medical Center and VA Tennessee Valley Healthcare System

Timothy Dresselhaus
VA San Diego Healthcare System

Objective: This study adapted ecological momentary assessment methods to: (a) examine differences in work stress between nurses and physicians, and (b) to study relationships between work stress, work activity patterns, and sleep. Design: A total of 185 physicians and 119 nurses (206 women, 98 men) working in four teaching hospitals participated in an observational study of work stress. Main Outcome Measures: Participants carried handheld computers that randomly prompted them for work activity, patient load, and work stress information. Results: Participants completed more than 9,500 random interval surveys during the study (an average of 30.8 surveys per person-week). Approximately 85% of all surveys were completed in full (73.3%) or partially (11.6%). Emotional stress scores among physicians were nearly 50% higher (26.9[19.0]) than those of nurses (18.1[14.9], r[302] = .37, p < .001). Direct and indirect care activities were associated with higher stress reports by both clinician groups (rs[159] = .14-.26, ps < .01). Sleep quality and quantity were predictors of work stress scores (ps < .05). Finally, higher work stress and lower sleep quality were also associated with poorer memory performance (r[302] = -.12,.17, ps < .05). Conclusions: The findings identify patterns of work stress in relationship to work activities, sleep habits, and provider differences that may be used to assist ongoing hospital work reform efforts.

Keywords: work stress, ecological momentary assessment, emotional stress, physicians, nurses

The daily work of physicians and nurses in the hospital setting is inherently stressful (Firth-Cozens, 2003; Lee & Wang, 2002; Thomas, 2004). Long workdays, high case loads, time pressures, poor sleep habits, and high performance expectations contribute to the well-documented rates of burnout, depression, job dissatisfaction, and workplace fatigue among nurses and physicians (Clever, 2002; Geurts, Rutte, & Peeters, 1999; Hillhouse & Adler, 1997; Inama, Nakao, Tsuchiya, Kuroda, & Katoh, 2004; Papp, Stoller, Sage, Aikens, Owens, Avidan et al., 2004; Rose, Ware, Kolm, & Risser, 2000; Shanafelt, Bradley, Wipf, & Back, 2002). The consequences of these characteristics may include decreased patient satisfaction, diminished care quality, and increased rates of medical errors (DiMatteo, Sherbourne, Hays, Ordway, Kravitz, McGlynn et al., 1993; Fletcher, Davis, Underwood, Mangrulkar, McMahon, & Saint, 2004 Grol, Mokkink, Smits, van Eijk, Beek, Mesker et al., 1985; Tarnow-Mordi, Hau, Warden, & Shearer, 2000; Weinger & Ebden, 2002). Prior studies have tended to focus on specific clinician subgroups—most often medical residents—and less is known concerning how often or under what situations elevated work stress is present. The job requirements for nurses and physicians also differ substantially. Assessing these clinician groups separately can allow for a more precise understanding of the factors contributing to work-related stress.

A limitation of nearly all previous hospital stress research is the reliance upon cross-sectional or retrospective survey methods. The validity of survey results is compromised by factors such as the single time point measurements, recall biases, difficulties determining the direction of temporal relationships, and the inability to link responses to the context in which they occurred (Kahneman, Krueger, Schkade, Schwarz, Stone, 2004; Kamarck, Debkski, & Manuck, 2002; Stone, Schwartz, Neale, Shiffman, Marco, Hickcox et al., 1998). Ecological momentary assessment methods (EMA), in which dynamic person experiences are captured in real-time, can improve upon these limitations and are a well-established methodology for naturalistic observation (Schwartz & Stone, 1998; Stone & Shiff-
Despite these advantages, however, EMA methods are rarely utilized in medical environments because of perceived logistical barriers (Whitman, Sereika, & Dachille, 2003).

This study is perhaps the first to use EMA technology to directly assess relationships between work stress and behavioral and contextual factors in the work environment. In this study, we hypothesized work stress in the hospital environment to be a product of both extrinsic (work activity, patient load, shift schedule, etc.) and intrinsic (sleep, cognitive, and memory function, etc.) factors. Figure 1 illustrates the conceptual model underlying this study, designed to identify associations between extrinsic and intrinsic factors with rates of adverse medical events using EMA methods. This figure describes work stress in terms of dimensions such as poor mood, increased perceived task demand, reduced cognitive performance, and fatigue, and postulates that these stress factors are affected by contextual variables including patient load, type of work activity, and recent sleep habits, among others.

The objective of this paper was to describe an application of EMA methods to the study of emotional stress, perceived workload, and hospital work activities across the 24-hr hospital workday in a sample of physicians and nurses. Participants were sampled using handheld computers programmed to collect real-time stress and work information from users at random intervals across the workday, and the data collection interface included three measures of emotional stress, perceived workload, and hospital work activities. Our specific goals were to: (a) Assess the feasibility of EMA methods for measuring emotional characteristics and job-related stress in a clinician sample; (b) Examine the pattern of work stress across different types of hospital work activities, with the expectation that those involving more direct patient care would be associated with higher work stress levels; and (c) Investigate relationships between daily sleep patterns and work stress characteristics, wherein we projected that work stress scores would increase in relation to less sleep and poorer quality sleep in the previous 24 hours.

Method

Participants

Study participants included 304 nurses and physicians caring for patients in pediatric or internal medicine inpatient settings in the four largest, core training sites in San Diego: Department of Veterans Affairs San Diego Health care System (VASDHS), University of California San Diego (UCSD) Hillcrest, UCSD Thornton, and Rady Children’s Hospital San Diego (RCHSD). Data collection took place over an 18-month period beginning April 2004. Participants were recruited through announcements at conferences, rounds and meetings, and by direct contact with the primary investigator, coinvestigators, or specially trained research assistants. Recruitment selectively targeted high-risk areas for adverse medication events based upon locally recorded hospital error statistics from 2003, including emergency care units, intensive care units, spinal cord injury units, and direct observation units. The physician sample comprised teams of attending physicians, residents, and interns. Physician teams at all sites were similarly structured with attending, senior resident, and 1 to 2 interns or two postgraduate year-2 residents constituting a team. Call frequency was no more than every fourth night at all sites. The research team included designated research coordinators working in each of the hospitals to direct recruitment efforts and to be responsible for data collection. IRB approval was obtained at each hospital and all participants provided written informed consent.

EMA Assessment

Study participants carried dedicated handheld computers for 1 week, during which they were sampled randomly over 90-min

---

**Flowchart of adverse drug events resulting from work processes and psychological factors in the workplace**

PH – palmheld
BTA – behavioral task analysis

*Figure 1.* Conceptual model addressing relationships between hospital activities, work stress, and performance in physicians and nurses.
The DHST collected three types of information from participants: (a) an initial, baseline sign-on that established an identification number and queried demographic information including demographic data, worksite, cognitive function as measured by the two 15-item logical memory II subtests from the Wechsler Memory Scale (Wechsler, 1987), provider type and history, and other static variables; (b) a daily sign-on that prompted participants at the beginning of each workday for information such as the participant’s previous night’s sleep and number of patients currently under their care; and (c) the random repeating daily surveys, that assessed dynamic characteristics such as current work activities and location, emotional stress characteristics, and perceived workload. Internal consistency for each of the two 15-item portions of the adapted memory test were good to very good ($r = .48-.85$, $p < .001$), and did not differ significantly by clinician group.

Intervals throughout each workday. The handheld computers and custom software, labeled the Dynamic Handheld Survey Tool (DHST), were designed for maximal flexibility to permit deployment across the highly variable schedules and demands of both nurses and physicians. Participants could enter start and end of workday information, manually turn off the devices during off-duty times, and delay information prompts when they occurred during critical patient care activities. Participants were instructed to use the DHST during the week while physically working at a hospital facility. No sampling, for example, was collected from users during on-call periods in which they were not actually present in the hospital. The interface allowed them to terminate random surveys at the end of their workdays (manual termination). The software was also programmed to deactivate automatically after periods of disuse (auto termination).

The DHST collected three types of information from participants: (a) an initial, baseline sign-on that established an identification number and queried demographic information including demographic data, worksite, cognitive function as measured by the two 15-item logical memory II subtests from the Wechsler Memory Scale (Wechsler, 1987), provider type and history, and other static variables; (b) a daily sign-on that prompted participants at the beginning of each workday for information such as the participant’s previous night’s sleep and number of patients currently under their care; and (c) the random repeating daily surveys, that assessed dynamic characteristics such as current work activities and location, emotional stress characteristics, and perceived workload. Internal consistency for each of the two 15-item portions of the adapted memory test were good to very good ($r > .80$), and correlated moderately in both clinician samples ($r = .27, .40$, for nurses and physicians, respectively, $p < .001$). Development and refinement of the DHST occurred over an extended pilot test period, including clinician participants from the targeted hospitals. Figure 2 provides a sample illustration of the DHST platform.

## Work Activities

The work activity assessment consisted of a branching logic series of questions to determine the location of the provider, participation in education, direct or indirect patient care, outpatient care, transit, personal, or unspecified (other) activities (Dresselhaus, Luck, Wright, Spragg, Lee, & Bozzette, 1998). Because of the many possible kinds of workplace activities for hospital providers, the branching logic questions produced hundreds of unique permutations. Assigned team providers then categorized these combinations into the above activity categories using previously published methods (Dresselhaus et al., 1998).

## Perceived Workload

Perceived workload and work demand characteristics were assessed in our provider sample using a computer-adapted form of the National Aeronautics and Space Administration-Task Load Index (NASA-TLX), a validated six-item measure of perceived workload and demand in the workplace (NASA Human Performance Research Group, 1987). The NASA-TLX scale has been used in dozens of previous studies addressing human factors in relationship to accidents, errors, and workload characteristics (Di Nocera, Fabrizi, Terenzi, & Ferla, 2006). For example, differences in NASA-TLX ratings were found to correspond to low and high traffic density settings among air traffic controllers (Di Nocera, Fabrizi, Terenzi et al., 2006). The NASA-TLX items measure perceptions of level of physical and mental demand, frustration, effort, perceived time pressure, and level of performance. The disparate domains measured by the NASA-TLX effectively capture the broad range of activities faced by nurses and physicians, whose jobs include important physical (e.g., nurses moving patients), mental (interacting with patients), time (brief patient visits), and performance pressure (little or no margin of error in clinical performance) components that are likely to affect effort and frustration levels. The NASA-TLX items appeared in the DHST with exactly the same wording and measurement format (10-point Likert scale) as in the paper version. Participants responded to each Likert question using a slide bar built into the DHST software to rate their perceptions on a 0 (low) to 9 (high) scale. Possible scores can range from 0 to 54. The internal consistency value for the NASA-TLX (.81) was very good. NASA-TLX subscales correlated moderately to strongly with the total scale score ($r = .48-.85$, $p < .001$), and did not differ significantly by clinician group.

## Emotional Stress Characteristics

The emotional stress assessment portion of the repeating surveys included a study-developed instrument containing 10 Likert-scale questions. The tool was developed using three items from the Diary of Ambulatory Behavioral States (Kamarck, Shiffman, Smithline, Goodie, Paty, Gny, et al., 1998), which also assessed emotional states in a palm computer format, and further expanded from pilot testing using emotion item domains identified from the large job strain and burnout among clinicians literatures (Geurts, Rutte, & Peeters, 1999; Hillhouse & Adler, 1997; Imai, Nakao, Tsuchiyu, Kuroda, & Katoh, 2004; Papp, Stoller, Sage, Aikens, Owens et al., 2004; Shanafelt, Bradley, Wipf, & Back, 2002; Thomas, 2004). The final emotion scale included items assessing...
participants’ feelings of tension, alertness, anger, sadness, fatigue, stress, unhappiness, tired, worried, and upset. Each item was rated from 0 to 9 to parallel the NASA-TLX display, and with higher scores indicating a greater magnitude of the characteristic in question. Total scores can range from 0 to 90. A time stamp (date, hour, minute) was recorded with each random survey. The internal consistency of the emotional stress scale (.90) was excellent.

Procedure

Participants met with the local research coordinator to receive DHST units and training. In most cases, DHST units were deployed for data collection carried out over periods of 4 to 6 days. Participants returned units the following week to the research coordinator, who downloaded information to a hospital study database on a secured server. Units were then recharged overnight for redeployment the following day. Each hospital site was provided with backup DHST units to permit continuous use in the case of malfunction or delayed return.

After receiving the DHST units and training, participants completed their initial (one time) sign-on. This was followed by the first daily sign-on, and thereafter by activity and work stress prompts that occurred randomly within blocks of 90-min sampling intervals across the day. In a 9-hr workday, six random surveys would occur, wherein any two surveys could be separated by as short as a few minutes or up to three hours depending on the randomization of the timing within the 90-min interval. In cases in which participants failed to respond to the random prompts, the DHST software reactivated after 5-min. If there were no response across three consecutive random interval prompts, the units automatically deactivated and restarted the following morning (in which case the three missed prompts were not counted against compliance rates). Participants could manually delay a response in the case of involvement in activities that could not be interrupted or to complete an interrupted survey; this action resulted in a new prompt after 5 minutes had passed. The software stored incomplete as well as complete survey responses. Participants were permitted to enroll in the study for multiple data collection weeks.

Equipment

We chose the Palm Zire 21 for data collection based on a balance of factors including size, weight, battery life, screen clarity/size, memory capacity, cost, and audibility of alarms. Software was written using Visual Basic and the AppForg operating system. A HotSync (HotSync for Palm OS) conduit allowed transfer of data from each hospital study database to the centralized server database.

Statistical Analyses

Reported statistics include descriptive findings of the demographic and psychosocial characteristics of the nurse and physician samples. We used chi-square tests for nurse-physician comparisons of proportions and percentages (e.g., survey completion rates). Analysis of variance (ANOVA) and t tests were used to assess differences in means. Group comparisons of work stress were calculated as between-groups tests using person-means aggregating across each participant’s days of use for each of the 304 nurses and physicians. Comparisons between work activities were completed using paired t tests of aggregated person-means. We reduced the work activities to three categories—direct care, indirect care, and combined other (merging the separate other, personal, transit, outpatient, communication, and education activities)—for these comparisons to create more hypothesis driven tests and reduce the number of total comparisons.

We used mixed effects linear regression to examine relationships between sleep (quality and quantity) and work stress scores. Covariates included age, gender, familiarity with patients, average patient load, and number of admissions in 24 hours. Direct patient care was coded as a binary variable comparing stress levels during direct patient care versus all other categories. Sleep predictors were modeled as fixed effects. A random (subject-specific) intercept was added to the regression models to account for within-patient dependence. Described correlations between memory scores and sleep and work stress were based on the 304 independent initial sign-on days for clinicians, as memory was assessed only one time. All tests were interpreted using a two-tailed alpha of .05 as the minimum criterion for significance. Separate inferential tests for nurses and physicians are presented in cases in which combined analyses indicated different results patterns. Analyses were completed using SPSS for windows, version 15.0 (www.spss.com).

Results

The study sample is described in Table 1, consisting of 185 physicians (age range 25–47), and 119 nurses (age range 23–59), completing a total of 2,291 independent days of reporting. Although physicians as a group completed more total daily surveys than nurses (5,673 vs. 3,688), the average number of daily surveys completed [4.4 (or 6.2 for the more common weekday only participant) on average per day for a 7-day week by both nurses and physicians] did not differ. Physicians were more likely than nurses to complete the random daily surveys (75.8% vs. 69.3%, respectively, \( \chi^2[1] = 48.0, p < .001 \)); however, they were more likely to miss a random survey as well (16.4% vs. 13.2%, \( \chi^2[1] = 13.6, p < \).001.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Nurses (N = 119)</th>
<th>Physicians (N = 185)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age**</td>
<td>39.3 (9.8)</td>
<td>30.2 (4.5)</td>
</tr>
<tr>
<td>Gender (% female)**</td>
<td>89%</td>
<td>54%</td>
</tr>
<tr>
<td># Patients*</td>
<td>2.7 (1.4)</td>
<td>7.0 (4.2)</td>
</tr>
<tr>
<td>Work day length</td>
<td>10.7 (2.6)</td>
<td>10.1 (1.8)</td>
</tr>
<tr>
<td>Sleep hours</td>
<td>6.4 (1.4)</td>
<td>6.4 (2.1)</td>
</tr>
<tr>
<td>Sleep quality*</td>
<td>1.9 (2.0)</td>
<td>1.9 (2.1)</td>
</tr>
<tr>
<td>Total random surveys</td>
<td>3688</td>
<td>5673</td>
</tr>
<tr>
<td>Survey response time (seconds)**</td>
<td>110.0 (19.0)</td>
<td>86.8 (14.7)</td>
</tr>
<tr>
<td>% completed responses**</td>
<td>69.3</td>
<td>75.8</td>
</tr>
<tr>
<td>% Partial responses*</td>
<td>17.5</td>
<td>7.8</td>
</tr>
<tr>
<td>% Missing responses*</td>
<td>13.2</td>
<td>16.4</td>
</tr>
</tbody>
</table>

* Sleep quality rated on a scale from 0 (very bad) to 3 (very good).
Mean values were compared using \( t \) test, tests of proportions completing using chi-square test of proportions.

\( p < .01 \). ** \( p < .001 \).
.001). Partially completed surveys accounted for the remainder (see Table 1). The occurrence of partially completed and missed surveys was not associated with time of day, work activity, or NASA-TLX scores (rs [302] < .10, NS). As evidence of convergent validity, the emotional stress scale scores were moderately correlated with NASA-TLX ratings for both clinician groups (.50 for nurses, .59 for physicians). Between-hospital comparisons of clinicians yielded no evidence of differences regarding usage patterns on the DHST or clinical characteristics (results not shown).

**NASA-TLX and Emotional Stress in Nurses and Physicians**

Although both clinician groups showed evidence of frequent elevations in work stress, physician’s emotional stress scores were approximately 50% higher than those reported by nurses (26.9 [19.0] vs. 18.1 [14.9], respectively, t[302] = .24, p < .001). Summarized in Table 2, individual item comparisons on the emotional stress scale indicated that physicians, relative to nurses, reported feeling less alert and more worried, tense, fatigued, unhappy, tired, upset, and stressed.

The clinician groups showed similar total NASA-TLX scores (26.5 [10.5] vs. 26.9 [12.1]) for physicians and nurses, respectively, t[302] = .02, p > .5), but did differ on specific item dimensions. Relative to physicians, nurses significantly higher levels of high physical demand (t[302] = .33, p < .001) and performance (t[302] = .18, p = .001), and lower levels of frustration (t[302] = .33, p < .001).

**Work Activities and Work Stress Reports**

Table 3 describes aggregated person-means of work activity categories by provider category and work stress scale. Paired t test results indicated that, among physicians comparisons of NASA-TLX scores, direct patient care was associated with higher ratings compared to either indirect care (t = .14, p = .002) or all other work activities (t = .24, p < .001). Indirect care values also exceeded other work activity ratings (t = .23, p < .001). Using emotional stress scores, direct and indirect care activities did not differ (t = .01, ns), but each of the latter remained reliably higher compared to other work activity averages (rs = .15, ps = .001).

Among nurses, direct patient, indirect care, and other activities differed systematically on the NASA-TLX. Direct care produced higher perceived demand ratings than indirect care (t = .14, p = .002), and all other work activities (t = .26, p < .001). Indirect care NASA-TLX ratings were also higher compared to other work activities (t = .25, p < .001). Emotional stress ratings also differed by work activity. Direct (t = .14, p = .001) and indirect (t = .17, p < .001) care activities were each associated with significantly higher emotional stress ratings compared to other categories. However, direct and indirect care categories were similar to each other for nurses (t = .02, ns).

**Sleep, Work Stress, and Cognitive Performance**

We observed a generally linear relationship between a clinician’s previous night of sleep and their work stress reports the following day, based upon 2,291 days of reporting (1,509 physician-days, and 782 nurse-days). A total of 22.4% of daily samples from physicians and 22.5% from nurses, indicated five or fewer hours of sleep the previous night. After adjusting for demographic and clinical characteristics, reported sleep quantity remained a significant predictor of emotional stress scores for both nurses (r = .09, p = .02) and physicians (r = .12, p < .001). Sleep quality was also a reliable predictor for the clinician groups (rs = .09 and .11, ps = .02, < .001). Adjusted relationships between sleep measures and NASA-TLX scores were not significant.

Finally, based on the 304 independent initial sign-on days by clinicians, better reported sleep quality from the previous night was positively associated with cognitive function as measured by logical memory scores, and higher emotional stress ratings were negatively associated with memory scores (rs[302] = .12, —.17, with sleep quality and emotional stress ratings, respectively, ps = .03, .01).

**Discussion**

This study employed EMA methods to assess work and emotional characteristics among physicians and nurses on inpatient medical and pediatric units. Although EMA is an increasingly common technique for obtaining self-report information in naturalistic environments (Kamarck, Debski, & Manuck, 2000), the real-time nature of EMA data collection presents potential difficulties in environments in which participants may be consistently too busy to respond. Contrary to these expectations, however, providers in this multisite hospital study were able to reliably respond to more than 80% of the random assessment prompts. This response rate suggests that EMA tools can efficiently capture multidimensional information from practicing clinicians with an acceptable response burden.

The results from this study extend previous research in several ways. First, the use of EMA methods permitted repeated within-person assessments of stress, workload, and activity patterns across extended periods, typically a full work week or more. This represents a substantial improvement from one-time retrospective questionnaire surveys used in many previous studies (Stone,

---

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Nurse-Physicians Means(SD) on the Emotional Stress Scale* and NASA-TLX*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse (N = 119)</td>
<td>Physician (N = 185)</td>
</tr>
<tr>
<td>Emotional stress scale**</td>
<td>18.1 (14.9)</td>
</tr>
<tr>
<td>Alert**</td>
<td>7.2 (2.0)</td>
</tr>
<tr>
<td>Anger**</td>
<td>1.2 (1.8)</td>
</tr>
<tr>
<td>Fatigued**</td>
<td>3.0 (2.5)</td>
</tr>
<tr>
<td>Sad**</td>
<td>1.0 (1.6)</td>
</tr>
<tr>
<td>Stressed**</td>
<td>2.8 (2.6)</td>
</tr>
<tr>
<td>Tension**</td>
<td>2.4 (2.3)</td>
</tr>
<tr>
<td>Tired**</td>
<td>3.2 (2.6)</td>
</tr>
<tr>
<td>Unhappy**</td>
<td>1.3 (1.8)</td>
</tr>
<tr>
<td>Upset**</td>
<td>1.4 (2.0)</td>
</tr>
<tr>
<td>Worried**</td>
<td>1.5 (2.0)</td>
</tr>
<tr>
<td>NASA-TLX</td>
<td>26.9 (12.1)</td>
</tr>
<tr>
<td>Mental demand</td>
<td>5.2 (2.7)</td>
</tr>
<tr>
<td>Time pressure</td>
<td>4.5 (2.7)</td>
</tr>
<tr>
<td>Effort</td>
<td>5.0 (2.7)</td>
</tr>
<tr>
<td>Frustration**</td>
<td>2.5 (2.5)</td>
</tr>
<tr>
<td>Physical demand**</td>
<td>3.6 (2.7)</td>
</tr>
<tr>
<td>Performance**</td>
<td>6.7 (2.3)</td>
</tr>
</tbody>
</table>

* Items scored on a 0–9 Likert-scale, 0 = low  ** Groups differ, p < .05.
Schwartz, Neale et al., 1998). Second, this study included a multisite sample of nurses and physicians, allowing an examination of work stress across separate clinician groups. The handheld computers used in the study included measures of stress and perceived workload measures adapted to the hospital sample. These measures assessed a broad array of distinct work stress characteristics in a continuous fashion, which may be more informative for designing workplace modifications than more generalized conditions such as burnout (e.g., Shanafelt et al., 2002). Finally, our approach imbedded work stress questions in a repeating survey that also included questions about important contextual factors such as work activity and sleep patterns in a far more sophisticated temporal pattern than available from previous research. Using separate measures of perceived work demand and emotional stress, we observed evidence of consistent higher work stress among physicians relative to nurses. Although our emotional stress scale was novel, the psychometric properties of the instrument (internal consistency and convergent validity with the NASA-TLX) were strong, supporting the value of the instrument. Differences in stress levels by work activity were also observed, with strikingly similar patterns reported by nurses and physicians regarding both differences and magnitude of effects. For both clinician groups, patient care activities (direct and indirect) were associated with higher perceived demand on the NASA-TLX and emotional stress scores compared to other work activities. Direct care ratings on the NASA-TLX also exceeded those of indirect care, however, the emotional stress levels described in these two work categories were similar. Effect size differences were in the small range for direct care-indirect care comparisons, and moderate for comparisons of patient care versus all other activities. To our knowledge, this is the first study to compare work stress dimensions in hospital providers according to their work activities.

In our results, sleep was a robust predictor of emotional stress ratings, even after adjusting for a variety of clinician and workload factors. Although our methods and analyses were more sophisticated than most previous sleep-hospital stress studies, the findings are consistent with previous literature. Using a variety of survey methods, sleep impairment has been associated with an increased rate of accidents, impaired cognitive performance, and burnout among medical providers (e.g., Dyrbye, Thomas, & Shanafelt, 2006; Gill, Haerich, Westcott, Godenick, & Tucker, 2006; Philibert, 2005) in previous studies. In a fast-moving clinical environment, fatigue and stress can compromise a provider’s memory, attention and information processing abilities, raising the risk of impaired clinical decisions and actions (Landrigan, Barger, Cade, Ayas, & Czeisler, 2006). We likewise observed that both sleep quantity and sleep quality were independent predictors of emotional stress ratings, suggesting unique value in these dimensions. The widespread presence of job stress and sleep deprivation in this study—despite recently mandated reductions in medical student workload hours (Accreditation Council of Graduate Medical Education, 2002)—suggests that sleep and fatigue-related factors remain important elements of work stress in the hospital environment.

**Study Limitations**

This descriptive study has several important limitations. It wasn’t possible to randomly select clinicians for participation. Although we have no evidence to suggest that volunteers differed systematically from nonparticipants, generalization of the findings to the broader population of nurses and physicians must be made with caution, especially to those outside of the higher risk clinics sampled here. Some missing surveys may not have been random, possibly occurring more often in situations with greater work demand, emotional stress, or both. Because of the nature of the sample and hospital setting, it wasn’t possible to enforce or record participant compliance with the device directly. Our software had many built in features to maximize compliance, such as repeatedly prompting users, encouraging and providing flexibility for responding through the interface, and automating the activation and deactivation process each day whereas allowing for manual overrides. This same built in flexibility, however, also provided the opportunity for a participant to turn off the device even while at work if they were inclined. This is a limitation of our EMA methodology with no obvious solution. Although our EMA approach included random sampling, we were constrained to a maximum of about six surveys per day to maintain a reasonable response burden. We cannot determine whether a more or less frequent sampling approach may have altered the emotional stress or work activity reporting patterns observed in this study. The assessment of emotional stress, including the classification of moderate and higher stress described in Table 2, was based upon a locally developed scale with only partial validation from previous research. Although we are unaware of any existing instruments designed to broadly assess emotional stress in the manner performed here that have been validated with handheld computers, it is important to recognize this limitation and encourage additional research. We didn’t collect ethnicity data on participants, who were primarily White and Asian, or health information that could have strengthened the implications of the job stress data. Lastly, use of the DHST during the workday may have induced...
some level of stress for participants. Although there is no evidence to suggest that the DHST was differentially more challenging to use for different clinician groups, it is important to acknowledge that the pattern of results may have been somewhat different using a less reactive methodology such as behavioral observation or retrospective questionnaires.

Conclusions

This study employed real-time EMA methods in a multisite population of physicians and nurses with three primary objectives: (a) to demonstrate the feasibility of EMA methods in a clinician sample; (b) to examined relationships between work activity categories and work stress responses, and (c) to study relationships between sleep patterns and work stress. Overall, physicians reported substantially higher levels of work stress compared to nurses, and both clinician groups showed elevated work stress during patient care activities compared to activities such as education, transit, or communication. Approximately 20% of the sampled days were completed by clinicians reporting less than 6 hours of sleep the night before, and reduced sleep quantity and quality were associated with significantly higher levels of emotional stress and perceived workload. These results offer a novel approach to studying the workplace demands faced by physicians and nurses, and offer evidence that can help guide ongoing efforts to improve the training and work environment for clinicians.

References


