Association of Resident Fatigue and Distress With Perceived Medical Errors

Colin P. West, MD, PhD
Angelina D. Tan, BS, BA
Thomas M. Habermann, MD
Jeff A. Sloan, PhD
Tait D. Shanafelt, MD

MEDICAL ERRORS AND PATIENT SAFETY continue to be an important concern for patients and physicians, especially since the Institute of Medicine reported in 1999 that between 48,000 and 98,000 Americans die each year due to preventable adverse events. As many as 50% of hospitalized patients may be affected by medical errors, and the human and monetary costs of these events are great. Numerous reports have implicated fatigue and sleepiness as contributors to medical errors. In separate studies, resident distress has also been shown to be an important factor in self-reported major medical errors and medication errors.

This research on fatigue and distress has informed the 2008 Institute of Medicine report on resident duty hours calling for prevention or mitigation of fatigue and promotion of resident well-being. Specific recommendations have primarily focused on reducing fatigue by limiting shift lengths, reducing the frequency of overnight work, and protecting time off duty. The costs to implement these recommendations would be great and their effectiveness is unknown, but there is little doubt that the proposed changes would transform the modern training environment. As such changes are considered, it is important to note that the independent contributions of fatigue and distress to medical errors are unknown because to our knowledge, fatigue and distress have not been assessed together in prior published research. Since fatigue and distress are related but distinct entities, assessment of their joint effects on patient safety outcomes is a critical part of efforts to improve patient care. To address this

Context Fatigue and distress have been separately shown to be associated with medical errors. The contribution of each factor when assessed simultaneously is unknown.

Objective To determine the association of fatigue and distress with self-perceived major medical errors among resident physicians using validated metrics.

Design, Setting, and Participants Prospective longitudinal cohort study of categorical and preliminary internal medicine residents at Mayo Clinic, Rochester, Minnesota. Data were provided by 380 of 430 eligible residents (88.3%). Participants began training from 2003 to 2008 and completed surveys quarterly through February 2009. Surveys included self-assessment of medical errors, linear analog self-assessment of overall quality of life (QOL) and fatigue, the Maslach Burnout Inventory, the PRIME-MD depression screening instrument, and the Epworth Sleepiness Scale.

Main Outcome Measures Frequency of self-perceived, self-defined major medical errors was recorded. Associations of fatigue, QOL, burnout, and symptoms of depression with a subsequently reported major medical error were determined using generalized estimating equations for repeated measures.

Results The mean response rate to individual surveys was 67.5%. Of the 356 participants providing error data (93.7%), 139 (39%) reported making at least 1 major medical error during the study period. In univariate analyses, there was an association of subsequent self-reported error with the Epworth Sleepiness Scale score (odds ratio [OR], 1.10 per unit increase; 95% confidence interval [CI], 1.03-1.16; P=.002) and fatigue score (OR, 1.14 per unit increase; 95% CI, 1.08-1.21; P=.001). Subsequent error was also associated with burnout (ORs per 1-unit change: depersonalization OR, 1.09; 95% CI, 1.05-1.12; P=.001; emotional exhaustion OR, 1.06; 95% CI, 1.04-1.08; P<.001; lower personal accomplishment OR, 0.94; 95% CI, 0.92-0.97; P<.001), a positive depression screen (OR, 2.56; 95% CI, 1.76-3.72; P<.001), and overall QOL (OR, 0.84 per unit increase; 95% CI, 0.79-0.91; P<.001). Fatigue and distress variables remained statistically significant when modeled together with little change in the point estimates of effect. Sleepiness and distress, when modeled together, showed little change in point estimates of effect, but sleepiness no longer had a statistically significant association with errors when adjusted for burnout or depression.

Conclusion Among internal medicine residents, higher levels of fatigue and distress are independently associated with self-perceived medical errors.

Author Affiliations: Division of General Internal Medicine (Dr West); Division of Biomedical Statistics and Informatics, Department of Health Sciences Research (Drs West and Sloan and Ms Tan); Department of Medicine (Drs West, Habermann, and Shanafelt); Division of Hematology (Drs Habermann and Shanafelt), Mayo Clinic, Rochester, Minnesota.

Corresponding Author: Colin P. West, MD, PhD, Division of General Internal Medicine and Biomedical Statistics and Informatics, Department of Medicine and Health Sciences Research, Mayo Clinic, 200 First St, SW, Rochester, MN 55905 (west.colin@mayo.edu).
knowledge gap, the prospective longitudinal Mayo Internal Medicine Well-being Study was extended to assess the independent contributions of fatigue and distress to self-reported medical errors when considered simultaneously, as well as to independently confirm other reports on the association of fatigue and sleepiness with errors using validated metrics and to update the previous report on the associations of quality of life (QOL), burnout, and symptoms of depression with self-perceived errors.10

METHODS
Participants
All categorical and preliminary internal medicine trainees at the Mayo Clinic Rochester Internal Medicine Residency program between July 2003 and February 2009 were eligible to participate. These residents attended 163 different US and international medical schools. Curricular structure and study enrollment procedures have been described previously.10 Current duty hour regulations were in effect for the entirety of this study. Participation in this study was voluntary and written informed consent was obtained from all participants. The Mayo Clinic Institutional Review Board approved this study.

Data Collection
Residents were electronically surveyed every 3 months throughout their training beginning in 2003. Surveys were administered by the Mayo Clinic Survey Research Center via e-mail link to an electronic form, and automated e-mail reminders were sent to enhance response rates. Participants were given approximately 10 days to complete each survey. Surveys were administered quarterly in July-August, October-November, January-February, and April-May, although exact survey timing differed slightly from quarter to quarter.

Surveys included questions about demographic characteristics, current rotation characteristics, coping strategies for dealing with stress, and report of self-perceived medical errors. Validated survey tools were used to measure fatigue, QOL, burnout, and symptoms of depression. Self-reported medical errors, QOL, and linear analog self-assessment of fatigue were assessed quarterly, while burnout and symptoms of depression were evaluated every 6 months. Quarterly recording of the Epworth Sleepiness Scale began in July 2007. Data were analyzed through February 2009. No study responses or identifying information for individual participants were accessible to the Mayo Clinic Department of Medicine.

Study Measures
Self-reported Medical Errors. Perceived medical errors were evaluated by self-report every 3 months by asking residents, “Are you concerned you have made any major medical errors in the last 3 months?” As discussed previously,10 self-reported errors in this study represent major medical errors as perceived by each individual resident.

Fatigue and Sleepiness. Fatigue and sleepiness are overlapping but different concepts.14,15 Fatigue may reflect a broader sense of weariness and depleted energy, while sleepiness refers to drowsiness and decreased alertness. In this study, fatigue was measured beginning in 2003 using a standardized linear analog self-assessment question. Respondents indicated their level of fatigue during the past week according to their own definition of the term on a 0 (“As bad as it can be”) to 10 (“As good as it can be”) scale. Therefore, worsening fatigue is indicated by a decrease in fatigue score. Beginning in July 2007, the Epworth Sleepiness Scale was added to the quarterly surveys. This instrument assesses an individual’s recent level of daytime sleepiness using 8 scenarios scored on a Likert scale from 0 (“No chance of dozing”) to 3 (“High chance of dozing”).16,17 A score of at least 10 is considered indicative of excessive daytime sleepiness.

QOL, Burnout, and Depression. QOL was measured by a single-item linear analog self-assessment. This instrument measured overall QOL on a 0 to 10 scale with the same anchors as the fatigue question. This scale has been validated across a wide range of medical conditions and populations.18-20

Burnout is a syndrome encompassing 3 domains (depersonalization, emotional exhaustion, and a sense of low personal accomplishment) associated with decreased work performance.21 Burnout was measured using the Maslach Burnout Inventory21 in which respondents rate the frequency of experiencing various feelings or emotions on a 7-point Likert scale with response options ranging from never to daily. Higher values of depersonalization and emotional exhaustion and lower values of personal accomplishment indicate burnout. This instrument has been used in previous studies of physicians.22-25

Depression screening used the 2-question approach described by Spitzer et al26 and validated by Whooley et al.27 This instrument has been used in a variety of patient populations26,27 including studies of physicians.10,22 This tool includes questions about depressed mood and anhedonia: (1) “During the past month, have you often been bothered by feeling down, depressed, or hopeless?” and (2) “During the past month, have you often been bothered by little interest or pleasure in doing things?” A positive screen for depression is indicated by a yes response to either question. As discussed previously,10 this screening instrument compares favorably with other depression screening instruments reported in the literature.27,28

Statistical Analyses
Standard univariate statistics were used to characterize the sample. Comparisons between residents reporting errors and residents reporting no errors were initially made using summary statistics, collapsing responses within each individual into a single mean outcome.29 These were analyzed using the
Of 430 eligible residents, there were 380 participants (88%) with no statistically significant differences in age, sex, or program type between participants and nonparticipants. The demographic characteristics of study participants are shown in Table 1. Of the participants, 356 (93.7%) completed at least 1 survey and 122 (32.1%) completed all surveys during the study period with a mean response rate to individual surveys of 67.5% (range, 52.2%-88.2%). In total, 2951 surveys were administered. A mean of 134 residents (range, 68-214) were surveyed in each quarter and 120 729 of 189 489 participants (88%) with no statistically significant differences in age, sex, program type, or parental status.

Univariate associations between fatigue and distress at each time point and a self-perceived error in the subsequent 3 months are shown in Table 3. Increased fatigue and sleepiness were associated with increased odds of reporting an error in the subsequent 3 months. Each 1-point increase in fatigue or Epworth Sleepiness Scale score was associated with a 14% and 10% increase in this odds, respectively. Diminished QOL, higher levels of burnout or symptoms of depression are shown in Table 2. Consistent with our previous report, residents reporting at least 1 error during the study period had significantly lower overall QOL (difference, −0.41; P = .001) and higher levels of burnout as evidenced by increased depersonalization (difference, 3.49; P < .001), increased emotional exhaustion (difference, 5.33; P < .001), and a lower sense of personal accomplishment (difference, −2.25; P = .001). In aggregate, 92 of 134 residents (68.7%) reporting an error screened positive for depression at least once during the study period, compared with 82 of 188 residents (43.6%) reporting no errors (odds ratio [OR], 2.83; P < .001). Residents who reported errors experienced greater fatigue as indicated by lower scores on the fatigue scale (difference, −0.54; P = .006).

Errors were reported in 279 of 1950 resident-quarters (14.3%). Overall, 139 study participants (39%) reported at least 1 major medical error during the study period, and 127 of 301 residents (42%) completing at least 1 year of training reported errors. Self-perceived error rates did not vary significantly by age, sex, program type, amount of student loan debt, relationship status, or parental status.

Summary measures to identify general associations between self-perceived errors and resident fatigue, QOL, burnout, and symptoms of depression are shown in Table 2. Consistent with our previous report, residents reporting at least 1 error during the study period had significantly lower overall QOL (difference, −0.41; P = .002) and higher levels of burnout as evidenced by increased depersonalization (difference, 3.49; P < .001), increased emotional exhaustion (difference, 5.33; P < .001), and a lower sense of personal accomplishment (difference, −2.25; P = .001). In aggregate, 92 of 134 residents (68.7%) reporting an error screened positive for depression at least once during the study period, compared with 82 of 188 residents (43.6%) reporting no errors (odds ratio [OR], 2.83; P < .001). Residents who reported errors experienced greater fatigue as indicated by lower scores on the fatigue scale (difference, −0.54; P = .006).

Univariate associations between fatigue and distress at each time point and a self-perceived error in the subsequent 3 months are shown in Table 3. Increased fatigue and sleepiness were associated with increased odds of reporting an error in the subsequent 3 months. Each 1-point increase in fatigue or Epworth Sleepiness Scale score was associated with a 14% and 10% increase in this odds, respectively. Diminished QOL, higher levels of burnout in all domains, and positive screening for depression were also each associated with increased odds of reporting an error in the subsequent 3 months. Each 1-point

Wilcoxon-Mann-Whitney test for continuous variables and the Fisher exact test for proportions.

To accommodate the repeated measures study design, the association of self-perceived errors with QOL, burnout, depression, and fatigue was evaluated using generalized estimating equations—an extension of generalized linear models that accounts for correlated repeated measurements within individuals.29,30 Analyses were performed examining the association of distress and fatigue with the likelihood of a self-perceived error during the following 3 months as reported at the subsequent survey time point. Thus, the assessment of all distress variables preceded the self-reported errors.

Multicollinearity among distress variables required that each model include self-reported errors and no more than 2 distress variables. Because the multivariable models did not yield significantly different estimates of effect for any variable, we reported the simpler models with each containing only a single distress variable. Each model accounting for the effect of fatigue or sleepiness also included only 1 distress variable and either fatigue or sleepiness for the same reason. To properly calculate variance terms for repeated measures analyses, the generalized estimating equations method requires that a correlation structure be specified. Selecting the correct correlation structure for generalized estimating equations analyses does not in general affect parameter estimation, but does allow more precise estimates. An exchangeable correlation structure was specified for these analyses, and correlations between variables across time were evaluated.

With a conservative assumption of a 75% response rate, this repeated-measures study had 80% power to detect a small Cohen effect size of 0.15 for variables collected beginning in 2003 and an effect size of 0.22 for variables collected beginning in 2007. Statistical analyses were conducted using SAS version 9.1 (SAS Institute Inc, Cary, North Carolina). Statistical significance was set at the .05 level, and all tests were 2-tailed.

RESULTS

Of 430 eligible residents, there were 380 participants (88%) with no statistically significant differences in age, sex, or program type between participants and nonparticipants. The demographic characteristics of study participants are shown in Table 1. Of the participants, 356 (93.7%) completed at least 1 survey and 122 (32.1%) completed all surveys during the study period with a mean response rate to individual surveys of 67.5% (range, 52.2%-88.2%). In total, 2951 surveys were administered. A mean of 134 residents (range, 68-214) were surveyed in each quarter and 120 729 of 189 489 participants (88%) with no statistically significant differences in age, sex, program type, or parental status.

Univariate associations between fatigue and distress at each time point and a self-perceived error in the subsequent 3 months are shown in Table 3. Increased fatigue and sleepiness were associated with increased odds of reporting an error in the subsequent 3 months. Each 1-point increase in fatigue or Epworth Sleepiness Scale score was associated with a 14% and 10% increase in this odds, respectively. Diminished QOL, higher levels of burnout in all domains, and positive screening for depression were also each associated with increased odds of reporting an error in the subsequent 3 months. Each 1-point
increase in overall QOL and personal accomplishment was associated with a 16% and 6% decrease in this odds, respectively. Each 1-point increase in depersonalization or emotional exhaustion was associated with a 9% and 6% increase in this odds, respectively. A positive depression screen was associated with a 2.56-fold increased odds of a self-reported error in the following 3 months.

Analyses modeling distress variables together with the fatigue score showed persistent statistical significance of all variables and little change in point estimates of effect, with only 1 exception (Table 4). In the model incorporating both emotional exhaustion and fatigue, fatigue no longer had a statistically significant association with subsequent errors. Analyses modeling distress variables together with the Epworth Sleepiness Scale showed similarly modest changes in point estimates of effect (Table 4). However, sleepiness was not significantly associated with errors when adjusted for burnout or depression. The correlation between fatigue and sleepiness was modest at 0.32, and the correlations of emotional exhaustion with fatigue and sleepiness were moderate at 0.47 and 0.42, respectively.

The results were not significantly altered by the addition to these models of other potential confounding or interfering factors (categorical or pre-
Quality of life

Depression

Burnout

Table 4. Adjusted Association of Fatigue, Sleepiness, QOL, Burnout, and Depression Symptoms With a Self-perceived Major Medical Error in the Following 3 Months.

<table>
<thead>
<tr>
<th>Variable by Metric</th>
<th>No.</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASA overall QOL</td>
<td>356</td>
<td>0.89 (0.82-0.96)</td>
<td>.003</td>
</tr>
<tr>
<td>LASA fatigue</td>
<td>356</td>
<td>1.09 (1.02-1.15)</td>
<td>.009</td>
</tr>
<tr>
<td>Burnout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depersonalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-DP</td>
<td>321</td>
<td>1.08 (1.04-1.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LASA fatigue</td>
<td>321</td>
<td>1.10 (1.01-1.19)</td>
<td>.02</td>
</tr>
<tr>
<td>Emotional exhaustion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-EE</td>
<td>321</td>
<td>1.05 (1.03-1.07)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LASA fatigue</td>
<td>321</td>
<td>1.04 (0.95-1.14)</td>
<td>.35</td>
</tr>
<tr>
<td>Personal achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-PA</td>
<td>321</td>
<td>0.95 (0.92-0.97)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LASA fatigue</td>
<td>321</td>
<td>1.14 (1.04-1.23)</td>
<td>.003</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any positive 2-item depression screen</td>
<td>322</td>
<td>2.22 (1.50-3.28)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LASA fatigue</td>
<td>322</td>
<td>1.10 (1.01-1.19)</td>
<td>.03</td>
</tr>
<tr>
<td>Sleepiness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASA overall QOL</td>
<td>171</td>
<td>0.86 (0.75-0.99)</td>
<td>.03</td>
</tr>
<tr>
<td>ESS</td>
<td>171</td>
<td>1.07 (1.01-1.14)</td>
<td>.02</td>
</tr>
<tr>
<td>Burnout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depersonalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-DP</td>
<td>146</td>
<td>1.05 (0.98-1.12)</td>
<td>.15</td>
</tr>
<tr>
<td>ESS</td>
<td>146</td>
<td>1.08 (0.98-1.18)</td>
<td>.10</td>
</tr>
<tr>
<td>Emotional exhaustion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-EE</td>
<td>146</td>
<td>1.07 (1.03-1.10)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ESS</td>
<td>146</td>
<td>1.03 (0.95-1.12)</td>
<td>.51</td>
</tr>
<tr>
<td>Personal achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBI-PA</td>
<td>146</td>
<td>0.94 (0.89-0.99)</td>
<td>.02</td>
</tr>
<tr>
<td>ESS</td>
<td>146</td>
<td>1.08 (0.99-1.17)</td>
<td>.07</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any positive 2-item depression screen</td>
<td>146</td>
<td>2.13 (1.11-4.10)</td>
<td>.02</td>
</tr>
<tr>
<td>ESS</td>
<td>146</td>
<td>1.08 (0.99-1.17)</td>
<td>.08</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; ESS, Epworth Sleepiness Scale; LASA, linear analog self-assessment; MBI-DP, Maslach Burnout Inventory-depersonalization; MBI-EE, Maslach Burnout Inventory-emotional exhaustion; MBI-PA, Maslach Burnout Inventory-personal accomplishment; OR, odds ratio; QOL, quality of life.

1Results are grouped to reflect separate adjusted models containing 1 distress variable and either the LASA fatigue score or the Epworth Sleepiness Scale score.
2Scales: LASA, 0 to 10; MBI-DP, 0 to 30; MBI-EE, 0 to 54; MBI-PA, 0 to 48; and ESS, 0 to 24.
3Odds ratio of a self-reported error in the following 3 months associated with a 1-unit increase in the score for each metric (except for decrease in LASA fatigue, to reflect worsened fatigue).
4Using generalized estimating equation models adjusted for time.
5See footnote to Table 2 for scale classifications.
clarify the unique roles fatigue and sleepiness may play in patient safety.

This study has several limitations. First, the extent to which the self-perceived errors reported in this study accurately reflect the frequency of medical errors and whether these perceived errors actually affected patient outcomes cannot be determined. No single method of measuring errors is ideal in all settings, but previous work has suggested that self-reported adverse events may be more likely to represent preventable medical errors. Self-reported errors have also been shown to have good overall correlation with events documented in the medical record.

Second, the generalizability of these results from a single academic medical center to other training programs is unknown. However, the participation and survey response rates were high relative to other physician surveys, and the error rates, burnout scores, rates of a positive depression screen, and self-reported errors have also been shown to have good overall correlation with events documented in the medical record.

Third, it is possible that feelings of distress or fatigue or experience with prior perceived errors might affect retrospective error reporting, although it is unclear whether such feelings would make reporting of errors more likely or less likely. This is an area worthy of additional study.

Fourth, the Epworth Sleepiness Scale measures daytime sleepiness, and the effect of daytime sleepiness on errors occurring at night is unclear. Additionally, it is unclear how acute fatigue may differ from chronic fatigue in its effect on error occurrence, and this study does not allow direct assessment of specific factors such as extended work shifts.

Fifth, the depression screening instrument cannot be used alone to diagnose depression. Although the positive likelihood ratio for this instrument is similar to that of other accepted depression screening tools, additional clinical evaluation would be necessary to diagnose depression in participants with positive screening scores.

Sixth, due to multicollinearity there was limited ability to separate the effects of all individual well-being variables from one another, or to separate the effects of fatigue and sleepiness. Larger sample sizes may be necessary to better resolve the associations among these variables. Because of these limitations, these results should be interpreted as associations rather than as definitive evidence of causation.

In summary, this study suggests that fatigue, sleepiness, burnout, depression, and reduced QOL are independently associated with an increased risk of future self-perceived major medical errors. In addition to the national efforts to reduce fatigue and sleepiness, well-designed interventions to prevent, identify, and treat distress among physicians are needed. Additional research is necessary to determine the most effective strategies for accomplishing these goals. Changes to the process of physician training should address both resident fatigue and distress in an effort to improve resident and patient safety.

Author Contributions: Dr West had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: West, Habermann, Sloan, Shanafelt. Acquisition of data: West, Tan, Shanafelt. Analysis and interpretation of data: West, Sloan, Shanafelt. Drafting of the manuscript: West, Shanafelt. Critical revision of the manuscript for important intellectual content: West, Tan, Habermann, Sloan, Shanafelt. Statistical analysis: West. Obtained funding: West, Habermann, Shanafelt. Administrative, technical, or material support: Tan, Habermann, Sloan, Shanafelt. Financial Disclosures: None reported. Funding/Support: This work was supported by the Mayo Clinic Department of Medicine Program on Physician Well-being.

Role of the Sponsor: The Mayo Clinic Department of Medicine Program on Physician Well-being played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation of the manuscript.

REFERENCES

33. Philibert I, Friedmann P, Williams WT; ACGME Work Group on Resident Duty Hours; Accreditation Council for Graduate Medical Education. New requirements for resident duty hours. *JAMA.* 2002;288(9):1112-1114.

The redwoods, once seen, leave a mark or create a vision that stays with you always. No one has ever successfully painted or photographed a redwood tree. The feeling they produce is not transferable. From them comes silence and awe. It’s not only their unbelievable stature, nor the color which seems to shift and vary under your eyes, no, they are not like any trees we know, they are ambassadors from another time.

—John Steinbeck (1902-1968)