Evaluating the Association of Multiple Burnout Definitions and Thresholds With Prevalence and Outcomes

D. Brock Hewitt, MD, MPH, MS; Ryan J. Ellis, MD, MS; Yue-Yung Hu, MD, MPH; Elaine O. Cheung, PhD; Judith T Moskowitz, PhD, MPH; Gaurava Agarwal, MD; Karl Y. Bilimoria, MD, MS

IMPORTANCE Physician burnout is a serious issue, given its associations with physician attrition, mental and physical health, and self-reported medical errors. Burnout is typically measured in health care by assessing the frequency of symptoms in 2 domains, emotional exhaustion and depersonalization. However, the lack of a clinically diagnostic threshold to define burnout has led to considerable variability in reported burnout rates.

OBJECTIVE To estimate the prevalence of burnout using a range of definitions (ie, requiring symptoms in both domains or just 1) and thresholds (ie, requiring symptoms to occur weekly vs a few times per year) and examine the strength of the association of various definitions of burnout with suicidal thoughts and thoughts of attrition among general surgery residents.

DESIGN, SETTING, AND PARTICIPANTS A cross-sectional national survey of clinically active US general surgery residents administered in conjunction with the 2019 American Board of Surgery In-Training Examination assessed burnout symptoms, thoughts of attrition, and suicidal thoughts during the past year. Multivariable logistic regression models were used to assess the association of burnout symptoms with thoughts of attrition and suicidal thoughts. Values of $R^2$ and C statistic were used to evaluate multivariable model performance.

EXPOSURES Burnout was evaluated with a 6-item, modified, abbreviated Maslach Burnout Inventory for 2 burnout domains: emotional exhaustion and depersonalization.

MAIN OUTCOMES AND MEASURES The primary outcome was prevalence of burnout. Secondary outcomes were thoughts of attrition and suicidal thoughts within the past year.

RESULTS Among 6956 residents (a 85.6% response rate; including 3968 men [57.0%] and 4041 non-Hispanic White individuals [58.1%]) from 301 surgical residency programs, 2329 (38.6%) reported at least weekly symptoms of emotional exhaustion, and 1389 (23.1%) reported at least weekly depersonalization symptoms. Using the most common definition, 2607 general surgery residents (43.2%) reported weekly burnout symptoms on either subscale. Subtle changes in the definition of burnout selected resulted in prevalence estimates varying widely from 3.2% (159 residents; most stringent: daily symptoms on both subscales) to 91.4% (5521 residents; least stringent: symptoms a few times per year on either subscale). In multivariable models, all measures of higher burnout symptoms were associated with increased thoughts of attrition (depersonalization: $R^2$, 0.097; C statistic, 0.717; emotional exhaustion: $R^2$, 0.137; C statistic, 0.758; both: $R^2$, 0.138; C statistic, 0.761) and suicidal thoughts (depersonalization: $R^2$, 0.102; C statistic, 0.750; both: $R^2$, 0.106; C statistic, 0.751) among general surgery residents (all $P < .001$).

CONCLUSIONS AND RELEVANCE In a national evaluation of general surgery residents, prevalence estimates of burnout varied considerably, depending on the burnout definition selected. Frequent burnout symptoms were strongly associated with both thoughts of attrition and suicide, regardless of the threshold selected. Future research on burnout should explicitly include a clear description and rationale for the burnout definition used.
Burnout is a multifaceted condition of overwhelming exhaustion, interpersonal detachment or cynicism toward one’s job, and a sense of reduced professional efficacy, driven by long-term workplace stress. Burnout has garnered attention in the medical community because of its reported association with physician attrition, mental and physical health, and self-reported medical errors. Each year, physician burnout–driven attrition and reduced clinical hours cost the US health care system approximately $4.6 billion. In a national study, more than half of practicing physicians reported at least 1 burnout symptom, almost twice the rate of the general population. However, exact prevalence estimates of burnout are unclear.

Burnout lacks validated clinical cutoffs, and assessment methods and measurement thresholds vary widely across studies. Three commonly used tools for burnout measurement provided by The National Academy of Medicine Action Collaborative on Clinician Well-Being and Resilience (the Maslach Burnout Inventory [MBI], the Oldenburg Burnout Inventory, and the Physician Work-Life Study’s single item) each evaluate burnout with different scales and dimensions. Regarding burnout definitions, a recent systematic review found 142 unique definitions for meeting burnout criteria (ie, being burned out), and overall prevalence estimates ranged from 0% to 80.5%, depending on the specialty of medicine examined and the criteria chosen. For general surgery residents alone, estimates of burnout prevalence range from 43% to 69%.

With the most common tool, the MBI, the frequency is measured on a 7-point Likert scale as symptoms experienced daily, a few times a week, once a week, a few times a month, once a month, a few times per year, or never. Prior studies have set their burnout threshold at different frequencies and required symptoms in 1 or multiple domains. The wide variation could be attributable to the cohorts examined, differing response rates, and/or these different definitions and thresholds of burnout used.

Burnout is typically measured in health care by assessing the frequency of symptoms in 2 domains, emotional exhaustion and depersonalization. To better understand the variation in burnout prevalence estimates among surgical residents and test varying definitions and thresholds of burnout, a comprehensive national survey of all US general surgery residents was conducted using an abbreviated MBI. The objectives of this study were to (1) estimate the prevalence of burnout using a range of definitions and thresholds and (2) examine the strength of the association between various definitions and thresholds of burnout with 2 important outcomes, thoughts of attrition and suicidal thoughts, among clinically active US general surgery residents.

Methods

Study Design and Participants

In collaboration with the American Board of Surgery, a voluntary, multiple-choice survey was administered to all examinees from Accreditation Council for Graduate Medical Education–accredited general surgery training programs immediately following the January 2019 American Board of Surgery In-Training Examination (ABSITE). The ABSITE is a computer-based, multiple-choice examination administered annually to US general surgery residents to measure knowledge and management of surgical pathology. The survey was preceded by a statement explaining that the purpose of the survey was research, the data would be deidentified prior to analysis, and program directors or chairs would never have access to individual responses. There were no incentives to participate in the survey. Responses were collected by the American Board of Surgery and deidentified before being transferred to Northwestern University for analysis. Excluded from analysis were 778 residents who were clinically inactive (ie, taking dedicated time to conduct research). After review of this study, including the survey content, the Northwestern University institutional review board office determined that it did not meet the definition of human subjects research. Survey completion constituted participant consent.

Survey Instrument

The 2019 survey items were adapted from previously published and validated tools. Pretest cognitive interviews were conducted with general surgery residents to assess overall survey coherence, balance, and clarity. The survey was then iteratively revised and retested with a larger sample of general surgery residents from multiple institutions.

A modified, abbreviated MBI–Human Services Survey for Medical Personnel was used to assess burnout symptoms. The 6-item instrument assessed 2 burnout domains: emotional exhaustion (3 items) and depersonalization (3 items) on a 7-point Likert scale (categorized as never, a few times a year or less, once a month, a few times a month, once a week, a few times a week, or every day). Individual subscale and overall burnout frequencies were reported using multiple criteria. Defining burnout includes 2 main points of clarification: the number of burnout domains involved and the symptom frequency dichotomization threshold (eg, every day, once a week, once a month). The full MBI includes 3 domains: emotional exhaustion, depersonalization, and personal accomplishment. In this study, we examined depersonalization and emotional exhaustion, because previous studies have found that these 2 domains most consistently describe a clinical burnout syndrome. Two categorizations of burnout domains were used. First, to meet the criteria for burnout, the resident had to report symptoms of only 1 domain,
either emotional exhaustion or depersonalization. The second, more stringent definition required a resident to report symptoms of both domains. Within these categories, multiple subscale thresholds were examined.

The most stringent subscale dichotomization burnout threshold was reporting symptoms daily, a threshold that included residents reporting burnout symptoms every day vs residents reporting symptoms a few times a week, once a week, a few times a month, once a month, a few times a year or less, and never. The most sensitive subscale threshold was reporting symptoms at least yearly, a dichotomization threshold including residents reporting symptoms every day, a few times a week, once a week, a few times a month, once a month, or a few times a year or less vs reporting symptoms as never occurring. Finally, each domain was also evaluated as a continuous variable, separately and as a combined overall burnout variable (a continuous variable of emotional exhaustion plus depersonalization).

Residents were also asked if they had considered leaving their program or taking their own life within the last year. Suicidal thoughts were assessed with the question, “During the past 12 months, have you had thoughts of taking your own life?” This question was immediately followed by a webpage providing the National Suicide Prevention Lifeline and urging respondents to reach out to their program directors if they have had such thoughts. No active outreach was possible because all data were deidentified, and confidentiality was assured as a precondition of survey completion.

**Statistical Analysis**

Raw and cumulative emotional exhaustion and depersonalization symptom frequencies were calculated. Separate multivariable logistic regression models were constructed to examine the association between varying definitions or thresholds of burnout and each resident outcome of interest: thoughts of attrition and suicidal ideation. Models were adjusted for all available resident demographics (eg, sex, race/ethnicity, and marital status) and program characteristics (eg, geographic location, type [academic or community or military], and size). Values of $R^2$ and C statistics were used to evaluate multivariable model performance. All models were estimated with robust SEs accounting for residents clustering within programs. Multivariable analyses were limited to individuals with complete survey responses. All tests were 2-sided, with significance set at .05. Statistical analyses were performed with Stata 14.1 (StataCorp).

**Results**

Of the 8129 eligible surgical residents taking the 2019 ABSTRE, 6956 had at least partial survey responses (response rate, 85.6%). Most residents were male (3968 [57.0%]), non-Hispanic White (4041 [58.1%]), married or in a relationship (5111 [73.5%]), and training in an academic program (4014 [57.7%]) (Table 1). Additional participant demographics and program characteristics are listed in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Characteristics of Residents From 301 Surgical Residency Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td>Total No. of residents</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>Non-Hispanic</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Other/unknown</td>
</tr>
<tr>
<td>Relationship status</td>
</tr>
<tr>
<td>Married or in a relationship</td>
</tr>
<tr>
<td>No relationship</td>
</tr>
<tr>
<td>Divorced or widowed</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Program size (total No. of residents), quartile</td>
</tr>
<tr>
<td>1 (&lt;24)</td>
</tr>
<tr>
<td>2 (24–34)</td>
</tr>
<tr>
<td>3 (35–49)</td>
</tr>
<tr>
<td>4 (≥50)</td>
</tr>
<tr>
<td>Program type</td>
</tr>
<tr>
<td>Academic</td>
</tr>
<tr>
<td>Community or military</td>
</tr>
<tr>
<td>Program location</td>
</tr>
<tr>
<td>Northeast</td>
</tr>
<tr>
<td>Southeast</td>
</tr>
<tr>
<td>Midwest</td>
</tr>
<tr>
<td>Southwest</td>
</tr>
<tr>
<td>West</td>
</tr>
</tbody>
</table>

Using the most common definition of burnout in the literature, 2607 general surgery residents (43.2%) reported weekly burnout symptoms on either subscale. Emotional exhaustion symptoms were reported by 510 residents (8.5%) as daily, 2329 (38.6%) as at least weekly (a cumulative grouping of symptoms daily, few times a week, and once a week), and 4379 (72.6%) as at least monthly (a cumulative grouping of daily symptoms to symptoms once a month); 579 residents (9.5%) reported never experiencing symptoms of emotional exhaustion (Table 2). Depersonalization symptoms were reported by 312 residents (5.2%) as daily, 1389 (23.1%) as at least weekly, and 3145 (52.3%) as at least monthly; 1474 (24.4%) reported never experiencing symptoms of depersonalization. Emotional exhaustion symptoms were reported in a more normal distribution, with a few times a month the most common frequency reported. Depersonalization symptoms were reported less often; almost half of the residents surveyed...
reported experiencing symptoms of personalization never (1410 [23.3%]) or a few times a year or less (1474 [24.4%]).

The first categorization used to evaluate burnout was reporting symptoms on only 1 subscale—either emotional exhaustion or personalization. Using the most sensitive threshold of at least yearly burnout symptoms, 5521 residents (91.4%) reported symptoms of either emotional exhaustion or personalization and met criteria for burnout (Table 2). The most stringent dichotomization threshold requiring daily burnout symptoms on either domain resulted in a burnout prevalence rate of 11.0% (in 663 residents). Changing the threshold resulted in an absolute difference of 80.4% in the burnout prevalence rate of 11.0% (in 663 residents). Changing the threshold resulted in an absolute difference of 80.4% in the burnout prevalence rate of 3.2% (in 159 residents), an absolute difference of 86.5%.

In multivariable models adjusting for resident demographics and program characteristics, all definitions of burnout, including analysis of burnout on a continuous scale, were significantly associated with both thoughts of attrition and suicidal thoughts. Using R² and C statistic to evaluate model fit, the continuous overall burnout score (ie, summation of both emotional exhaustion and personalization scores) performed better to anticipate resident thoughts of attrition (R², 0.138; C statistic, 0.761) and suicidal thoughts (R², 0.106; C statistic, 0.751), as did the individual continuous subscales of emotional exhaustion (thoughts of attrition: R², 0.137; C statistic, 0.758; suicidal thoughts: R², 0.102; C statistic, 0.750) and depersonalization (thoughts of attrition: R², 0.097; C statistic, 0.717; suicidal thoughts, R², 0.077; C statistic, 0.718), compared with any of the dichotomized burnout definitions or thresholds (thoughts of attrition: R² range, 0.082-0.095; C statistic range, 0.707-0.720; suicidal thoughts: R² range, 0.051-0.060; C statistic range, 0.686-0.700) (P < .001 for all comparisons; Table 3).

The 2 burnout subscales, emotional exhaustion and personalization, were associated with thoughts of attrition and suicidal thoughts; more frequent symptoms were associated with higher rates of thoughts of attrition and thoughts of suicide, indifferent to the threshold assessed (Figure). However, a clear inflection point was not observed for overall burnout, emotional exhaustion, or depersonalization scores for either wellness outcome.

Discussion

This comprehensive national study of US general surgery residents demonstrated considerable variability in burnout prevalence among general surgery residents, depending on the burnout definition or threshold selected. Available burnout assessment instruments lack a clinically validated dichotomization...
Burnout lacks a consensus definition and clear measurement standard both in medicine and other disciplines; however, the 3-domain definition developed by Maslach et al.² and the MBI are the most common definition and measurement standard. The World Health Organization classifies burn-out in the International Classification of Diseases, 11th Revision as a multidimensional occupational phenomenon, not a medical condition.²³ In addition, burnout lacks a Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) definition. This diagnostic ambiguity has led to numerous burnout assessment tools and published thresholds without a single gold standard. Accepted burnout assessment instruments range from simple single-item evaluations in which individuals use their own definition of burnout (eg, the American Medical Association Mini Z) to more complex multidimensional inventories, such as the Copenhagen and Oldenburg Burnout Inventories and the MBI. Furthermore, the intent of these instruments is to evaluate burnout academically (ie, for statistical associations) and not clinically (ie, for a diagnosis); by design, the tools evaluate burnout on a continuous scale and do not recommend or provide a clinically validated threshold or cutoff signifying a burnout diagnosis.

Using several published burnout categorizations and thresholds on the same population, we found that burnout prevalence estimates ranged widely from 3.2% to 91.4%, with the most common definition estimating that 43.2% of general surgery residents are experiencing weekly burnout symptoms. Accepting that burnout is a multidimensional construct and distinct from mental illnesses, such as depression, there are 2 important considerations with defining burnout: frequency of symptoms and the association between the burnout subscales, emotional exhaustion, and depersonalization in this study. Because there is no accepted frequency that is considered burnout, researchers must decide where to draw the line: daily, weekly, monthly, or yearly symptoms. Clearly, if the occurrence of even 1 symptom in a year is considered burnout, the prevalence is much higher than 1 in which burnout is considered to require daily symptoms. In addition, researchers have disagreed about whether burnout requires the occurrence of symptoms from more than 1 burnout subscale (eg, both emotional exhaustion and depersonalization) or having symptoms of 1 subscale (eg, emotional exhaustion or depersonalization) is substantial enough to deem someone to be experiencing burnout. Thus, these differences in how to define burnout lead to large differences in prevalence estimates that make comparisons difficult between studies. This work highlights the need for burnout research to specify and justify the definition of burnout used.

Previous studies have demonstrated an association between frequent burnout symptoms and poor wellness outcomes.³⁻⁵
We found that increasing frequency of burnout symptoms were significantly associated with both thoughts of attrition and suicidal thoughts among general surgery residents. Importantly, this association was observed regardless of the burnout definition selected, and the model diagnostics were similar among the various definitions or thresholds tested. Thus, the specific definition or threshold may not be particularly important, but the definition must be specified. In general, individuals experience burnout symptoms differently,24 such that weekly symptoms for 1 individual may be significant enough to trigger suicidal thoughts, whereas another individual may experience burnout symptoms daily and not have suicidal thoughts. More important than the actual prevalence estimates are the significant associations between frequent burnout symptoms and poor wellness outcomes. Future studies examining interventions should focus on reducing the frequency of burnout symptoms instead of the prevalence of burnout.

Limitations

This study has several potential limitations. First, as a cross-sectional study, we could only explore associations and could not identify causes. Second, the timing of the survey immediately following the ABSITE may have affected resident responses, but the direction of the bias is uncertain because both examination-associated distress and postexamination relief may be occurring simultaneously. Third, because this was a cross-sectional study and survey responses were completely anonymous, we were unable to follow up or connect actual attrition and suicide rates among residents in this study. Finally, since survey questions involved exposures over the past year, recall bias may exist. Nonetheless, this study offers the opportunity to examine burnout measurement in a national evaluation of surgical residents with a response rate that is substantially higher than that of prior studies.

Conclusions

In a national evaluation of general surgery residents, burnout prevalence varied considerably depending on the definition selected. These results emphasize the need for clear reporting of the criteria used for burnout assessment and a justification for the rationale for using that approach.

ARTICLE INFORMATION

Accepted for Publication: May 21, 2020.


Author Contributions: Drs Bilimoria and Hewitt had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Hewitt, Ellis, Cheung, Agarwal, Bilimoria.

Drafting of the manuscript: Hewitt, Ellis, Bilimoria.

Acquisition, analysis, or interpretation of data: Hewitt, Ellis, Hu, Moskowitz, Agarwal, Bilimoria.

Statistical analysis: Hewitt, Ellis, Bilimoria.

Obtained funding: Bilimoria.

Administrative, technical, or material support: Ellis, Hu, Supervision: Hu, Agarwal, Bilimoria.

Conflict of Interest Disclosures: Dr Bilimoria reported grants from American College of Surgeons, Association for Graduate Medical Education, and American Board of Surgery during the conduct of the study. Dr Agarwal reported having received honoraria for lectures on burnout and well-being. No other disclosures were reported.

Funding/Support: This study is supported by funding from the American College of Surgeons, Accreditation Council for Graduate Medical Education, and the American Board of Surgery. Drs Hewitt and Ellis were supported by postdoctoral research fellowships from the Agency for Healthcare Research and Quality (grant 5T32HS000708). Dr Cheung was supported by a postdoctoral research fellowship from the National Science Foundation (grant 1714952).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES


In their study in this issue of JAMA Surgery, Hewitt et al have compared definitions and thresholds of burnout in general surgery residents, providing important insights into strategies to measure physician burnout. The study offers multiple strengths: a large and representative sample (N = 6956), an excellent survey response rate (85.6%), the use of a validated gold standard measure of burnout, and exploration of prevalence based on inconsistencies in the literature. The authors have asserted that the Maslach Burnout Inventory is scored and interpreted inconsistently in the literature regarding the time frame and number of symptoms needed for a respondent to be classified as having burnout. In fact, much of the recent prevalence research has required an elevation (using defined cutoffs) in only 1 of 3 dimensions (emotional exhaustion, depersonalization, or a decreased sense of personal accomplishment) for an individual to be considered burned out.\(^1\)\(^2\)\(^3\) Such an approach may not capture burnout as a complex and dynamic construct.

We propose shifting the focus from burnout measurement issues to a broader perspective that includes known correlates of burnout, many of which already have scores to indicate meaningful levels. Such constructs include depression, fatigue, perceived stress, mindfulness, self-compassion, meaning and engagement in work, quality of life, and job satisfaction.\(^4\)\(^5\) In fact, a brief validated screening index is available that assesses these multiple dimensions and offers threshold scores to indicate physicians in distress.\(^6\) Two single-item burnout measures are also available that correlate highly with the full scale and other meaningful outcomes (eg, suicidal ideation, medical errors).\(^7\)

Through a cultural shift, we can each play an active role in building an environment of collegiality and connectedness to counter burnout. For example, taking the time to ask simple questions can create important connections and raise awareness of struggles our peers may silently experience. These may include, “How are you? What are you doing to take care of yourself? What does burnout mean to you? What matters most to you? How would you describe your ability to work, enjoy life, and prioritize your relationships?” In addition, drivers of burnout vs engagement (workload, efficiency of work processes, flexibility or control, work-home integration, alignment of individual and organizational values, social support or sense of community, meaning in work, work ambivalence, and leadership culture) point to ways to prevent and address burnout.\(^4\)\(^5\)

A recent article offered a resilience challenge to physicians (the so-called 10% factor): “Even if you believe that ‘they’ (organizations, regulators, resistant patients, leaders, loved ones) are 90% of the problem, what 10% are you willing to own in order to foster your personal resilience?\(^8\)\(^9\)\(^10\)\(^11\) Together, organizational efforts, individually focused interventions or attitude changes (eg, mindfulness training, small-group programs to promote community and connectedness, personal boundaries, time for leisure activities, focus on relationships), and leadership behaviors (eg, seeking input on work-associated improvements, communicating transparently, offering recognition for work well done) can all effectively decrease physician burnout and maximize performance and engagement.\(^2\)^\(^3\)^\(^4\)^\(^5\)


**REFERENCES**