

What They Are Not Telling Us

Analysis of Nonresponders on a National Survey of Resident Well-Being

Chelsea F. Cardell, MD, MS,*†‡ Tarik K. Yuce, MD, MS,* Tiannan Zhan, MS,* Josh S. Eng, PhD,*§ Elaine O. Cheung, PhD,* Caryn D. Etkin, PhD, MPH,* Daniela Amortegui, MA,*§ Andrew Jones, PhD,|| Jo Buyske, MD,|| Karl Y. Bilimoria, MD, MS,§ and Yue-Yung Hu, MD MPH*¶

Objectives: To characterize nonrespondents to a national survey about trainee well-being, examine response patterns to questions of sensitive nature, and assess how nonresponse biases prevalence estimates of mistreatment and well-being.

Background: Surgical trainees are at risk for burnout and mistreatment, which are discernible only by self-report. Therefore, prevalence estimates may be biased by nonresponse.

Methods: A survey was administered with the 2018 and 2019 American Board of Surgery In-Training Examinations assessing demographics, dissatisfaction with education and career, mistreatment, burnout, thoughts of attrition, and suicidality. Responders in 2019 were characterized as survey “Completers,” “Discontinuers” (quit before the end), and “Selective Responders” (selectively answered questions throughout). Multivariable logistic regression assessed associations of respondent type with mistreatment and well-being outcomes, adjusting for individual and program characteristics. Longitudinal survey identifiers linked survey responses for eligible trainees between 2018 and 2019 surveys to further inform nonresponse patterns.

Results: In 2019, 6956 (85.6%) of 8129 eligible trainees initiated the survey, with 66.5% Completers, 17.5% Discontinuers, and 16.0% Selective Responders. Items with the highest response rates included dissatisfaction with education and career (93.2%), burnout (86.3%), thoughts of attrition (90.8%), and suicidality (94.4%). Discontinuers and Selective Responders were more often junior residents and racially/ethnically minoritized than Completers. No differences were seen in burnout and suicidality rates between Discontinuers, Selective Responders, and Completers. Non-White or Hispanic residents were more likely to skip questions about racial/ethnic discrimination than non-Hispanic White residents (21.2% vs 15.8%; odds ratio [OR], 1.35; 95% confidence interval [CI], 1.19–1.53), particularly when asked to identify the source. Women were not more likely to omit questions regarding gender/gender identity/sexual orientation discrimination (OR, 0.91; 95% CI, 0.79–1.04) or its sources (OR, 1.02; 95% CI, 0.89–1.16). Both Discontinuers and Selective Responders more frequently reported physical abuse (2.5% vs 1.1%; $P = 0.001$) and racial discrimination (18.3% vs 13.6%; $P < 0.001$) on the previous survey (2018) than Completers.

Conclusions: Overall response rates are high for this survey. Prevalence estimates of burnout, suicidality, and gender discrimination are likely minimally impacted by nonresponse. Nonresponse to survey items about racial/ethnic discrimination by racially/ethnically minoritized residents likely results in underestimation of this type of mistreatment.

Keywords: burnout, diversity, equity, and inclusion, mistreatment, nonresponse, surgical education, well-being

INTRODUCTION

Surgeons are at high risk for burnout, particularly during their training years.^{1–3} Physician burnout is associated with medical errors, reduced job effort, and career attrition, impacting individual physicians, patients, and the health care system.^{4,5} In surgical residency, a substantial portion of burnout is attributable to mistreatment (eg, discrimination, harassment, abuse).² While mistreatment has been observed across all surgical

trainees, certain groups are at higher risk: women and racially/ethnically minoritized residents more frequently experience mistreatment.^{2,3,6}

We have previously estimated prevalence rates for burnout (38.5%) and mistreatment (gender discrimination 79.8%, racial/ethnic discrimination 23.7%, bullying 66.9%) in general surgery residents.^{2,3,7,8} These data derive from an annual national cross-sectional survey of all surgical residents training in Accreditation Council for Graduate Medical

From the *Northwestern Quality Improvement, Research, & Education in Surgery (NQUIRES), Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL; †Department of Surgery, Loyola University Medical Center, Maywood, IL; ‡Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL; §Department of Surgery, Indiana University, Indianapolis, IN; ||The American Board of Surgery, Philadelphia, PA; and ¶Division of Pediatric Surgery, Ann & Robert H. Lurie Children’s Hospital of Chicago, Feinberg School of Medicine, Northwestern University, Chicago, IL.

Disclosure: The authors declare that they have nothing to disclose. The Surgical Education Culture Optimization through targeted interventions based on National comparative Data Trial is supported by the American College of Surgeons, the Accreditation Council for Graduate Medical Education, and the American Board of Surgery. T.K.Y. was supported by a postdoctoral research fellowship ([Agency for Healthcare Research and Quality] 5T32HS000078).

Reprints: Yue-Yung Hu, MD, MPH, Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery, Feinberg School of Medicine, Northwestern University, 633N. St Clair St, 20th Floor, Chicago, IL 60611. Email: yueyunghu@luriechildrens.org.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Annals of Surgery Open (2022) 4:e228

Received: 6 November 2022; Accepted: 10 November 2022

Published online 12 December 2022

DOI: 10.1097/AS9.0000000000000228

Education-accredited programs following the American Board of Surgery In-Training Examination (ABSITE), which achieves very high response rates (85.6%–99.3%), particularly compared to other literature examining physician burnout (17.9%–76.4%).^{9–15} However, as burnout and mistreatment are sensitive topics discernable only by self-report, prevalence estimates may be substantially biased by nonresponse.^{1,15,16}

In initial iterations, survey response rates were near complete. In subsequent years, changes were made to survey instructions and administration to emphasize the optional nature of the survey, which led to modest declines in response rates, both for the survey overall and for specific questions. We sought to leverage this variation in response rate to: (1) characterize the nonrespondents, (2) examine whether any subgroups of trainees and/or training programs were less likely to respond to questions of sensitive nature, and (3) assess how nonresponse biases prevalence estimates of mistreatment and/or well-being obtained via this survey mechanism.

METHODS

Study Development and Administration

The data source for this study included 2018 and 2019 responses from a national, multiple-choice survey administered following the ABSITE. Development and validation of the survey instrument has been previously described; the survey is constructed from existing survey instruments that are adapted through the cognitive interviews and pilot testing with general surgery trainees from multiple programs with iterative refinement.^{1,9–12,14,17–33} The bulk of the survey remains unchanged year-to-year; however, small changes are occasionally made based on our evolving understanding of wellness. All trainees see the questions in the same order. This survey is administered as part of the Surgical Education Culture Optimization through targeted interventions based on National comparative Data (SECOND) Trial (ClinicalTrials.gov number, NCT03739723).

The survey is preceded by written (displayed on each computer screen) and oral statements (read by on-site proctors) detailing that the purpose of data collection is for research only, participation is voluntary and has no impact on ABSITE score, responses are de-identified, and programs have no access to individual survey responses.¹⁸ This content was consistent between 2018 and 2019. However, in response to feedback, in 2019, these instructions were modified to further emphasize that survey participation is optional and confidential; the American Board of Surgery (ABS) added a preamble stating, “This concludes the 2019... ABSITE. The following section contains a survey... The survey is not mandatory and has no effect on your ABSITE scores. To complete the survey, click next. To end your assessment, click exit,” and the SECOND Trial preamble changed from “All data will be de-identified with respect to program and individual for analyses and reporting,” to, “The survey responses are never associated with your personal identity. All data are de-identified for analyses and reporting. Your program will not have access to your individual responses.” Additionally, the survey administration software was altered to emphasize that the survey could be exited at any time. All data are de-identified by the ABS prior to transmission to Northwestern for analysis. Based on the de-identified nature of the data, the Northwestern University Institutional Review Board determined this study was exempt from human subjects review.

Resident and Program Characteristics

Race/ethnicity was queried in 2019 and categorized into non-Hispanic White, non-White and/or Hispanic, or prefer not to say. Marital status was queried in both years and categorized as no relationship (single/not in a relationship, divorced/

separated, widowed) or married/in a relationship. The ABS provided data on resident gender, clinical postgraduate level (categorized as postgraduate year [PGY] 1, 2/3, or 4/5), and ABSITE score percentile (quartiled). The ABS performed longitudinal linkage of survey responses for individual trainees who were clinically active and therefore eligible for both the 2018 and 2019 surveys prior to de-identification and transmission to Northwestern for analysis.

The ABS also provided program characteristics including size (quartiled: 2–25, 26–30, 37–50, >50 residents), type (university-based, independent, military), and geographic location (Northeast, Southeast, Midwest, Southwest, West).

Outcome Measures

In both 2018 and 2019, the survey included questions about dissatisfaction with career choice, time for rest, and quality of overall resident education; thoughts of attrition from the residency program within the last academic year; and suicidality within the last 12 months.^{17,29,31,33} Dissatisfaction (with education, rest time, and career choice) outcomes were dichotomized into neutral/dissatisfied/very dissatisfied versus satisfied/very satisfied. Thoughts of leaving the program were dichotomized into neutral/disagree/strongly disagree versus agree/strongly agree. Burnout was assessed using a modified abbreviated Maslach Burnout Inventory-Human Services Survey for Medical Professionals and was defined as any symptom of emotional exhaustion or depersonalization experienced at least weekly.^{34,35}

In 2018, survey respondents were asked to report the frequency (never, now and then, monthly, weekly, daily) with which they experienced various forms of mistreatment: (1) bullying or verbal/emotional abuse, (2) sexual harassment, (3) discrimination on the basis of gender, or (4) discrimination on the basis of race/ethnicity. In 2019, specific behaviors corresponding to each category were given (eg, “crude/sexually demeaning or explicit remarks, stories, or jokes” or “unwanted physical sexual attention (e.g., inappropriate or uncomfortable touching; attempts to touch, fondle or kiss)” for sexual harassment; the Short Negative Acts Questionnaire for bullying)²⁰; if respondents reported experiencing any behavior in a category, then they were considered to have experienced that form of mistreatment (ie, bullying, sexual harassment, discrimination on the basis of gender/sexual orientation/gender identity, discrimination on the basis of race/ethnicity/religion).^{3,6,7,40} In both years, trainees were then asked to identify a primary source of each form of mistreatment (2018) or behavior (2019) experienced.^{3,6,7} Mistreatment outcomes were dichotomized by frequency (never vs any).

Response rates were calculated for the overall survey as well for individual questions. Response rates for questions that were thematically grouped (ie, dissatisfaction with education, rest time, and career choice; burnout) are reported as a single average response rate across the related questions. Questions with conditional branching (ie, those which are only prompted by affirmative answers to preceding questions) were excluded from this analysis.

Characterization of Nonrespondents

All clinically active survey respondents in 2019 were categorized into three groups based on their patterns of missing item responses: Completers answered all questions continuously to the end of the survey, Discontinuers stopped answering questions at some point during the survey and did not resume, and Selective Responders selectively answered questions throughout the survey until the end and answered one of the four items in the final question block, indicating that they had viewed all survey questions. Those who selectively answered questions before dropping out of the survey prior to the final question block (ie,

who were both Discontinuers and Selective Responders) were categorized as Discontinuers. Because there were few residents who were categorized as Selective Responders in 2018, data from Discontinuers and Selective Responders in that cohort are reported together as Partial Responders.

Statistical Analysis

Descriptive statistics were calculated for each respondent type. The χ^2 tests were used to compare resident and program-level characteristics between the three respondent groups in 2019. Multivariable multinomial logistic regression was used to assess the association of respondent type (with Completers as comparison group) with resident and program-level characteristics, adjusting for clustering of responses at the program level. Additional multivariable logistic regression models assessed associations of respondent type with well-being outcomes (satisfaction, burnout, thoughts of attrition, suicidality), adjusting for individual and program characteristics, and accounting for clustering of responses at the program level. Multivariable logistic regression was also used to assess association between resident characteristics and nonresponse to potentially sensitive questions. The χ^2 tests were used to compare demographic characteristics and frequency of burnout and mistreatment between Partial Responders in 2018 and 2019.

All statistical analyses were performed using Stata software, version 14.1 (StataCorp, College Station, TX) and SAS version 9.4 (SAS Institute, Cary, NC).

RESULTS

A total of 6956 of 8129 eligible trainees responded at least in part to the 2019 survey (response rate 85.6%). After exclusion of conditional branching questions, a total of 102 questions were included for analysis. Of the 6595 respondents, 4628 (66.5%) residents completed all questions (ie, Completers), 1217 (17.5%) answered some early questions then exited the survey before the end (ie, Discontinuers), and 1111 (16.0%) intermittently answered questions through the end of the survey (ie, Selective Responders).

Selective Responders on average completed 91.1% of questions, and Discontinuers on average completed 43.7% of questions before leaving the survey. Well-being items were answered with high fidelity: 93.2% of residents answered questions about dissatisfaction with education and career, 86.3% completed the abbreviated Maslach Burnout Inventory, 90.8% responded to the question about thoughts of attrition, and 94.4% responded to the question about suicidality. Of all survey takers, 85.% answered questions about sexual harassment, 82.7% answered questions about discrimination based on gender/sexual orientation/gender identity, 81.5% answered questions about discrimination based on race/ethnicity, and 88.7% answered questions about bullying.

Demographic characteristics differed significantly between the three respondent groups, with junior trainees, non-White and/or Hispanic trainees, and trainees not in a relationship more highly represented in both nonresponse groups (Table 1). After adjusting for other resident and program characteristics,

TABLE 1.
Demographic Characteristics of 2019 Survey Respondent Groups

Demographic Characteristics	Respondent Type (N = 6956)			P
	Completers (n = 4628)	Discontinuers (n = 1217)	Selective Responders (n = 1111)	
	n (%)	n (%)	n (%)	
Gender				0.45
Male	2641 (58.4)	699 (60.3)	628 (58.0)	
Female	1881 (41.6)	460 (39.7)	454 (42.0)	
PGYlevel				<0.001
Intern (PGY 1)	1072 (23.2)	443 (36.4)	294 (26.5)	
Junior (PGY 2/3)	1845 (39.9)	441 (36.2)	461 (41.5)	
Senior (PGY 4/5)	1711 (37.0)	333 (27.4)	356 (32.0)	
Marital status				<0.001
Relationship	3503 (75.7)	825 (71.3)	783 (72.1)	
No relationship	1125 (24.3)	332 (28.7)	303 (27.9)	
Race/ethnicity				<0.001
Non-Hispanic White	2892 (62.5)	554 (49.4)	595 (54.3)	
Black/Hispanic/Asian/other	1153 (33.6)	460 (41.0)	435 (39.7)	
Prefer not to say	183 (4.0)	107 (9.6)	65 (5.9)	
ABSITEScore				0.02
Quartile 1 (lowest)	1059 (22.9)	334 (27.4)	284 (25.6)	
Quartile 2	1179 (25.5)	316 (26.0)	290 (26.1)	
Quartile 3	1227 (26.5)	281 (23.1)	267 (24.0)	
Quartile 4 (highest)	1163 (25.1)	286 (23.5)	270 (24.3)	
Program size (number of residents)				0.02
Quartile 1 (2–25)	1208 (26.1)	307 (25.2)	284 (25.6)	
Quartile 2 (26–30)	1163 (25.1)	272 (22.4)	279 (25.1)	
Quartile 3 (37–50)	1214 (26.2)	302 (24.8)	270 (24.3)	
Quartile 4 (>50)	1043 (22.5)	336 (27.6)	278 (25.0)	
Program type				0.06
University-based	2665 (57.6)	701 (57.6)	648 (58.3)	
Independent	1818 (39.3)	490 (40.3)	444 (40.0)	
Military	145 (3.1)	26 (2.1)	19 (1.7)	
Program location				0.01
Northeast	1498 (32.4)	441 (36.2)	389 (35.0)	
Southeast	944 (20.4)	205 (16.8)	203 (18.3)	
Midwest	1051 (22.7)	257 (21.1)	229 (20.6)	
Southwest	508 (11.0)	144 (11.8)	147 (13.2)	
West	627 (13.6)	170 (14.0)	143 (12.9)	

Percentages may not add up to 100 due to rounding.

TABLE 2.
Resident and Program Characteristics Associated With Survey Nonresponse

Resident & Program Characteristics	Respondent Type*	
	Discontinuers	Selective Responders
	OR (95% CI) N = 1106	OR (95% CI) N = 1074
PGYlevel		
Intern (PGY 1)	1.00	1.00
Junior (PGY 2/3)	0.59 (0.50–0.69)	0.91 (0.77–1.08)
Senior (PGY 4/5)	0.49 (0.41–0.58)	0.77 (0.64–0.91)
Marital status		
Relationship	1.00	1.00
No relationship	1.07 (0.92–1.24)	1.12 (0.96–1.31)
Race/ethnicity		
Non-Hispanic White	1.00	1.00
Black/Hispanic/Asian/other	1.45 (1.25–1.67)	1.30 (1.13–1.50)
Prefer not to say	2.87 (2.20–3.75)	1.60 (1.18–2.17)
ABSITEScore		
Quartile 1 (lowest)	1.00	1.00
Quartile 2	0.91 (0.75–1.10)	0.94 (0.78–1.14)
Quartile 3	0.82 (0.67–0.99)	0.84 (0.69–1.01)
Quartile 4 (highest)	0.88 (0.72–1.06)	0.91 (0.75–1.10)
Program size (number of residents)		
Quartile 1 (2–25)	1.00	1.00
Quartile 2 (26–30)	1.01 (0.80–1.26)	1.02 (0.84–1.25)
Quartile 3 (37–50)	1.12 (0.87–1.45)	0.96 (0.77–1.20)
Quartile 4 (>50)	1.36 (1.02–1.82)	1.10 (0.86–1.40)
Program type		
Academic	1.00	1.00
Community	1.16 (0.94–1.44)	1.04 (0.86–1.24)
Military	0.83 (0.48–1.42)	0.61 (0.37–1.01)
Program location		
Northeast	1.00	1.00
Southeast	0.81 (0.64–1.02)	0.88 (0.72–1.08)
Midwest	0.88 (0.70–1.09)	0.88 (0.73–1.07)
Southwest	1.05 (0.79–1.40)	1.18 (0.93–1.48)
West	0.93 (0.72–1.21)	0.89 (0.71–1.11)

*Multinomial logistic regression accounts for clustering of responses at the program level. Reference group is Completers (respondents who completed the survey in its entirety).

TABLE 3.
Association of Wellness With Survey Nonresponse

Well-Being Outcomes	Respondent Type*	
	Discontinuers	Selective Responders
	OR (CI)	OR (CI)
Dissatisfaction with:		
Quality of overall resident education	1.17 (0.98–1.40)	1.18 (1.01–1.36)
Time for rest	1.19 (1.00–1.41)	1.12 (0.98–1.29)
Decision to become a surgeon	1.28 (1.05–1.55)	1.13 (0.95–1.35)
Thoughts of leaving the program	0.97 (0.74–1.29)	1.24 (1.03–1.49)
Burnout	1.10 (0.86–1.39)	1.04 (0.91–1.19)
Suicidality	0.92 (0.61–1.37)	1.14 (0.83–1.57)

Six separate logistic regression analyses regressing dissatisfaction with education, rest time, career choice, attrition, burnout, suicidality outcomes on respondent type with controls for gender, race, marital status, PGY level, ABSITE score quartile, program size, program type, and program location, and accounting for clustering of responses at program level.

*Reference group is Completers.

PGY 4/5 residents were less likely to be Discontinuers (odds ratio [OR], 0.49; 95% confidence interval [CI], 0.42–0.57) or Selective Responders (OR, 0.77; 95% CI, 0.65–0.91) when compared to interns (Table 2). Compared to non-Hispanic Whites, racially/ethnically minoritized residents were more likely to be Discontinuers (OR, 1.46; 95% CI, 1.27–1.67) or Selective Responders (OR, 1.30; 95% CI, 1.13–1.50), as were trainees who declined to identify their race or ethnicity (Discontinuers:

OR, 2.85; 95% CI, 2.23–3.64 and Selective Responders: OR, 1.60; 95% CI, 1.18–2.15) (Table 2). Trainees in the largest training programs (quartile 4, >50 trainees) were more likely to be Discontinuers (OR, 1.41; 95% CI, 1.04–1.91). Resident gender, marital status, ABSITE score, program location and program type (university-based, independent, military) were not significantly associated with being a Discontinuer or Selective Responder on multivariable analysis.

Non-White or Hispanic residents were more likely to omit questions regarding racial/ethnic/religious discrimination (21.2% vs 15.8%; OR, 1.35; 95% CI, 1.18–1.54), especially regarding the sources (20.5% vs 14.6%; OR, 1.43; 95% CI, 1.24–1.64). Female trainees were not more likely to omit questions regarding gender/gender identity/sexual orientation discrimination (17.7% vs 18.9%; OR, 0.91; 95% CI, 0.79–1.04) or its sources (18.0% vs 17.7%; OR, 1.02; 95% CI, 0.89–1.16).

The patterns of the missingness for the Discontinuers and Selective Responders as they progressed through the survey are displayed in Figure 1. Discontinuers missed 3–90 questions, with a mean of 57 (SD 23). Selective Responders skipped 1–75 questions, with a mean of 9 (SD 13). After adjusting for resident and program characteristics, Discontinuers who completed the survey through dissatisfaction questions were more likely to be dissatisfied with choice to be a surgeon (OR, 1.28; 95% CI, 1.05–1.55) and time for rest (OR, 1.19; 95% CI, 1.00–1.41) (Table 3). Selective Responders who completed dissatisfaction and attrition questions were more likely to be dissatisfied with residency education (OR, 1.18; 95% CI, 1.01–1.36) and have thoughts of attrition (OR, 1.24; 95% CI, 1.03–1.49) (Table 3).

Seven thousand four hundred fifteen total eligible trainees responded to the 2018 survey (response rate, 99.3%). Of 4821 clinically active trainees eligible for both the 2018 and 2019 survey, 4230 (87.7%) responded at least in part to both, and 591 (12.3%) responded to the 2018 but declined the 2019 survey (Table 4). Compared to those who responded to both surveys, those who declined the 2019 survey entirely were more often interns (35.0% vs 31.1%; $P = 0.001$), not in a relationship (30.6% vs 27.1%; $P = 0.004$), and low ABSITE scorers (26.1% vs 21.9% in lowest quartile; $P = 0.001$) in 2018 (Table 4). Additionally, they reported higher proportions of physical abuse (4.2% vs 1.5%; $P < 0.001$), sexual harassment (12.4% vs 9.1%; $P = 0.02$), and racial discrimination (18.6% vs 15.0%; $P = 0.03$) in 2018. Partial Responders in 2019 (ie, both Discontinuers and Selective Responders) reported physical abuse (2.5% vs 1.1%; $P = 0.001$) and racial discrimination (18.3% vs 13.6%; $P < 0.001$) at higher rates in 2018 than Completers of the 2019 survey. Aside from more Partial Responders seen at larger training programs (program size quartile 5, >50 residents, 20.0% vs 26.0%; $P = 0.001$), no program characteristics were seen to impact 2019 nonresponses.

DISCUSSION

Mistreatment and poor well-being remain substantial problems in surgical training, and measurement of these problems relies upon self-report. Overall, the response rates on our national survey of well-being are high. After increasing the number of survey exit points, the majority of residents (86%) still chose to respond. For comparison, physician well-being surveys more typically have response rates in the 18%–32% range.^{15,16} Although one-third of respondents belonged to a partial respondent group (Selective Responders and Discontinuers), key questions about well-being had response rates of 85%–95%, and sensitive questions about mistreatment were answered at rates of 82%–85%. Few other reports have been able to capture such a comprehensive picture of physician well-being, even by anonymous survey.^{36,37}

Belonging to a nonrespondent group was independently associated with measures of dissatisfaction with education, rest time, and career choice, suggesting nonresponse has a root cause (eg,

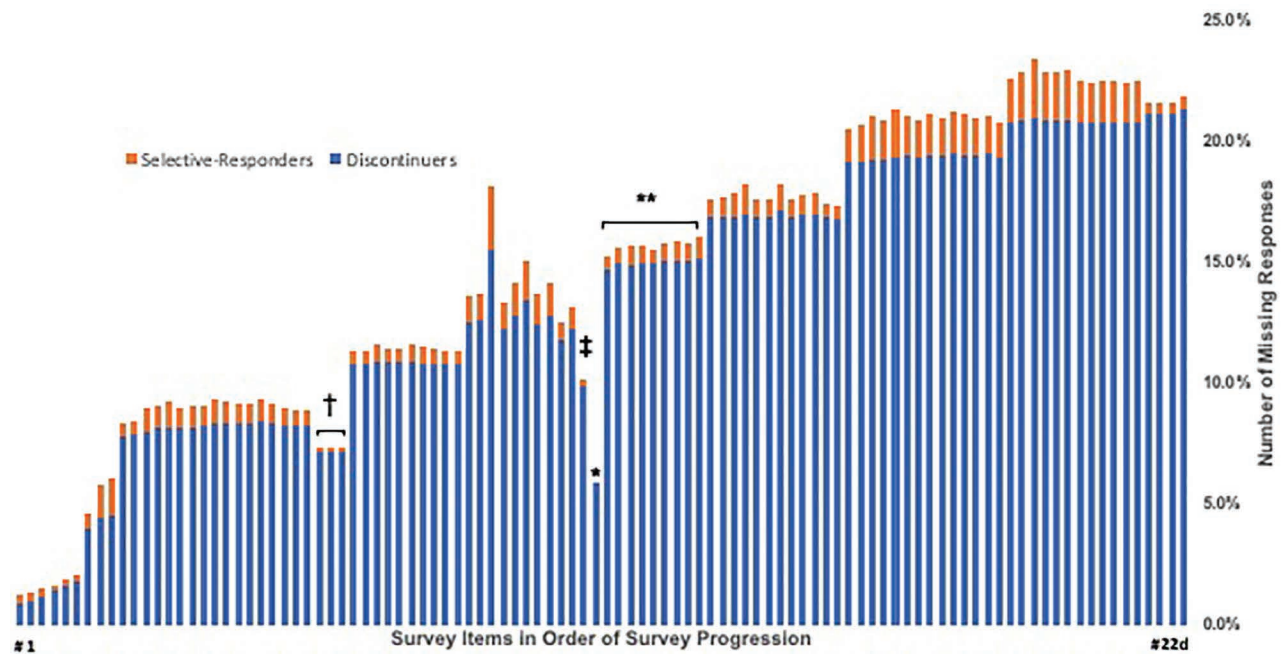


FIGURE 1. Patterns of missingness for survey respondents. All respondents were presented survey questions in the same order, reflected by the x axis with question 1 on the left and the final question on the right. †Dissatisfaction response rate (93.2%). ‡Suicidality response rate (94.4%). *Thoughts of attrition response rate (90.8%). **Burnout response rate (86.3%).

TABLE 4.
Demographic Characteristics of 2018 Survey Respondent Groups

Demographic Characteristics	Respondent Type (N = 4821)		P
	Total Nonresponse to 2019 Survey (n = 591)	Responded to 2019 Survey (n = 4230)	
	n (%)	n (%)	
Gender			0.004
Male	234 (39.6)	1722 (40.7)	
Female	349 (59.1)	2493 (58.9)	
PGYlevel			0.001
Intern (PGY 1)	207 (35.0)	1314 (31.1)	
Junior (PGY 2/3)	276 (46.7)	1842 (43.5)	
Senior (PGY 4/5)	108 (18.3)	1074 (25.4)	
Marital status			0.004
Relationship	410 (69.4)	825 (71.3)	
No relationship	181 (30.6)	1145 (27.1)	
ABSITEScore			0.001
Quartile 1 (lowest)	154 (26.1)	925 (21.9)	
Quartile 2	178 (30.1)	1097 (25.9)	
Quartile 3	141 (23.9)	1107 (26.2)	
Quartile 4 (highest)	118 (20.0)	1101 (26.0)	
Program size (number of residents)			0.018
Quartile 1 (2–25)	172 (29.1)	1128 (26.7)	
Quartile 2 (26–30)	123 (20.8)	1125 (26.6)	
Quartile 3 (37–50)	145 (24.5)	1030 (24.3)	
Quartile 4 (>50)	151 (25.5)	947 (22.4)	
Program type			0.381
Academic	341 (57.8)	2417 (57.4)	
Community	229 (38.8)	1689 (40.1)	
Military	20 (3.4)	104 (2.5)	
Program location			0.172
Northeast	176 (29.8)	1358 (32.1)	
Southeast	136 (23.0)	857 (20.3)	
Midwest	124 (21.0)	931 (22.0)	
Southwest	83 (14.0)	497 (11.7)	
West	72 (12.2)	587 (13.9)	

a distrust of confidentiality or overall disengagement) that transcends identification with any particular demographic group. However, no differences were seen between nonrespondents in

burnout or suicidality. As expected, based on prior validation work of these instruments,^{3,6,7,34,35} our findings indicate that these measures are stable to circumstance. Prior research on

survey bias suggest that physician responses may be less susceptible to nonresponse bias than the general public.³⁶

Although we know that racial/ethnic discrimination is common in the physician workplace,^{6,13} particularly for visibly (non-White) minoritized people,⁸ non-White and Hispanic residents were less likely to respond to questions about racial/ethnic discrimination. While we can only speculate about the reason behind nonresponse to these questions regarding race and ethnicity, fear of loss of confidentiality is a plausible explanation. Black and Latinx Americans are severely underrepresented in medical school compared to their share of the US population, and they are even further underrepresented in surgical residency,³⁸ as such, race/ethnicity alone would theoretically be sufficient information to identify them. The SECOND Trial seeks to gather this information to inform and generate the national discussion about the lived experience of trainees and has implemented multiple layers of protections for resident confidentiality: (1) we receive only de-identified data from the ABS; (2) we provide only aggregated and benchmarked data to participating programs, with each metric reported as a performance quartile, compared to other programs in the country, rather than percentages of residents (eg, of all programs in the country, yours performs in the worst quartile for suicidality); and (3) we withhold data for which there are four or fewer susceptible residents (eg, data on racial/ethnic discrimination for racially/ethnically minoritized trainees). However, residents may not be familiar with all of these safeguards, particularly if their program elects to not share their SECOND Trial data with them. Additional research is needed to further understand nonresponse in non-White or Hispanic residents.

We did not find differences in prevalence between nonrespondent groups for sexual harassment or discrimination based on gender/sexual orientation/gender identity. These findings suggest that previously published prevalence estimates for these types of mistreatment are accurate.^{2,3,7,8,40} We posit that because surgical residencies are approaching gender parity, (1) identification based upon gender alone is becoming increasingly challenging, and (2) visibility of these issues is increasing. As such, the sense of disempowerment among women facing these issues may be decreasing, resulting in fewer nonresponses.

Nonrespondents were found more frequently at larger training programs. We hypothesize that the local sense of safety and community within the training program may influence differential response rates. Small programs may have the ability to develop close-knit communities in which residents feel comfortable voicing their concerns, whereas large programs, particularly ones in which residents rotate across multiple hospital sites, may struggle to create the same sense of community or engagement. Additionally, larger programs frequently have more preliminary residents who, by definition, lack of long-term continuity within their programs, and who may feel less compelled to report on the survey because they are unlikely to reap any long-term benefits of doing so.

Our study has several potential limitations. First, we make inferences from nonresponses based upon demographics and historical responses of the nonresponders; however, we cannot know with certainty how they would have responded to unanswered questions. Second, all surveys are subject to recall bias. Whenever possible, we used validated survey items (eg, burnout, suicidality) that are known to be stable over the time period being measured. For all other items (eg, mistreatment), we tried to limit the impact of recall bias by restricting the period of recall to the current academic year (ie, starting 6 months prior to survey administration). Third, the sensitive content of the questions may result in social desirability bias (ie, respondents may misrepresent their true answers to better fit a desired social norm). We believe that nonresponse may have been a mitigation strategy for the stigma associated with certain responses (eg, racial/ethnic discrimination). Nevertheless, this bias should result in

conservative estimations of the prevalence of stigmatized outcomes (ie, mistreatment, burnout, thoughts of attrition, and suicidality). Fourth, longitudinal tracking of ABSITE scores and program enrollment in the SECOND Trial (ClinicalTrials.gov number NCT03739723), which provides anonymous, aggregated feedback to programs, may have led to resident concerns of confidentiality and subsequently missing or biased reporting, particularly in programs with few non-White or Hispanic residents, despite assurances of confidentiality in the survey preamble. Again, this fear would be expected to bias our prevalence estimates in the conservative direction. Finally, the questions appeared in the same order for all respondents with mistreatment questions appearing near the end of the survey; therefore, we cannot be definitively determine whether these questions were skipped due question sensitivity versus survey fatigue.

Understanding the scope of mistreatment and poor well-being is critical. Prior estimations of burnout, suicidality, and gender discrimination are not likely to be significantly influenced by nonresponse. However, nonresponses to questions about racial/ethnic discrimination likely results in underestimation. Our findings speak to the crucial need for creating inclusive environments in which concerns may be safely voiced and meaningfully addressed. Nonetheless, this survey mechanism remains a high-fidelity way to capture the voice of residents in a standardized fashion across the country.

REFERENCES

- Dyrbye LN, Burke SE, Hardeman RR, et al. Association of clinical specialty with symptoms of burnout and career choice regret among US resident physicians. *JAMA*. 2018;320:1114–1130.
- Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med*. 2019;381:1741–1752.
- Zhang LM, Ellis RJ, Ma M, et al. Prevalence, types, and sources of bullying reported by US general surgery residents in 2019. *JAMA*. 2020;323:2093–2095.
- Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Inter Med*. 2017;177:1826–1832.
- Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. *JAMA*. 2017;317:901–902.
- Yuce TK, Turner PL, Glass C, et al. National evaluation of racial/ethnic discrimination in US surgical residency programs. *JAMA Surg*. 2020;155:526–528.
- Schlick CJR, Ellis RJ, Etkin CD, et al. Experiences of gender discrimination and sexual harassment among US general surgery residents. *J Am Coll Surg*. 2020;231:S245.
- Yuce TK, Bilimoria KY, Hu YY. Discrimination in US surgical training programs—reply. *JAMA Surg*. 2020;155:1084–1085.
- Baldwin DC Jr, Daugherty SR, Rowley BD. Racial and ethnic discrimination during residency: results of a national survey. *Acad Med*. 1994;69:S19–S21.
- Cook DJ, Liutkus JF, Risdon CL, et al. Residents' experiences of abuse, discrimination and sexual harassment during residency training. *McMaster University Residency Training Programs*. *CMAJ*. 1996;154:1657–1665.
- Fnais N, Soobiah C, Chen MH, et al. Harassment and discrimination in medical training: a systematic review and meta-analysis. *Acad Med*. 2014;89:817–827.
- Nagata-Kobayashi S, Maeno T, Yoshizu M, et al. Universal problems during residency: abuse and harassment. *Med Educ*. 2009;43:628–636.
- Nunez-Smith M, Pilgrim N, Wynia M, et al. Race/ethnicity and workplace discrimination: results of a national survey of physicians. *J Gen Intern Med*. 2009;24:1198–1204.
- Paice E, Aitken M, Houghton A, et al. Bullying among doctors in training: cross sectional questionnaire survey. *BMJ*. 2004;329:658–659.
- Shanafelt TD, Balch CM, Bechamps GJ, et al. Burnout and career satisfaction among American surgeons. *Ann Surg*. 2009;250:463–471.
- West CP, Dyrbye LN, Sinsky C, et al. Resilience and burnout among physicians and the general US working population. *JAMA Netw Open*. 2020;3:e209385.
- Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. *N Engl J Med*. 2016;374:713–727.

18. Bilimoria KY, Chung JW, Hedges LV, et al. Development of the Flexibility in Duty Hour Requirements for Surgical Trainees (FIRST) trial protocol: a national cluster-randomized trial of resident duty hour policies. *JAMA Surg.* 2016;151:273–281.
19. Carr PL, Ash AS, Friedman RH, et al. Faculty perceptions of gender discrimination and sexual harassment in academic medicine. *Ann Intern Med.* 2000;132:889–896.
20. Conway PM, Høgh A, Nabe-Nielsen K, et al. Optimal cut-off points for the short-negative act questionnaire and their association with depressive symptoms and diagnosis of depression. *Ann Work Expo Health.* 2018;62:281–294.
21. Crebbin W, Campbell G, Hillis DA, et al. Prevalence of bullying, discrimination and sexual harassment in surgery in Australasia. *ANZ J Surg.* 2015;85:905–909.
22. Fitzgerald LF, Gelfand MJ, Drasgow F. Measuring sexual harassment: theoretical and psychometric advances. *Basic App Soc Psychol.* 1995;17:425–445.
23. Greytak E, Gill A, Conron K, et al. *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys.* Williams Institute, 2014.
24. Jaggi R, Griffith KA, Jones R, et al. Sexual harassment and discrimination experiences of academic medical faculty. *JAMA.* 2016;315:2120–2121.
25. Liebschutz JM, Darko GO, Finley EP, et al. In the minority: black physicians in residency and their experiences. *J Natl Med Assoc.* 2006;98:1441–1448.
26. Richman JA, Flaherty JA, Rospenda KM. Perceived workplace harassment experiences and problem drinking among physicians: broadening the stress/alienation paradigm. *Addiction.* 1996;91:391–403.
27. Richman JA, Flaherty JA, Rospenda KM, et al. Mental health consequences and correlates of reported medical student abuse. *JAMA.* 1992;267:692–694.
28. Sandler BJ, Tackett JJ, Longo WE, et al. Pregnancy and parenthood among surgery residents: results of the first nationwide survey of general surgery residency program directors. *J Am Coll Surg.* 2016;222:1090–1096.
29. Shanafelt TD, Balch CM, Dyrbye L, et al. Special report: suicidal ideation among American surgeons. *Arch Surg.* 2011;146:54–62.
30. Siller H, Tauber G, Komlenac N, et al. Gender differences and similarities in medical students' experiences of mistreatment by various groups of perpetrators. *BMC Med Educ.* 2017;17:1–8.
31. Sullivan MC, Yeo H, Roman SA, et al. Surgical residency and attrition: defining the individual and programmatic factors predictive of trainee losses. *J Am Coll Surg.* 2013;216:461–471.
32. Torres-Harding SR, Andrade AL Jr, Romero Diaz CE. The Racial Microaggressions Scale (RMAS): a new scale to measure experiences of racial microaggressions in people of color. *Cultur Divers Ethnic Minor Psychol.* 2012;18:153–164.
33. Yang AD, Chung JW, Dahlke AR, et al. Differences in resident perceptions by postgraduate year of duty hour policies: an analysis from the flexibility in duty hour requirements for surgical trainees (FIRST) trial. *J Am Coll Surg.* 2017;224:103–112.
34. McManus IC, Winder BC, Gordon D. The causal links between stress and burnout in a longitudinal study of UK doctors. *Lancet.* 2002;359:2089–2090.
35. Riley MR, Mohr DC, Waddimba AC. The reliability and validity of three-item screening measures for burnout: evidence from group-employed health care practitioners in upstate New York. *Stress Health.* 2018;34:187–193.
36. Kellerman SE, Herold J. Physician response to surveys: a review of the literature. *Am J Prev Med.* 2001;20:61–67.
37. Wallace JE. Mental health and stigma in the medical profession. *Health.* 2012;16:3–18.
38. Butler PD, Longaker MT, Britt LD. Major deficit in the number of underrepresented minority academic surgeons persists. *Ann Surg.* 2008;248:704–711.
39. Bruce AN, Battista A, Plankey MW, et al. Perceptions of gender-based discrimination during surgical training and practice. *Med Educ Online.* 2015;20:25923.
40. Heiderscheid EA, Schlick CJR, Ellis RJ, et al. Experiences of LGBTQ+ residents in US general surgery training programs. *JAMA Surg.* 2021;157:23–32.