UC Irvine
UC Irvine Previously Published Works

Title
An Empirical National Assessment of the Learning Environment and Factors Associated With Program Culture.

Permalink
https://escholarship.org/uc/item/3kk514z9

Journal
Annals of Surgery, 270(4)

Authors
Ellis, Ryan
Hewitt, D
Hu, Yue-Yung
et al.

Publication Date
2019-10-01

DOI
10.1097/SLA.0000000000003545

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Ann Surg. Author manuscript; available in PMC 2020 October 01.

Published in final edited form as:

Ann Surg. 2019 October ; 270(4): 585–592. doi:10.1097/SLA.0000000000003545.

An Empirical National Assessment of the Learning Environment and Factors Associated With Program Culture

Ryan J. Ellis, MD, MS[*,†], **D. Brock Hewitt, MD, MS, MPH**[*,‡], **Yue-Yung Hu, MD, MPH**[*], **Julie K. Johnson, PhD**[*], **Ryan P. Merkow, MD, MS**[*,†], **Anthony D. Yang, MD, MS**[*], **John R. Potts III, MD**[§], **David B. Hoyt, MD**[†], **Jo Buyske, MD**[¶], **Karl Y. Bilimoria, MD, MS**[*,†,✉]

[*]Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery and Center for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, IL [†]American College of Surgeons, Chicago, IL [‡]Department of Surgery, Thomas Jefferson University Hospital, Philadelphia, PA; [§]Accreditation Council for Graduate Medical Education (ACGME), Chicago, IL; and [¶]American Board of Surgery, Philadelphia, PA.

Abstract

Objectives: To empirically describe surgical residency program culture and assess program characteristics associated with program culture.

Summary Background Data: Despite concerns about the impact of the learning environment on trainees, empirical data have not been available to examine and compare program-level differences in residency culture.

Methods: Following the 2018 American Board of Surgery In-Training Examination, a cross-sectional survey was administered to all US general surgery residents. Survey items were analyzed using principal component analysis to derive composite measures of program culture. Associations between program characteristics and composite measures of culture were assessed.

Results: Analysis included 7387 residents at 260 training programs (99.3% response rate). Principal component analysis suggested that program culture may be described by 2 components: Wellness and Negative Exposures. Twenty-six programs (10.0%) were in the worst quartile for both Wellness and Negative Exposure components. These programs had significantly higher rates of duty hour violations (23.3% vs 11.1%), verbal/physical abuse (41.6% vs 28.6%), gender discrimination (78.7% vs 64.5%), sexual harassment (30.8% vs 16.7%), burnout (54.9% vs 35.0%), and thoughts of attrition (21.6% vs 10.8%; all $P < 0.001$). Being in the worst quartile of both components was associated with percentage of female residents in the program ($P = 0.011$), but not program location, academic affiliation, size, or faculty demographics.

Conclusions: Residency culture was characterized by poor resident wellness and frequent negative exposures and was generally not associated with structural program characteristics. Additional qualitative and quantitative studies are needed to explore unmeasured local social dynamics that may underlie measured differences in program culture.

✉ k-bilimoria@northwestern.edu.

The authors report no conflicts of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

duty hours; general surgery; residency; training environment

Poor physician wellness, often characterized by burnout and career dissatisfaction, has been linked to attrition and poor mental health.[1–4] Trainees, especially those within the surgical subspecialties, may be particularly at risk.[5] As a result, the Accreditation Council for Graduate Medical Education (ACGME) has highlighted the need to address the growing problems with stress, burnout, and depression among trainees and develop interventions to improve wellness overall.[6]

The training environment itself may contribute significantly to the development of burnout and poor wellness. Reports of trainee mistreatment have existed for decades,[7–9] and behaviors such as abuse, discrimination, and harassment may be more common in surgical training.[10–12] The workload itself can be particularly daunting in surgery despite modifications to work hour restrictions.[13–15]

Despite efforts to understand training stressors and resident wellness, it is unclear how much program-level variation exists in the residency learning environment (eg, verbal abuse, discrimination, burnout). Moreover, studies to date evaluating residency training program culture have generally been limited to single parameters examined individually (eg, bullying, burnout), which has precluded development of composites for global residency program assessment. Consolidation of multiple measures of the training environment into composites describing program culture would aid internal assessment and could be used to evaluate wellness interventions. Additionally, such composite tools could be used to determine if structural elements (eg, program type, size, or faculty demographics) are related to overall training program culture. The objectives of this study were to 1) describe residency program culture using indicators pertaining to resident wellness and environmental exposures, 2) develop empirically derived composites to describe the residency training environment, and 3) evaluate program-level factors associated with program culture.

METHODS

Study Design and Participants

A voluntary, multiple-choice survey was administered to general surgery residents following the January 2018 American Board of Surgery In-Training Examination (ABSITE). Responses were deidentified prior to analysis. The study population was limited to clinically active residents. All programs with at least 5 residents overall (≥1 per class) and at least 1 female resident were considered in the analysis (3 excluded programs). The Northwestern University Institutional Review Board office determined that this study constitutes non-human subjects research.

Resident Survey

The survey instrument included content related to resident wellness, duty hour adherence, and experience with workplace gender discrimination, verbal and physical abuse, and sexual

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

harassment. Survey items were adapted from previously published and validated surveys.[16–18] Pretest cognitive interviews were conducted with general surgery residents to evaluate survey coherence and clarity, followed by iterative revisions.[15,17]

Resident Exposures and Measures of Wellness

All resident-level exposures were aggregated to the program level. Residents reported the number of months that they had violated the ACGME 80 hours per week rule (defined as averaging >80h/wk over a 4-wk period; dichotomized for analysis: 0–2 vs 3+ mo). Residents also reported experiences with discrimination based on gender, abuse (either verbal or physical), and sexual harassment since they began residency training. The Perceived Stress Scale 4 was used to evaluate resident stress over the last month.[19] In this study, residents more than 1 standard deviation above the normative mean in perceived stress were considered to experience severe stress.[20] Residents were also asked to rate their satisfaction with being a surgeon and their overall wellbeing on a 5-point Likert scale (Very Dissatisfied to Very Satisfied) which was dichotomized for analyses (Very Dissatisfied, Dissatisfied, or Neutral vs Satisfied or Very Satisfied).

Symptoms of burnout were assessed using a modified, abbreviated Maslach Burnout Inventory Human Services Survey for Medical Personnel.[21–23] The instrument assessed emotional exhaustion and depersonalization with 3 questions for each domain. To facilitate data interpretation and presentation of the 2 burnout subscales, subscale scores were dichotomized into 2 groups. Burnout was defined as reporting symptoms of *either* emotional exhaustion or depersonalization at least weekly. Residents were also asked if they agreed with the following statement, “I have considered leaving my program in the last year” on a 5-point Likert scale (strongly agree to strongly disagree). Responses of agree or strongly agree were considered “thoughts of attrition.” Suicidal thoughts were assessed using a validated,[24–26] single survey question, “During the past 12 months, have you had thoughts of taking your own life?” Affirmative responses were immediately provided with the information necessary to contact the National Suicide Prevention Lifeline and a behavioral health professional.

Program Characteristics

Program characteristics collected included program size (total number of surgical residents broken into quartiles: 6–25, 26–37, 38–51, 58–81 residents), program type (academic or community/ military), program location (Northeast, Southeast, Midwest, Southwest, West), and percentage of female residents (separated into quartiles). The genders of each program’s chair and program director were ascertained, and the Association of American Medical Colleges provided demographic information regarding surgery faculty for each US medical school participating in their Faculty Roster in 2016. From this, proportions of female and non-White attending surgeons were calculated. Programs were categorized into quartiles based on these proportions. Programs with more than 1 medical school affiliation were assigned to the school listed as the primary affiliate.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Constructing Composite Measures of Program Environment

Principal component analysis (PCA) is a statistical method for data reduction that, in this study, was used to examine interrelated survey items and group variables into a smaller number of composite variables. In this study, PCA reduced numerous measures of resident wellness and the residency program learning environment into a small number of composite factors. Program-level rates of 8 resident exposure variables were included in the PCA: duty hour violations, gender discrimination (among female residents only), verbal/physical abuse, sexual harassment (among female residents only), severe stress, dissatisfaction with surgical career, burnout, and thoughts of attrition. Components with eigenvectors ≥1 were retained for use in the main statistical analysis and orthogonally rotated to ease interpretation. Discrimination between high- and low-performing programs by PCA was validated using 2 wellness outcome variables not included in the PCA derivation: overall wellbeing and suicidal thoughts.

Statistical Analysis

The Mann–Whitney *U* test was used for comparison of non-normally distributed data. Bivariate associations were examined using chi-square tests. Missing data were rare (<1%) and were excluded from analyses as noted in the tables. Level of significance was set to 0.05. Data analyses were performed at Northwestern University. Statistical analyses were performed using STATA 15.1 (StataCorp LP, College Station, TX).

RESULTS

Of 7464 clinically active residents across 263 ACGME-accredited programs, 7413 provided survey responses (99.3% response rate). One new program with fewer than 5 residents and the 2 programs that had no female residents were excluded from the analyses, yielding 7387 residents at 260 programs for the final analysis.

Program Characteristics and Variation

The largest number of programs were located in the Northeast (32.7%), and nearly half were academic (46.1%). Fewer than 10% of programs had a female Department Chair, while 19.6% of programs had a female program director. Additional details regarding program characteristics can be found in Table 1.

Substantial program-level variation was observed in both resident exposure and wellness variables (Table 2). The median program-level rate of frequent duty hour violations was 11.9% (IQR: 5.8%–19.2%), while the median program-level rate of verbal/physical abuse was 30.0% (IQR: 20.8%–38.3%). Among female residents, the median program-level rate of gender discrimination was 66.7% (IQR: 50.0%–76.8%) and 16.7% for sexual harassment (IQR: 9.1%–28.6%). Similar distributions were observed for program-level rates of severe stress (median 13.3%, IQR: 8.3%–19.7%), dissatisfaction with surgical career (22.9%, IQR: 16.5%–29.4%), burnout (36.6%, IQR: 28.6%–46.9%), and thoughts of attrition (11.3%, IQR: 6.3%–16.3%).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Principal Component Analysis (PCA)

The Bartlett test for sphericity ($P < 0.001$) and Kaiser–Meyer–Olkin statistic (0.799) indicated that the 8 program-level variables were appropriately intercorrelated for principal component analysis. Two components were extracted with eigenvalues ≥1, which together captured 53.8% of the variation of the 8 input variables. Patterns of loadings suggested that Component 1 (generalized as "*Program Wellness*") is predominantly derived from program-level stress, career satisfaction, burnout, and thoughts of attrition, while Component 2 (generalized as "*Negative Exposures*") is predominantly derived from program-level verbal/physical abuse, gender discrimination, and sexual harassment (Table 3).

Program Wellness and Negative Exposures were subsequently dichotomized as highest (worst) quartile programs compared with the bottom 3 quartile programs. There were significant differences in in median values of input variables that were used in the PCA across quartile groupings (Table 4). Programs in the worst quartile (n = 65) of Program Wellness had significantly higher rates of duty hour violations, verbal/physical abuse, gender discrimination, stress, dissatisfaction with being a surgeon, burnout, and thoughts of attrition ($P < 0.001$ for all) compared with programs in the best 3 quartiles (Table 4). There were not significantly differences rates of sexual harassment based on the Program Wellness component ($P = 0.514$).

Programs in the worst quartile (n = 65) of Negative Exposures had significantly higher rates of duty hour violations, verbal/physical abuse, gender discrimination, sexual harassment, stress, burnout, and thoughts of attrition compared with programs in the other 3 quartiles (Table 4). There were no significant differences in rates of dissatisfaction with being a surgeon ($P = 0.708$) or thoughts of attrition (0.061) based on the Negative Exposures component. Programs in the highest quartile of both Program Wellness and Negative Exposures (n = 26) had significantly higher median values for all 8 variables ($P < 0.001$; Table 4). Programs in the worst quartile of Wellness, Negative Exposures, also had higher rates of poor wellbeing and suicidal thoughts (all $P < 0.005$).

Associations With Program Factors

Program-level analysis demonstrated no association between being in the worst quartile of Wellness based on location ($P = 0.139$), program type ($P = 0.426$), gender of department chair ($P = 0.651$) or program director ($P = 0.058$), percentage of female faculty ($P = 0.292$). There were significant differences in program wellness based on percentage of female residents (ranging from 15.2% in the first quartile to 37.3% in the fourth quartile; $P = 0.007$), and non-monotonic differences were observed between Program Wellness and size ($P = 0.024$) and percentage of non-White faculty ($P = 0.047$, Table 5). Only geographic location was associated with Negative Exposures (ranging from 34.1% of programs in the Northeast to 16.1% of programs in the Southeast; $P = 0.043$). There was also a statistically significant but non-monotonic association between percentage of female residents and programs in the worst quartile of both Program Wellness and Negative Exposures (ranging from 1.5% in the second quartile to 18.6% in the fourth quartile; $P = 0.011$, Table 5).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DISCUSSION

In this study, a national survey of general surgery residents was used to characterize residency program culture as measured by duty hour violations, resident mistreatment, resident wellness, and career satisfaction. Principal component analysis identified 2 distinct domains of program-level culture that reflected Wellness and Negative Exposures. Results of the PCA were validated by demonstrating that programs with poor scores on the 2 components also had high rates of poor overall wellbeing and suicidal ideation. Resident Negative Exposures appeared to vary by geographic region, while program wellness varied based on number of female residents. Other program factors such as size, academic affiliation, and demographics of leadership and faculty were not associated with program culture. To our knowledge, these results provide the most comprehensive empiric analysis of program-level variation in training culture at US surgery programs performed to date.

This work adds significantly to previous studies of the medical and surgical training environment. Previous studies on trainee wellness and environmental exposures have focused on medical students[27,28] and residents[5,11,29,30] with consistent demonstration of relatively high rates of both poor wellness and negative exposures in both populations. However, these studies and subsequent meta-analyses are limited by relatively small sample sizes and low survey response rates, which make them vulnerable to nonresponse bias and preclude robust institutional comparisons. Moreover, these studies have generally focused on single exposures or wellness measures (ie, the study only examines burnout in isolation without consideration of negative exposures). The high response rate and breadth of this study make it a much more comprehensive characterization of the training environment.

A striking result of this study is the variation in program-level rates of all measured variables, including some programs with nearly every female resident reporting gender-based mistreatment and more than half of all residents considering attrition. Perhaps more importantly, this distribution demonstrates that a similar number of programs have very *low* rates of reported outcomes such as burnout. This implies that good program-level outcomes are achievable and that there is nothing inherent in the surgical training environment that obligates tolerance of high rates of poor wellness.

The contours of program-level residency culture are complicated. The results indicate that, while some programs have poor rates of many or all studied indicators, there are unlikely to be uniformly “good” or “bad” programs. Even among those in the worst quartile of both Wellness and Negative Exposures, there are programs with relatively low rates of some exposures (eg, verbal/physical abuse). Conversely, some programs outside of the highest quartile for either component may have had relatively high rates of 1 or 2 exposures. Thus, while the program-level variation indicates that low rates of exposures are possible, the inconsistent relationship between variables implies that good program-level outcomes in 1 variable do not guarantee good culture overall. This finding highlights the need for composite assessment of the training environment, as individual factors may not reflect the environment as a whole.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Regarding those programs with uniformly high rates of measures of poor culture, it is notable that there were relatively few program factors associated with being in the highest quartile of either PCA component, including academic affiliation and demographics of leadership and faculty. The only significant structural factors associated with program culture were percentage of female residents, program size, and percentage of non-White faculty (associated with Program Wellness) and geographic location (Program Negative Exposures). These results are challenging to interpret in isolation, especially as variation was often not monotonic within these associations. Similarly, observed associations between geographic location and Negative Exposures may be driven by regional differences in behavior or reporting, or could be the result of confounding variables not captured in this study. Mechanisms underlying these finding are not clear, and additional studies may be required to validate these associations. However, there were some trends (eg, very few programs highest quartile of both components had female department chairs) that did not reach statistical significance due to sample size. Regarding faculty demographics, it is important to note that recent literature indicates that the “tipping point” in social convention (eg, the point at which a more diverse faculty begins to change discrimination patterns within a program) is approximately 30%. Because relatively few programs have reached this level of diversity, it may be that any effects of increasing diversity have not yet been realized.

The program-level variation, paired with the paucity of associated program-level structural factors, indicates that there are significant unmeasured cultural variables (eg, social dynamics) that account for the observed differences. These local influences, which may include factors such as resident autonomy, education versus service workloads, and program engagement in trainee wellness, must be explored to further define local influences associated with program culture. Moreover, these results imply that uniform, global strategies to improve resident wellness (eg, national changes in duty hours) may have heterogeneous effects between programs. Targeted strategies focusing on fostering a healthy learning environment and improving on areas specific to the local environment may be more successful. Development of such interventions will require both qualitative and additional quantitative investigations of the local environment. Such steps are planned in the upcoming SECOND trial, which will develop and test a best practices toolkit for improving the training environment.

This study must be interpreted in light of its limitations. First, the survey being administered in conjunction with the ABSITE examination may influence the results. However, we do not believe there would be predictable directionality to this influence (eg, individuals could be either elated or distressed at the end of the examination). Second, it is impossible to differentiate between actual differences in behavior within a program and differences in reporting rates between programs. It is possible that some programs with very healthy and open cultures would have high rates on some indicators due to resident comfort in discussing wellness and answering the survey questions on sensitive topics (eg, duty hour violations, burnout). However, we do not believe that this would happen in a systematic way that would lead to significant directional bias in the results because these programs are unlikely to uniformly share structural characteristics. Third, principal component analysis does not categorize programs a priori based on their data patterns. Thus, any group of programs with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

"poor culture" derived from these results are the consequence of defining a cutoff in the PCA results (eg, top quartile). Finally, while PCA results effectively condense data (in this case from 8 variables to 2), the PCA components do not capture all of the variance in the underlying data (in this case just over 50%). However, we believe that development of empiric, digestible composite measures to assess programs is desirable and do not believe that any variance sacrificed by this method would qualitatively change the results.

CONCLUSION

Measures of program culture such as duty hour violations, verbal/physical abuse, discrimination, and burnout are related and may cluster at the program-level. However, programs were rarely high outliers in all measurements, and high program-level prevalence of these issues was not associated with program characteristics. This implies that uniform strategies to improve the learning environment may have inconsistent results, and that interventions must account for local cultural factors and social dynamics. Qualitative and quantitative studies further examining local practices are needed to guide targeted interventions to improve trainee wellness.

Acknowledgments

This study is supported by funding from the American Board of Surgery (ABS), American College of Surgeons (ACS), and Accreditation Council for Graduate Medical Education (ACGME). RJE and DBH were supported by postdoctoral research fellowships (Agency for Healthcare Research and Quality [AHRQ] 5T32HS000078). RPM is supported by the Agency for Healthcare Quality (K12HS023011). ADY is supported by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (K08HL145139).

DISCUSSANTS

Dr Mary Klingensmith (St. Louis, MO):

Dr Ellis, you are to be congratulated along with your coauthors on a well-planned study which attempts to empirically describe surgical residency program culture and its characteristics. This work is another valuable contribution by Dr Bilimoria's group, and I'm glad to see that he and his collaborators are continuing to work in these high-impact areas.

Interestingly, wellness seems to quickly be becoming the seventh competency area for the ACGME, and perhaps that's appropriate. I do believe that your paper adds meaningfully to the discussion by giving voice to nearly all residents who took the 2018 ABSITE by surveying them at the conclusion of that exam. And your use of the principal component analysis or PCA to determine the variables that are important in both program wellness and program mistreatment will provide us with some insight for targeted intervention, as you mentioned, through your second trial.

I do have 3 questions for you.

First, I would like to take issue with your use of the conclusion of the ABSITE at the time of the survey. While I understand that you feel that program comparison is likely maintained through this, I'm concerned about using the end of the ABSITE to administer lengthy surveys regarding wellness and burnout. While the survey was described as optional, the fact

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that 99.3% of all eligible residents completed the survey suggests to me that the residents themselves did not believe this to be an optional survey.

The timing, after having endured a many-hour test, which for many is the culmination of a stressful preparation period of study, and after a test for many of whom it has become a high-stakes experience with fellowship matching success at least somewhat dependent on that outcome. I do have concerns that many of your residents were not exactly feeling happy about their career choice in that very moment. And I do believe that this could have skewed your findings and overestimated their negative feelings toward their program and our profession.

Other than to engage a captive audience, was there a compelling reason to give this survey at that time, and have you considered perhaps another time to administer such a survey?

Second, I note that you are collaborating with the ACGME on this work, and you are possibly aware that last month a little over half of all accredited general surgery programs received a citation from the Surgery RRC for duty-hour violations. It's clear that the hour violations have become a never event for the Surgery RRC, but your findings suggest that only 20% of residency programs are duty-hour violators. Can you correlate your findings with those of the RRC? And do you have confirmation of these findings that are represented in your survey? Similarly, the ACGME administers an annual program survey to residents, and I wonder if some of the survey data that you collected correlates with their findings.

Put differently, where does the truth really lie with regard to these program culture variables that you are trying to measure?

Lastly, I wonder if you can correlate program culture and patient outcome. As was mentioned in the discussion of the last paper, I wonder if departments which have poor culture and high rates of resident mistreatment are also providing poor quality care as measured by NSQIP. If this is true, I think it could be a very powerful motivator for hospitals and ACGME consortia to invest in improvements in program culture as another step in the quality improvement process.

Again, congratulations on a beautifully presented paper. I look forward to your thoughts.

Dr Ryan J. Ellis:

Thank you, Dr Klingensmith, for the kind words and the thoughtful questions.

Regarding your first question on the timing of the ABSITE survey, this is not the first generation of this survey, and there certainly have been long-standing debates on the timing of the survey. We believe that the tradeoff between nonresponse bias and potential biases that might come from the acute stress favors utilizing that exact time, in terms of getting an estimation of the global audience as opposed to just those people who open and take surveys that are distributed in different ways. Moreover, many of these measures, specifically the perceived stress scale 4 measuring stress and the Maslach Burnout Inventory, are validated instruments that have been shown to be robust acute stressors. Moreover, the fact that our program cultural measures were not different based on ABSITE scores gives me some

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

confidence in the analysis insofar as somebody who just did very well on the ABSITE did not appear to answer questions differently than someone who may have done more poorly on the ABSITE. Finally, there may also some relief or elation at the exam that may balance out the initial stress of the exam.

In terms of our response rate being 99.3% and the question of whether the survey is truly optional, there are opt-out buttons all along the survey, and written and verbal instructions at the outset that clearly tell the residents that it's optional. We have been exploring alternative times to administer similar surveys going forward because data monitoring will be paramount moving forward with the SECOND trial, but for now we believe that the post-ABSITE timing remains the best option.

Second, you asked about the possibility of linking these data with the RRC or ACGME data. First, I would like to clarify the definition of duty-hours violations used in this study.We dichotomized responses based on the number of duty-hour violations they reported, so it was not a comparison zero duty hour violations versus one or more. It was actually a comparison of less than or greater than 3 violations of the 80 hour weekly average out of the last 6 last months, so a linear comparison of our rates versus things from the RRC might be a little challenging. In terms of linking these data with the RRC or ACGME, we are actively in conversation to either retrospectively link responses or perhaps take ACGME survey questions and put them in this survey in the future to assess concordance.

Finally, you asked about potentially linking program culture with patient outcomes. This is especially notable because of previously demonstrated links between burnout and self-reported medical errors. I can say that we actually have done a similar analysis with data from the 2017 survey, in a study that is currently being written up. We compared program-level burnout with program-level NSQIP outcomes and demonstrated no statistically significant association between program burnout and failure to rescue, mortality, or death and serious morbidity. I do think that burnout as a single measure to stratify the programs might be a little narrow, and I do look forward to hopefully using these composites to do a similar analysis in the future.

Dr Barry Inabnet (New York, NY):

I have no disclosures. I would like to congratulate you on attempting to create an objective measure of residency culture.

Two questions. First, there are often many other factors at play, particularly in the era of health care mergers, acquisitions, and so forth. Could you comment on the role and influence of cultural change that can lead to leadership instability at both the department level and at the institution level that may have trickle-down impact culture on culture and residency education?

Secondly, do you have any data on the size of residency programs and its impact on culture? I hypothesize that at the extremes, that is, smaller and larger programs, there may be challenges and opportunities inherent in the number of trainees. Thank you.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dr Ryan J. Ellis:

Thank you Dr Inabnet. I'll start with the second question. Based on the size of the program, there was no difference in these cultural composites. We have done a lot of different analyses on these data, and there are some slight differences in individual variables based on program size. For example, patterns in thoughts of attrition between smaller and larger programs look a little different. But when focusing specifically on these cultural composites, and there were no significant differences.

Regarding the myriad other factors that were not captured in the survey, we have been lucky enough to continue to administer the survey at the end of the ABSITE, and most recently have added some of the institutional factors you mentioned. Things like mentorship, administrative support that the residents receive, operative autonomy, and other parts of culture that we were not able to measure in the past.

We hope to ascertain information on more structural factors, such as leadership stability, longevity of program directors, and longevity of department chairs through a separate branch of the SECOND trial which surveys program directors themselves. There, the program directors will be given the opportunity to comment on the state of their program, program leadership, and how they assess problems, and will provide data from that angle. So we will hopefully be able to incorporate a lot of those factors, but we don't have those data yet.

Dr James Korndorffer (Stanford, CA):

It's that former program director part that gives me some concern, particularly given the granularity of information you have: program type, size, location, chair, demographics, et cetera. It would take a little effort to identify the exact programs that have that culture. And considering some of the findings you identified, burnout and such have been linked to things such as increased suicide rate, do you believe that you have an ethical responsibility, quite frankly, to mention and talk to those low-performing program directors about these concerning findings to allow them to improve? Essentially, perhaps it ought to be "You can't improve what you don't know."

Dr Ryan J. Ellis:

Thank you Dr Korndorffer for that very important question. After consultation with the APDS leadership, that actually is one of the major reasons that we have planned to structure the SECOND trial the way that we have, such that both arms are given some of the most important, actionable information. Every program participating in the SECOND trial will get feedback on the global psychological status of their residents in terms of burn out and thoughts of attrition. We actually do ask about suicidal ideation directly on the survey, and all programs will get that feedback as well. So, we have been very cognizant of the sensitivity of the data and will be giving some data back to all participating sites. I think that that does help with that issue.

In terms of the data and programs being identifiable, yes, in theory, a motivated individual might be able to identify programs based on those characteristics in our dataset. But there

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are no plans to make such data public, and all the appropriate steps are taken to make sure the data are deidentified and that they are protected.

Dr Douglas E. Wood (Seattle, WA):

The corollary to that last question is that there are a lot of chairs and program directors in the room, all wondering where they are on that graph. At least I am. I imagine I am speaking for others.

Besides waiting for the Second Trial, is there a way to get feedback of where is our program in that graph? Where are our weak points? What things could we be paying attention to, to address problems within our own institution?

Dr Ryan J. Ellis:

Thank you, Dr Wood, for that question. I have talked about the survey results a few times and I often get a similar response, which is, “We need to know now, we would like to know exactly where we are now.” In terms of the trial, the interventions and all the things we are deriving from doing visits across the country and describing best practices, will be delayed temporarily. I can assure you we are moving as quickly as possible, and hope to have data fed back to programs within a year.

Besides waiting for the SECOND trial, there would be opportunities for self-assessment. We won’t be able to give final reports obviously, but every question that was used to derive these cultural composites is publicly available and is, in total, 14 or 15 questions long. The survey that we administer at the end of the ABSITE is much longer, but these composites are derived from that small 14 or 15 questions subset. So in theory, yes, it is very possible for local investigations to occur while you wait for SECOND trial data to sort of get a gestalt of where you are as a program, which you could then compare to the national distributions of program level rates that will be published in this study.

Dr Karl Bilimoria:

I just want to follow-up on that. The data are coming very soon. We hope to have them by fall 2019 for those programs that enroll in the SECOND trial. So we are moving very quickly as we did with the FIRST trial. Hopefully, you will get that answer very soon.

REFERENCES

1. Shanafelt TD, Balch CM, Bechamps GJ, et al. Burnout and career satisfaction among American surgeons. Ann Surg. 2009;250:463–471. [PubMed: 19730177]
2. Dyrbye LN, Thomas MR, Power DV, et al. Burnout and serious thoughts of dropping out of medical school: a multi-institutional study. Acad Med. 2010;85:94–102. [PubMed: 20042833]
3. Shanafelt TD, Balch CM, Dyrbye L, et al. Special report: suicidal ideation among American surgeons. Arch Surg. 2011;146:54–62. [PubMed: 21242446]
4. West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. JAMA. 2011;306:952–960. [PubMed: 21900135]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. 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. [PubMed: 30422299]
6. Vassar L ACGME seeks to transform residency to foster wellness. May 28,2015 Available at: https://www.ama-assn.org/residents-students/resident-student-health/acgme-seeks-transform-residency-foster-wellness. Accessed March 4, 2019.
7. Lomis KD, Carpenter RO, Miller BM. Moral distress in the third year of medical school; a descriptive review of student case reflections. Am J Surg. 2009;197:107–112. [PubMed: 19101252]
8. Daugherty SR, Baldwin DC Jr, Rowley BD. Learning, satisfaction, and mistreatment during medical internship: a national survey of working conditions. JAMA. 1998;279:1194–1199. [PubMed: 9555759]
9. Huang Y, Chua TC, Saw RPM, et al. Discrimination, bullying and harassment in surgery: a systematic review and meta-analysis. World J Surg. 2018;42:3867–3873. [PubMed: 29971462]
10. Frank E, Brogan D, Schiffman M. Prevalence and correlates of harassment among US women physicians. Arch Intern Med. 1998;158:352–358. [PubMed: 9487232]
11. 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. [PubMed: 24667512]
12. Nagata-Kobayashi S, Maeno T, Yoshizu M, et al. Universal problems during residency: abuse and harassment. Med Educ. 2009;43:628–636. [PubMed: 19573185]
13. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. Ann Surg. 2010;251:995–1000. [PubMed: 19934755]
14. Krug MF, Golob AL, Wander PL, et al. Changes in resident well-being at one institution across a decade of progressive work hours limitations. Acad Med. 2017;92:1480–1484. [PubMed: 28353505]
15. 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. [PubMed: 26836220]
16. Bilimoria KY, Quinn C, Dahlke AR, et al. Utilization and underlying reasons of duty hour flexibility in the flexibility in duty hour requirement for surgical trainees (FIRST) trial. J Am Coll Surg. 2017;224:118–125. [PubMed: 27884805]
17. 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. [PubMed: 26720622]
18. Maslach CJ, Susan E, Leiter. et al. Maslach Burnout Inventory Manual. 4th ed Menlo Park, CA: Mind Garden, Inc; 2016.
19. Cohen S, Williamson GM. Perceived stress in a probability sample of the United-States. Clar Symp. 1988;31–67.
20. Warttig SL, Forshaw MJ, South J, et al. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). J Health Psychol. 2013;18:1617–1628. [PubMed: 24155195]
21. McManus IC, Gordon D, Winder BC. Duties of a doctor: UK doctors and good medical practice. Quality Health Care QHC. 2000;9:14–22. [PubMed: 10848365]
22. 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. [PubMed: 12086767]
23. 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. [PubMed: 28524379]
24. Meehan PJ, Lamb JA, Saltzman LE, et al. Attempted suicide among young adults: progress toward a meaningful estimate of prevalence. Am J Psychiatry. 1992;149:41–44. [PubMed: 1728183]
25. Dahlin M, Joneborg N, Runeson B. Stress and depression among medical students: a cross-sectional study. Med Educ. 2005;39:594–604. [PubMed: 15910436]
26. Dyrbye LN, Thomas MR, Massie FS, et al. Burnout and suicidal ideation among U.S. medical students. Ann Intern Med. 2008;149:334–341. [PubMed: 18765703]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

27. Castillo-Angeles M, Watkins AA, Acosta D, et al. Mistreatment and the learning environment for medical students on general surgery clerkship rotations: what do key stakeholders think? Am J Surg. 2017;213:307–312. [PubMed: 28131325]

28. Castillo-Angeles M, Calvillo-Ortiz R, Acosta D, et al. Mistreatment and the learning environment: a mixed methods approach to assess knowledge and raise awareness amongst residents. J Surg Educ. 2019;76:305–314. [PubMed: 30318301]

29. Lussiez A, Bevins J, Plaska A, et al. General surgery resident satisfaction on cardiothoracic rotations. J Surg Educ. 2016;73:95–100. [PubMed: 26531743]

30. Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of burnout among physicians: a systematic review. JAMA. 2018;320:1131–1150. [PubMed: 30326495]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 1.

Surgical Residency Program-level Characteristics

	Overall n (%)
Overall	260
Location	
Northeast	85 (32.7)
Southeast	56 (21.5)
Midwest	55 (21.2)
Southwest	28 (10.8)
West	36 (13.8)
Program type	
Academic	120 (46.1)
Community or military	138 (53.1)
Unknown	2 (0.8)
Program size (number of residents)	
≤25	115 (44.2)
26–37	61 (23.5)
38–51	50 (19.2)
≥52	34 (13.1)
Percentage of female residents	
Quartile 1 (≤31.8%)	66 (25.4)
Quartile 2 (32.3%–38.9%)	65 (25.0)
Quartile 3 (39.0%–46.7%)	70 (26.9)
Quartile 4 (≥46.8%)	59 (22.7)
Department chair	
Male	198 (76.2)
Female	22 (8.5)
Unknown	40 (15.4)
General surgery program director	
Male	209 (80.4)
Female	51 (19.6)
Program ABSITE performance Quartile 1	
Quartile 1 (≤42.0)	66 (25.4)
Quartile 2 (42.1–48.4)	66 (25.4)
Quartile 3 (48.5–55.4)	63 (24.2)
Quartile 4 (≥55.5)	65 (25.0)
Percentage of female faculty	
Quartile 1 (≤17.9%)	60 (23.1)
Quartile 2 (18.0%–22.8%)	59 (22.7)
Quartile 3 (22.9%–26.4%)	61 (23.5)
Quartile 4 (≥26.5%)	58 (22.3)

	Overall n (%)
Unknown	22 (8.5)
Percentage of non-White faculty	
Quartile 1 (≤26.7%)	60 (23.1)
Quartile 2 (26.8%–32.9%)	60 (23.1)
Quartile 3 (33.6%–42.0%)	62 (23.9)
Quartile 4 (≥42.1%)	56 (21.5)
Unknown	22 (8.5)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2.

Program-level Variation in Reported Exposures

	Program-level Rate of Residents Reporting
Frequent duty hour violations	
Median (IQR)	11.9% (5.8%–19.2%)
Range	0.0%–41.7%
Gender discrimination *	
Median (IQR)	66.7% (50.0%–76.8%)
Range	0.0%–100.0%
Verbal/physical abuse	
Median (IQR)	30.0% (20.8%–38.3%)
Range	0.0%–66.7%
Sexual harassment *	
Median (IQR)	16.7% (9.1%–28.6%)
Range	0.0%–100.0%
Severe stress	
Median (IQR)	13.3% (8.3%–19.7%)
Range	0.0%–42.9%
Dissatisfaction with being a surgeon	
Median (IQR)	22.9% (16.5%–29.4%)
Range	0.0%–53.8%
Burnout	
Median (IQR)	36.6% (28.6%–46.9%)
Range	6.3%–73.1%
Thoughts of attrition	
Median (IQR)	11.3% (6.3%–16.3%)
Range	0.0%–66.7%

* Among female residents (n = 2935).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 3.

Principal Component Analysis Component Loading

Cultural Factor (Program-level)	Component 1 Wellness	Component 2 Negative Exposures
Frequent duty hour violations	0.313	0.207
Gender discrimination *	0.085	0.513
Verbal/physical abuse	0.237	0.398
Sexual harassment *	−0.143	0.670
Severe stress	0.429	0.025
Dissatisfaction with being a Surgeon	0.505	−0.253
Burnout	0.399	0.125
Thoughts of attrition	0.470	−0.086

* Component loadings presented as values between −1.0 and +1.0, with higher absolute values indicating higher influence on the value of the component.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 4.

Differences in Psychosocial Stressors and Outcomes Between Low- and High-performing Programs Across Two Dimensions of Program Culture (n = 260)

	Component 1 Program Wellness Median % (IQR)			Component 2 Program Negative Exposures Median % (IQR)			Worst Quartile of Both Components Median % (IQR)		
	Top Quartile n = 65	Other Programs	*P* Value	Top Quartile n = 65	Other Programs	*P* Value	Top Quartile* n = 26	Other Programs	*P* Value
Frequent duty hour violations	21.4 (13.8–28.6)	9.3 (4.5–15.6)	<0.001	17.2 (10.3–25.0)	10.5 (5.0–17.4)	<0.001	23.3 (17.1–30.4)	11.1 (5.3–17.9)	<0.001
Gender discrimination*	69.2 (62.5–81.3)	63.2 (50.0–75.0)	<0.001	80.0 (69.2–90.0)	60.0 (45.5–70.8)	<0.001	78.7 (68.4–88.9)	64.5 (50.0–75.0)	<0.001
Verbal/physical abuse	36.8 (28.6–43.9)	27.3 (17.1–36.0)	<.001	38.5 (32.6–45.5)	25.9 (18.2–35.3)	<0.001	41.6 (35.1–48.0)	28.6 (20.0–36.4)	<0.001
Sexual harassment*	16.7 (10.0–30.0)	16.7 (8.7–28.6)	0.514	35.7 (27.3–44.4)	14.3 (0.0–20.0)	<0.001	30.8 (26.3–40.0)	16.7 (7.7–25.0)	<0.001
Severe stress	22.2 (17.7–26.9)	11.1 (6.7–15.8)	<0.001	16.1 (10.5–24.1)	12.5 (7.7–18.2)	0.003	23.9 (17.9–28.3)	12.5 (7.7–18.2)	<0.001
Dissatisfaction with being a surgeon	32.0 (26.9–41.7)	21.1 (14.3–26.7)	<0.001	22.7 (15.9–29.0)	22.9 (16.7–30.0)	0.708	29.2 (24.1–36.0)	22.2 (15.9–29.2)	<0.001
Burnout	48.8 (42.9–56.1)	33.3 (25.0–40.0)	<0.001	46.8 (33.3–55.0)	34.8 (26.7–42.9)	<0.001	54.9 (46.3–60.0)	35.0 (26.9–44.4)	<0.001
Thoughts of attrition	19.4 (13.8–28.6)	9.1 (5.6–14.0)	<0.001	12.8 (7.0–20.0)	11.1 (6.1–15.6)	0.061	21.6 (13.8–26.1)	10.8 (5.9–15.4)	<0.001
Poor overall wellbeing†	60.0 (56.1–66.7)	41.7 (33.3–50.0)	<0.001	56.3 (44.2–61.5)	44.2 (33.3–54.5)	<0.001	60.8 (58.6–66.7)	44.4 (34.5–55.6)	<0.001
Suicidal thoughts†	5.4 (2.6–8.7)	3.8 (0.0–6.7)	0.004	5.3 (2.3–8.7)	3.7 (0.0–6.7)	0.004	7.7 (3.4–9.7)	4.0 (0.0–6.9)	0.004

*Top quartile of both component 1 and component 2.

†Not included in derivation of PCA components, used to validate component separation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 5.

Factors Associated With Program Culture (Wellness, Negative Exposures, or Worst Quartile on Both) (n = 260 Programs)

	Worst Quartile Wellness		Worst Quartile Negative Exposures		Overall Program Culture (Worst Quartile on Both Components)	
Overall Rate	**25.0%**	***P* Value**	**25.0%**	***P* Value**	**10.0%**	***P* Value**
Location		0.137		0.043		0.354
Northeast	24.7%		34.1%		8.2%	
Southeast	23.2%		16.1%		8.9%	
Midwest	18.2%		16.4%		7.3%	
Southwest	21.4%		21.4%		10.7%	
West	41.7%		33.3%		19.4%	
Program type		0.426		0.172		0.105
Academic	27.5%		28.3%		13.3%	
Community or military	23.2%		21.0%		7.3%	
Program size (number of residents)		0.024		0.935		0.110
≤25	16.5%		23.5%		5.2%	
26–37	36.1%		24.6%		16.4%	
38–51	26.0%		28.0%		12.0%	
≥52	32.4%		26.5%		11.8%	
Percentage of female residents		0.007		0.543		0.011
Quartile 1 (≤31.8%)	15.2%		27.3%		7.6%	
Quartile 2 (32.3%–38.9%)	16.9%		18.5%		1.5%	
Quartile 3 (39.0%–46.7%)	31.4%		28.6%		12.9%	
Quartile 4 (≥46.8%)	37.3%		25.4%		18.6%	
Department chair		0.651		0.757		0.313
Male	27.3%		25.8%		11.6%	
Female	31.8%		22.7%		4.6%	
General surgery program director		0.058		0.928		0.639
Male	22.5%		24.9%		9.6%	
Female	35.3%		25.5%		11.8%	
Program ABSITE Performance		0.706		0.290		0.930

Ann Surg. Author manuscript; available in PMC 2020 October 01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	Worst Quartile Wellness		Worst Quartile Negative Exposures		Overall Program Culture (Worst Quartile on Both Components)	
Overall Rate	**25.0%**	***P* Value**	**25.0%**	***P* Value**	**10.0%**	***P* Value**
Quartile 1 (≤42.0)	25.8%		28.8%		9.1%	
Quartile 2 (42.1–48.4)	28.8%		22.7%		12.1%	
Quartile 3 (48.5–55.4)	25.4%		17.5%		9.5%	
Quartile 4 (≥55.5)	20.0%		30.8%		9.2%	
Percentage of Female Faculty		0.292		0.360		0.110
Quartile 1 (≤17.9%)	20.0%		23.3%		6.7%	
Quartile 2 (18.0%–22.8%)	20.3%		17.0%		5.1%	
Quartile 3 (22.9%–26.4%)	31.2%		24.6%		13.1%	
Quartile 4 (≥26.5%)	31.0%		31.0%		17.2%	
Percentage of Non-White Faculty		0.047		0.578		0.696
Quartile 1 (≤26.7%)	20.0%		18.3%		8.3%	
Quartile 2 (26.8%–32.9%)	35.0%		25.0%		13.3%	
Quartile 3 (33.6%–42.0%)	16.1%		29.0%		8.1%	
Quartile 4 (≥42.1%)	32.1%		23.2%		12.5%	