

A National Survey of Preliminary versus Categorical General Surgery Residents' Personal and Professional Experiences

Sarah Lund

University of Michigan Medicine: University of Michigan Michigan Medicine

Kaiser Sadiq

George Washington University Medical Center

Xinyi Luo

Tulane University Health Sciences Center: Tulane University School of Medicine

Rebecca Moreci

Louisiana State University

Justine Broecker

justinebroecker@gmail.com

UCSF: University of California San Francisco <https://orcid.org/0000-0002-2134-1600>

Research Article

Keywords:

Posted Date: July 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10035330/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose

Preliminary general surgery residents represent a substantial share of surgical trainees but remain understudied. We compared the personal and professional training experiences—spanning quality of life, mentorship, professional trajectory, research opportunity, workplace mistreatment, and burnout—of preliminary and categorical general surgery residents in a national sample, and examined whether the two preliminary pathways (designated and non-designated) differ from one another.

Methods

In a cross-sectional anonymous electronic survey fielded nationally over two cycles (2024 and 2025), PGY-1 and PGY-2 categorical and preliminary general surgery residents rated six prespecified, co-equal outcomes—quality-of-life, trajectory, and mentorship satisfaction; burnout (two-item Maslach Burnout Inventory); workplace mistreatment (a status-referenced composite); and perceived research disadvantage. Comparisons used the Benjamini-Hochberg false-discovery-rate correction as a robustness check, and multivariable logistic regression adjusted for gender, international medical graduate status, citizenship, race, and region. Among preliminary residents, we examined alignment between reported initial goals and anticipated next-year trajectory as one element of professional satisfaction. Comparisons of this alignment between demographic subgroups (for example, by medical-school background) were exploratory and not powered a priori; given small subgroup sizes, they are reported descriptively as hypothesis-generating.

Results

Of 152 respondents, 81 were categorical and 69 preliminary. Preliminary residents reported lower satisfaction with quality of life (45% vs 68%), professional trajectory (58% vs 79%), and mentorship (25% vs 57%); higher workplace mistreatment (93% vs 47%); and greater perceived research disadvantage (53% vs 35%); all five differences were significant and held under false-discovery-rate correction. After demographic adjustment, preliminary status remained independently associated with all five, especially mistreatment. Burnout did not differ by status. The two preliminary pathways did not differ significantly (an underpowered comparison), both faring worse than categorical residents.

Conclusions

Preliminary surgical residents reported worse personal and professional experiences than categorical residents, especially workplace mistreatment, mentorship, and academic opportunity, persisting after

demographic adjustment. Designated and non-designated residents reported similar experiences in an underpowered comparison; how their needs differ warrants further study.

Introduction

Preliminary general surgery residency positions remain a major component of the surgical training workforce; in the 2025 Main Residency Match, approximately 1,778 categorical and 1,194 preliminary PGY-1 general surgery positions were offered nationally[1]. Preliminary residents are a heterogeneous group—designated residents are tied to a planned transition into another specialty, while non-designated residents have no such commitment—and are disproportionately international medical graduates and non-citizens. The preliminary year has long been viewed in two opposing ways—as a golden opportunity to enter surgery, or as a form of indentured servitude that mainly serves institutional needs—yet despite their numbers and central role in surgical service, how preliminary residents experience training relative to their categorical peers is largely unknown. Whether the position serves the trainees who fill it or the system that relies on them is hard to judge without first describing that experience.

The existing evidence defines the problem in two ways. Burnout, mistreatment, and inadequate support are well documented among surgical residents generally[2, 3, 4, 5], yet little is known about how they affect preliminary residents specifically. Prior work on this population has largely tracked career destinations[6, 7, 8] rather than the training experience; the one substantial comparison of training experience, by Sullivan and colleagues, drew on a 2009 survey and focused on professional and training attitudes—program satisfaction, collegiality, and operative experience—rather than personal wellbeing or workplace mistreatment, which it did not examine as distinct constructs[9]. No contemporary national study has characterized the personal and professional experiences of preliminary residents, leaving it unknown whether their experiences differ from those of categorical residents, and whether any differences are widespread or local.

This gap matters because the quality of the training environment varies substantially across programs. Programs have begun to invest in resident well-being initiatives[10], but rates of adverse experiences such as mistreatment and burnout differ markedly between programs[11]—so whether a resident trains in a supportive environment depends substantially on which program they enter. National, accreditation-supported efforts have emerged to address this variation by benchmarking programs against comparative data, as in the FIRST[12] and SECOND trials conducted with the ACGME, ACS, and ABS[11]; in these, however, preliminary status has not been a variable of interest. Comparable national data on preliminary residents do not yet exist, and characterizing their experiences at scale is therefore a necessary first step.

We therefore conducted a national survey of preliminary and categorical general surgery residents to provide the first contemporary comparison of their personal and professional experiences—extending beyond the training-satisfaction focus of earlier national work to encompass wellbeing, mentorship, and workplace mistreatment. To organize what “training experience” encompasses, we drew on the learning-

environment framework of Gruppen and colleagues[13], mapping our domains to its personal (quality of life, burnout), social (mentorship, mistreatment), and organizational (research, trajectory) components. We asked whether preliminary and categorical residents differ across these domains nationally and, secondarily, whether the designated and non-designated pathways differ. Because professional satisfaction depends partly on whether a trainee is progressing toward their intended career, we also examined, among preliminary residents, how their anticipated next-year trajectory aligned with the goals they reported on entering residency. The study was not powered to detect differences in this alignment between demographic subgroups, so such comparisons are exploratory.

Methods

Study Design and Participants

An anonymous, voluntary electronic survey was distributed nationally to PGY-1 and PGY-2 general surgery residents during May–June 2024 and 2025 through individual residency programs and professional surgical societies. Both categorical and preliminary residents were eligible. Preliminary residents were further categorized as designated or non-designated by self-report; this classification was applied only to residents whose current position was preliminary, yielding 18 designated and 47 non-designated. Because it was disseminated anonymously through multiple programs and society channels rather than to an enumerated list, the number of residents reached—and thus a precise response rate—could not be determined; 152 residents responded.

Because the survey was fielded in two consecutive years, the questionnaire asked respondents whether they had completed it in a prior year, allowing repeat participation to be identified and tracked rather than excluded. Five respondents reported prior-year completion; given this small number, these responses were retained and all responses were treated as independent.

The study was approved by the central Institutional Review Board; informed consent was implied by voluntary completion of the anonymous survey. Two respondents who did not complete the consent item or specify training position were excluded from comparative analyses.

Conceptual Framework

To organize and interpret the measured domains, we drew on the learning-environment framework of Gruppen and colleagues, which characterizes the training environment in terms of personal, social, and organizational components[13]. This framework is descriptive and organizing rather than explanatory; the specific mapping of our domains onto its components is given under Survey Measures. Social identity theory[14] is invoked only in the Discussion, as one interpretive lens on the workplace mistreatment findings.

Survey Measures

Six outcomes were specified a priori as co-equal endpoints, with no single primary outcome; they were organized by the learning-environment components described above. The personal component comprised (1) quality of life satisfaction and (2) burnout; the social component comprised (3) mentorship satisfaction and (4) workplace mistreatment; and the organizational component comprised (5) perceived access to research opportunities and (6) professional trajectory satisfaction. Each is defined below.

Prior to fielding, the instrument was piloted with preliminary and categorical residents and surgical faculty to assess clarity and face and content validity, and items were revised based on their feedback; pilot responses were not included in the analytic sample.

Burnout was assessed with the validated two-item Maslach Burnout Inventory, whose emotional-exhaustion and depersonalization items correlate strongly with the full subscales[15, 16]; quality of life used a single global self-rating shown to perform comparably to multi-item instruments[17]. Mentorship, trajectory, and research access used investigator-developed Likert items informed by their respective constructs[18].

Quality of life, professional trajectory, and mentorship satisfaction were dichotomized as satisfied (satisfied or very satisfied) versus not satisfied (neutral, unsatisfied, or very unsatisfied).

The primary burnout outcome was prespecified as symptoms “most” or “all of the time”; broader thresholds were examined only as prespecified sensitivity analyses and labeled as such.

Workplace mistreatment was assessed with five items asking how often in the past year the respondent experienced specific negative interactions “because of my preliminary vs categorical status” (being treated unwelcomingly, having one’s opinion overlooked, being ignored, having one’s work assumed inferior, and being treated differently by status). Items were adapted from validated workplace-incivility instruments—the Workplace Incivility Scale and the Uncivil Workplace Behavior Questionnaire—reworded to reference training status[19, 20]. Identical items were administered to both groups; responses were frequency-based (0, 1–3, 4–6, 7–9, or ≥ 10 times), and the primary definition was endorsement of any item at least once. A higher-frequency definition (endorsement of any item ≥ 4 –6 times) was analyzed as a pre-specified sensitivity definition, because identical items were administered to both groups and a substantial minority of categorical residents endorsed at least one item. Formal measurement invariance across training-status groups was not tested, because the per-group sample sizes (69 preliminary, 81 categorical) are too small to support a stable multi-group factor analysis of a five-item ordinal scale; an underpowered test would risk a misleading null.

Reduced research access (referred to throughout as “perceived research disadvantage”) was defined as agreeing or strongly agreeing that preliminary status negatively impacted research opportunities.

Because self-reported disadvantage can be dismissed as one group's subjective complaint, we also asked whether categorical residents independently recognized the same disadvantages, providing an external check: agreement would corroborate that the disadvantage is real and visible, whereas a gap would indicate it is felt mainly by those experiencing it and underrecognized by others. For four domains (rotations, research opportunities, sense of community, and mentorship), parallel items asked all respondents—preliminary and categorical—to rate how much preliminary status affects preliminary residents, on the same five-point agree/disagree scale; comparing the two groups' ratings yielded a cross-group perception measure, with agreement defined as "agree" or "strongly agree." These items were investigator-developed, informed by the content of validated workplace-climate and incivility instruments.

Two additional items asked of preliminary residents captured career intentions: their initial goal on entering residency (a categorical general surgery position, another categorical surgical specialty, a categorical non-surgical specialty, or other) and their plan for the following year (categorical general surgery, another categorical surgical or non-surgical specialty, a further preliminary year, research, or other). We defined goal-trajectory alignment as a match between a resident's stated initial goal and their reported next-year destination. Both reflect self-reported intentions rather than verified appointments.

Missing Data

Item-level missingness was low and approximately uniform (1.3% for most items, up to 3.9% for the research-access item), consistent with incidental non-response rather than a systematic mechanism. Two respondents (1.3%) did not record a training position and were excluded from comparative analyses. Analyses used available cases per outcome; given the few missing values, imputation was not performed, and per-comparison denominators are reported in the tables.

Statistical Analysis

Descriptive statistics were calculated for all variables. Group comparisons used Pearson chi-square tests, with Fisher's exact test where any expected cell count was below five.

The study was designed a priori to detect a between-group difference of approximately 20 percentage points at 80% power and a two-sided alpha of 0.05, corresponding to a target of roughly 50 to 95 respondents per group depending on baseline rate; the survey was fielded over two successive cycles to accrue toward this target. To aid interpretation, we also examined the study's sensitivity to the observed effects. For the primary comparison (81 categorical, 69 preliminary), the sample gave approximately 80–100% power for the observed differences in quality of life, trajectory, mentorship, and mistreatment, but was underpowered for burnout (~ 30%) and research disadvantage (~ 60%). For the designated-versus-non-designated subgroup (20 vs 48), power was low for all outcomes (< 20%); detecting a 15-point difference at 80% power would require roughly 120 designated residents. Subgroup comparisons are therefore exploratory and not treated as evidence of equivalence.

Univariate logistic regression used preliminary versus categorical status as the exposure. Multivariable models adjusted for a single prespecified covariate set applied uniformly across all six outcomes—gender, international medical graduate (IMG) status, citizenship, race, and geographic region—each selected because it both differed between groups and has an established association with the outcomes, the two conditions defining a confounder. In large national samples, women report higher rates of mistreatment and burnout, residents report racial discrimination, and mistreatment varies across programs[18]; IMG and citizenship status are likewise associated with differences in trainee experience. Applying the same set to every outcome keeps the models comparable; a secondary change-in-estimate screen ($\geq 10\%$ change in the preliminary-status estimate) yielded consistent conclusions. Because mistreatment was highly prevalent (93%), its odds ratios are large with wide confidence intervals, so we emphasize the association’s direction and consistency over its magnitude.

Because multiple residents may have come from the same residency program, and residents in a given program share the same training environment, their responses may not be fully independent; the survey also spanned two cycles. Accounting for this statistically would require knowing each respondent’s program, but—given the sensitivity of the mistreatment items and the small number of preliminary residents at any single program—program was deliberately not collected in order to protect anonymity, so this program-level clustering could not be modelled.

The six outcomes were prespecified as co-equal endpoints and all reported regardless of result; we therefore treat each as its own hypothesis rather than imposing a multiplicity penalty on a small, predefined family. As a robustness check, we additionally applied the Benjamini-Hochberg false-discovery-rate procedure across the six outcomes, reporting both raw and adjusted (q) values; demographic comparisons (Table 1) were descriptive and not corrected, and outcomes with q near 0.05 are flagged as more fragile.

Table 1
Respondent characteristics by training status.

Characteristic	Categorical (n = 81)	Preliminary (n = 69)	p-value
PGY-1	46 (57%)	50 (72%)	0.046
Male sex	29 (36%)	40 (58%)	0.007
White race	56 (69%)	43 (62%)	0.380
Hispanic ethnicity	9 (11%)	10 (14%)	0.535
International medical graduate	9/80 (11%)	26 (38%)	< 0.001
Non-US citizen	10/80 (12%)	22 (32%)	0.004
Academic program	79 (98%)	68 (99%)	1.000
Region			0.613
Midwest	25 (31%)	22 (32%)	
Northeast	18 (22%)	14 (20%)	
Southeast	20 (25%)	22 (32%)	
Southwest	10 (12%)	9 (13%)	
Northwest	5 (6%)	1 (1%)	
Other	3 (4%)	1 (1%)	
<p><i>Values are n (%). Denominators are 81 (categorical) and 69 (preliminary) unless otherwise noted. "Non-US citizen" denotes any respondent not reporting US citizenship (including permanent residents and those who preferred not to answer); a narrower visa-holder-only definition yields 26% versus 10%. P-values from Pearson chi-square, or Fisher's exact test where any expected cell count was < 5 (academic program). Two respondents who did not specify training status were excluded.</i></p>			

The designated and non-designated subgroups are distinct NRMP appointment categories, prespecified rather than data-driven. We report rates for all three groups descriptively and test the designated-versus-non-designated contrast directly with Fisher exact tests; a single three-group test is not presented, as it would chiefly re-express the primary comparison. Multivariable subgroup models were not estimated, because the small designated subgroup (n = 18, among whom mistreatment was universal) would not support stable adjusted estimates; these comparisons are interpreted as exploratory and hypothesis-generating.

For the workplace mistreatment composite, content validity rested on its derivation from validated incivility instruments (above); internal-structure validity was assessed by ordinal Cronbach's alpha and, because the items are ordinal, by exploratory factor analysis of the polychoric correlation matrix with KMO and Bartlett tests, on the 150 complete cases. For adjustment, IMG status was defined by non-US

medical school, non-US citizenship included visa and green-card holders, and race was dichotomized as white versus non-white. Analyses used Python.

A generative AI tool (Claude, Anthropic) was used to assist with editing manuscript text. All data, statistical results, and references were independently verified by the authors, who reviewed and edited all content and take full responsibility for the work.

Results

Respondent Characteristics

A total of 152 residents responded, including 81 categorical residents (53%) and 69 preliminary residents (45%); 2 respondents (1%) did not specify training position and were excluded from comparative analyses. Among the 69 preliminary residents, 18 (26%) were designated, 47 (68%) were non-designated, and 4 (6%) did not specify a designation; designation reflects self-reported current position. Preliminary residents were significantly more likely to be male (58% vs 36%, $p = 0.007$), non-US graduates (38% vs 11%, $p < 0.001$), and non-US citizens (32% vs 12%, $p = 0.004$). No significant differences were observed in race or geographic region (Table 1). Notably, 12 of the 81 current categorical residents (15%) reported a preliminary initial match (10 non-designated, 2 designated), indicating that a meaningful share had converted from a preliminary year into their categorical position.

Resident Outcomes

Compared with categorical residents, preliminary residents reported significantly lower quality of life, professional trajectory, and mentorship satisfaction, higher workplace mistreatment, and greater perceived research disadvantage (Table 2). These five prespecified outcomes were each significant on their own terms, and all five also remained significant after Benjamini-Hochberg correction across the six-outcome family (all $q \leq 0.031$), with perceived research disadvantage the most fragile ($q = 0.031$). Burnout did not differ significantly under either the primary definition (43% vs 32%, $p = 0.18$) or a broader threshold (90% vs 80%, $p = 0.11$).

Table 2
Personal and professional outcomes by training status.

Outcome	Categorical	Preliminary	p-value	BH q
Quality of life satisfaction	55/81 (68%)	31/69 (45%)	0.005	0.008
Professional trajectory satisfaction	64/81 (79%)	40/69 (58%)	0.005	0.008
Mentorship satisfaction	46/81 (57%)	17/69 (25%)	< 0.001	< 0.001
Workplace mistreatment (any, $\geq 1\times$)	38/81 (47%)	64/69 (93%)	< 0.001	< 0.001
Perceived research disadvantage	27/78 (35%)	36/68 (53%)	0.026	0.031
Burnout (most/all of the time)	26/81 (32%)	29/68 (43%)	0.184	0.184
Burnout (sometimes or more)	65/81 (80%)	61/68 (90%)	0.111	—
<p><i>Values are n/N (%). Satisfaction outcomes dichotomized as satisfied/very satisfied versus other. The any-endorsement mistreatment composite is the primary definition; the $\geq 4-6\times$ composite (Table 3) is a prespecified sensitivity definition. Raw p-values from Pearson chi-square. BH q = Benjamini-Hochberg false-discovery-rate-adjusted value across the six prespecified co-equal outcomes (the broader burnout definition is a sensitivity analysis and not part of the family). Five of six outcomes remain significant at $q < 0.05$.</i></p>				

Cross-Group Perceptions of Preliminary Disadvantage

Both groups rated the degree to which preliminary status affects preliminary residents across four domains, allowing comparison of how each group perceived the same disadvantage. Preliminary residents consistently perceived greater impact than categorical residents did: they more often agreed that preliminary status adversely affects mentorship (67% vs 43%, $p = 0.005$), sense of community (62% vs 46%, $p = 0.048$), and research opportunities (53% vs 35%, $p = 0.030$); ratings of effects on rotations were similar between groups (72% vs 65%, $p = 0.38$). Thus, for three of four domains, the group experiencing preliminary status perceived its disadvantages more acutely than the comparison group perceived them.

Reliability and Validity Assessment

The workplace mistreatment composite demonstrated excellent internal consistency (ordinal Cronbach $\alpha = 0.94$ across the five frequency-coded items). Because the items are ordinal, construct validity was assessed using the polychoric correlation matrix: KMO testing indicated strong sampling adequacy (0.91), and Bartlett's test of sphericity was significant ($\chi^2 = 977$, $df = 10$, $p < 0.001$), supporting factorability. Exploratory factor analysis demonstrated a coherent one-factor structure, with the first factor accounting for the large majority of common variance and strong standardized loadings across all five items (0.89–0.96), supporting construct validity.

Multivariable Logistic Regression

After adjustment for the prespecified covariate set (gender, IMG status, citizenship, race, and geographic region), preliminary status remained independently associated with (Table 3):

Table 3
Univariate and multivariable association of preliminary status with each outcome.

Outcome (preliminary vs categorical)	Univariate OR (95% CI)	Adjusted OR (95% CI)
Quality of life satisfaction	0.39 (0.20–0.75)	0.30 (0.14–0.64)
Professional trajectory satisfaction	0.37 (0.18–0.75)	0.41 (0.19–0.91)
Mentorship satisfaction	0.25 (0.12–0.50)	0.19 (0.09–0.44)
Perceived research disadvantage	2.18 (1.13–4.22)	2.45 (1.13–5.35)
Workplace mistreatment (any, $\geq 1\times$)	14.5 (5.3–39.7)	15.8 (5.2–48.3)
Workplace mistreatment ($\geq 4-6\times$)	4.75 (2.38–9.48)	4.06 (1.90–8.67)
Burnout (most/all of the time)	1.53 (0.79–2.97)	1.71 (0.81–3.61)
<i>OR, odds ratio. All estimates are from standard logistic regression; adjusted models include gender, international medical graduate status, citizenship, race (white vs non-white), and geographic region. Because workplace mistreatment was endorsed by 93% of preliminary residents, its odds ratios are large with wide confidence intervals and should be interpreted as indicating direction rather than precise magnitude.</i>		

- lower quality of life satisfaction (aOR 0.30, 95% CI 0.14–0.64)
- lower professional trajectory satisfaction (aOR 0.41, 95% CI 0.19–0.91)
- lower mentorship satisfaction (aOR 0.19, 95% CI 0.09–0.44)
- markedly increased workplace mistreatment (adjusted OR ~ 15 , 95% CI excludes 1; estimate imprecise owing to the high prevalence)
- greater perceived research disadvantage (aOR 2.45, 95% CI 1.13–5.35)

In the secondary per-outcome change-in-estimate screen, the covariates that materially altered the estimate differed by outcome: quality of life and research access required no adjustment; professional trajectory was confounded chiefly by IMG status; mentorship by sex and citizenship; burnout by sex; and mistreatment by IMG status and citizenship. The uniform and outcome-specific models produced consistent conclusions.

Burnout was not independently associated with preliminary status. The mistreatment odds ratio is large but imprecise because mistreatment was endorsed by nearly all preliminary residents; we therefore emphasize its direction and consistency rather than the exact value. Under a higher-frequency sensitivity definition (any item $\geq 4-6$ times; prevalence 67% vs 30%, $p < 0.001$), the association remained strong and significant (adjusted OR ~ 4 , 95% CI excludes 1), confirming that the mistreatment finding is not an artifact of the lenient primary threshold.

Trajectory Analysis Across Preliminary Pathways

Self-reported next-year intentions confirmed that preliminary residents pursue heterogeneous paths rather than a single route forward (Table 5): plans spanned categorical general surgery, another categorical surgical or non-surgical specialty, a further preliminary year, and research. This heterogeneity is consistent with the structure of preliminary training—non-designated residents hold no confirmed onward position, while designated residents are contracted toward another specialty—and underscores that preliminary residents are not a single group uniformly seeking categorical general surgery. Because designation status and next-year plans were both self-reported, we report these intentions descriptively and do not draw inferential comparisons between pathways from them. A small number of responses were internally inconsistent: a few residents who identified as designated—meaning they hold a confirmed position in another specialty—nonetheless reported entering a categorical general surgery position the following year, which should not occur by definition. The most likely explanation is misinterpretation of the designation terminology or misclassification of training status rather than a genuine plan, and these few responses do not affect the primary preliminary-versus-categorical findings.

Table 5
Self-reported next-year intentions of preliminary residents.

Intended next-year position	All preliminary (n = 69)
Categorical general surgery position	25 (36%)
Other categorical non-surgical specialty	14 (20%)
Continued (non-designated) preliminary year	10 (14%)
Other categorical surgical specialty	9 (13%)
Other	8 (12%)
Research position	3 (4%)
<i>Values are n (%) of all preliminary residents (n = 69) reporting each intended next-year position. “Categorical general surgery position” denotes securing a categorical general surgery spot for the following year. Intentions are shown for preliminary residents overall rather than split by designation, because designation status and stated next-year plans were self-reported and a small number of responses were internally inconsistent. Percentages may not sum to 100 owing to rounding. Preliminary residents pursued heterogeneous paths, with only about a third intending to enter a categorical general surgery position.</i>	

Initial career goals largely aligned with these next-year plans, supporting the validity of the intentions measure. Among preliminary residents who reported a specific categorical goal on entering residency, 58% (34/59) reported entering that same type of position the following year. Attainment varied by goal: of those who wanted categorical general surgery, 63% (22/35) reported entering it, and of those who wanted another categorical surgical specialty, 40% (8/20) reported entering that specialty. Goal attainment differed by background. For any categorical position, international and US graduates entered residency with comparable goals (92% vs 81%, $p = 0.30$), yet US graduates were substantially more likely

to report securing one the following year (86% vs 42%, $p < 0.001$). For categorical general surgery specifically, among the smaller subgroup who reported that goal, US graduates were again more likely than international graduates to report securing it (13/14, 93% vs 9/21, 43%; Fisher exact $p = 0.004$). Because this latter comparison rested on small subgroups, was not powered a priori, and relied on self-reported plans, it should be regarded as hypothesis-generating rather than confirmatory.

Table 4 displays outcome rates for categorical residents and the two preliminary pathways. Both pathways sat well below categorical residents on the favorable outcomes and above them on mistreatment and research disadvantage, consistent with the primary findings. The two pathways, however, resembled each other: in direct comparisons, designated and non-designated residents did not differ significantly on any outcome (all $p \geq 0.59$) or demographic characteristic (all $p > 0.40$), with between-pathway gaps averaging under five percentage points. Because the designated subgroup was small ($n \approx 20$) and underpowered ($< 20\%$ power across outcomes), these comparisons are inconclusive rather than evidence of equivalence; we present the three groups descriptively rather than as a single three-group test, which would chiefly re-express the primary preliminary-versus-categorical comparison.

Table 4
Trajectory analysis: outcomes across categorical and the two preliminary pathways.

Outcome	Categorical (n = 81)	Designated prelim (n = 18)	Non-designated prelim (n = 47)	D vs ND p [¶]
Quality of life satisfaction	68%	44%	47%	0.79
Professional trajectory satisfaction	79%	67%	55%	0.59
Mentorship satisfaction	57%	17%	28%	1.00
Workplace mistreatment (any, $\geq 1\times$)	47%	100%	91%	1.00
Perceived research disadvantage	35%	65%	49%	0.79
Burnout (most/all of the time)	32%	39%	41%	0.79
<i>Values are % endorsing each outcome and are descriptive. [¶]The p-value column reports the direct comparison of designated versus non-designated preliminary residents (Fisher exact); none differed significantly. The categorical column is shown for reference—its consistent separation from both preliminary pathways reflects the primary preliminary-versus-categorical findings reported above—but no three-group test is presented, as it would chiefly re-express that primary comparison. Four preliminary residents who did not specify a designation are not shown.</i>				

Discussion

Across this national sample, preliminary status was independently associated with lower satisfaction with quality of life, mentorship, and trajectory, more mistreatment, and greater perceived research disadvantage—a less supportive training environment on nearly every dimension measured. These results echo and extend Sullivan and colleagues, who more than a decade ago found preliminary residents less satisfied with their programs, less collegially integrated, and less satisfied with their operative experience[9]; their work centered on professional and training attitudes, whereas our data extend the comparison into quality of life, mentorship, research opportunity, and interpersonal mistreatment. A sobering observation is one of continuity. Resident wellbeing has drawn sustained national attention over the past decade—the 2017 ACGME Common Program Requirements[21] and initiatives such as the FIRST and SECOND trials[11, 12]—yet none examined preliminary status. Against that backdrop, the preliminary-categorical disadvantage Sullivan described over a decade ago persists in our data across multiple domains, even as the broader field has turned to trainee wellbeing.

Mistreatment was the standout finding. Its adjusted odds ratio is large but imprecise (mistreatment was near-universal among preliminary residents), so we emphasize direction over magnitude; it persisted after demographic adjustment, pointing to training status itself, not demographic identity, as the driver. Social identity theory offers an explanation[14]: preliminary status functions as a salient social category, and the items—being treated as inferior, ignored, or overlooked because of that status—describe classic status-based out-group treatment.

The five mistreatment items correlated strongly and reflected a single dimension, supporting their combination into one measure. One caveat: because the items ask about being treated poorly specifically because of one's status, they may not mean the same thing to both groups, and our sample was too small to test this formally without risking a misleading result. This is distinct from the fact that some categorical residents also reported mistreatment, which is simply the expected baseline of a comparison group and does not undermine the between-group difference.

Preliminary residents also reported lower mentorship satisfaction and more research barriers, both holding after adjustment—mirroring the documented importance of belonging and mentorship to surgical trainee well-being[22, 23]. These patterns are consistent with structural disadvantages attached to the preliminary role, although—because the data are cross-sectional and capture status and outcomes at a single point—we cannot establish that this structure is their cause rather than a correlate.

A separate observation concerns perception. Preliminary residents rated their disadvantage in mentorship, community, and research more acutely than categorical residents did, paralleling a documented pattern in which advantaged groups underestimate others' disadvantage[24]. This invisibility may itself matter: a disadvantage unrecognized by those around it is less likely to be named or addressed—though, having measured perceptions and outcomes separately, our data cannot test this.

Burnout was the exception, not differing by training status, though it ran high in every group, consistent with prior data[2, 3]. This null should be read cautiously: the comparison was underpowered, and because female sex is an established burnout risk factor and preliminary residents were

disproportionately male (58% vs 36%), composition could have masked a true difference. Burnout may also be a more universal feature of surgical residency—endorsed at high rates by both groups—and thus less sensitive to training status than relational and structural measures such as mentorship and opportunity, where baseline rates varied more and differences were pronounced. Our data cannot distinguish this from simple lack of power. Where we did detect clear disadvantages—in mistreatment, mentorship, and opportunity—the differences were large; the burnout comparison was simply too underpowered to determine whether a similar difference exists. We therefore cannot say whether burnout is genuinely less affected by preliminary status or whether our sample was merely too small to detect a difference. Prior work has tracked where these trainees land[6] rather than what the year costs them; our findings begin to address the latter.

We detected no significant differences between the two preliminary pathways on any outcome, but this comparison was underpowered (about 18 designated residents, < 20% power), so we draw no conclusions from it. Designated residents remain poorly characterized in the literature, which has centered on non-designated trainees seeking to convert to categorical positions[6]; whether the two groups have distinct needs is an open question our data cannot answer.

Our findings complicate both common characterizations of the preliminary year—as either a valuable opportunity or an exploitative one. For many it is a genuine bridge forward: in prior national data, roughly 40% of all preliminary residents ultimately obtain a categorical general surgery position and a majority some categorical position across specialties[6]. Our sample was consistent: about 70% of preliminary residents were entering some categorical position the following year, of whom about half were continuing in general surgery and half moving into other specialties (so categorical general surgery accounted for roughly a third of all preliminary residents). Goal attainment was substantial but incomplete: among residents with a specific categorical goal, 58% entered that same type of position and 69% some categorical position—for most, the preliminary year advanced their aims, while a meaningful minority did not reach the position they sought. For those it serves, the role serves an important function, and calling it simply exploitative would understate that. At the same time, preliminary residents carry real clinical service load yet reported less support and more mistreatment than their categorical peers, and the bridge is uneven: prior work indicates that those least likely to transition are disproportionately international graduates, visa holders, and residents from groups already underrepresented in surgery[6]. These same groups were over-represented in the preliminary role in our sample (preliminary residents were far more likely than categorical residents to be international graduates, 38% vs 12%, $p < 0.001$), and this disadvantage appeared to extend to outcomes: international and US graduates entered residency with comparable categorical aspirations, yet international graduates were significantly less likely to report securing any categorical position the following year (42% vs 86%, $p < 0.001$). Comparable aspirations did not yield comparable destinations. The preliminary year thus helps many residents advance while its burdens fall most heavily on those already marginalized and least likely to move on, raising a question of equity rather than overall benefit: the role appears to serve some residents well and others far less so.

These disparities raise a question about how the preliminary position is structured. Preliminary residents carry real clinical service load yet reported less mentorship, more mistreatment, and lower satisfaction than categorical residents alongside them, suggesting the supportive elements of training may not track service contribution. We can describe this pattern but not its cause; our cross-sectional data cannot speak to intent. Still, the disparities make several questions worth asking: what the preliminary position is for, what it owes the residents who fill it, and whether its current form is equitable. To the extent these questions point to structural features of the role, supporting these residents may require attention to the position itself, not only added wellness resources.

Because we observed these disparities across a national, multi-program sample rather than at a single site, they appear widespread rather than confined to particular programs. Individual programs can and do address resident experience, but leaving this to local discretion risks uneven treatment—the disadvantages preliminary residents face would then depend on which program they enter. Preliminary residents therefore merit explicit attention at a system level: bodies such as the ACGME and the American Board of Surgery could extend the well-being benchmarking already applied to residents broadly^[11] to preliminary residents as a distinct group, so their experience is monitored consistently rather than absorbed into program-level aggregates.

Limitations

This study has several limitations. As a cross-sectional survey, it cannot establish causal relationships, and self-reported outcomes are subject to recall and perception bias. Participation was voluntary, and without a documented response rate, selection and nonresponse bias cannot be quantified. As noted, responses may not be fully independent (program-level clustering could not be modelled, and five reported prior-year participation). Because mistreatment was near-universal among preliminary residents, its adjusted odds ratio indicates direction rather than magnitude, and the sample was underpowered for smaller effects, particularly burnout (~ 30% power), leaving that null inconclusive. The six outcomes were prespecified and reported regardless of result, so the analysis was not exploratory fishing, though we applied false-discovery-rate correction as a safeguard.

Several domains used investigator-developed single-item measures, which may not capture these multidimensional constructs and cannot be evaluated for internal-consistency reliability; formal reliability testing was possible only for the multi-item mistreatment composite, and even there we did not test measurement invariance across groups. Burnout used the validated two-item MBI rather than the full inventory, capturing fewer dimensions of burnout. The study's principal strengths lie in its national, multi-institutional sample, the validated mistreatment composite, and the prespecified analytic approach.

These limitations point to several directions for future work. Because a cross-sectional survey can document that disparities exist but not why they arise, qualitative study—resident interviews or focus groups—would help explain the mechanisms behind these findings, such as why preliminary residents perceive less support across the domains we measured and how the experiences of designated and

non-designated residents differ. Additional multicenter and single-center studies could test whether these patterns hold across other settings and program cultures, and longitudinal or outcome-linked data could clarify whether differences in intended career trajectory translate into differences in realized ones. The difference we observed in categorical transition between international and US graduates (discussed above) was based on self-reported next-year plans rather than verified appointments, and warrants dedicated study with outcome-linked follow-up. More broadly, this study characterized preliminary residents' experiences during the preliminary year rather than tracking their career trajectories, and was neither focused on nor powered for analyses of where they ultimately go. Questions such as whether the type of categorical position secured (for example, surgical versus non-surgical) differs by background cannot be answered here—the relevant subgroups were small (only 11 international graduates entered a categorical position), and the apparent patterns should be treated as hypotheses. Adequately powered studies of preliminary residents' career trajectories and destinations, ideally with verified outcome data and explicit attention to equity by background, are an important future direction. Finally, evaluating the impact of specific initiatives intended to support preliminary residents would help establish what actually improves their experience.

Conclusions

Preliminary general surgery residents reported significantly worse personal and professional experiences than categorical residents, particularly with respect to workplace mistreatment, mentorship, and academic opportunity. These differences persisted after adjustment for demographic factors. Although the cross-sectional design precludes causal inference, the findings are consistent with structural disadvantages associated with training status and warrant further study. Multi-institutional and longitudinal work is needed to clarify the mechanisms underlying these differences—including how intended career goals shape resident experience—and to identify interventions that promote equity between preliminary and categorical surgical trainees.

References

1. National Resident Matching Program. Results and Data: 2025 Main Residency Match. Washington, DC: National Resident Matching Program; 2025.
2. Elmore LC, Jeffe DB, Jin L, et al. National survey of burnout among US general surgery residents. *J Am Coll Surg*. 2016;223(3):440–51.
3. Dimou FM, Eckelbarger D, Riall TS. Surgeon burnout: a systematic review. *J Am Coll Surg*. 2016;222(6):1230–9.
4. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med*. 2019;381(18):1741–52. <https://doi.org/10.1056/NEJMsa1903759>.
5. Shanafelt TD, Balch CM, Bechamps GJ, et al. Burnout and career satisfaction among American surgeons. *Ann Surg*. 2009;250(3):463–71.

6. Hardy WA, Litts CS Jr, White JA. Career outcomes of nondesignated preliminary surgery residents: a systematic review and meta-analysis. *J Surg Educ*. 2023;80(12):1773–80.
7. Ahmad R, Mullen JT. Career outcomes of nondesignated preliminary general surgery residents at an academic surgical program. *J Surg Educ*. 2013;70(6):690–5.
8. Yoo PS, Kozol R, Reilly P, Seashore JH, Duffy A, Chandawarkar R, Longo WE. Fate of non-designated preliminary general surgery residents seeking a categorical residency position. *Am J Surg*. 2009;198(5):593–5.
9. Sullivan MC, Yeo H, Roman SA, Jones AT, Bell RH Jr, Sosa JA. Discrepancies in training satisfaction and program completion among 2662 categorical and preliminary general surgery residents. *Ann Surg*. 2013;257(6):1174–80.
10. Salles A, Liebert CA, Esquivel M, Greco RS, Henry R, Mueller C. Perceived value of a program to promote surgical resident well-being. *J Surg Educ*. 2017;74(6):921–7.
11. Bilimoria KY, Hu YY, Cheung EO et al. Surgical Education Culture Optimization through targeted interventions based on National comparative Data (SECOND) Trial. *ClinicalTrials.gov* identifier NCT03739723; 2019–2024.
12. 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(8):713–27. <https://doi.org/10.1056/NEJMoa1515724>.
13. Gruppen LD, Irby DM, Durning SJ, Maggio LA. Conceptualizing learning environments in the health professions. *Acad Med*. 2019;94(7):969–74. <https://doi.org/10.1097/ACM.0000000000002702>.
14. Tajfel H, Turner JC. The social identity theory of intergroup behavior. In: Worchel S, Austin WG, editors. *Psychology of Intergroup Relations*. 2nd ed. Nelson-Hall; 1986. pp. 7–24.
15. West CP, Dyrbye LN, Sloan JA, Shanafelt TD. Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. *J Gen Intern Med*. 2012;27(11):1445–52.
16. de Boer AGEM, van Lanschot JJB, Stalmeier PFM, van Sandick JW, Hulscher JBF, de Haes JCJM, Sprangers MAG. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual Life Res*. 2004;13(2):311–20.
17. Martin RJ, Hine DW. Development and validation of the Uncivil Workplace Behavior Questionnaire. *J Occup Health Psychol*. 2005;10(4):477–90.
18. Salles A, Lin D, Liebert C, et al. Belonging and well-being in surgical training. *Ann Surg*. 2021;273(2):202–8.
19. Maslach C, Jackson SE, Leiter MP. *Maslach Burnout Inventory Manual*. 4th ed. Mind Garden; 2018.
20. Salles A, Cohen GL, Mueller CM. The relationship between belonging and burnout among general surgery residents. *J Surg Educ*. 2019;76(2):321–8.
21. Accreditation Council for Graduate Medical Education. *Common Program Requirements (Residency), Section VI: The Learning and Working Environment*. Chicago, IL: ACGME; 2017.

22. World Health Organization. WHOQOL: Measuring Quality of Life. Geneva: World Health Organization; 1997.
23. Cortina LM, Magley VJ, Williams JH, Langhout RD. Incivility in the workplace: incidence and impact. *J Occup Health Psychol*. 2001;6(1):64–80.
24. Kraus MW, Onyeador IN, Daumeyer NM, Rucker JM, Richeson JA. The misperception of racial economic inequality. *Perspect Psychol Sci*. 2019;14(6):899–921.
<https://doi.org/10.1177/1745691619863049>.