

# UC Irvine

## UC Irvine Previously Published Works

### Title

Cumulative Effect of Flexible Duty-hour Policies on Resident Outcomes: Long-term Follow-up Results From the FIRST Trial.

### Permalink

<https://escholarship.org/uc/item/4rf8v561>

### Journal

Annals of Surgery, 271(5)

### Authors

Khorfan, Rhami

Yuce, Tarik

Love, Remi

et al.

### Publication Date

2020-05-01

### DOI

10.1097/SLA.0000000000003802

Peer reviewed



Published in final edited form as:

*Ann Surg.* 2020 May ; 271(5): 791–798. doi:10.1097/SLA.0000000000003802.

## Cumulative Effect of Flexible Duty-hour Policies on Resident Outcomes:

### Long-term Follow-up Results From the FIRST Trial

Rhami Khorfan, MD, MS<sup>\*</sup>, Tarik K. Yuce, MD<sup>\*</sup>, Remi Love, MA<sup>\*</sup>, Anthony D. Yang, MD<sup>\*</sup>, Jeanette Chung, PhD<sup>\*</sup>, David B. Hoyt, MD<sup>†</sup>, Frank R. Lewis, MD<sup>‡</sup>, Karl Y. Bilimoria, MD, MS<sup>\*,†,✉</sup>

<sup>\*</sup>Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL

<sup>†</sup>American College of Surgeons, Chicago, IL

<sup>‡</sup>American Board of Surgery, Philadelphia, PA

### Abstract

**Objective:** To investigate the long-term effect of flexible duty-hour policies on resident outcomes

**Background:** The Flexibility in Duty Hour Requirements for Surgical Trainees trial showed no significant difference in overall resident well-being between flexible and standard duty-hour policies after 1 year. However, long-term exposure to flexibility could have adverse effects on duty-hour violations, resident satisfaction, and well-being.

**Methods:** In 2014, 117 programs were randomized to flexible or standard duty-hour policy. Residents were surveyed annually following the American Board of Surgery In-Training Exam (~3750 residents/yr; response rate 99%). Four-year trends within the flexible policy arm were analyzed using cluster-corrected Chi-squared tests. Differences between study arms were modeled using multivariable logistic regression.

**Results:** Over time, there was a trend toward fewer 80-hour work week violations in the flexible arm (19.8%–17.0%,  $P=0.06$ ), and increased satisfaction with flexible duty-hours (91.9%–94.3%,  $P<0.05$ ). Although well-being decreased over time (85.1%–81.5%,  $P=0.01$ ), this was seen globally with no difference between study arms [odds ratio (OR) 0.96 (0.74–1.25)]. Likewise, at the end of the study period, there was no association between flexible policy and duty-hour violations [OR 1.25 (0.95–1.61)] or satisfaction with duty hours [OR 0.80 (0.55–1.19)] compared to standard policy. Residents in flexible duty-hour programs reported significantly fewer lapses in continuity than standard policy residents, until all programs transitioned to flexibility by 2018.

✉ k-bilimoria@northwestern.edu.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site ([www.annalsurgery.com](http://www.annalsurgery.com)).

**Conclusion:** Cumulative time under flexible duty-hour policies had no detrimental effects on duty-hour violations or resident well-being. After multiple years of flexibility, residents continue to report a high rate of satisfaction and positive effects on continuity of care.

### Keywords

duty hour; surgical residency training; well-being

The Flexibility in Duty Hour Requirements for Surgical Trainees (FIRST) trial examined the impact of flexible duty-hour policies (ie, eliminated restrictions on daily shift hour limits and time off between shifts) versus standard 2011 duty-hour policies on patient outcomes and resident well-being. The FIRST Trial demonstrated no significant difference in resident satisfaction with overall well-being and education between flexible and standard duty-hour policies after 1 year.<sup>1</sup> A follow-up analysis in 2017 demonstrated a decrease in negative perceptions of flexible duty-hour policies 1 year later.<sup>2</sup> In response to these and other findings, the Accreditation Council for Graduate Medical Education (ACGME) transitioned all programs to flexible duty-hour policy in July 2017, by waiving limits on individual shift length while maintaining the 80-hour-per-week limit.<sup>3</sup>

However, there are concerns that long-term exposure to flexible duty-hour policies may have cumulative adverse effects on resident duty-hour violations and well-being.<sup>4</sup> Although the majority of residents using flexible duty-hours reported doing so voluntarily, and generally for the purposes of educational opportunities or to minimize lapses in patient care, it is possible that over time resident duty-hour violations have increased and resident satisfaction decreased.<sup>5,6</sup> In addition, during the FIRST Trial, residents in flexible duty-hour programs were more likely to report negative effects on time for family/friends, rest, and extracurricular activities.<sup>1</sup> It is unknown whether these effects are potentially worsened with increasing time under flexibility.

To address these concerns, we conducted a 4-year follow-up analysis of the FIRST Trial examining longitudinal effects of flexible duty-hour policies on resident outcomes. Our objectives were (1) to investigate the effect of multiple years of flexible duty-hour policies on duty-hour violations, resident satisfaction, continuity of care, and perceived negative effects on well-being, and (2) to evaluate how these outcomes differed between flexible and standard duty-hour policy groups over time.

## METHODS

In 2014, 117 general surgery residency programs were recruited and randomized to either standard duty-hour policies (n = 59) or flexible duty-hour policies (n = 58) as part of the FIRST Trial. Programs in the flexible duty-hour arm were required to adhere to the 80-hour work week limit (averaged over 4 weeks), but were not required to adhere to limits on maximum shift lengths and time off between shifts. After the trial concluded, programs were granted waivers from the ACGME to continue their assigned study conditions. This continued until ACGME national policy changed in 2017 to accommodate flexible duty-hour regulations, which resulted in all programs transitioning to less restrictive duty-hour policies similar to the flexible duty-hour arm of the FIRST Trial. The design of the

prospective, cluster-randomized FIRST trial has previously been described in more detail.<sup>1,7</sup> The trial was designated as non-human subjects research by the Northwestern University IRB.<sup>8</sup>

### Survey Tool

Resident outcomes for the FIRST Trial were collected in collaboration with the American Board of Surgery, as part of a survey administered to all general surgery residents nationally immediately following the American Board of Surgery In-Training Examination. Although the survey was administered to all residents, we included only self-reported clinically-active residents in this study as this is the cohort for whom duty-hour regulations are most relevant. This survey was developed by a multidisciplinary team and underwent iterative testing and revision by the team, with cognitive interviews and multiple rounds of pilot testing by residents from multiple institutions.<sup>1,7</sup>

When evaluating outcomes in the standard policy arm, it is important to keep in mind the change in ACGME policy that transitioned these programs to more flexible duty-hour policies in 2017. Therefore, the longitudinal analysis of outcomes in this arm can be divided into the standard policy period (2015–2017) and the post-transition period (2018).

### Duty-hour Violations

A primary outcome for this analysis was frequency of violation of the 80-hour work week regulation. We focused specifically on violations of the 80-hour rule, as opposed to other duty-hour violations, as this was applicable to both the standard and flexible duty-hour groups. Although residents were asked about violations of the 80-hour rule in 2016, 2017, and 2018, a new wording of the question was introduced in 2017. Given this change in 2017, we are unable to directly compare duty-hour violations from 2016 to 2018.

In 2016 and 2017, residents were asked, “During your most recent typical month on a general surgery rotation, approximately how many times did you exceed more than 80 hours in a week?” As in previous studies, the frequency of reported duty-hour violations was dichotomized to more than 2 times per month versus 2 or fewer times per month for this question, given that ACGME regulations average hours over a 4-week period.<sup>5</sup> This question was used to assess changes in the frequency of violation within each study arm over time, and differences between study arms, before the change in ACGME duty-hour requirements.

In 2017 and 2018, residents were asked “Overall, how many months have you violated the 80-hour duty-hour limit during this academic year?” This outcome was dichotomized to 0 or 1 month versus more than 1 month. These responses were used to compare frequency of duty-hour violations between the flexible and standard policy arms before the change in 2017, and afterward in 2018, and examine the continued effect of flexibility on duty-hour violations within the flexible policy arm.

### Resident Satisfaction

Resident-reported satisfaction with various aspects of training was collected each year. This included satisfaction with overall well-being, program duty-hour regulations, continuity of

care, patient safety, work hours and scheduling, hand-offs, overall resident education, time for rest, decision to become a surgeon, and job satisfaction. Residents responded using a 5-point Likert scale (very dissatisfied, dissatisfied, neutral, satisfied, or very satisfied). This scale was dichotomized to dissatisfied (dissatisfied or very dissatisfied) or satisfied (neutral, satisfied, very satisfied) for analysis.

### **Continuity of Care**

Residents were also asked how often in the past month they had experienced a lapse in their education or clinical continuity of care. These specific instances included leaving/missing a conference, handing off an active patient care issue, handing off care of a critically ill patient, leaving during an operation, missing an operation, returning to the hospital after leaving to operate on a patient, and returning to the hospital after leaving to provide other patient care. Response options for these items were either 0 times, 1–2 times, 3–5 times, or >5 times. These were dichotomized to 0 times or at least 1 time for analysis.

### **Perceived Negative Effects on Professional and Personal Factors**

The final group of questions elicited resident's perceptions of whether their program's duty-hour regulations had a positive, neutral, or negative effect on various personal and professional factors. These included how duty-hours affected time for family/friends, time for hobbies, resident's health, how well-rested residents felt, ability to attend educational conferences, availability for cases, and resident morale. Responses were dichotomized to either negative effect or neutral/positive effects. Of note, the wording of these questions was changed in the 2018 survey to "satisfaction with" instead of "effects of" these factors. Therefore, the responses are not directly comparable, and only 2015–2017 responses were used to assess these outcomes.

### **Statistical Analysis**

Frequency of duty-hour violations was analyzed both between study arms and over time within each study arm. These bivariate analyses were done using Rao-Scott design-corrected Chi-square tests for survey data, with responses clustered by residency program.<sup>9</sup> This methodology allowed adjustment for clustering of responses within a residency program to account for nonindependence of residents within a program. Satisfaction, continuity, and perception outcomes were analyzed within the flexible arm over the 4-year period using the same methodology.

In the multivariable analyses, the association between duty-hour policy arm and each of the outcomes was modeled using multivariable logistic regression with robust clustered standard errors. A model was constructed for each of the dichotomous outcomes described above, with study arm designation and survey year as the primary predictors of interest. Models were adjusted for covariates including program type (academic, community, or military), geographic region (Northeast, Southeast, Midwest, Southwest, or West), resident sex, and postgraduate year (PGY) level (1–5).

To assess whether study arm had different effects on outcomes over time, we included interaction terms between study arm and survey year. We then used linear combinations of

the coefficients to isolate the individual effect of duty-hour arm designation at each year. This difference-in-differences analysis was performed to test the independent effect of being in the flexible duty-hour arm every year, whereas adjusting for secular trends. Analyses were performed using SAS version 9.4 and STATA version 13.1 with a significance level of  $P < 0.05$  for all analyses.

## RESULTS

In 2015, 3767 residents from 117 programs participating in the FIRST trial responded to the survey. Similarly, there were 3807 respondents in 2016, 3789 in 2017, and 3741 in 2018. Survey response rates were consistently high across the 4-year period (98.5%, 99.2%, 99.3%, and 99.3%, respectively). About 49% of respondents were from flexible duty-hour programs every year. Distributions of resident sex, geographic region, program type, and PGY-level were similar across study years (Table 1).

### Duty-hour Violations

Within the flexible arm, there was a trend toward fewer 80-hour work week violations reported from 2016 to 2017 (Table 2, 19.8%–17.0%,  $P = 0.06$ ). Similarly, from 2017 to 2018, responses to the updated question showed a nonsignificant decrease in violations (26.5% in 2017 to 24.7% in 2018,  $P = 0.21$ ). Although the rate of residents reporting working >80 hours in more than 2 weeks was higher in the flexible arm than in the standard arm in 2016 (19.8% vs 16.1%,  $P = 0.03$ ), there was no difference between the 2 arms in 2017 and 2018 (Table 2). Specifically, 17.0% of flexible policy residents and 14.5% of standard policy residents reported working >80 hours more than 2 weeks of the last month in 2017 ( $P = 0.15$ ). When asked how many months they had violated the 80-hour work week rule, 26.5% of flexible-policy residents and 23.3% of standard-policy residents reported a violation in more than 1 month in 2017 ( $P = 0.21$ ). This difference decreased in 2018, to 24.7% of flexible policy residents and 24.4% of standard policy residents in 2018 ( $P = 0.91$ ).

### Resident Satisfaction

Resident satisfaction under flexible policy was consistently high across most outcomes, with no significant decline in satisfaction over the 4-year span (Fig. 1; also see table in Supplemental Digital Content 1, <http://links.lww.com/SLA/B998>). Resident satisfaction with their program's flexible duty-hour regulations showed a small but significant increase over time (from 91.9% in 2015 to 94.3% in 2018,  $P < 0.05$ ). In addition, the proportion of residents reporting satisfaction with their job grew from 87.3% in 2015 to 93.2% in 2018 ( $P < 0.01$ ), as did satisfaction with the decision to become a surgeon (90.8% in 2015 to 94.0% in 2018,  $P < 0.01$ ). The percent of residents reporting satisfaction with overall well-being was the only outcome that declined, from 85.1% in 2015 to 81.5% in 2018 ( $P = 0.01$ ), but this also declined in the standard-policy arm (88.0%–77.9%,  $P < 0.01$ ).

Throughout the standard policy period (2015–2017), flexible policy residents had significantly higher satisfaction with continuity of care [Table 3; odds ratio (OR) 2.09 (1.42–3.06),  $P < 0.01$  in 2017] and handoffs [OR 1.76 (1.36–2.28),  $P < 0.01$  in 2017] compared to standard policy residents. Satisfaction with overall well-being was no different between the

2 arms in the standard policy period, but significantly higher in the flexible arm in the post-transition period [OR 1.36 (1.09–1.71),  $P = 0.01$ ].

### Continuity of Care

Residents in the flexible policy arm reported low rates of lapses in continuity of care at baseline in 2015 (Fig. 1, also see table in Supplemental Digital Content 1, <http://links.lww.com/SLA/B998>). These rates remained low across the study years, and in many instances decreased. There were no differences in the proportion of flexible policy residents reporting leaving during an operation (7.0% in 2015 vs 6.4% in 2018,  $P = 0.60$ ) or missing a conference (13.3% in 2015 vs 15.4% in 2018,  $P = 0.22$ ). From 2015 to 2018, flexible policy residents reported significantly fewer instances of handing off active patient care (32.0%–28.5%,  $P < 0.01$ ), handing off critically ill patients (24.6%–21.7%,  $P < 0.01$ ), and missing an operation (29.9%–23.6%,  $P < 0.01$ ). However, there were large decreases in the frequencies of residents returning to the hospital to operate on a patient (45.5%–29.5%,  $P < 0.01$ ) and returning to the hospital to provide other patient care (50.4%–33.4%,  $P < 0.01$ ).

At baseline in 2015, flexible policy residents had a significantly lower likelihood of missing conference [OR 0.51 (0.39–0.65),  $P < 0.01$ ], handing off active patient care [OR 0.51 (0.42–0.61),  $P < 0.01$ ], handing off critical patients [OR 0.59 (0.48–0.72),  $P < 0.01$ ], leaving an operation [OR 0.47 (0.32–0.68),  $P < 0.01$ ], and missing an operation [OR 0.56 (0.44–0.70),  $P < 0.01$ ] compared to standard policy residents. These differences persisted throughout the standard policy period (2015–2017) until the standard arm transitioned to flexible duty-hours (Fig. 2; also see table in Supplemental Digital Content 2, <http://links.lww.com/SLA/C3>). Post-transition, the rates of lapses in continuity converged, with the standard policy arm now having almost identical rates with the flexible arm with no significant difference in any of the measured continuity outcomes in 2018.

### Perceived Negative Effects

In the FIRST trial, about 25% of residents under flexible duty-hour policy reported that duty-hour policies had negative effects on personal factors such as feeling well-rested and time for family. However, after 2 additional years under flexibility, there were significant decreases in perceived negative effects of duty-hours on residents' health (18.3%–12.0%,  $P < 0.01$ ), time for family and friends (24.8%–14.0%,  $P < 0.01$ ), time for hobbies (25.7%–14.5%,  $P < 0.01$ ), feeling well-rested (26.4%–14.4%,  $P < 0.01$ ), and resident morale (16.5%–11.4%,  $P < 0.01$ ) (Fig. 3; also see table in Supplemental Digital Content 1, <http://links.lww.com/SLA/B998>). Perceived negative effects on educational and training factors, whereas low, did increase somewhat over the study period. More residents reported negative effects on availability for elective cases (14.8%–19.2%,  $P < 0.01$ ) and availability for urgent cases (14.9%–19.2%,  $P < 0.01$ ), whereas there was no change in residents reporting negative effects on educational conference attendance (12.3%–10.0%,  $P = 0.09$ ).

Although perceived negative effects of duty-hour policies on personal factors were higher among flexible policy residents compared to standard policy residents, this difference decreased over time (Fig. 3). This can be seen, for example, in the association between flexible policy and perceived negative effects on health from 2015 [OR 2.99 (2.24–3.97),  $P <$

0.01] to 2017 [OR 1.34 (1.02–1.76),  $P = 0.04$ ]. There were similar trends in time for family and friends, time for hobbies, and feeling well-rested (Table 3). Negative effects on resident morale did not differ between flexible and standard policy residents throughout the study period [OR 0.84 (0.63–1.13),  $P = 0.25$  in 2017]. Being in the flexible policy arm was consistently associated with fewer perceived negative effects on conference attendance [OR 0.53 (0.40–0.69),  $P < 0.01$ ] and availability for urgent [OR 0.47 (0.39–0.57),  $P < 0.01$ ] and elective cases [OR 0.59 (0.48–0.74),  $P < 0.01$ ] compared to the standard arm.

## DISCUSSION

Although the FIRST Trial demonstrated that flexible duty-hour policies did not lead to worse patient or resident outcomes after 1 year, it was important to investigate the long-term effects of flexible policies on surgical residents, as cumulative time under flexibility could have a cumulative impact on residents. In this 4-year follow-up analysis of the FIRST Trial, we found no evidence of increased duty-hour violations, decreased satisfaction, or decreased well-being attributable to flexible duty hours among surgical residents over time. Furthermore, we found that the benefits of flexibility when compared to standard policy persisted, whereas most of the negative effects of flexibility diminished over time.

Our group previously reported that residents in the flexible policy group used flexibility to work longer shifts selectively to benefit patient care or resident training; however, the flexible group also more frequently reported working greater than 80 hours per week.<sup>5</sup> In our current study, we demonstrate that violations of the 80-hour rule in the flexible-policy group decreased over time. This may be due to residents and programs adapting to and managing flexibility differently over time. Furthermore, we found that these rates did not differ between the standard and flexible group in 2017 and 2018. This is especially notable given the transition of standard policy programs to flexibility in 2017, a change that did not result in more frequent duty-hour violations in 2018. In any case, we found no evidence that multiple years under flexibility led to widespread violations of the 80-hour work week requirement.

It was previously shown that residents had a high rate of satisfaction with flexible duty-hours, and that the majority of surgical residents would choose a program with flexible duty-hours if given the choice.<sup>10</sup> We found that resident satisfaction with flexible duty hours further increased with longer time under flexibility, to a peak of 94.3% being satisfied with flexibility in 2018. In addition, there was no significant decrease in satisfaction with duty-hour regulations in the standard policy group after those programs transitioned to flexibility in 2017. Satisfaction with patient safety and continuity of care were higher among flexible policy residents in 2015–2017, with the standard policy group achieving similar rates of satisfaction post-transition as expected.

One concerning finding was the decrease in resident satisfaction with overall well-being and time for rest. Importantly, this decrease from 2015 to 2018 was seen in all residents regardless of their program's duty-hour regulations, flexible or standard policy. Therefore, this is unlikely to be related to duty-hour policy and may be a reflection of other factors influencing well-being during this time period. Indeed, there has been a significant focus on



physician and resident wellness in the past several years, especially in the field of general surgery.<sup>11–13</sup> Thus, the increasing rates may be due to increased awareness and recognition of these important problems, an actual worsening in well-being, or some combination of these and other unmeasured factors. These findings were the impetus for a deeper investigation of surgical resident well-being<sup>14</sup> and led to the development of the SECOND trial assessing tools to improve program culture and enhance resident well-being (NCT03739723). When comparing these outcomes in 2018, the standard policy residents reported lower satisfaction after transitioning to flexibility than flexible policy residents. This is in contrast to other satisfaction and continuity outcomes, which were almost identical in 2018. This may indicate an initial adjustment period after transitioning to flexibility, or that the benefits of flexibility differ between the group that started with flexibility compared to those who transitioned later.

One of the most notable benefits of flexibility identified in the FIRST Trial emphasized by resident trainees was the improvement in patient care and continuity.<sup>1,15</sup> We found these results to be durable over time, with flexible policy residents consistently reporting lower rates of lapses in continuity of care. This is important from a patient care perspective, as it minimizes the hand-offs and transitions of care where medical errors can occur.<sup>16,17</sup> It is also essential from a training standpoint, as flexible policy residents were significantly less likely to miss valuable educational time engaged in patient care, operating, or attending conferences. Fortunately, we found that after the change in ACGME policy in 2017, programs previously randomized to standard policy had continuity of care rates equivalent to those of the flexible group in 2018. Before this, many of the outcomes in the standard arm show a trend of slowly approaching the flexible arm in 2016–2017. Although reasons for this are unclear, it may represent a “cross-over” effect, with some residents or programs in the standard arm adopting some of the features of the flexible policy in the years after the FIRST trial, but before the ACGME change.

In general, residents perceived negative effects on educational and training factors associated with standard policy, whereas negative effects on personal factors were more associated with flexible policy. Surgical trainees previously identified the negative impact of flexible duty-hour policy on perceived personal health, and this was particularly concerning.<sup>15</sup> However, we found that these effects decreased over time, with perceived personal health approaching the same level as standard policy residents by 2017. Similarly, negative effects on other personal factors, including time for family and feeling well-rested, decreased significantly over time among flexible residents by 10%–12%. On the other hand, the negative effects on training factors such as availability for cases and conference attendance were persistently worse throughout for the standard policy group compared to flexible policy.

Our study has certain limitations given its data source and study design. In this analysis, resident survey responses were collected from the same 117 programs annually. However, responses were de-identified and therefore individual resident responses could not be tracked over time, thus limiting our ability to estimate the effects of cumulative exposure to flexible duty-hour policy for individual residents. This precluded our ability to perform a true longitudinal, repeated-measures analysis, which would require linking a resident’s responses over repeated surveys. Another limitation is the change in survey question wording over time

regarding duty-hour violations, changes that were made to address different objectives at each point in time. This is notable due to the change in the time frame (past month up to 2017 vs past year in 2018), and in the explicit inclusion of the word “violation” in the second question and not the first. As such, we were unable to assess changes in the frequency of 80-hour violations over the entire 4-year span; however, inclusion of both questions in 2017 allows us to compare between study arms after 4 years of flexibility.

## CONCLUSIONS

Our study demonstrates that cumulative time under flexible duty hours appeared to have no detrimental effects on duty-hour violations or resident satisfaction with well-being. After multiple years of flexibility, residents continue to report a high rate of satisfaction with duty-hour policies and other aspects of training. The positive effects of flexibility on continuity of care are robust and durable over time, with an immediate improvement in programs transitioning from standard to flexible policy. Perceived negative effects, on the other hand, seem to diminish with increasing time under flexibility. We did; however, identify a decrease in resident perceived well-being regardless of duty-hour policy, which should continue to be a target for further work to improve the training environment of general surgery residency.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

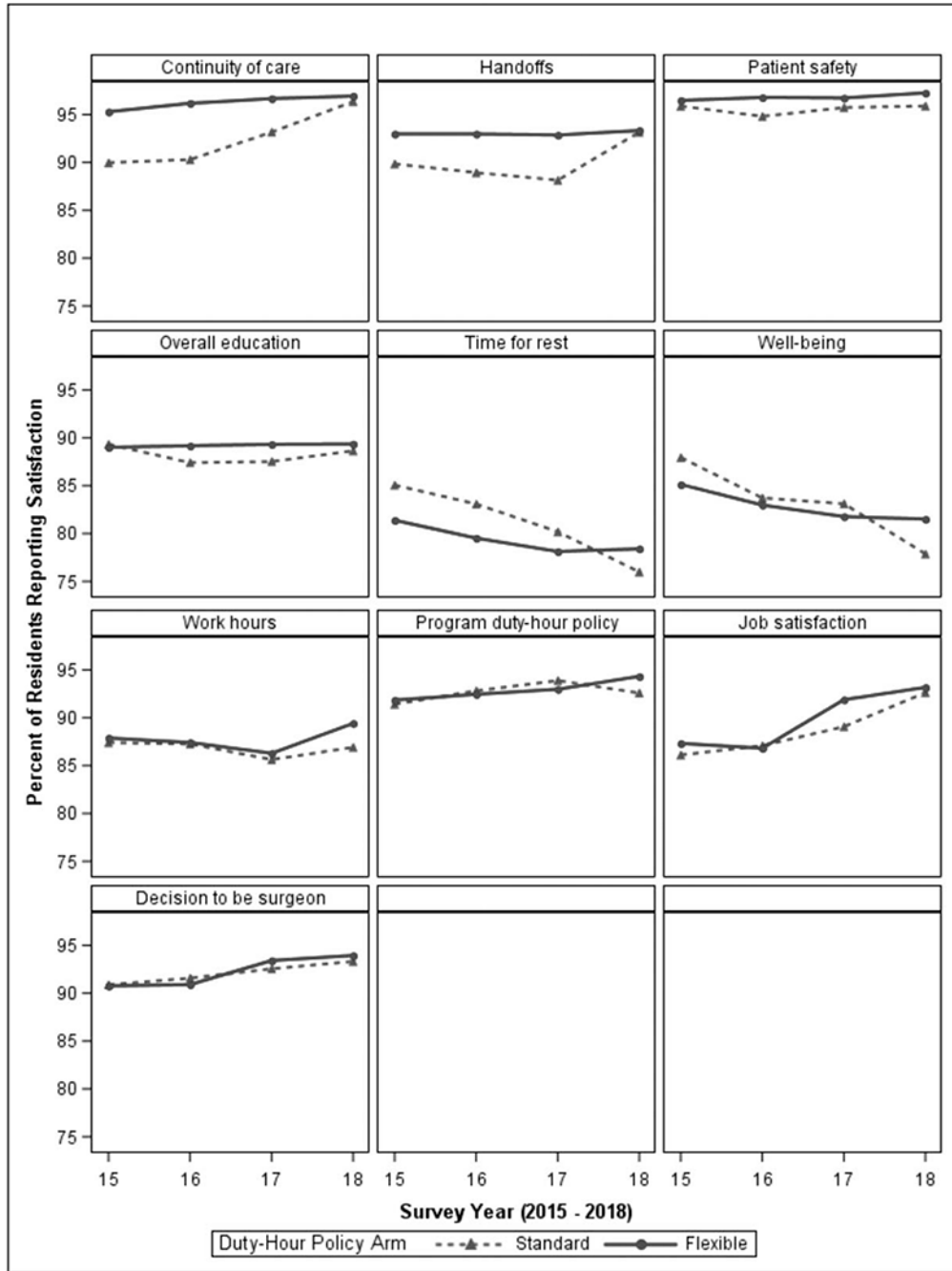
## Acknowledgments

This work was supported by the American Board of Surgery, the American College of Surgeons, and the Accreditation Council for Graduate Medical Education as a continuation of the FIRST Trial. RK is partially supported by the National Institutes of Health (5T32HL094293). TY is partially supported by the Agency for Healthcare Research and Quality (5T32HS000078). AY is partially supported by the National Heart Lung and Blood Institute of the National Institutes of Health (K08HL145139).

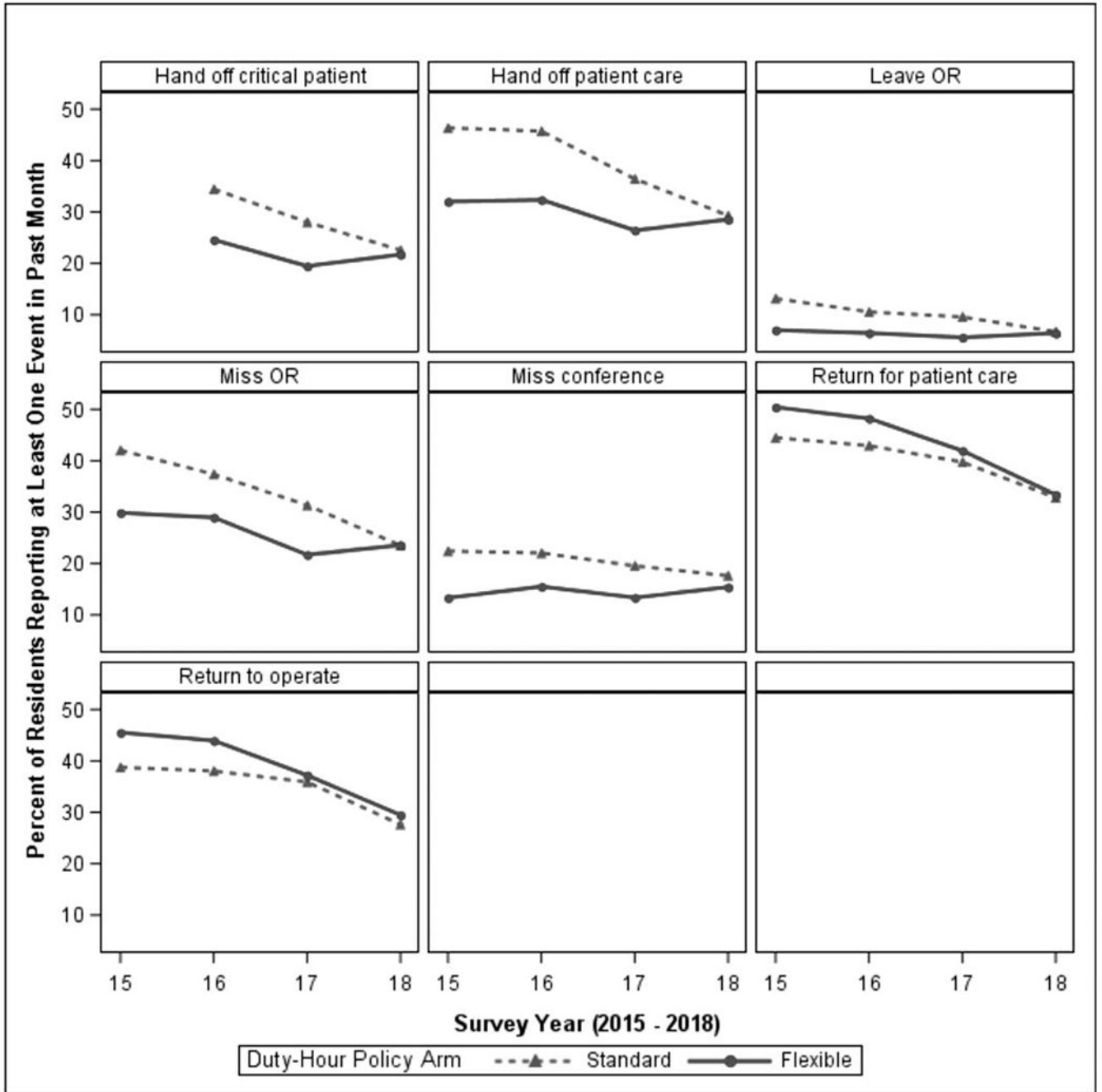
## REFERENCES

1. 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]
2. Dahlke AR, Quinn CM, Chung JW, et al. Surgical residents’ work hours and well-being in year 2 of the FIRST trial. *N Engl J Med*. 2017;377:192–194. [PubMed: 28700847]
3. Accreditation Council for Graduate Medical Education. Common Program Requirements. 2017 Available at: <https://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements> Accessed 11 November 2019.
4. McHill AW, Czeisler CA, Shea SA. Resident physician extended work hours and burnout. *Sleep*. 2018;41:zsy112.
5. Bilimoria KY, Quinn CM, Dahlke AR, et al. Use and underlying reasons for duty hour flexibility in the flexibility in duty hour requirements for surgical trainees (FIRST) trial. *J Am Coll Surg*. 2017;224:118–125. [PubMed: 27884805]
6. Blay E Jr, Engelhardt KE, Hewitt DB, et al. Evaluation of reasons why surgical residents exceeded 2011 duty hour requirements when offered flexibility: a FIRST trial analysis. *JAMA Surg*. 2018;153:860–862. [PubMed: 29898203]
7. 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]

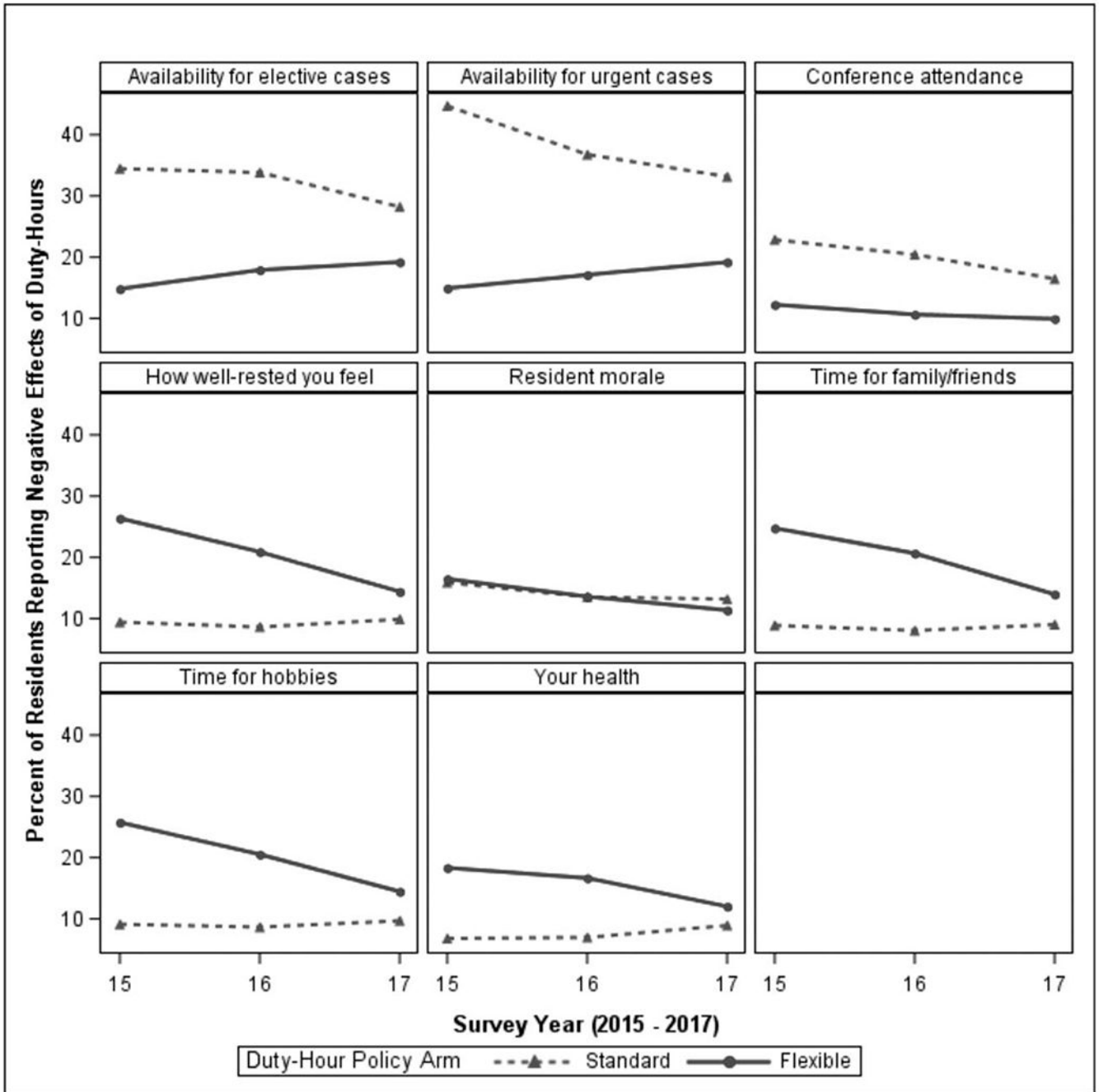
8. Minami CA, Odell DD, Bilimoria KY. Ethical considerations in the development of the flexibility in duty hour requirements for surgical trainees trial. *JAMA Surg.* 2017;152:7–8. [PubMed: 27732709]
9. SAS Institute Inc. *The SURVEYFREQ Procedure SAS/STAT 9.2 User's Guide, Second Edition* Cary, NC, USA: SAS Institute Inc; 2009
10. Yang AD, Chung JW, Dahlke AR, et al. Differences in resident perceptions by postgraduate year of duty hour policies: an analysis from the flexibility in duty hour requirements for surgical trainees (FIRST) trial. *J Am Coll Surg.* 2017;224:103–112. [PubMed: 27825914]
11. 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]
12. Elmore LC, Jeffe DB, Jin L, et al. National survey of burnout among US general surgery residents. *J Am Coll Surg.* 2016;223:440–451. [PubMed: 27238875]
13. Raj KS. Well-being in residency: a systematic review. *J Grad Med Educ.* 2016;8:674–684. [PubMed: 28018531]
14. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med.* 2019;381:1741–1752. [PubMed: 31657887]
15. Mouawad NJ, Kulaylat A, Hoffman RL. Perspective of surgical resident trainees on accreditation council for graduate medical education changes in resident work environment and duty hours. *JAMA Surg.* 2017;152:907–908. [PubMed: 28679005]
16. Jensen AM, Sanders C, Doty J, et al. Characterizing information decay in patient handoffs. *J Surg Educ.* 2014;71:480–485. [PubMed: 24776880]
17. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med.* 2014;371:1803–1812. [PubMed: 25372088]



**FIGURE 1.** Resident satisfaction outcomes by duty-hour policy arm, by year. Percent of residents reporting satisfaction (very satisfied/satisfied/neutral vs dissatisfied/very dissatisfied) with each of the outcomes. Number of responses differ by year and survey question. In the standard duty-hour policy arm: N = 1888 in 2015, N = 1890 in 2016, N = 1949 in 2017, and N = 1903 in 2018. In the flexible duty-hour policy arm: N = 1782 in 2015, N = 1784 in 2016, N = 1838 in 2017, and N = 1834 in 2018.



**FIGURE 2.** Continuity outcomes by duty-hour policy arm, by year. Number of responses differ by year and survey question. In the standard duty-hour policy arm: N = 1944 in 2015, N = 1959 in 2016, N = 1949 in 2017, and N = 1904 in 2018. In the flexible duty-hour policy arm: N = 1821 in 2015, N = 1847 in 2016, N = 1838 in 2017, and N = 1836 in 2018.



**FIGURE 3.** Perceived negative effects of duty-hour policies by duty-hour policy arm, by year. Percent of residents reporting a negative effect of duty-hours (negative effect vs neutral/positive effect) on each of the outcomes. Number of responses differ by year and survey question. In the standard duty-hour policy arm: N = 1892 in 2015, N = 1885 in 2016, and N = 1949 in 2017. In the flexible duty-hour policy arm: N = 1783 in 2015, N = 1780 in 2016, and N = 1838 in 2017.

**TABLE 1.**

**Resident and Program Characteristics by Survey Year**

	2015		2016		2017		2018		P
	N	%	N	%	N	%	N	%	
Overall	3767	24.9%	3807	25.2%	3789	25.1%	3741	24.8%	0.93
Duty-hour am									
Standard	1945	51.6%	1960	51.5%	1951	51.5%	1905	50.9%	
Flexible	1822	48.4%	1847	48.5%	1838	48.5%	1836	49.1%	
Sex									0.24
Male	2273	60.3%	2305	60.6%	2257	59.9%	2177	58.5%	
Female	1494	39.7%	1502	39.5%	1512	40.1%	1547	41.5%	
Geographic region									0.99
Northeast	1041	27.6%	1057	27.8%	1058	27.9%	1022	27.3%	
Southeast	827	22.0%	820	21.5%	838	22.1%	855	22.9%	
Midwest	965	25.6%	979	25.7%	980	25.9%	970	25.9%	
Southwest	464	12.3%	473	12.4%	448	11.8%	442	11.8%	
West	470	12.5%	478	12.6%	465	12.3%	452	12.1%	
Program type									0.98
Academic	2713	72.0%	2720	71.5%	2690	71.0%	2668	71.3%	
Community	1014	26.9%	1,048	27.5%	1059	28.0%	1033	27.6%	
Military	40	1.1%	39	1.0%	40	1.1%	40	1.1%	
PGY level									0.97
1	1126	29.9%	1120	29.4%	1092	28.8%	1054	28.2%	
2	793	21.1%	786	20.7%	779	20.6%	793	21.2%	
3	666	17.7%	677	17.8%	673	17.8%	662	17.7%	
4	606	16.1%	630	16.6%	637	16.8%	635	17.0%	
5	576	15.3%	594	15.6%	608	16.1%	597	16.0%	

**TABLE 2.**

Duty-hour Violations by Duty-hour Policy Arm and Survey Year, 2016–2018

	2016	2017	2018
>2 80-h wk in past month			
Standard	16.1%	14.5%	N/A
Flexible	19.8%	17.0%	N/A
	<i>P</i> = 0.03		
	<i>P</i> = 0.15		
>1 80-h monthly violation in past year			
Standard	N/A	23.3%	24.4%
Flexible	N/A	26.5%	24.7%
	<i>P</i> = 0.41		
	<i>P</i> = 0.21		
	<i>P</i> = 0.21		
	<i>P</i> = 0.91		

Reported *P*-values are from Rao-Scott design-corrected Chi-square tests with responses clustered by residency program. Not applicable; question was not asked on the survey during that year.



**TABLE 3.**

Association of Flexible Duty-hour Policy with Resident Outcomes by Survey Year

	2015			2016			2017			2018		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
<b>Satisfaction outcomes</b>												
Well-being	0.84	0.65-1.08	0.17	1.01	0.82-1.24	0.94	0.96	0.74-1.25	0.78	1.36	1.09-1.71	0.01
Continuity	2.23	1.65-3.02	<0.01	2.66	1.93-3.67	<0.01	2.09	1.42-3.06	<0.01	1.20	0.81-1.77	0.37
Patient safety	1.24	0.79-1.95	0.36	1.71	1.13-2.59	0.01	1.38	0.88-2.18	0.16	1.65	1.11-2.47	0.01
Work hours	1.08	0.83-1.40	0.56	1.04	0.81-1.33	0.75	1.08	0.81-1.43	0.60	1.36	1.04-1.77	0.02
Handoffs	1.53	1.16-2.02	<0.01	1.68	1.31-2.16	<0.01	1.76	1.36-2.28	<0.01	1.07	0.78-1.47	0.68
Overall education	1.00	0.72-1.39	0.99	1.21	0.89-1.65	0.22	1.20	0.85-1.69	0.30	1.11	0.82-1.50	0.49
Time for rest	0.82	0.64-1.06	0.13	0.84	0.69-1.06	0.14	0.93	0.73-1.18	0.56	1.25	1.00-1.56	<0.05
Program duty hours	1.01	0.74-1.39	0.93	0.91	0.66-1.25	0.55	0.80	0.55-1.19	0.27	1.32	0.95-1.82	0.10
Job	1.09	0.86-1.39	0.47	0.96	0.76-1.21	0.73	1.37	1.00-1.88	0.05	1.09	0.80-1.48	0.57
Decision to be a surgeon	1.00	0.77-1.30	0.99	0.93	0.73-1.18	0.55	1.14	0.80-1.63	0.47	1.15	0.82-1.62	0.42
<b>Continuity outcomes</b>												
Miss conference	0.51	0.39-0.65	<0.01	0.62	0.48-0.81	<0.01	0.60	0.45-0.80	<0.01	0.83	0.61-1.12	0.22
Handoff active patient care	0.51	0.42-0.61	<0.01	0.54	0.44-0.66	<0.01	0.60	0.48-0.74	<0.01	0.92	0.74-1.14	0.43
Handoff critical patient care	0.59	0.48-0.72	<0.01	0.59	0.48-0.72	<0.01	0.59	0.49-0.71	<0.01	0.91	0.72-1.14	0.42
Leave operation	0.47	0.32-0.68	<0.01	0.55	0.37-0.82	<0.01	0.53	0.35-0.80	<0.01	0.91	0.61-1.35	0.63
Miss an operation	0.56	0.44-0.70	<0.01	0.65	0.52-0.82	<0.01	0.58	0.46-0.72	<0.01	0.97	0.77-1.22	0.80
Return to hospital to operate	1.37	1.03-1.83	0.03	1.35	1.04-1.76	0.03	1.11	0.84-1.48	0.45	1.14	0.87-1.49	0.35
Return to hospital to provide patient care	1.36	0.96-1.92	0.09	1.34	0.95-1.87	0.09	1.18	0.86-1.61	0.30	1.08	0.79-1.48	0.61
<b>Perceived negative effects</b>												
Time for family/friends	3.44	2.59-4.55	<0.01	2.97	2.35-3.76	<0.01	1.65	1.27-2.15	<0.01			
Time for hobbies	3.51	2.67-4.62	<0.01	2.72	2.15-3.45	<0.01	1.57	1.23-2.01	<0.01			
Health	2.99	2.24-3.97	<0.01	2.57	1.99-3.30	<0.01	1.34	1.02-1.76	0.04			
Feeling well-rested	3.46	2.68-4.45	<0.01	2.76	2.15-3.54	<0.01	1.50	1.15-1.97	<0.01			
Ability to attend educational conferences	0.44	0.34-0.56	<0.01	0.43	0.34-0.55	<0.01	0.53	0.40-0.69	<0.01			
Availability for elective cases	0.32	0.25-0.40	<0.01	0.42	0.34-0.52	<0.01	0.59	0.48-0.74	<0.01			
Availability for urgent cases	0.21	0.17-0.26	<0.01	0.35	0.28-0.42	<0.01	0.47	0.39-0.57	<0.01			

	2015			2016			2017			2018		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Resident morale	1.03	0.81–1.32	0.81	1.00	0.78–1.27	0.99	0.84	0.63–1.13	0.25			

Results derived from separate multivariable logistic regression models with robust clustered standard errors for each year and survey question, adjusted for resident sex, postgraduate year level, geographic region, and program type.

CI indicates confidence interval; OR, odds ratio.