NBER WORKING PAPER SERIES

RETURNS TO TRAINEES IN A HIGHLY REGULATED LABOR MARKET

Michael R. Richards
Jonathan Seward
Christopher M. Whaley

Working Paper 34365
http://www.nber.org/papers/w34365

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2025, Revised April 2026

The authors thank Tom Koch for valuable feedback on earlier versions of this work and the organizers for the 2025 ASHEcon and 2025 Stata Microeconometrics conferences for opportunities to present this work. The authors also thank Brendan Rabideau for assisting with information gathering on physician residency programs. The authors have no conflicts of interest to disclose, and any remaining errors are their own. Access to Medicare data was provided through the RAND Center of Excellence on Health System Performance, which is funded through a cooperative agreement (1U19HS024067-01) with the Agency for Healthcare Research and Quality. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w34365

Returns to Trainees in a Highly Regulated Labor Market
Michael R. Richards, Jonathan Seward, and Christopher M. Whaley
NBER Working Paper No. 34365
October 2025, Revised April 2026
JEL No. I11, I18, J24, J44

ABSTRACT

Human capital formation enhances worker productivity and firm performance. However, firms may not fully internalize the financial benefits from investing in employee skill-building, which can disincentivize worker training and suggest roles for government support. We examine these tradeoffs in the context of physician training through "residency" programs. Using the universe of Medicare inpatient and outpatient records from 2014-2019 and differences-in-differences analyses, we show that hospitals' productivity and revenues grow by 10-20% after underwriting a procedure-oriented residency—though heterogeneity exists. The hospitals benefiting from trainee labor conservatively net over $2 million per year, revealing substantive private incentives to finance physician training.

Michael R. Richards
Cornell University
and NBER
michael.richards@cornell.edu

Christopher M. Whaley
Brown University
christopher_whaley@brown.edu

Jonathan Seward
Baylor University
Jonathan_Seward@baylor.edu

A online appendix is available at http://www.nber.org/data-appendix/w34365

I. Introduction

Physician training (i.e., "residency") programs comprise a multi-billion-dollar Graduate Medical Education—GME—enterprise that is funded through a mixture of public and private financing. GME also has a roughly century-long history and plays an influential role in the nearly $5 trillion US healthcare economy.[1] Residencies are structured as multi-year apprenticeship-styled programs for aspiring physicians following their completion of medical school, and at the conclusion of the minimum training years, graduates become eligible for licensing and billing credentials at the state level and specialty board certification at the national level.[2] The underlying rationale is that these residency programs further the educational experience for medical school graduates to better ensure downstream patient safety by providing extensive "hands-on" and specialized clinical training—typically over the course of three to seven years (Mowery 2015). While these are paid training positions, the salary levels are standardized, regulated, and low. Resident salaries average approximately $65,000 per year and have remained flat in real terms over several decades (Chandra, Khullar, and Wilensky 2014; Medscape 2021).

Beyond being immune to market forces, trainee salaries are also a fraction of a given hospital's cost of acquiring a fully trained physician (e.g., see Gottlieb *et al*. 2025). The fixed duration of the relevant GME program additionally provides teaching hospitals near certainty in terms of how many trainees they will have within a given specialization and for how long. Thus, while workers from a variety of occupations are known to accept lower wages when anticipating a future acceleration in earnings (Rosen 1972), this special and highly regulated labor market context affords hospitals added advantages: a guaranteed low pay rate and

---

[1] National US healthcare spending statistics are available from the Centers for Medicare and Medicaid Services (CMS) and can be found here: https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical.

[2] Of note, states vary on how many years of GME training must be completed to become eligible for a license. State-level descriptions can be found here: https://www.fsmb.org/step-3/state-licensure/.

negligible risk of trainee attrition.[3] Teaching hospitals are also permitted to bill payers at market rates for services directly rendered by their trainees under the supervision of attending (i.e., fully-trained) physicians.[4] Yet, the resulting balance of hospital costs relative to revenues from trainee labor inputs has been previously unquantified.

Despite this substantive unknown, economics research has a long-running interest in the human capital accumulation and labor market functioning tied to the medical profession. For example, economists have highlighted that medical societies' control over the accreditation of GME training creates perverse incentives to restrict supply and prop up higher wages for practicing physicians (Friedman and Kuznets 1945; Kessel 1958; Leffler 1978; Nicholson and Propper 2012). Estimates for physicians' lifetime returns to their human capital likewise demonstrate that the choice of residency program, and hence specialization, has profound implications (Nicholson 2002; Bhattacharya 2005) and that trainees' choices can respond to market signals (Chen, Lo Sasso, and Richards 2018). Economic research has also examined skill variation across residency programs (Epstein and Nicholson 2009; Doyle, Ewer, and Wagner 2010; Epstein, Nicholson, and Asch 2016) as well as variation in team composition and performance within a residency program (Chan 2021). However, broader effects on healthcare delivery, health outcomes, and economic impacts remain largely unknown (IOM 2014).

Importantly, almost no work has investigated the impact of GME on the hospitals themselves. Only a small literature has examined teaching hospitals preferences over physician trainees according to immigration status (e.g., Richards, Chou, and Lo Sasso 2009; Lo Sasso 2021). Rabideau, Richards, and Whaley (2024) offer a novel examination of the hospital-level effects of emergency medicine and obstetrics programs in Florida on treatment approaches and case volumes—finding treatment intensity to be lower and suggestive evidence for increased capacity

[3] Trainee quit rates are low, and residency program switching is very costly (e.g., trainee would typically have to restart the multiyear process).

[4] The Medicare public insurance program also provides an upward payment adjustment for "teaching hospital" status—i.e., those participating in GME (Sloan 2021).

following GME program debut. Yet, we are unaware of any contemporary research directly capturing the productivity and financial implications for sponsoring hospitals.

To shed light in this area, we leverage the universe of Traditional Medicare claims data from 2014 through 2019 and differences-in-differences (DD) analyses to estimate the impact of launching a general surgical residency training program on hospitals' inpatient and outpatient procedural productivity and revenues from the Medicare program. Medicare is the largest payer in the US healthcare system and the single largest source of funds for physician training (discussed in Section II below). We focus on general surgery programs since more hospitals have recently debuted such programs, and they are typically the largest surgical training programs in terms of size (i.e., number of participating residents). Additionally, surgical care is often one of the most profitable and therefore financially influential service lines for hospitals. Quantifying procedural output and associated revenues is also straightforward. The key services and payments are contained within the observed inpatient or outpatient procedural case, as opposed to the more nebulous long-run care management (e.g., primary care or cardiology medical practice), which may involve a host of ancillary services, supporting specialist referrals, multiple care sites, and difficult to define treatment bundle endpoints.

We ultimately find that hospitals newly introducing a surgical training program during our study period experience positive impacts on their overall procedural productivity across both the inpatient and outpatient settings. The trainee effects are dynamic and grow larger with time—increasing output and revenues by 10-20% within three years of program launch. We also show that our estimates are robust to the chosen method of DD estimation in the context of staggered treatment timing and that the increased throughput appears more consistent with business stealing than market expansion. Additionally, we demonstrate important heterogeneity in the effects, which at least suggest that the effects may be non-linear in program size. The observed effects are exclusively found among relatively larger training programs among treatment group hospitals. Hospitals investing in smaller programs demonstrate no differential increases in their procedure output or revenue streams. Importantly, baseline comparisons indicate that relatively larger program sites do not have meaningfully greater bed capacity or nurse staffing—both in levels and intensity—than the smaller surgical training programs, which indicates an influence of program size that is separate from hospital size. We also find that surgical trainee labor appears complementary with fully-trained surgeon labor, rather than operating as a substitute. Specifically, as hospitals become more productive over time, supplementary data suggests that they are also utilizing more fully trained surgeons to pair with their new trainees. A similar pattern is not evident for surgical capacity investments. We do observe modest growth in total patient beds; however, the growth is present for relatively larger and smaller new training programs, which is not consistent with driving more surgical procedures overall for the larger training programs, as indicated in our main findings. Taken together, the results for these other hospital operational margins suggest that the observed productivity and revenue changes following a surgical residency program debut are not simply manifestations of a broader strategy and effort to increase hospital throughput and revenue streams.

A simple return-on-investment (ROI) exercise shows that in steady-state, the marginal improvements in Medicare revenues are likely insufficient to cover the costs of launching and sustaining a new program. However, with some conservative assumptions around additional *non*-Medicare revenue that could materialize from expanded surgical human capital and procedural capacity enabled by trainee-attending physician complementarities among affected hospitals, net positive returns become achievable. Assuming only weakly favorable bargaining terms with respect to non-Medicare payers suggests more than $2 million in positive cash flow for the average hospital investing in a sufficiently sized surgical trainee program. While federal and state governments have long used public funds to subsidize physician training programs (discussed in Section II), and regularly debate the merits of increasing such funds, our results demonstrate scope for the net financial returns to sponsoring hospitals to justify their private investments in further trainee expansions—at least among traditionally profitable service lines for the industry. Shifting more of the annual GME underwriting to certain hospitals and/or certain specialties within hospitals could allow public funds to be better targeted toward hospitals or programs unable to sustainably self-finance physician training. That said, if hospital-level gains mostly reflect recaptured volume rather than market expansion, the net welfare effects will depend on relative costs and quality differences across gaining and losing providers over the long-run.

II. Brief Background on US Graduate Medical Education

Modern US healthcare requires physicians to be licensed, which includes the successful completion of medical school and an accredited residency training program (AMA 2018; Sloan 2021). Residency programs are typically hospital-based and accredited by the Accrediting Council of Graduate Medical Education (ACGME), which is comprised of multiple professional organizations, including those involved in US medical school accreditation as well (Ludmerer and Johns 2005; ACGME 2022). The economic tools of market design built the now well-known two-sided match between hospitals and physician trainees that takes place each year (Roth 1984; Roth and Peranson 1999). The Centers for Medicare and Medicaid Services (CMS) is also a key sponsor for physician training programs. CMS spends approximately \$15 billion annually on physician residency subsidies, with the average training hospital collecting roughly \$15 million per year to support its GME programs (GAO 2021). A hospital hosting an accredited training program can expect direct GME payments to cover resident and faculty salaries—including overhead—as well as indirect payments to offset other costs of being a teaching hospital (Iglehart 1999; Wynne *et al*. 2006; Cox *et al*. 2009; IOM 2014). Though, the GAO estimates that approximately 70% of GME hospitals use internal funds to support their residency programs to some degree (ACGME 2021; AHA 2021).

Due to ballooning physician training expenditures for the Medicare program since its inception, Congress capped GME funding at 1996 levels via the Balanced Budget Act of 1997 (Iglehart 1999; IOM 2014). This cap has forced residency programs launched after BBA ‘97 to both seek formal approval *and* often self-finance the endeavor, rather than rely on public subsidies from Medicare.

Figure 1 captures the ebbs and flows for residency program founding years across all physician specialties from the early-20th century to present day. The slowdown after BBA '97 is evident; however, the late 2010s demonstrate some of the most active years for new residency launches.

Importantly, the federal government has recently appropriated funding for expanding publicly supported GME positions throughout the US—the first instance of such policy action in more than two decades. However, the magnitude has been modest (i.e., 1,000 positions), with further increases under consideration for future policy (Boyle 2020; Hughes 2021; Hellmann 2022). These current and potential future initiatives tied to physician training, coupled with various warnings of physician shortages, have also reinvigorated the debate on how much GME the US healthcare system needs and how should any eventual expansions be financed.

III. Empirical Strategy

A. MEDICARE CLAIMS DATA

We rely on the universe (i.e., 100%) of Traditional Medicare fee-for-service (FFS) inpatient and outpatient hospital claims spanning 2014 through 2019.[5] As previously noted, the Medicare program is the most prominent financier of physician GME training programs and is therefore assumed by policymakers—at least implicitly—to represent a patient population expecting to strongly benefit from training programs' existence and expansion. Medicare beneficiaries also commonly require higher complexity and procedural-focused care compared to other (younger) patients, and this already large segment of the US population is only growing as the country's demography shifts older.

Within the Medicare claims data, we focus on the quantities and associated hospital payments tied to procedure-oriented care—typically surgical services, such as breast surgery, colorectal surgery, or hip/knee replacements for Medicare beneficiaries. We focus on procedures and surgeries primarily for two reasons: 1) output, and hence productivity, is easy to track, and 2) these service lines are some

[5] Usefully, 2019 is the last full calendar year that was not impacted by the COVID-19 pandemic, which dramatically disrupted hospitals' operations, including for procedural and surgical services.

of the most profitable activities hospitals engage in— so much so that they often cross-subsidize other, loss-making hospital units and forms of care delivery (e.g., maternity care, trauma care, and charity care). All relevant inpatient and outpatient procedural services are identified via the corresponding Berenson-Eggers Type of Service (BETOS) codes used by the Center for Medicare & Medicaid Services (CMS) as part of standard Medicare billing. Specifically, services in the "P" BETOS category, with the exception of "P0" (anesthesia), "P7A/B" (oncology), and "P9A/B" (dialysis), are considered valid procedural and surgical care for our analytic purposes. After identifying all procedural and surgical Medicare claims and corresponding hospital payments, we aggregate to the hospital-half-year level to capture the total volume of procedures performed in each setting (i.e., inpatient and outpatient) for the Medicare market and the resulting Medicare revenue (in nominal dollars) flowing to the hospital over time.

B. ESTIMATION

Our information on US physician residency program founding dates is from publicly provided information reported on the physician social networking platform, Doximity.[6] The platform is both popular among physicians and advertises that roughly 90% of medical students engage with its "Residency Navigator" tool. From the publicly available information, we could gather a given program's founding date, specialization, size, and affiliated hospitals—including the estimated percent of time a typical resident could expect to work at a particular hospital. One drawback of this information is that we cannot observe residency programs that have been terminated; however, this is likely to be a rare event in the US healthcare system and has no material bearing on our empirical findings.

We ultimately identify 26 general acute care hospitals that introduced a new, general surgery residency program in either July of 2015, 2016 or 2017. The 26 participating hospitals belong to 22 different hospital referral regions (HRRs)

[6] Of note, the platform requires a user account and login for much of its features. However, no account was required to recover the program founding dates needed for our analyses—i.e., only public facing information available to any internet user was referenced.

spanning 14 states—including larger states such as, California, Florida, Michigan, and New York. We chose to focus on general surgery programs since they are the most common surgical training programs, tend to be larger in terms of cohort size, and expose trainees to a wide variety of procedural care, which they begin to take part in relatively early in the surgical residency experience.

Appendix Figure 1 demonstrates that the general surgery GME programs exhibit a spike in launches during the late 2010s that is consistent with the broader trend across all physician specialties previously observed (Figure 1). We exclude from the analytic sample any hospitals debuting similar residency programs after 2017 (due to limited post-period observations) or hospitals participating in a relevant residency program but only capturing a small minority of the training physicians' allocated clinical time over the duration of the program.[7] We also excluded a few hospitals that introduced a concurrence or sequence of surgical residency programs during our study period (e.g., general surgery in one year and then followed by a subspecialized field, such as orthopedic surgery, the following year). Our control group hospitals are those not launching a surgery residency program during our analytic period. As part of our robustness tests (described below), we also ensure that our findings are insensitive to further restricting the control group to out-of-market hospitals (i.e., those located in different HRRs than the 26 treatment group hospitals).

Table 1 displays the half-year summary statistics for our 26 treatment group hospitals and 2,933 control group hospitals for the 2014 baseline year. On average, the hospitals launching new general surgery residency training programs perform more inpatient and outpatient procedures at baseline and consequently receive greater revenue flows from the Medicare program. They also have about twice as many non-training physicians performing procedural cases at the hospital at baseline.

---

[7] Residency programs often span multiple hospitals. Our information on individual residency programs lists the specific hospitals where resident physicians will rotate as well as the percentage of time each resident is expected to spend at a given hospital participating in the residency program. We only considered hospitals capturing at least 30% of the residents' training time for treatment group inclusion.

We estimate the impacts of starting a new procedure-oriented residency program via a straightforward differences-in-differences (DD) framework. Our initial estimations rely on the standard two-way fixed effects specification since residency programs uniformly begin in July of each year and all treatment group hospitals are restricted to those with 2015, 2016, or 2017 debut dates. Additionally, 16 general surgery programs begin in 2016, specifically, with another 8 beginning in 2017. Our donor pool of control group hospitals is also large. These contextual features inherently limit the scope for empirical issues tied to the differential timing in treatment (e.g., see Goodman-Bacon 2021). We further address such estimation concerns below in Section IIIC.

We adapt our DD estimation to an event study model because treatment dynamics are expected in this analytic setting—given that programs will be adding residents over multiple years—and we wish to formally test for pre-launch parallel trending between our treatment group hospitals and those in the control, comparison group.

$$Y_{ht} = \sum_{\substack{j=-4 \\ j\neq-1}}^{-1} \alpha_j \left[\mathbf{1}(t - D_h = j)\right] + \sum_{j=0}^{6} \delta_j \left[\mathbf{1}(t - D_h = j)\right] + \theta_h + \lambda_t + \varepsilon_{ht}$$

Equation (1) captures our outcomes of interest (Y_{ht}) at the hospital half-year level and include hospital (θ) and quarter-year (λ) fixed effects. The D_h indicator variable is set to one for hospitals belonging to the treatment group and zero otherwise. The α_j coefficients formally test for differential trending in the two years preceding the procedural residency program launch while the series of δ_j coefficients flexibly allow for differential changes in outcomes for our treated hospitals up to three years following the start of the hospital's new physician surgical training program. We also create pooled DD estimates by using the standard two-way fixed effects estimation found in Equation (2) below:

$$Y_{ht} = \beta Implemented_{ht} + \theta_h + \lambda_t + \varepsilon_{ht}$$

The binary treatment variable, $Implemented_{ht}$, turns on when the relevant hospital first starts its general surgery GME program during the second half of 2015, 2016, or 2017. The standard errors are clustered at the hospital level throughout.

C. ROBUSTNESS

We also implement a series of robustness checks to validate our inferences from the standard two-way fixed effects estimations previously described in Section IIIB. First, we test the sensitivity of our main findings and inferences to the use of alternative DD estimators (Cengiz *et al*. 2019; Borusyak, Jaravel, and Spiess 2021; Callaway and Sant'Anna 2021) and present the resulting event study results side-by-side for ease of comparison. As previously mentioned, there is expectedly lower risk of bias from differential treatment timing in our setting due to 16 of 26 treatment events occurring in a single year (2016) and another 8 treatment events occurring in the immediately following year (2017). Nevertheless, we show that our findings are robust to these other DD analytic strategies. Second, we refine the control group hospitals to only include those that operate in markets (i.e., HRRs) separate from the 22 HRRs where our treatment group hospitals are located. This ensures that any local market spillover effects do not contaminate our DD estimates and hence preserves the SUTVA requirement. And finally, given the disparate surgical procedure volumes noted in Table 1, we rerun our main analyses when requiring the control group hospitals to have performed at least 150 inpatient surgical procedures at baseline and thus be no more than one standard deviation below the treatment group mean (Table 1). As presented and discussed in Section IVA, none of these modifications to our estimation approach or analytic sample construction have a material impact on our empirical findings or interpretations.

IV. Results

A. MAIN FINDINGS

Figure 2 presents our key result. Treatment group hospitals' procedural output trends closely track the control group hospitals over the two years leading up to the surgical training program launch date. Then, during the subsequent years after the new residency program is in place, treatment group hospitals demonstrate a differential increase in inpatient procedure volume (Panel A), which begins with the residency launch and grows over time. By the third year following the residency debut, affected hospitals generate 20-25 more inpatient procedures per half-year, which translates to an 8-10% relative increase over their baseline levels (Table 1). Changes for outpatient procedural volume (Figure 2, Panel B) are more equivocal, but there is at least a suggestive increase in productivity that culminates in approximately 130 more outpatient procedures per half-year by the end of our study period—roughly 12% more than their baseline level reported in Table 1. Figure 3 shows that our event study estimates and accompanying inferences are not meaningfully changed by the use of alternative DD estimators, which again is unsurprising within our analytic context. Appendix Figure 2 reproduces the event study findings when excluding any within-market (i.e., same HRR) control group hospitals. The results are virtually identical to Figure 2. Likewise, when imposing that control group hospitals have to be relatively higher volume at baseline (Appendix Figure 3), the findings are largely indistinguishable from Figure 2, which further support our prior interpretations.

Next, we move to our revenue-focused event study results in Figure 4. Increased productivity may not necessarily translate to greater revenue flows, if, for example, incorporation of trainee labor leads to more case volumes with a simultaneous shift to lower complexity—and hence lower reimbursing—procedures. Figure 4 makes clear that this is not the case. Following their training program launches, treatment group hospitals are differentially adding between \$200,000 and \$700,000 per half year in Medicare inpatient procedural revenue (Panel A, Figure 4) or 5-19% (Table 1) within the first three years of program launch. Similarly, in Panel B of Figure 4, we see a rise in Medicare outpatient revenue—albeit slower to materialize. Three years after the surgical residency launch, the average treatment group hospital gains approximately $200,000 (18% over baseline in Table 1) in Medicare outpatient procedural revenue per half-year. The commensurate, if not greater, increase in revenues following the observed change in productivity suggests that the expansion of surgical cases is not leading hospitals to focus on a more select set of services (e.g., smaller, uncomplicated procedures). The upshot is that over a half-million dollars in combined additional revenue from the Medicare market accrues to the hospital every six months within two years of its surgical training program's inception. Just as before, our findings are insensitive to use of alternative DD estimators to address differential timing in treatment (Figure 5).

Table 2 provides the pooled post-period DD estimates using Equation (2). Consistent with what is evident in Figures 2 and 4, the pooled DD estimates are at an intermediate level when comparing the short- and long-run dynamics observed in the event study findings. The pooled DD estimates are also more precise for the inpatient procedural volumes and revenues, which aligns with the previously displayed event study estimates.

B. BUSINESS STEALING VERSUS MARKET EXPANSION

Section IVA established our key findings from the introduction of a procedure-oriented physician training program—namely, a hospital-level productivity increase that simultaneously enhances the hospital's inpatient and outpatient Medicare revenue streams. Yet, our previous results do not reveal if affected hospitals' improvements in procedure output are due to cases no longer being diverted to a competing hospital (i.e., business stealing) or from a general increase in the supply of procedural services to Medicare patients overall (i.e., market expansion). The latter could be indicative of latent demand—and hence an improvement in access for Medicare beneficiaries. However, it could also be consistent with supplier-induced demand, which would suggest inefficient care delivery (i.e., "low-value" services) and possibly worsening consumer and social welfare.

To shed light on where the marginal cases are coming from, we move to a market-level DD analyses where the HRRs home to our treatment group hospitals from Section IV are considered treated while all other HRRs across the US are used as controls. The estimation then proceeds similarly as to what was described in Section IIIB, with the exception of an HRR-half-year procedural output aggregation and the use of market (m) level fixed effects and clustering.

$$Y_{mt} = \sum_{\substack{j=-4 \\ j\neq -1}}^{-1} \alpha_j \left[\mathbf{1}(t - D_m = j)\right] + \sum_{j=0}^{6} \delta_j \left[\mathbf{1}(t - D_m = j)\right] + \mu_m + \lambda_t + \varepsilon_{mt}$$

Appendix Figure 4 offers the corresponding event study results. Hospital markets exposed to a new surgical residency program debut are trending with unexposed markets before the training program launch date, with no clear indications of aggregate procedure volume increases in the post-period. The event study estimates for inpatient procedural cases are smaller (by 70-80%) in magnitude than what is observed in our hospital-level DD analyses and lack precision as well as the previously documented dynamic pattern. The estimates cannot rule out some degree of inpatient surgery market expansion in Appendix Figure 4, but the event study pattern is not strongly compelling. Moreover, the outpatient procedural volumes tend to not only be statistically insignificant at conventional levels, but also the post-period coefficients are negatively signed throughout. Taken together, there is no clear evidence that the Medicare procedural care market expands after new physician training programs are introduced—at least not in the short-run. The findings from Appendix Figure 4, coupled with those belonging to Figure 2, seem likely to indicate that hospitals are able to steal business from local competitors after launching the relevant residency program.

Given that healthcare consumers are often weakly aware (or unaware) of hospital staffing nuance and complexity, it is perhaps unlikely that a new GME program positively would appreciably shift demand towards the hospital. In fact, if consumers are aware of the new training program, one might expect a *negative* shift in demand since patients may wish to avoid less experienced care providers, especially surgeons. A more plausible explanation is that the introduction of new trainee labor relaxes previous capacity constraints for the hospital, which allows the hospital to avoid forgoing cases it would otherwise accept in the presence of sufficient staffing (i.e., the necessary physician labor inputs).

C. HETEROGENEITY AND COMPLEMENTARY LABOR INPUTS

As previously noted, physician GME residency programs vary in their size and in the amount of effective resident labor time available to a given hospital as residents rotate through multiple hospitals over the duration of their training programs. In this section, we explore this potential source of heterogeneity to see if there are "scale effects" for the positive contributions of physician trainee labor to hospitals' productivity and payments.

To do so, we simply multiply the steady-state size of each general surgery residency program in our treatment group by each hospital's corresponding claim on a given resident's work time. This calculation produces a full-time equivalent (FTE) measure for each treatment group hospital's general surgery GME investment. We then stratify our treatment group by the median value (9.6 FTE general surgery training physicians) and re-estimate Equation (1) and Equation (2) twice—one for each half of the treatment group hospitals, with the set of control group hospitals fixed across estimations. However, before estimating, we also benchmark baseline capacity among treated hospitals: high-FTE sites are not meaningfully larger in terms of bed capacity and, if anything, these hospitals rely on lower nurse staffing—both in levels and intensity—relative to low-FTE training program sites.[8]

Figure 6 shows that larger programs drive the gains despite no advantage in baseline size or nurse intensity, as noted immediately above. Inpatient volumes are

[8] The specific descriptive statistics are as follows: Beds (mean [SD]): Low-FTE 357.25 [147.83] vs High-FTE 380.93 [160.95]. RN FTE: 728.50 [288.11] vs 658.36 [399.44]. RN FTE per bed: 2.07 [0.42] vs 1.68 [0.34]. RN FTE per 1,000 adjusted patient days: 4.67 [0.79] vs 3.75 [0.75]. N=26 treated hospitals, 2014 baseline; split at the median trainee FTE.

flat for smaller programs (Panel C), and outpatient results are at most suggestive (Panel D). Thus, within Figure 6, the findings suggest that the relatively larger surgical programs are key drivers of our previous results—and hence witness clear productivity benefits from the new training program investment. The most conservative interpretation is that relatively smaller programs do not generate higher procedural volumes after incorporating surgical physician trainees into their workforce. The findings in Figure 7 also closely align with those from Figure 6. The Medicare revenue gains are predominantly, if not exclusively, found among the top half of our treatment group training programs in terms of size. Table 3 offers the pooled DD estimates when focusing on this specific subset of treated hospitals. Consistent with Figures 6 and 7, the pooled DD estimates are statistically significant across all four outcome measures, and when compared to these hospitals' baseline levels, the DD estimates reflect productivity and revenue gains of approximately 8-10% and 13-16%, respectively.

While Section IVA shows productivity and revenue benefits materializing for hospitals following the debut of a new surgical physician training program, the findings in Figure 6, Figure 7, and Table 3 demonstrate that the impact is potentially dependent on program size and may not necessarily scale linearly. This is potentially important nuance when weighing the costs and benefits of public versus private financing of GME expansions as well as when contemplating where subsidies should be targeted. But even for the relatively larger programs that appear to drive the overall effects previously observed, the presence of new physician trainee labor could also generate greater indirect costs for hospitals, which would need to be balanced against the previously estimated gains.

More specifically, the presence of relatively lower skilled physicians, among this high-skill profession, could necessitate the hiring of additional fully-trained physicians to oversee, teach, and collaborate with the new trainees. Put differently, training and non-training physician labor could be complementary inputs engaged in joint production on the margin—at least when hospitals launch a novel training program. If true, this could drive higher labor costs for the hospital overall when incorporating the new trainees (i.e., beyond the direct labor costs from internally financing the new residency positions).

We test for evidence consistent with labor complementarities, and hence higher labor costs for hospitals, by re-estimating our original, hospital-level DD event study specification on the relatively larger residency programs with the outcome of interest being the number of unique non-training physicians billing for procedural cases at the hospital over time. Non-training physicians are identified using the National Provider Identification (NPI) number attached to a given billing physician and then matched to the publicly available Physician Compare data from CMS. We use the latter information to recover the year of medical school graduation and subsequently exclude from the hospital-level physician count any physician that graduated after 2009 since more recent graduates could plausibly still be in a training program. This analytic step ensures that we do not conflate increases in non-training (or "attending") physicians with an overall increase following the arrival of physician trainees, which can and do bill the Medicare program under their own NPI. We also impose that any observed increase in non-training physician labor is *employed* labor for the hospital, rather than new community physicians bringing their cases to the hospital for the first time. Doing so allows for highly conservative hospital cost estimates from hiring additional, fully-trained physician labor.

Figure 8 displays the event study results for the number of non-training physicians working at the hospital and again stratified by the relative size of the training program. Panel A of Figure 8 reveals an increase in complementary (non-training) physician labor following the surgical residency debut, which seems to grow over the first three years of the program's existence. Conversely, Panel B of Figure 8 shows no differential changes in the quantity of non-training physicians (employed or otherwise) at the comparatively smaller new surgical residency programs from our treatment group sample. These programs neither experience clear gains in terms of productivity and revenues but also do not commit to new outlays in terms of hiring more fully trained ("attending") physicians. Because market-level procedure totals do not obviously expand over our study horizon, the hospital-level gains for comparatively larger training programs are best interpreted as recaptured volume rather than responses to new or latent demand.

The findings in Figure 8 therefore indicate that hospitals potentially bear more indirect costs from investing in procedure-oriented physician training programs through expanded hiring of fully trained physicians. Additionally, the observed larger program advantage is unlikely to reflect simple scale; rather, it is consistent with trainee–attending complementarities that raise the marginal product of attending time and peri-operative inputs within the hospitals positioned to deploy them.

V. Hospital Capacity and Capital Investments

Hospitals that launch new general surgery residency programs may do so as part of broader operational changes and strategic motivations—meaning a residency launch could coincide with expansion of inpatient capacity, increased utilization pressure, or shifts in healthcare workforce and staffing. The observed effects in Section IV would consequently capture more than the incremental impact of trainee labor alone. Distinguishing between these possibilities is important for interpreting the previously estimated changes in volume and revenue around training program adoption.

To help disentangle competing interpretations, we leverage supplementary information at the hospital-level drawn from the American Hospital Association (AHA) annual surveys over our study period. Our supplementary empirical exercise focuses on two key measures of capital investments: total beds and total operating rooms. These margins intentionally reflect substantive capital investments by hospitals and capture “hard” capacity. If residency adoption tends to occur alongside hospital-wide expansion, such as adding staffed capacity or substantially increasing surgical resources, then we would expect to observe discrete changes in these measures around the training program launch window. Conversely, if residency program adoption primarily reflects changes in the availability and organization of clinical labor, with capacity largely held fixed, then contemporaneous and meaningful shifts in these hospital-wide metrics would be less likely.

Figure 9 offers the event-study estimates for the AHA capacity outcome measures, with the treatment group again stratified by relative size of the newly launched training program. The total number of staffed beds does modestly increase in Figure 9 (Panels A and B); however, the pattern is present for relatively large and small training programs—as opposed to localizing to the larger programs where we previously documented procedural and revenue gains. There are no indications that the debut of a new general surgery GME program corresponds to an expansion in operating room capacity. The event-study evidence in Figure 9 therefore does not align with an interpretation that the procedure and revenue increases for the Medicare market are corresponding to greater hospital capacity overall. There is not a large-scale shift in hospital-wide capacity at the time of the surgical residency introduction, and even if the number of beds is increasing, it is also increasing among the smaller training programs, which show no differential change in their surgical care throughput and revenues.

VI. Hospital Return on Investment

We conclude our empirics by implementing a simple back-of-the-envelope exercise to provide a reasonable measure of the return on investment (ROI) for hospitals self-financing a new surgical training program. Understanding and quantifying such private incentives can usefully inform ongoing debates over the presence, degree, and targeting of public funds devoted to GME.

This exercise involves a few simplifying assumptions and analytic choices. First, we focus on the relatively larger programs from our treatment group hospitals since that subset demonstrated the stronger and clearer productivity and revenue effects (Section IVC). Second, we impose $400,000 as the assumed annual salary cost for each additional non-training physician hired by the hospital during our analytic time period.[9] This figure is based on recent academic work as well as trade

[9] To get the effective number of additional non-training hires, we multiply the final event study coefficient for number of non-training physicians (Figure 8) by the median percent of residents'

press information (e.g., see Whaley *et al*. 2021; Gottlieb *et al*. 2025; Taylor 2025) and may be an over-estimate due to known wage penalties and lower earnings trajectories over the working life course for employed physicians (e.g., Whaley *et al*. 2021). We also assume an additional 30% (relative to salary) in labor costs for both the additional resident and non-training physicians to cover non-wage benefits.[10] Third, we assume that productivity gains would extend beyond the Medicare market since there are no regulatory limits that restrict trainee labor contributions to just Medicare patients. Transaction prices are not readily available for other payer markets and typically depart from Medicare reimbursement rates in both directions (e.g., Gottlieb, Hale Shapiro, and Dunn 2018; Spivack *et al*. 2021; Meiselbach *et al*. 2023). To account for this, we perform the exercise under different "Medicare markup" scenarios whereby we consider the markup to be net of all non-Medicare payers, which is inherently a mix of positive (e.g., commercial payers) and negative (e.g., Medicaid) adjustments. Finally, we impose that the productivity and revenue gains stabilize at our final observed post-launch event study time period, as opposed to continuing to grow dynamically in subsequent (unobserved) years, and we ignore any other potential revenue streams tied to additional training labor and/or employed non-training physicians (e.g., physician billings, outpatient office visits, ancillary services such as labs and imaging, etc.). We believe this combination of analytic decisions should, at a minimum, provide an appropriately conservative ROI estimate—and perhaps offer a lower bound, at least among hospitals with sufficiently large new GME programs to generate higher procedural throughput.

Table 4 describes the annualized costs and reimbursement changes for these hospitals in steady-state (i.e., based on the final observed post-period and assumed to stabilize from then on). Hospitals experience over $1 million in training labor

time the hospital is expected to receive, which is roughly 75% among the larger programs in our treatment group. In other words, we assume that their hiring is proportional to the increase in FTE trainee labor.

[10] We consider resident salaries to be $65,000 during our analytic period, which aligns with known national figures and reflects that trainee salaries demonstrate limited growth over their programs' durations (Chandra, Khullar, and Wilensky 2014; Medscape 2021).

costs, which would be more than offset by the Medicare revenue gains of $2 million per year. However, the potential for large increases in non-training complementary labor costs (almost $4 million) overwhelm the Medicare revenue increases and leave the hospital with a nearly $3 million net negative cash flow, so long as the hospital is unable to generate additional revenue in any other (non-Medicare) payer markets. Given that Medicare is approximately 30% of these hospitals' total revenues, columns 2-4 of Table 4 explore alternative scenarios that assume the empirically observed increase in Medicare revenues (Section IV) are also only 30% of the total increases (i.e., across all payer markets) the hospital experiences. Under a no net markup over Medicare scenario (column 2), the hospital would realize a profit of $2.57 million in steady-state— a return of 51% relative to their annual training program outlay (i.e., the combined labor costs at the top of Table 4). With even modest net markups over Medicare due to more favorable negotiations with non-Medicare payers and/or a more favorable payer mix, the hospital's positive cash flow could exceed $3 million per year—again, among hospitals with sufficient scale for their surgical training program.

VII. Conclusions

With projections of looming physician shortages, some provider practices and related policy activity have targeted physician substitutes, albeit with mixed success (e.g., see Barnes *et al*. 2018; Martsolf *et al*. 2018; Traczynski and Udalova 2018; Alexander and Schnell 2019; Auerbach, Buerhaus, and Staiger 2020; Markowitz and Adams 2022; Barnes *et al*. 2023; Chen *et al*. 2023; Hoehn-Velasco *et al*. 2023; McMichael 2023; Dillender *et al*. 2024; Guo, Kilby, and Marks 2024; McMichael 2025). Despite these efforts, many industry stakeholders and policymakers view expanded physician supply as inevitable and necessary for the US healthcare system over the medium- to long-run. Achieving this aim would ultimately require a larger funnel of physician trainees entering GME programs; otherwise, the supply of fully trained physicians remains approximately fixed (i.e., the labor inflows and outflows maintain the status quo balance). But more residency training would also necessitate additional outlays to support an expanded GME system, with the incidence of the resulting financial burden open for debate.

Our study demonstrates that new residency programs have recently formed through significant private financing, and under certain conditions, hospitals appear to recoup their investments and ultimately profit from underwriting physician training. Launching a surgical training program generates 10-20% improvements in procedural productivity and revenues—seemingly via recaptured business from local competitors—within three years of program debut. However, sufficient program scale (in terms of number of physician trainee FTEs) could be important. The bottom half of our observed new surgical residencies fail to demonstrate any uptick in productivity, at least during the first few years following GME program implementation. On the other hand, under conservative assumptions, the hospitals investing in relatively larger training programs seem capable of generating $2 million or more per year of net income in steady-state, even after accounting for higher labor costs tied to complementary inputs (i.e., fully trained and supervising physicians). Such heterogeneous findings are at least consistent with an underlying critical size threshold—though not definitive—and baseline comparisons indicate that 'high-FTE' training sites are not meaningfully larger in terms of bed capacity and, if anything, rely on lower nurse staffing—measured in levels or intensity.[11]

Additionally, the hospitals already able to self-finance these recent program launches—and, relatedly, to successfully navigate the GME regulatory process—may have been in a stronger position to benefit from more training labor than the marginal hospital today. Diminishing returns could also materialize as more programs debut in a given market and fewer cases remain to compete away—although increasing demand from an aging population could counterbalance risks from local GME "congestion" and competition for patients.

[11] It remains a possibility that dichotomizing the treatment group along the physician trainee FTE dimension is also dividing the treatment group according to some other, highly correlated and influential factor. Though, baseline comparisons indicate that the 'high-FTE' hospitals are not markedly larger with respect to inpatient size and are actually less nurse-intensive with respect to staffing than 'low-FTE' training hospitals in our treatment group.

Economists have long remarked that investments in employees' human capital have clear benefits for workers but ambiguous effects on participating firms, including whether firms can internalize the downstream returns. Such tensions could limit privately funded training opportunities and suggest roles for government intervention and public subsidies (Pigou 1912; Rosenstein-Rodan 1943; Acemoglu and Pischke 1999). Given the historical precedent and our findings, a mixture of public and private financing for US physician training is likely to persist in this special and highly regulated labor market during the present and foreseeable future. However, our evidence suggests scope for potentially better targeting of public funds to certain hospitals and/or certain types of GME programs. We therefore view this work as a first step in the type of economic analyses that can better inform more nuanced discourse and decision-making around the accompanying tradeoffs in this area of health policy. Future empirical work can usefully do more to quantify both sides of the ledger for training hospitals—i.e., the private costs as well as the private benefits from investing in physician residency programs.

REFERENCES

Acemoglu, Daron, and Jörn-Steffen Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." *Economic Journal*, 109 (453): F112-142.

ACGME. 2021. "Data Resource Book." https://www.acgme.org/globalassets/pfassets/publicationsbooks/2020-2021_acgme_databook_document.pdf.

AHA. 2021. "GAO: 70% of Teaching Hospitals Are Self-Funding Residency Slots." June 22, 2021. https://www.aha.org/news/headline/2021-06-22-gao-70-teaching-hospitals-are-self-funding-residency-slots.

Alexander, Diane, and Molly Schnell. 2019. "Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health." *Journal of Health Economics*, 66: 145-162.

AMA. 2018. "Obtaining a Medical License." American Medical Association. May 15, 2018. https://www.ama-assn.org/residents-students/career-planning-resource/obtaining-medical-license.

Auerbach, David I., Peter I. Buerhaus, and Douglas O. Staiger. 2020. "Implications of the Rapid Growth of the Nurse Practitioner Workforce in the US." *Health Affairs*, 39 (2): 273-279.

Barnes, Hilary, Michael R. Richards, Matthew D. McHugh, and Grant Martsolf. 2018. "Rural and Nonrural Primary Care Physician Practices Increasingly Rely on Nurse Practitioners." *Health Affairs*, 37 (6): 908-914.

Barnes, Hilary, Alice J. Chen, Matthew D. McHugh, and Michael R. Richards. 2023. "Public Insurance Expansions and Labor Demand in Physician Practices." *Journal of Human Resources*, online ahead of print: https://doi.org/10.3368/jhr.0920-11176R2.

Bhattacharya, Jayanta. 2005. "Specialty Selection and Lifetime Returns to Specialization within Medicine." *Journal of Human Resources* XL (1): 115-143.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." arXiv Preprint: 2108.12419. https://doi.org/10.48550/arXiv.2108.12419.

Boyle, Patrick. 2020. "US Physician Shortage Growing." AAMC. June 26, 2020. https://www.aamc.org/news-insights/us-physician-shortage-growing.

Callaway, Brantly and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225 (2): 200-230.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *Quarterly Journal of Economics*, 134 (3): 1405-1454.

Chan, David C. 2021. "Influence and Information in Team Decisions: Evidence from Medical Residency." *American Economic Journal: Economic Policy*, 13 (1): 106-137.

Chandra, Amitabh, Dhruv Khullar, and Gail R. Wilensky. 2014. "The Economics of Graduate Medical Education." *New England Journal of Medicine*, 370 (25): 2357-2360.

Chen, Alice, Anthony T. Lo Sasso, and Michael R. Richards. 2018. "Supply-Side Effects from Public Insurance Expansions: Evidence from Physician Labor Markets." *Health Economics*, 27 (4): 690-708.

Chen, Alice J., Elizabeth L. Munnich, Stephen T. Parente, and Michael R. Richards. 2023. "Provider Turf Wars and Medicare Payment Rules." *Journal of Public Economics*, 218 (February): 104812.

Cox, Malcolm, David M. Irby, Molly Cooke, William Sullivan, and Kenneth M. Ludmerer. 2009. "American Medical Education 100 Years after the Flexner Report." *New England Journal of Medicine*, 10 (1): 131-133.

Dillender, Marcus, Anthony T. Lo Sasso, Brian J. Phelan, and Michael R. Richards. 2024. "Occupational Licensing and the Healthcare Labor Market." *Journal of Human Resources*, online ahead of print: https://doi.org/10.3368/jhr.0722-12450R2.

Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner. 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics*, 29 (6): 866-882.

Epstein, Andrew J., and Sean Nicholson 2009. "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections." *Journal of Health Economics*, 28 (6): 1126-1140.

Epstein, Andrew J., Sean Nicholson, and David A. Asch. "The Production of and Market for New Physicians' Skill." *American Journal of Health Economics*, 2 (1): 41-65.

Friedman, Milton, and Simon Kuznets. 1945. "Income from Independent Professional Practice." National Bureau of Economic Research.

GAO. 2021. "Caps on Medicare-Funded Graduate Medical Education at Teaching Hospitals." https://www.gao.gov/assets/gao-21-391.pdf.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225 (2): 254-277.

Gottlieb, Joshua D., Adam Hale Shapiro, and Abe Dunn. 2018. "The Complexity of Billing and Paying for Physician Care." *Health Affairs*, 37 (4): 619-626.

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2025. "The Earnings and Labor Supply of US Physicians." *Quarterly Journal of Economics*: 1243-1298.

Guo, Jiapei, Angela E. Kilby, and Mindy S. Marks. 2024. "The Impact of Scope-of-Practice Restrictions on Access to Medical Care." *Journal of Health Economics*, 94: 102844.

Hellmann, Jessie. 2022. "Biden Administration Announces $19 Million for New Residency Positions." Modern Healthcare. February 3, 2022. https://www.modernhealthcare.com.

Hoehn-Velasco, Lauren, Diana R. Jolles, Alicia Plemmons, and Adan Silverio-Murillo. 2023. "Health Outcomes and Provider Choice under Full Practice Authority for Certified Nurse Midwives." *Journal of Health Economics*, 92: 102817.

Hughes, Stacey. 2021. "AHA Letter to House and Senate Sharing Priorities in Upcoming Budget Reconciliation Legislation." AHA. September 2, 2021. https://www.aha.org.

Iglehart, John K. 1999. "Support for Academic Medical Center—Revisiting the 1997 Balanced Budget Act." *New England Journal of Medicine*, 341 (4): 299-304.

IOM. 2014. "GME Financing–Graduate Medical Education that Meets the Nation's Health Needs–NCBI Bookshelf." Graduate Medical Education that Meets the Nation's Health Needs, edited by J Eden, D Berwick, and G Wilensky. https://www.ncbi.nlm.nih.gove/books/NBK248024/.

Kessel, Reuben A. 1958. "Price Discrimination in Medicine." *Journal of Law and Economics*, 1 (October): 20-53.

Leffler, K.B. 1978. "Physician Licensure: Competition and Monopoly in American Medicine." *Journal of Law and Economics*, 21 (1): 165-186.

Lo Sasso, Anthony T. 2021. "Regulating High-Skilled Immigration: The Market for Medical Residents." *Journal of Health Economics*, 76 (March): 102436.

Ludmerer, Kenneth M., and Michael M.E. Johns. 2005. "Reforming Graduate Medical Education." *Journal of the American Medical Association*, 294 (9): 1083-1087.

Markowitz, Sara, and E. Kathleen Adams. 2022. "The Effects of State Scope of Practice Laws on the Labor Supply of Advanced Practice Registered Nurses." *American Journal of Health Economics*, 8 (1): 65-98.

Martsolf, Grant R., Hilary Barnes, Michael R. Richards, et al. 2018. "Employment of Advanced Practice Clinicians in Physician Practices." *JAMA Internal Medicine*, 178 (7): 988-990.

McMichael, Benjamin J. 2023. "Supply-Side Health Policy: The Impact of Scope-of-Practice Laws on Mortality." *Journal of Public Economics*, 222: 104901.

McMichael, Benjamin J. 2025. "The Impact of Nurse Practitioner Scope-of-Practice Laws on Preventable Hospitalizations." *Journal of Health Economics*, 103: 103044.

Meiselbach, Mark K., Yang Wang, Jianhui Xu, Ge Bai, and Gerard F. Anderson. 2023. "Hospital Prices for Commercial Plans Are Twice Those for Medicare Advantage Plans When Negotiated by the Same Insurer." *Health Affairs*, 42 (8): 1110-1118.

Mowery, Yvonne M. 2015. "A Primer on Medical Education in the United States through the Lens of a Current Resident Physicians." *Annals of Translational Medicine*, 3 (18): 270.

Nicholson, Sean. 2002. "Physician Specialty Choice under Uncertainty." *Journal of Labor Economics*, 20 (4): 816-847.

Nicholson, Sean, and Carol Propper 2012. "Medical Workforce." Handbook of Health Economics, edited by Mark Pauly, Thomas McGuire, and Pedro Barros.

Pigou, Arthur C. 1912. Wealth and Welfare. London: Macmillan.

Rabideau, Brendan, Michael R. Richards, and Christopher M. Whaley. 2024. "Training Labor and Treatment Behavior: Evidence from Physician Residency Programs." *Health Economics*, 33 (9): 2059-2087.

Richards, Michael R., Chiu-Fang Chou, and Anthony T. Lo Sasso. 2009. "Importing Medicine: A Look at Citizenship and Immigration Status for Graduating Residents in New York State from 1998-2007." *Medical Care Research and Review*, 66 (4): 472-485.

Rosen, Sherwin. 1972. "Learning and Experience in the Labor Market." *Journal of Human Resources*, 7: 326-342.

Rosenstein-Rodan, Paul. 1943. "Problems of Industrialisation of Eastern and Southern-Eastern Europe." *Economic Journal*, 53: 202-211.

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92 (6): 991-1061.

Roth, Alvin E. 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89 (4): 748-780.

Sloan, Frank A. 2021. "Quality and Cost of Care by Hospital Teaching Status: What Are the Differences?" *Milbank Quarterly*, 99 (1): 273-327.

Spivack, Steven B., Generva F. Murray, Hector P. Rodriguez, and Valerie A. Lewis. 2021. "Avoiding Medicaid: Characteristics of Primary Care Practices with No Medicaid Revenue." *Health Affairs*, 40 (1): 98-104.

Taylor, Mariah. 2025. "29 Physician Specialties Ranked by Pay." Becker's Hospital Review, July 9th, 2025. Available here: https://www.beckershospitalreview.com/compensation-issues/29-physician-specialties-ranked-by-pay-2/.

Traczynski, Jeffrey, and Victoria Udalova. 2018. "Nurse Practitioner Independence, Health Care Utilization, and Health Outcomes." *Journal of Health Economics*, 58 (March): 90-109.

Whaley, Christopher M., Daniel R. Arnold, Nate Gross, and Anupam B. Jena. 2021. "Physician Compensation in Physician-Owned and Hospital-Owned Practices." *Health Affairs*, 40 (12): 1865-1874.

Wynne, Barbara, Cassandra Guarina, Lindsey Morse, and Michelle Cho. 2006. "Alternative Ways of Financing Graduate Medical Education." https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101617515.pdf.

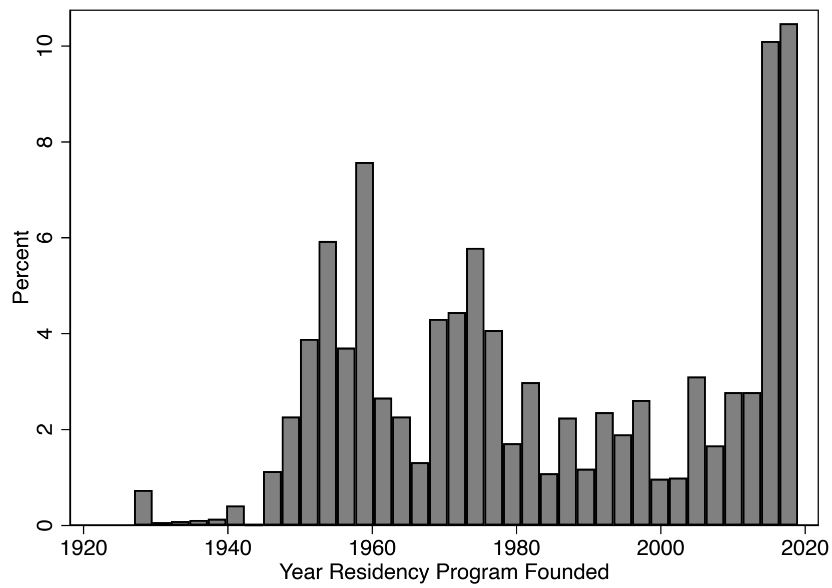


Figure 1
Annual Physician Residency Training Program Debuts (All Specialties)

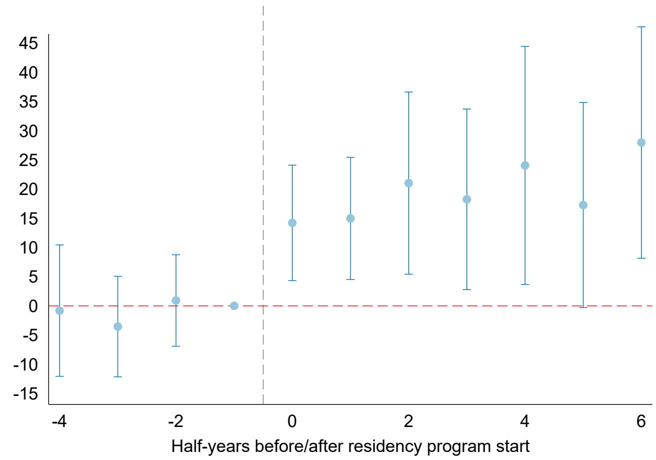
Notes: Historical residency information collected from web-based lookups as described in Section III.

Table 1

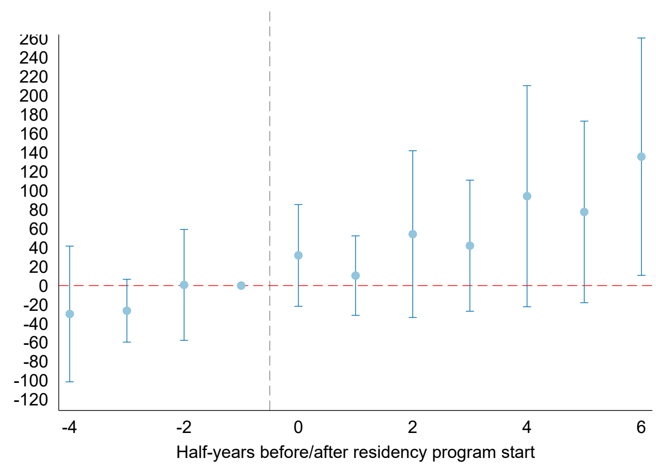
Summary Statistics for Hospitals' Medicare Market Procedural Care in 2014

	Treatment Group Hospitals	Control Group Hospitals
	Mean (SD)	Mean (SD)
Inpatient Procedures	244.8 (132.1)	98.55 (111.7)
Inpatient Revenue ('000)	3,705 (2,003)	1,443 (1,886)
Outpatient Procedures	1,045 (887.8)	442.8 (517.1)
Outpatient Revenue ('000)	1,085 (691.5)	461.8 (546.1)
Unique Physicians Performing Cases	157.2 (80.8)	77.90 (81.1)
Hospitals	26	2,933

Notes: Data are from 100% Medicare FFS claims and are aggregated to the half-year level. Revenues are in thousands of dollars and in nominal terms. Unique physicians are identified by their respective federal National Provider Identification (NPI) number reported in the claims data. Treatment group hospitals span 22 Hospital Referral Regions (HRRs) and include 14 states: AZ, CO (2), FL (4), MA, MD, MI (3), NC, NJ (3), NV (2), NY (4), OH, OK, PA, and SC.



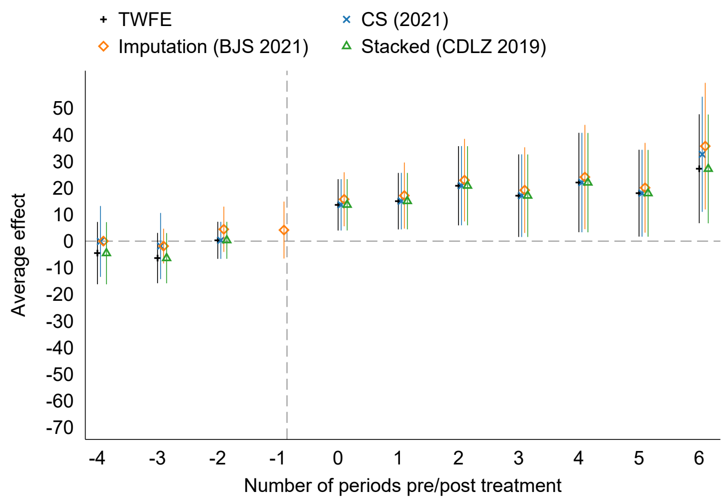
(a) Hospital Inpatient



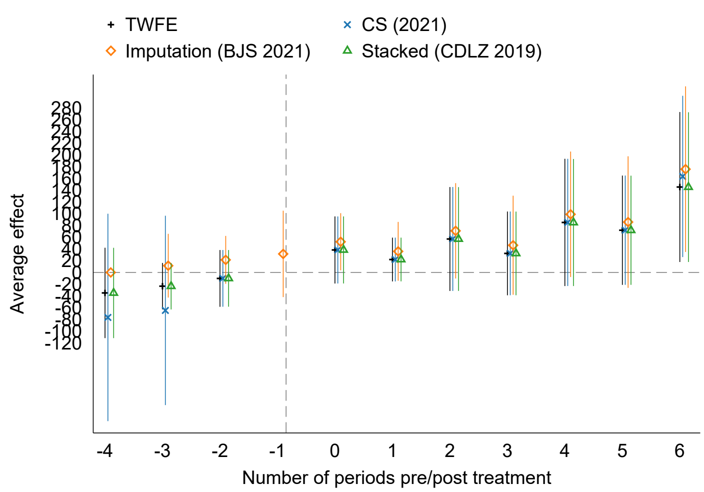
(b) Hospital Outpatient

Figure 2
Event Study Estimates for Effects of Trainee Program on Number of Medicare Procedures Stratified by Clinical Setting.

Notes: Procedure volumes are at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Estimation described in Section 3.



(a) Hospital Inpatient

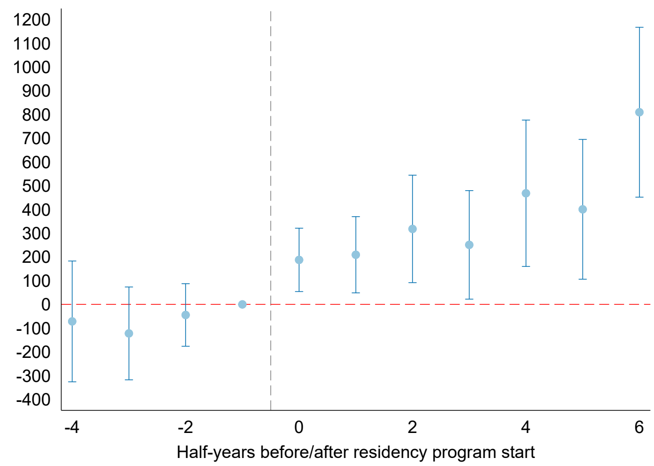


(b) Hospital Outpatient

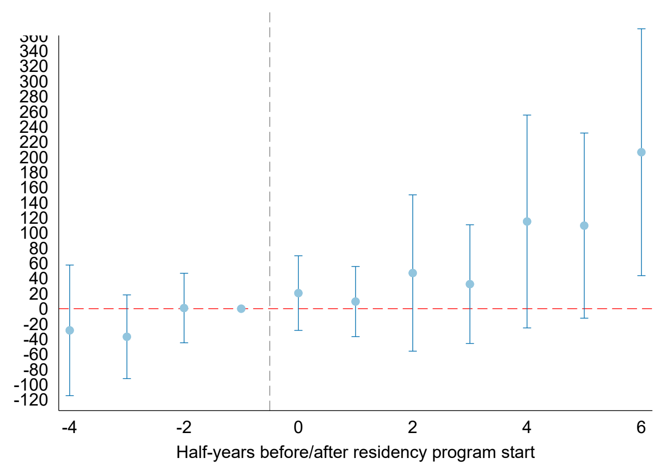
Figure 3

Sensitivity Tests for Effects of Trainee Program on Number of Medicare Procedures Stratified by Clinical Setting.

Notes: Procedure volumes are at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Estimation described in Section 3.



(a) Hospital Inpatient

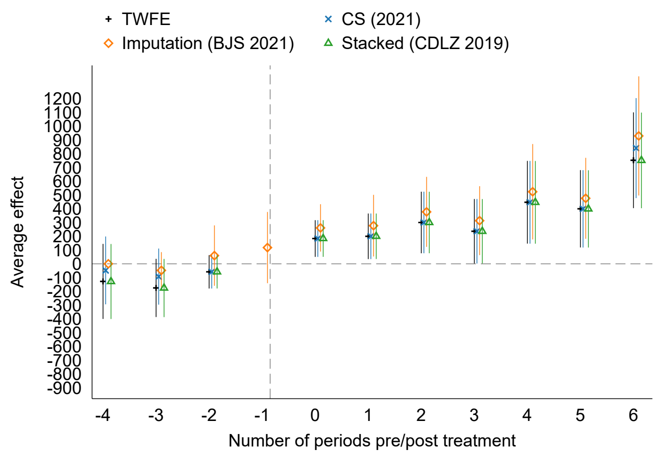


(b) Hospital Outpatient

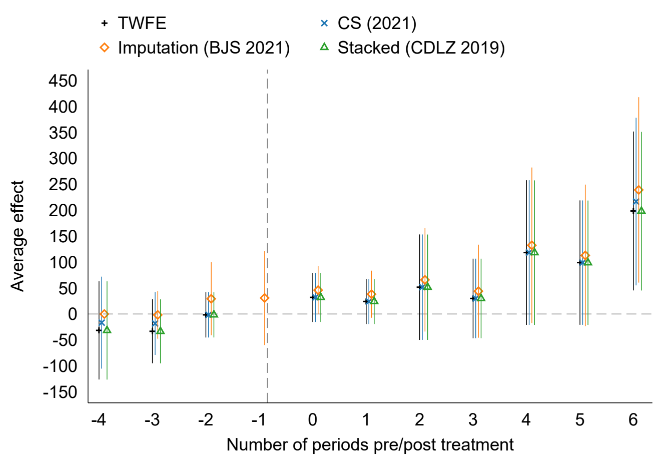
Figure 4

Event Study Estimates for Effects of Trainee Program on Total Revenue from Medicare Procedures Stratified by Clinical Setting.

Notes: Total revenues are in ‘000 nominal dollar terms and at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Estimation described in Section 3.



(a) Hospital Inpatient



(b) Hospital Outpatient

Figure 5

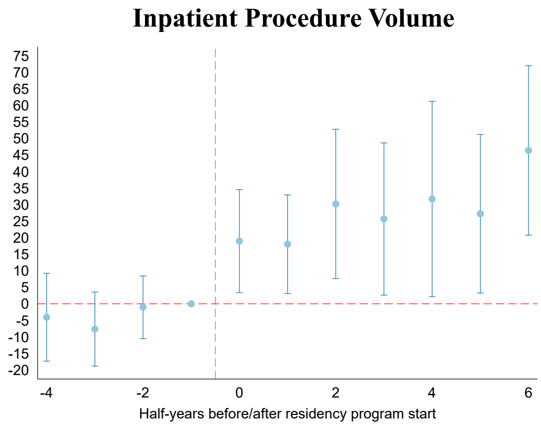
Sensitivity Tests For Effects of Trainee Program on Total Revenue from Medicare Procedures Stratified by Clinical Setting.

Notes: Total revenues are in ‘000 nominal dollar terms and at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Estimation described in Section 3.

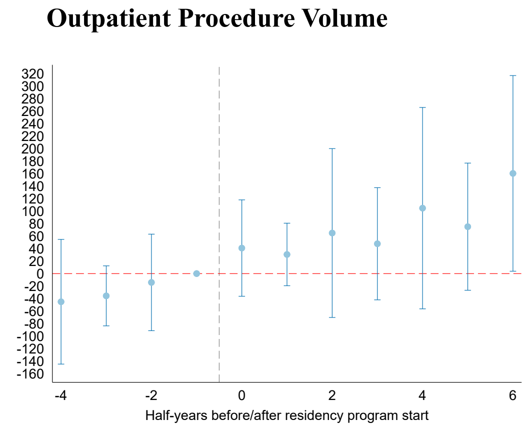
Table 2
TWFE Diff-In-Diff Estimates for Trainee Program Aggregated Post-Period Effects on Hospitals' Medicare Market Activity

	Inpatient Procedures	Outpatient Procedures	Inpatient Revenue	Outpatient Revenue
	(1)	(2)	(3)	(4)
Treated x Post	17.30**	76.95	421.2***	97.47
	(7.69)	(40.32)	(138.7)	(53.33)
Time Fixed Effects	Yes	Yes	Yes	Yes
Hospital Fixed Effects	Yes	Yes	Yes	Yes
Observations	35,508	35,508	35,508	35,508

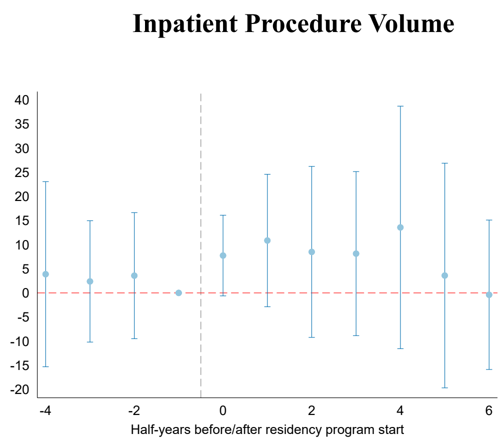
Notes: Standard errors clustered at the hospital level. *** P value at 0.01 ** P value at 0.05



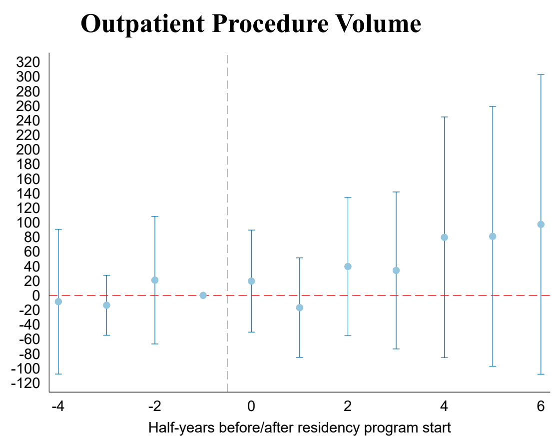
(a) Above Median Program Size



(b) Above Median Program Size



(c) Below Median Program Size

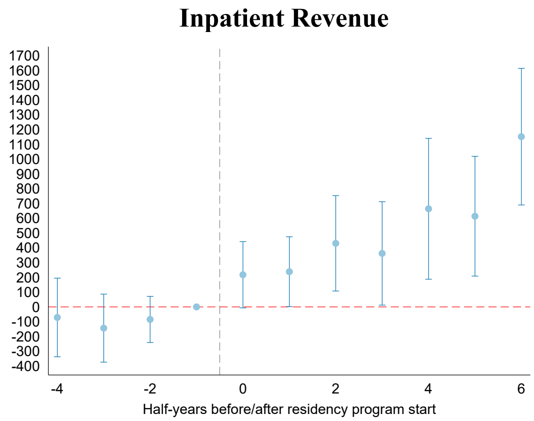


(d) Below Median Program Size

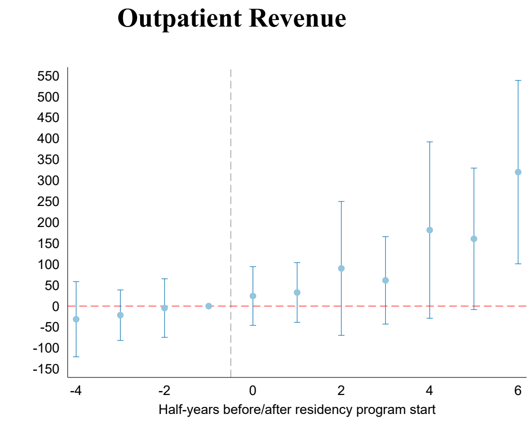
Figure 6

Event Study Estimates for Effects of Trainee Program on Number of Medicare Procedures Stratified by Program Size.

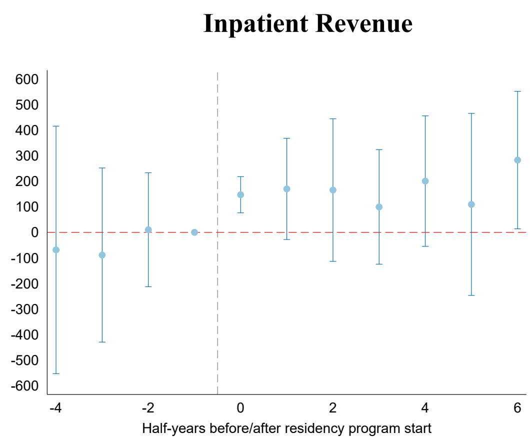
Notes: Procedure volumes are at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Graduate Medical Education (GME) program size reflects the effective number of full-time equivalent (FTE) surgical residents the hospital has at steady-state for the treatment group programs. The median value for our treatment group hospitals is 9.6 FTEs. Control group hospitals are consistent across estimations.



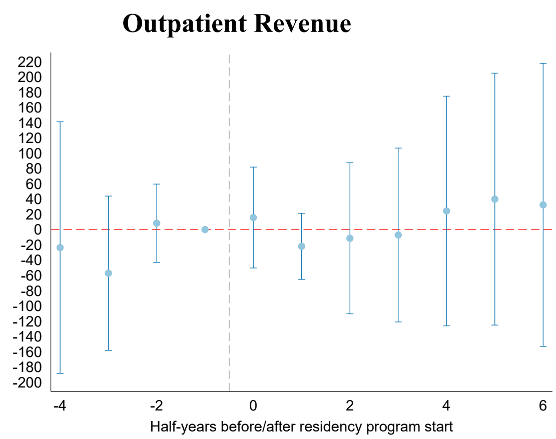
(a) Above Median Program Size



(b) Above Median Program Size



(c) Below Median Program Size



(d) Below Median Program Size

Figure 7

Event Study Estimates for Effects of Trainee Program on Total Revenue from Medicare Procedures Stratified by Program Size.

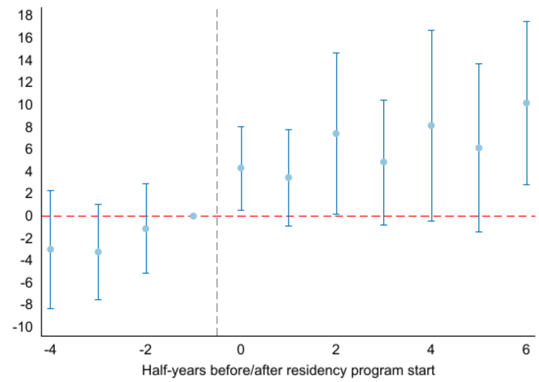
Notes: Total revenues are at the half-year level for the Medicare market. Zero time point is the half-year the residency program officially launched. Graduate Medical Education (GME) program size reflects the effective number of full-time equivalent (FTE) surgical residents the hospital has at steady-state for the treatment group programs. The median value for our treatment group hospitals is 9.6 FTEs. Control group hospitals are consistent across estimations.

Table 3

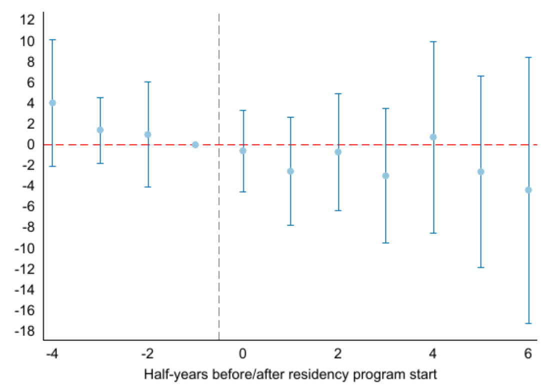
TWFE Diff-In-Diff Estimates for Trainee Program Aggregated Post-Period Effects on Hospitals' Medicare Market Activity for Larger Programs

	Inpatient Procedures	Outpatient Procedures	Inpatient Revenue	Outpatient Revenue
	(1)	(2)	(3)	(4)
Treated x Post	27.85***	96.79**	608.2***	158.8**
	(10.48)	(47.06)	(189.1)	(72.90)
Time Fixed Effects	Yes	Yes	Yes	Yes
Hospital Fixed Effects	Yes	Yes	Yes	Yes
Treated Baseline Mean	257.1	1,132	3,708	1,126
Observations	35,376	35,376	35,376	35,376

Notes: Treatment group is restricted to the general surgery residency programs at or above the median FTE value (9.6). Standard errors clustered at the hospital level. *** P value at 0.01 ** P value at 0.05 * P value at 0.10



(a) Fully Trained Physicians (High-FTE)

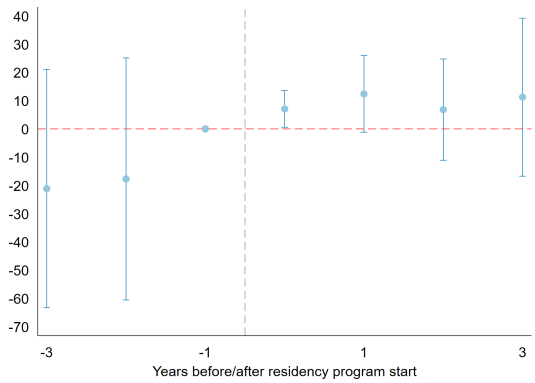


(b) Fully Trained Physicians (Low-FTE)

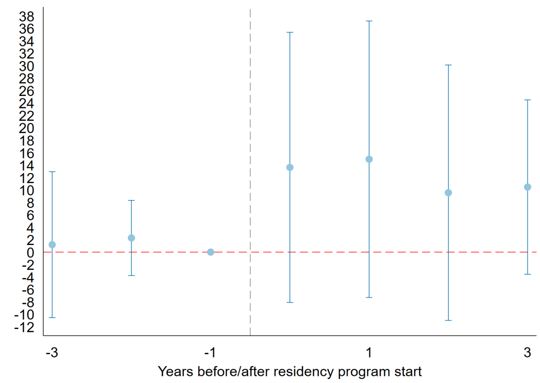
Figure 8

Event Study Estimates for Effects of Trainee Program on Number of Non-Training Surgeons.

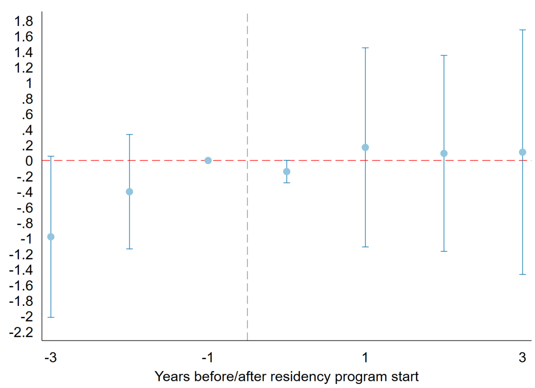
Notes: Physician counts are at the half-year level and capture unique physician NPIs as the rendering physician for one or more procedures in the relevant hospital. Zero time point is the half-year the residency program officially launched. Graduate Medical Education (GME) program size reflects the effective number of full-time equivalent (FTE) surgical residents the hospital has at steady-state for the treatment group programs. The median value for our treatment group hospitals is 9.6 FTEs. Control group hospitals are consistent across estimations.



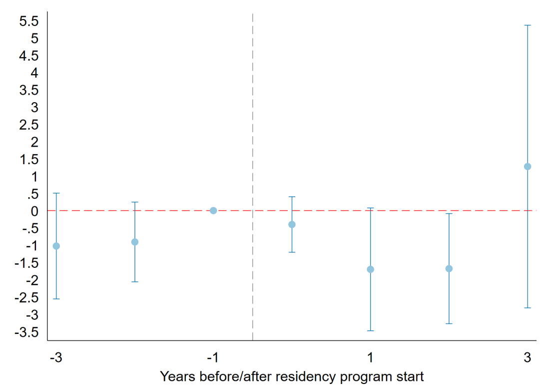
(a) Total Staffed Beds — High-FTE



(b) Total Staffed Beds — Low-FTE



(c) Operating Rooms — High-FTE



(d) Operating Rooms — Low-FTE

Figure 9
Event Study Estimates for Effects of Trainee Program on Hospital Capital Investment.

Notes: AHA annual survey data, 2010–2019. Total staffed beds (BDTOT) and number of operating rooms (OPRA) are measured at the hospital-year level. Zero time point is the half-year the residency program officially launched. Graduate Medical Education (GME) program size reflects the effective number of full-time equivalent (FTE) surgical residents the hospital has at steady-state for the treatment group programs. The median value for our treatment group hospitals is 9.6 FTEs. Control group hospitals are consistent across estimations.

Table 4
Projected Hospital Return on Investment (Annual '000 Dollar Terms) for Relatively Larger Training Programs in Steady State

	(1)	(2)	(3)	(4)
Trainee Labor Costs	(1,100)	(1,100)	(1,100)	(1,100)
Complementary Non-Training Physician Labor Costs	(3,900)	(3,900)	(3,900)	(3,900)
Additional Medicare Inpatient Revenue	1,800	1,800	1,800	1,800
Additional Medicare Outpatient Revenue	470	470	470	470
Additional Medicare Revenue (Total)	2,270	2,270	2,270	2,270
Percent of Operating Revenue Medicare	--	30%	30%	30%
Percent of Operating Revenue Non-Medicare	--	70%	70%	70%
Additional Non-Medicare Revenue (0% Markup)	--	5,300	--	--
Additional Non-Medicare Revenue (10% Markup)	--	--	5,830	--
Additional Non-Medicare Revenue (25% Markup)	--	--	--	6,625
Net Revenue Gain (Medicare + Non-Medicare), less labor costs)	(2,730)	2,570	3,100	3,895

Notes: We restrict to the DD effect estimates for the relatively larger (above median) surgical training programs from our main analysis treatment group. We assume 13 trainee FTEs (approximately the median in this subgroup) and use national reporting by AAMC on trainee salaries and national information on physician salaries and assume an additional 30% for non-wage benefits and costs (e.g., malpractice coverage). For complementary physician labor, we conservatively assume 10 additional fully trained physicians at 75% effort, based on the final post-period n_grad_2009 estimate from Figure 8. For the marginal increases in Medicare revenue, we use a conservative estimate based on the average of the final two observed half-year event study estimates from Figure 7 (i.e., the average of lag5 and lag6), annualized. Using HCRIS data, the average percent of operating revenues from the traditional Medicare program (inpatient and outpatient) is approximately 30% in 2014.