Physician Learning and Forgetting: Evidence from the US Military

Trevor Fitzgerald* Atul Gupta† Stephen D. Schwab‡

Abstract

Uniformed physicians in the military are quasi-randomly assigned combat deployments, leading to gaps of varying durations in their practice routines. We leverage this unique setting to provide novel evidence on skill depreciation (forgetting) and accumulation (learning) among physicians in the military healthcare system. We find robust evidence that readmission rates increase by more than 48% in the first two months after a surgeon returns from a deployment and that Emergency Department visits increase by 36%. Using two classes of physicians with different decisions, we compare outcomes for surgeons who are more likely to rely on physical skill with obstetricians who are more likely to rely on their own judgment when faced with the decision of whether to deliver through cesarean section. We also show that surgeons quickly recover physical skills but that obstetricians maintain elevated cesarean section rates for up to 9 months. Finally, we include a learning model where we find that 90 day volume has nearly twice the impact on outcomes for obstetricians than it does for surgeons.

*Alvarez College of Business, The University of Texas at San Antonio; United States Army; trevor.fitzgerald@my.utsa.edu
†The Wharton School, University of Pennsylvania; National Bureau of Economic Research; atulgup@wharton.upenn.edu
‡Alvarez College of Business, The University of Texas at San Antonio; stephen.schwab@utsa.edu

Career breaks are ubiquitous. According to a LinkedIn study of 23,000 workers, approximately 62% of workers have taken a break of more than a month off work during their careers, and 35% are interested in taking one in the future. The most common reasons for these breaks include parenting, professional development, and pursuing personal goals [Liu, 2023]. In addition to these planned breaks, 15% of US workers report having taken leave for a Family Medical Leave Act (FMLA) qualifying condition in the past year [U.S. Department of Labor, 2020]. The average length of leave for these individuals is 28 business days with the majority due to an employee's own serious health condition. Of note, the vast majority of employees return to their same position with the same employer after their leave. These numbers, however, are likely an underestimate of breaks in production as many employees receive accommodations while dealing with health conditions or injuries [Buchanan and Van Niekerk, 2022]. For example, a surgeon may sprain a wrist and be unable to perform surgery but still conduct consults and post-operative care.

Faced with these breaks in production, employers worry about skill depreciation. According to a poll of 400 hiring managers on LinkedIn, skill depreciation was among the top four reasons why employers view career breaks negatively [Tynan and Hedges, 2024, Dowd-Higgins, 2024]. In this paper we provide evidence regarding both the level of skill depreciation from breaks in production and the length of time it takes to return to baseline. We do this by analyzing performance outcomes for two types of physicians - surgeons and Obstetrician/Gynecologists (Ob/Gyns).

Healthcare is an ideal setting for this study for two primary reasons. First, most prior studies on individual skill decay have focused on the effects of inactivity on the completion of routine tasks with little worker discretion, such as those found in manufacturing [Globerson et al., 1989, Shafer et al., 2001]. While the speed and quantity of production are straightforward to measure in that type of setting, these employees are less likely to take long breaks due to financial constraints [U.S. Department of Labor, 2020]. In comparison, there is little evidence on skill depreciation and relearning among highly-educated workers with significant discretion. Productivity is more difficult to measure in these settings, as tasks are more complex and differentiated in nature, yet these are likely the workers most likely to take a career break [Boyens et al., 2022] as they are more likely to have both financial resources and access to paid leave and/or FMLA. Second, fluctuations in skill are arguably more consequential in the healthcare sector, since patient health outcomes may be sensitive to minor decreases in physician skill [Hockenberry and Helmchen, 2014]. However, several identification challenges have stymied progress in studying physician learning dynamics; for example, physicians typically control their work schedules, match endogenously with patients, and are typically paid on the margin for incremental services.

In this paper, we leverage the unique institutional characteristics of the US Military Health System (MHS) to overcome these identification challenges. Three features of the setting allow us to quantify physician learning and forgetting. First, uniformed physicians in the military are quasi-randomly assigned to military base hospitals to fulfill the needs

of the military. Similarly, physicians are also quasi-randomly deployed to combat zones for varying periods to support the military's combat operations. As we show, the procedures performed during deployment focus on critical care and stabilization and have little overlap with the procedures performed routinely at their home base, motivating the possibility of skill depreciation due to inactivity. Second, patients in the military health system have generous insurance coverage but do not have a meaningful choice among providers, alleviating concerns about physician-patient matching on unobserved factors. Finally, uniformed physicians are salaried employees and do not have an incentive to perform unnecessary or more highly reimbursed procedures.

Uniformed physicians in the military healthcare system, therefore, experience exogenously assigned gaps in their practice routines. These periods of relative inactivity for the deployed may lead to skill decay, which could manifest itself in a decline in the quality of care provided to their patients upon return from combat deployment. We follow the previous literature and refer to this process as "forgetting." Once they return to routine practice on the base, deployers should regain lost skills due to learning-by-doing and may return to their baseline performance levels.

Conceptually, there are two margins on which a gap in practice may affect patient care. First, physicians may lose the physical skill ("muscle memory") to conduct a specific procedure. We expect this will scale with the difficulty of the procedure. Second, physicians may lose confidence in their own decision-making and default to simpler or safer procedures. While we cannot formally test this, we provide suggestive evidence by considering the experience of two physician specialties. First, we consider surgeons and the outcomes of the surgeries they perform after returning from a deployment. Our primary finding is that 90-day readmission rates increase by about 6 percentage points (48%) but that this affect is short-lived with surgeons cutting this increase in half by month 3 and returning to baseline within 5-6 months. Second, we consider ob/gyns and the decision to perform either a Cesarean-section (C-section) or a vaginal delivery. The determination of whether to let a birthing-mother progress in natural labor or shift to a C-section is largely a judgment call based on the physician's assessment of the wellness of the fetus. We find that physicians increase their C-section rate by about 76% after returning from a deployment and that this effect endures for up to 9 months. Together these findings suggest that loss of confidence in one's own decision-making can have a comparable effect to the loss of muscle memory.

This paper contributes to two strands of literature. First, we provide, to our knowledge, the first causal evidence on learning and forgetting among physicians. Due to the unique institutional features of our setting and the richness of the data, we are able to relax several limitations faced by previous studies on this question. For example, two recent studies leveraged short gaps (of a few days) in the performance of complex surgeries to test whether surgeons experience skill decay and, surprisingly, found contradictory results. One examined heart bypass (CABG) surgeries in Pennsylvania and found that each additional day away from the operating room increases postoperative mortality rates, consistent

with forgetting [Hockenberry and Helmchen, 2014]. The other studied postoperative mortality rates for patients with hip fractures in England and concluded that similar short gaps improve mortality rates [Rachet-Jacquet, 2022]. In both cases, the source of variation in the timing and duration of the gaps is unclear and raises the question of whether these may be correlated with unobserved factors that affect surgeon performance. In addition to leveraging credibly exogenous variation in the timing and duration of deployments, we are also able to quantify the impacts of much longer gaps in practice. Finally, we analyze physicians' return to their baseline skill levels using a non-linear model, allowing us to estimate the rate of forgetting.

The paper also contributes to the growing literature studying individual-level learning-by-doing. Recent studies have used detailed data on an individual's job history and production to quantify learning in professions other than manufacturing. For example, Ost [2014] uses data from North Carolina to quantify learning and forgetting among elementary and middle school teachers. Haggag et al. [2017] trace a learning gradient for taxi drivers in New York City measured in terms of earnings per shift. Both studies explore the mechanisms underlying changes in skill and document the task specificity of skill appreciation. However, these studies also suffer from the lack of a quasi-experimental research design, raising concerns about the potential for bias.

The remainder of the paper is organized as follows. Section 1 describes the key characteristics of the institutional setting. Section 2 describes the data and provides descriptive statistics. Section 3 describes the empirical strategy, including the key identification assumptions. Section 4 presents the main results on learning and forgetting. Finally, Section 5 concludes.

1 Background

1.1 The Military Health System

We study the learning dynamics for physicians employed by the US Military Health System (MHS), which is the Department of Defense's integrated healthcare system. In addition to insuring over 9 million beneficiaries through its Tricare benefit, the MHS provides care for approximately 4.6 million enrollees in its direct care hospital system. This system consists of 45 inpatient hospitals and 373 clinics located throughout the United States, Europe, and East Asia and is staffed by a mix of military and civilian health care providers.

There are several unique aspects of the MHS that make it ideal for studying physician learning dynamics. Physicians are salaried and do not receive any income on the margin if they provide additional services. Neither is their performance evaluation based on patient volume. Hence, their clinical decisions are less likely to be influenced by financial incentives. A large fraction of uniformed doctors, like non-physician active-duty soldiers, are deployed at

some point in their careers and the timing of these deployments is unrelated to physician quality, prior experience, or personal preferences. Another unique aspect of care in the MHS is that patients in the default healthcare plan, known as Tricare Prime, typically cannot choose which physician on the clinic staff performs a procedure, unless they are referred outside of the military system. For example, the appointing system does not allow beneficiaries to request a specific physician. Although retirees and military family members can choose to disenroll from Tricare Prime, military service members are legally bound to remain in the default plan. Additionally, generally only Tricare Prime members have access to the Military hospital system. Furthermore, beneficiaries are only referred to civilian care if the procedure cannot be obtained in the direct care system within specified access-to-care thresholds.

1.2 Deployments

Military physicians practice in hospitals located on military bases when they are not needed for an operational mission (i.e., "deployment"). Decisions regarding who deploys are made differently by each of the military services; however, they are all based on a "next person up" mentality and not driven by physician quality or the physician's preferences.

During deployments, physicians can be assigned to what the military refers to as role 2 or role 3 care sites. Although support at role 3 is significantly higher than at role 2 sites, physicians at both types of sites focus on wound debridement, initial fixation, and wound vacuum-assisted closure (VAC) (role 3 only). Injured service members are evacuated outside of role 3 care within 24 hours of injury. See Malin and Brannan [2012] for a detailed analysis of the experience of an orthopedic surgeon in Afghanistan. Notably, Malin and Brannan write, "our unique capabilities were only required for a small percentage of the time." Similarly, obstetricians typically only provide primary care during deployment. In our analysis sample, both the mean and median deployment is around 6 months long, but we observe deployments as long as a year (see Figure **??**). Hence, deployments represent a relatively long break in the practice routine. In addition, service members who return from deployment receive up to 30 days of vacation time plus approximately two weeks of reintegration training, so that even in the unlikely case that a deployed physician did happen to perform some complex surgeries while deployed, they would have at least a 6-week break in practice. We aim to exploit these quasi-randomly assigned breaks in practice as shocks to physician skill and study learning dynamics.

2 Data

To study these potential changes in skill levels among physicians, we obtained access to detailed patient-level claims and personnel data from the US Military Health System Data Repository (MDR).

The MDR consists of multiple files for each patient encounter within the direct care system and medical claims for

encounters in the civilian "purchased care" system. The medical claims data are similar in structure, scope, and detail to standard datasets that economists may be familiar with, such as Medicare fee-for-service claims with each record including a complete accounting of the care provided. Inpatient records include the national provider ID (NPI) for the attending physician, hospital and ward identifiers, the diagnosis related group (DRG), and admission and disposition dates along with an indicator for whether admission was for medical or surgical reasons. Physician claims include the NPI for the treating physician, a code for the physician's specialty, common procedural terminology (CPT) codes for up to 10 procedures performed, and the date the procedure(s) was performed. Both files include demographic data on the patient, including their age, sex, and beneficiary status (i.e., active-duty military, retired military, or dependent). We focus on encounters over 2007–2017. It is not possible to identify physicians in the data before 2007 since the Center for Medicare and Medicaid Services (CMS) first issued NPIs in late 2006. We measure physician skill using a standard measure of care quality, the 90-day readmission rate for patients, which we compute using the claims data.

In addition, we have personnel data from the Defense Manpower Data Center (DMDC). DMDC is the military's central repository of personnel information, including demographic and assigned unit information for each service member and civilian employee, including rank, longevity in the military, assigned unit location, and the beginning and end date for operational deployments. This personnel data allows us to link uniformed physicians observed serving patients in the claims sample with information on their demographics and professional service history.

2.1 Sample Construction

We create our sample by focusing on index surgeries. We define an index surgery as one that is performed in an inpatient setting at least 90 days after any previous hospital admission. We use both military direct care data and civilian healthcare data to determine whether a surgery is an index and whether there is readmission within 90 days. We then further limit the sample to records that indicate that the same physician NPI was both the attending physician and the primary treating physician performing the surgery in the corresponding physician records. Finally, we limit the sample to any surgical procedure (DRG) which we observe at least 100 times in the data so that we can identify a DRG fixed effect.

Next, we use the DMDC data to determine the beginning and end data of all physician deployments. We modify these dates based on the final surgery before and first surgery after a deployment in order to get a more precise estimate of the total experience gap created by the deployment. We construct our analysis sample by creating a stacked sample of surgeons who deployed at some point in the data. In order to estimate a surgeon fixed effect, we limit the sample to those who have at least five surgeries prior to and after the deployment. Our control units are made up of physicians who either did not deploy during the sample period or did not deploy or return from a deployment within one year of

their matched treated surgeon's deployment. To match the restriction, we require the control surgeons to also have five surgeries before and after their matched surgeon's deployment.

3 Empirical Strategy

Our research design leverages quasi-random assignment of physicians to combat deployments of varying durations to provide identifying variation. We compare patterns in outcomes for patients served by deploying physicians in the six months before and after a deployment to those for patients served by similar, but nondeploying physicians. We match each deployed physician to a subset of nondeploying physicians observed practicing in the year of deployment by specialty, race, gender, and military rank, which is a proxy for experience. Each nondeployer can match to multiple deployers since the period she is observed in may overlap with multiple deployments.

We then vertically "stack" the datasets of each matched group to form the analysis sample used to estimate the regression models. Hence, this is a stacked difference-in-differences design that incorporates matching [Cengiz et al., 2019, Deshpande and Li, 2019]. This approach is superior to conventional two-way fixed effects (TWFE) estimation since it recovers the average treatment effect on the treated (ATT) in staggered treatment designs, while the TWFE estimator is biased in the presence of treatment effect heterogeneity [Baker et al., 2022]. The equation below presents the baseline estimating equation.

3.1 Descriptive Statistics

Table 1 describes the final analysis sample. Our sample construction approach, described above, retains 690 surgeons and 486 obstetricians of which 369 and 87 deploy respectively. They are observed managing 28,725 and 47,551 hospital stays, respectively. Panel A describes key patient attributes. Since patients are typically active duty service members or their family members, they tend to be young, healthy, and more racially diverse. The mean age of the surgical patients is 34 and the mean age of obstetric patients is 27. Elixhauser comorbidty socres are 0.51 and 0.25 respectively indicating a quite healthy sample.

Panel B describes characteristics of the hospital stay and treating physicians. The number of procedures is the mean number of difference procedure codes listed on the inpatient claim. An example would be an Obstetrician that both delivers a baby and performs an episiotomy would count as two procedures. Both specialties perform about two procedures per hospital stay. Next, we show the physician's experience levels based on time in the military which generally coincides with graduating medical school. The average surgeon has 11.39 years of experience while the average obstetrician has 9.1 years of experience.

Panel C shows average outcomes for the sample. The clinical and economics literature has used short-term mortality and readmission rates as the preferred measures of provider quality for a long time [Cutler, 1995, Jencks et al., 2009]. The Affordable Care Act subsequently mandated that Medicare introduce performance pay programs for hospitals, assessing their quality on both these measures [Gupta, 2021]. Since the patients in our sample are typically healthy, death is very rare and unsuited as a quality measure. We therefore use short-term rehospitalization or readmission rate as the primary quality metric for our analysis. We also consider whether a patient has a follow-up emergency department visit.

We define readmission as any subsequent hospital admission within 90 days of the index hospital stay. We chose 90 days as most hospital stays in our sample are for conditions or procedures paid for using a 90-day global payments. Global payments are lump sum payments intended to cover all care expenses incurred over a specified amount of time. Baseline readmission rates are over four times higher for surgical procedures compared to obstetric procedures (13% compared to 3%). We define ED visit post-discharge if a patient has an ED visit within 90 days after their discharge from the hospital. While Medicare penalizes hospitals based on 30-day readmission rates for some procedures, we use 90 days to be consistent with our readmission outcome. This is quite common in our sample with 20% of surgical patients and 15% of obstetric patients having a follow-up ED visit.

Finally, we consider C-section rates for obstetricians. We observe in the data whether a physician conducts a C-section or vaginal delivery based on the DRG of the procedure. Overall, we observe a 25% c-section rate for these physicians.

Using this sample, we estimate the following model:

$$Y_{ijbt} = \alpha_{bj} + \alpha_{jt} + \alpha_j + \beta\, D_j 1(Post)_{jt} + X_i'\delta + \epsilon_{ijbt} \tag{1}$$

where Y_{ijbt} denotes the outcome for patient i served by a physician in the matched group j at clinic (hospital-ward) b in year t. We include a vector of patient demographic controls (X_i) to account for the observed differences in patient complexity and demographics. We include matched group fixed effects (α_j), as well as a fully interacted set of fixed effects for each clinic, α_{bj}, and year, (α_{tj}). D_j is an indicator for the deploying physician within group j. $1(post)_{jt}$ is an indicator for patients served by the deploying physician after returning from a deployment. β is the coefficient of interest that captures the DiD estimate of deployment on patient outcomes.

4 Results

We begin this section by showing evidence that the critical assumptions embedded in our empirical strategy are met. These assumptions are 1) that deployments are as good as random, 2) that physicians who deploy would have remained on a parallel trend with those that do not deploy were it not for the deployment, and 3) that any effect is due to the break in production and not due to other aspects of the deployment.

Table 2 supports our first assumption that deployments are as good as random. Panel A shows a simple t-test of means between patients that see a deployer and those that see a non-deployer in the 6 months prior to a deployment. We separate these models by surgeons on the left (columns 2-4) and Obstetricians on the right (columns 5-7). Panel B shows the same t-test for hospital and physician characteristics of those that deploy and don't deploy. None of the patient characteristics in panel A are statistically different from each other. In panel B, only physician age is different at even a 10% level. There is no evidence that higher or lower quality physicians systematically deploy. We also report the joint significance F-statistic across from a model including all 11 variables in both panels. Both F-stats are less than 1 with P-values of 0.57 and 0.61 for surgeons and Obstetricians respectively.

Our parallel trends assumption is the primary assumption necessary for the validity of difference in differences design, and is inherently untestable. However, we provide suggestive evidence in the form of event studies. Figure 2a shows an event study of the probability of a readmission before and after a surgeon deploys and Figure 2b shows the same for Obstetrician C-section rates. Figure 3 displays trends in post-discharge ED visits for both groups. Trends in the pre-period would bring into question the validity of the parallel trends assumption. We don't observe any pre-trends providing suggestive evidence that any effect in the post-period is attributable to the deployment. We also plot both observed and predicted rates of readmissions or C-section rates based on patient characteristics. We note that while the predicted rates don't change, the observed rates are flat in the pre-period with a clear increase at the time of deployment.

A further threat to the parallel trends assumption would be if physicians see systematically different patients after returning from deployment. Table 3 applies our main equation to a series of variables that could call into question the validity of our design - changes in patient volume, changes in patients' predicted outcomes based on observable characteristics, patients' underlying health (elixhauser comorbidity index), and patient age. None of these are statistically different after the deployment for either specialty.

Our final assumption is needed in order to interpret the difference in differences coefficients as a result of the break in production rather than as a more wholistic impact of the deployment itself such as through a a change in the physician's mental health. Table 4 provides evidence that there is a break in production for deployed physicians. Panel A displays the top ten procedures in our data and how often they are done in a year in a deployed environment across all sites and

providers. Only two of the top ten most common procedures in our analysis sample are seen on an annual basis and the volume is extremely low resulting in a fraction of a procedure per surgeon. Panel B displays the top ten complex procedures performed in a deployed environment and the total volume across all sites and providers per year. Even the top ten are done at a extremely low-volumes. Surgeons spend most of their time ready if needed, but mostly conducting primary care and conducting very simple procedures (e.g. strapping/splinting or removing a toenail) for the Soldiers stationed near them.

Table 5 provides suggestive evidence that there is no change in physician mental health. Column 1 displays the results of five different markers of mental health - the probability that a physician is diagnosed with any mental health disorders, the probability that a physician is diagnosed with a mental health disorder included in the Elixhauser index, the probability that a physician is diagnosed with anxiety or stress, the probability that a physician is diagnosed with a mood disorder, and the probability that a physician is diagnosed with a substance abuse disorder. Columns 2 and 3 separate the analysis by Obstetricians and Surgeons. None of the analyses are statistically significant with a mix of small positive and negative coefficients. Combined tables 4 & 5 support our assumption that any affect of the deployment on patient outcomes is due to a break in production.

Table 6 displays our main results. Panel A shows the difference-in-differences results obtained by estimating equation 1 discussed above for surgeons and panel B for obstetricians. In Column 1 we display the coefficient estimate only for the first two months after returning from deployment. In column 2, we display both the cumulative impact on months 1-4 and the marginal effect during months 3 & 4. In column 3, we report the cumulative effects for 6 months and the marginal effect for months 5-6. We find that a surgical procedure performed by a physician in his first two months after a deployment is associated with a 48% (6.2 pp) increase in the probability of a readmission though that the marginal effect is no longer statistically significant by month 4. Given an average cost of an admission of seven thousand dollars, this increase is both clinically and economically meaningful.

In panel B, we report a 75% (18.7 pp) increase in the probability of a C-Section after an obstetrician returns from deployment. This effect does not begin to dissipate for at least 6 months with a return to baseline around month 9. An important context of this setting is that OB's deliver based on an "on-call" system so this is not related to their own provision of prenatal care. This is also evidenced by the flat predicted c-section rates shown in Figure 2b and in table 3 panel B, column 2.

4.1 Learning Model

To better understand the mechanisms underlying skill decay and relearning, we estimate a learning model that relates patient outcomes to measures of both recent and cumulative procedural experience. Table 9 presents the results. Our

primary focus is on the coefficient for the natural log of each physician's rolling 90-day volume, which captures how variation in a physician's recent procedural activity is associated with patient outcomes.

The results indicate that recent procedural experience plays an important role in shaping physician performance. For surgeons, a 1% increase in the number of procedures performed in the prior 90 days is associated with an estimated 0.022 percentage point decrease in the probability of a 90-day patient readmission. For obstetricians, a comparable 1% increase in recent procedural volume is associated with a larger estimated effect of roughly a 0.044 percentage point reduction in the likelihood of performing a c-section. While these marginal effects may appear small, they accumulate meaningfully over larger changes in procedural experience. For example, a 10% increase in recent volume corresponds to a reduction of approximately a quarter of a percentage point for surgeons and half a percentage point for obstetricians. In contrast, cumulative experience, measured using the natural log of rolling prior-year volume is not significantly related to patient outcomes for either group. This distinction highlights the importance of recency over accumulation, suggesting that physicians with extensive historical case volumes may still experience diminished performance if their recent exposure to procedures has been limited.

Taken together, these findings reinforce the earlier results showing a temporary spike in adverse outcomes following a gap in practice. The learning model provides additional evidence that the performance decline reflects a loss of procedural familiarity and that learning-by-doing is the dominant mechanism through which physicians recover. The larger effect of recent volume for obstetricians relative to surgeons suggests that performance in more judgment-intensive tasks, such as deciding whether to proceed with a Cesarean section, may be especially sensitive to breaks in practice and recency of experience.

4.2 Heterogeneity

The average effects are informative but may also conceal meaningful heterogeneity in how practice gaps affect physician performance. To investigate this, we re-estimate our main specification with an interaction term capturing multiple dimensions of interest, including procedure complexity, physician experience, and deployment length, using the following equation:

$$\begin{aligned} Y_{ijkt} = \alpha + X_{it}'\gamma_1 + \gamma_2 Z_{ijt} + \beta_1 \, 1(\text{post-deployment})_{jt} \\ + \beta_2 \, 1(\text{post-deployment})_{jt} Z_{ijt} + \lambda_{\text{DRG}} + \theta_j + \xi_k + \delta_t + \epsilon_{ijkt} \end{aligned} \tag{2}$$

Table 8 shows our results for each of these dimensions. We first examine whether deployments affect more com-

plex procedures differently. For surgeons, the estimated interaction effects are positive across all complexity thresholds, indicating larger performance declines for more complex cases. When defining complexity as more than three procedures, the effect is marginally significant, showing a roughly 5.7 percentage point larger increase in readmission rates compared to less complex cases. For procedures involving more than four procedures, the estimated effect is large in magnitude but imprecisely estimated, reflecting the smaller number of highly complex cases in the sample. While not statistically significant, the direction and size of this coefficient suggest that the consequences of deployment-induced skill decay may be amplified when task demands are highest. For obstetricians, the estimated effects across all complexity thresholds are close to zero and statistically insignificant. This result aligns with earlier findings that post-deployment changes in C-section decisions are less influenced by procedural difficulty and are instead likely tied to judgment-based confidence rather than technical execution.

We next test whether physician experience moderates the effects of deployment. Among surgeons, we find a statistically significant effect for those with less than 10 years of experience. This indicates that less experienced surgeons experience a roughly five percentage point larger increase in readmissions following deployment relative to more experienced peers.

5 Discussion and Conclusion

In this paper we empirically test whether long breaks in production lead to worse patient outcomes (”forgetting”) and how long it takes physicians to return to their past baseline (”learning by doing”). We use a difference in differences model in a military context where physicians are pulled from hospitals and sent overseas where they rarely use their skills and are extremely unlikely to perform any complex surgeries. By considering long breaks in production we are able to provide new information on the volume-outcome effect, specifically as it relates to skill depreciation and re-accumulation.

Our results indicate that temporary breaks in production lead to decreased quality of care and that the effects are quantitatively important. This is further confirmation of the well documented volume-outcome pattern that has been shown in other medical care settings by previous studies. There is continuing debate over whether this is a product of learning (‘practice makes perfect’) or selection of better quality physicians into larger hospitals. We believe the unique features of our setting rule out the latter explanation, and hence provides novel empirical evidence on physician learning.

Furthermore, by looking at two very different specialties we are able to differentiate physical skill from decision-making mechanisms. We find fairly short impact on physical skill and larger and longer-lasting effects on decision-

making.

Our results have implications for both policy and management. From a management perspective, workers returning from career breaks may want to either shadow another worker or begin with simple tasks prior to conducting more complex procedures. Given the recent emphasis on reducing worker burnout, managers may want to schedule routine breaks with a strong plan on how to reintegrate these employees in a manner that minimizes the effects of skills depreciation. Recent work by Rachet-Jacquet [2022] has shown that small breaks may actually improve outcomes which combined with this study indicates there may be a trade-off between skill depreciation and employee morale.

Beyond the health care setting, depreciation in human capital among complex tasks has large managerial implications. Managers not only allocate tasks, but approve broadening assignments, vacations, and other breaks that decrease volume of specific complex tasks. There are many benefits to these decisions, but they must be evaluated in light of the potential negative effects on performance. More so, managers may want to plan to assign moderately difficult tasks that allow a worker to ease back in and gain the depreciated human capital in a lower-stakes environment after a long break. Managers should specifically plan for how to reintegrate workers when their jobs entail judgement-based decision making.

In continuing work, we analyze the frequency and types of surgeries physicians perform in their return. While we focus on outcomes for complex surgeries, surgeons perform frequent minor surgeries between major surgeries. We intend to analyze whether surgeons who perform more minor surgeries prior to their major surgeries regain their skill faster than those who perform fewer minor surgeries.

References

Andrew C Baker, David F Larcker, and Charles CY Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022.

Chantel Boyens, Michael Karpman, and Jack Smalligan. Access to paid leave is lowest among workers with the greatest needs: Findings from the december 2021 well-being and basic needs survey. https://www.urban.org/sites/default/files/2022-07/Access%20to%20Paid%20Leave%20Is%20Lowest%20among%20Workers%20with%20the%20Greatest%20Needs.pdf, July 2022. Accessed: 2025-09-01.

Helen Buchanan and Lana Van Niekerk. Work transitions after serious hand injury: Current occupational therapy practice in a middle-income country. *Australian occupational therapy journal*, 69(2):151–164, 2022.

Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454, 2019.

David M Cutler. The incidence of adverse medical outcomes under prospective payment. *Econometrica*, 63(1):29–50, 1995.

Manasi Deshpande and Yue Li. Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy*, 11(4):213–248, 2019.

Caroline Dowd-Higgins. Hitting the pause button: 2024 guide to taking a career break and coming back. LinkedIn Pulse, 2024. URL https://www.linkedin.com/pulse/hitting-pause-button-2024-guide-taking-career-break-coming-bi1pc/. Accessed: 2025-09-01.

Shlomo Globerson, Nissan Levin, and Avraham Shtub. The impact of breaks on forgetting when performing a repetitive task. *IIE transactions*, 21(4):376–381, 1989.

Atul Gupta. Impacts of performance pay for hospitals: The readmissions reduction program. *American Economic Review*, 111(4):1241–1283, 2021.

Kareem Haggag, Brian McManus, and Giovanni Paci. Learning by driving: Productivity improvements by new york city taxi drivers. *American Economic Journal: Applied Economics*, 9(1):70–95, 2017.

Jason M. Hockenberry and Lorens A. Helmchen. The nature of surgeon human capital depreciation. *Journal of Health Economics*, 37:70 – 80, 2014. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2014.06.001.

Stephen F Jencks, Mark V Williams, and Eric A Coleman. Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428, 2009.

Jennifer Liu. Career breaks are common—so why are we still hiding them on résumés? *CNBC Make It*, November 2023. URL https://www.cnbc.com/2023/11/07/career-breaks-are-common-so-why-are-we-still-hiding-them-on-resumes.html. Accessed: 2025-09-01.

Andrew S Malin and Patrick S Brannan. The deployed military orthopaedic surgeon: experiences of a recent iowa graduate in afghanistan. *The Iowa Orthopaedic Journal*, 32:231, 2012.

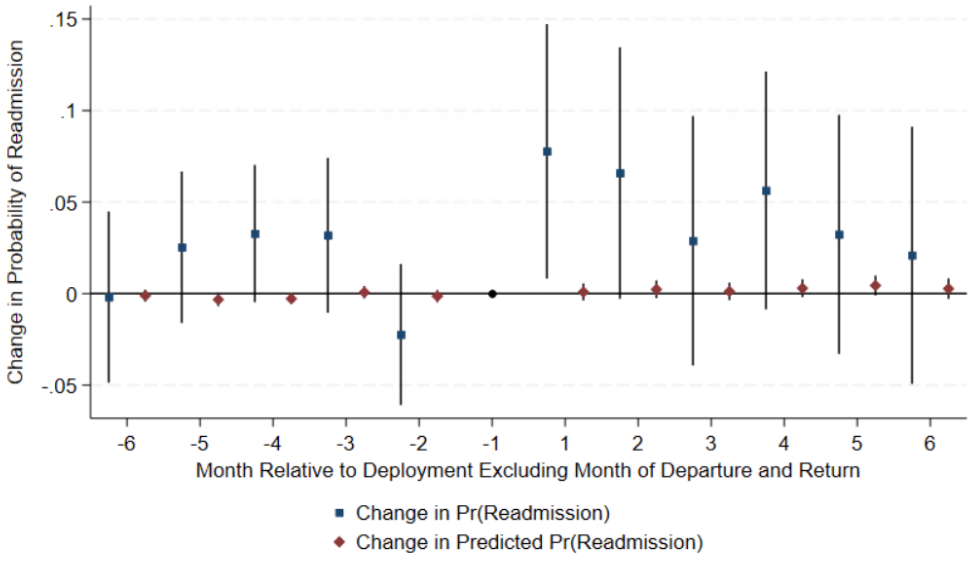
Ben Ost. How do teachers improve? the relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, 6(2):127–151, 2014.

Laurie Rachet-Jacquet. Do breaks from surgery improve the performance of orthopaedic surgeons? *Journal of Health Economics*, 85:102667, 2022. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2022.102667. URL https://www.sciencedirect.com/science/article/pii/S0167629622000832.

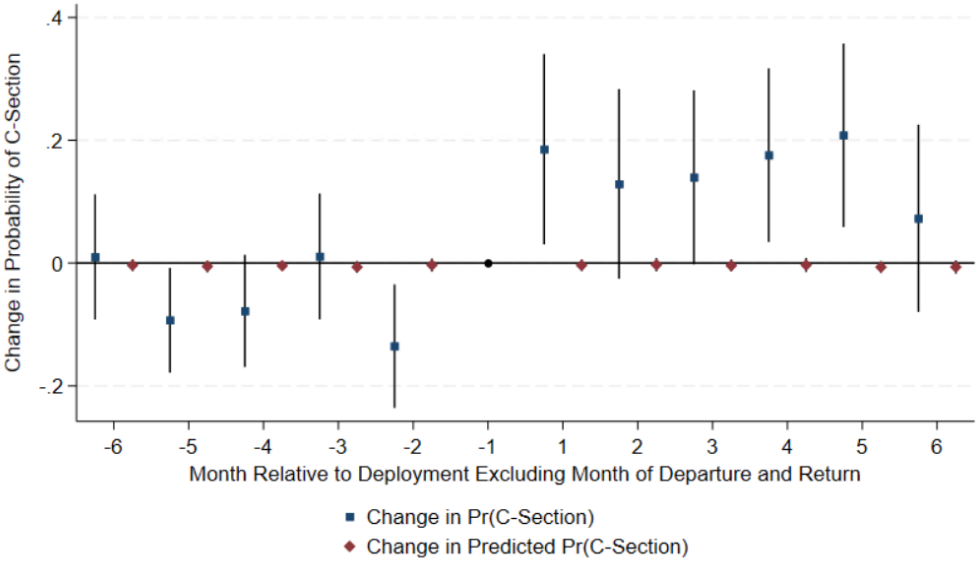
Scott M Shafer, David A Nembhard, and Mustafa V Uzumeri. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, 47(12):1639–1653, 2001.

Katy Tynan and Kristi Hedges. Research: Résumé gaps still matter. Harvard Business Review, July 2024. URL `https://hbr.org/2024/07/research-resume-gaps-still-matter`. Accessed: 2025-09-01.

U.S. Department of Labor. Family and medical leave act (fmla) 2018 employee and worksite surveys: Final report. `https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/WHD_FMLA2018SurveyResults_FinalReport_Aug2020.pdf`, August 2020. Accessed: 2025-09-01.



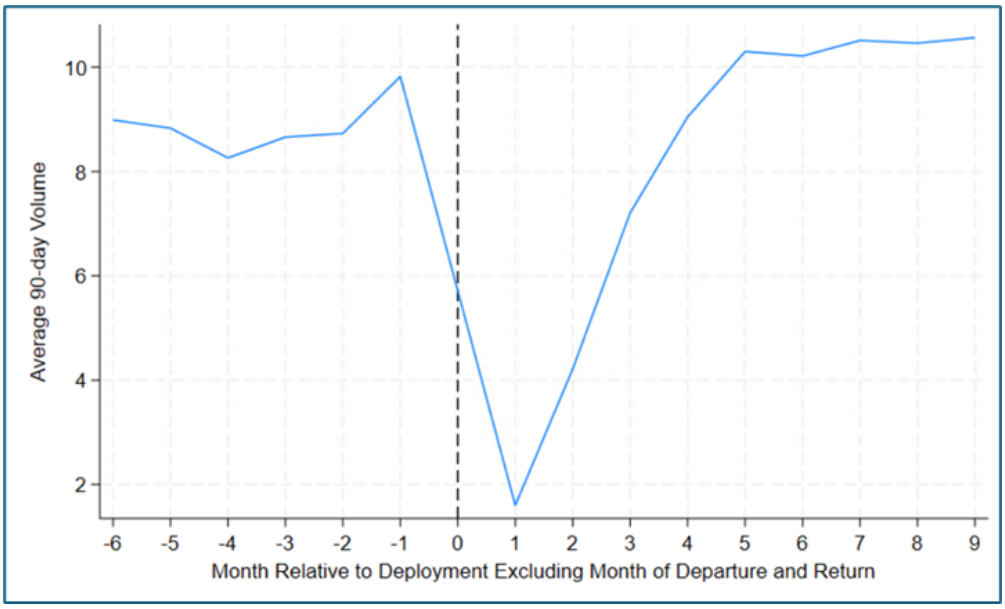
(a) Surgeon Readmission/Predicted Readmission Rate



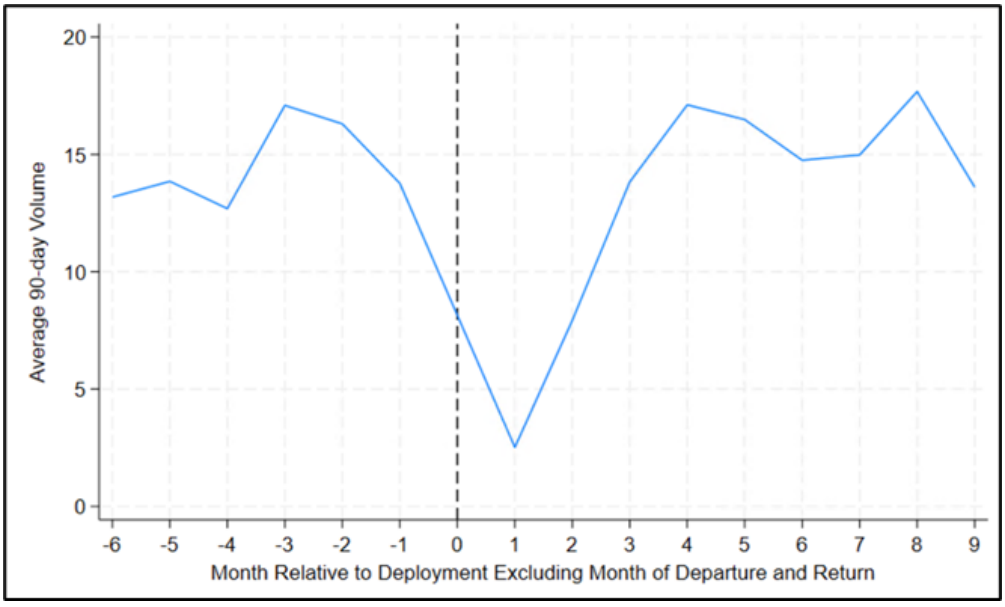
(b) Obstetric C-Section rate

Figure 1: Effect on Readmission Rates/C-sections of Major Procedures

Notes: The figure presents dynamic effects on 90-day readmissions (panel A) and C-section probability (panel B) for patients treated by deployers, relative to those treated by non-deployers. The coefficients were obtained by estimating Equation 1 using the patient-level analysis sample. Figures also show changes in predicted rates of each outcome based on observable patient characteristics. Error bars depict 95% confidence intervals. The standard errors are clustered by physician.



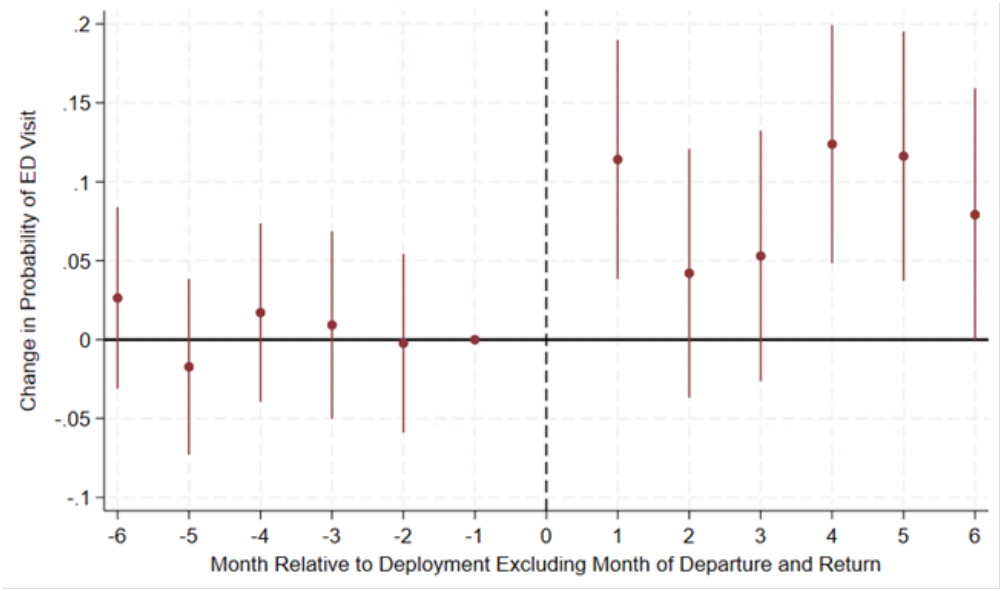
(a) Surgeon Rolling 90-Day Volume



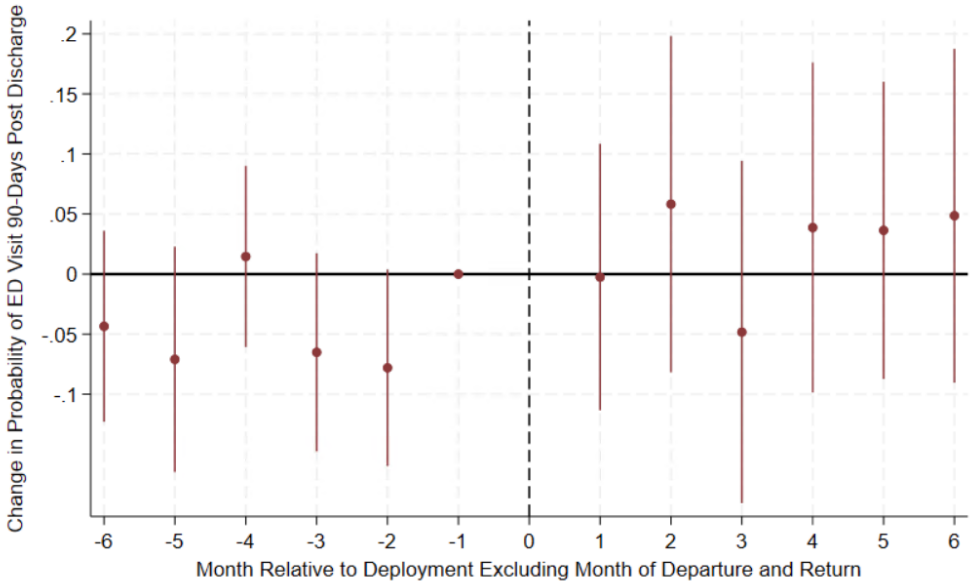
(b) Obstetric Rolling 90-Day Volume

Figure 2: Cumulative Rolling 90-Day Patient Volume

fnoteNotes: The figure presents dynamic changes in average 90 day volume for deploying physicians.



(a) Surgeon: Patient ED Visit



(b) Obstetric: Patient ED Visit

Figure 3: Effect on Patients Seeking Emergency Dept. Care within 90 days of Dishcarge

Notes: The figure presents dynamic effects on eac outcome for patients treated by deployers, relative to those treated by non-deployers. The coefficients were obtained by estimating Equation 1 using the patient-level analysis sample. Error bars depict 95% confidence intervals. The standard errors are clustered by physician.

Table 1:
Descriptive Statistics

	Surgical		Obstetrics	
Variable	Mean	Std Dev	Mean	Std Dev
Panel A - Patient Characteristics				
Age	34.12	17.52	26.95	5.07
Gender - Female	0.28	0.45	1.00	—
Race - White	0.62	0.48	0.52	0.49
Elixhauser Score	0.51	0.94	0.25	0.52
Panel B - Provider Characteristics				
Number of Procedures	1.82	1.55	2.33	1.40
Physician Experience (Years)	11.39	6.12	9.10	5.96
Panel C - Outcomes				
Readmission Rate	0.13	0.34	0.03	0.16
ED Visit Post-Discharge	0.20	0.40	0.15	0.36
C-Section Rate			0.25	0.43
N-Unique Hospital Stays	28,725		47,551	
N-Physicians	690		486	

Note: The table presents descriptive statistics on the analysis sample. Panel A shows characteristics for surgical and obstetric patients. Panel B displays the provider characteristics and panel C shows outcome measures for our three main outcome measures. Readmissions and ED Visits are calculated based on 90-day all-cause readmission or ED visits respectively. C-section rates are based on the MSDRG listed on the facility claim.

Table 2: Balance Table

Variable	Surgical			Obstetrics		
	Non-Deployer	Deployer	T-Stat	Non-Deployer	Deployer	T-Stat
Panel A - Patient Characteristics						
Age	33.88	33.19	-0.85	26.89	27.06	1.09
Gender - Female	0.28	0.29	0.51	0.99	0.99	-1.04
Race - White	0.63	0.63	0.53	0.52	0.52	0.07
Elixhauser Score	0.50	0.50	0.15	0.24	0.26	1.49
C-Sections				0.25	0.25	-0.16
N-Hospital Stays	40,795	9,375		68,082	2,865	
Panel B - Hospital Stay Characteristics						
Number of Procedures	1.81	1.78	-0.43	2.32	2.39	0.91
Readmission Rate	0.13	0.14	0.41	0.03	0.03	0.02
Gender - Female	0.14	0.09	-1.51	0.60	0.53	-0.72
Race - White	0.86	0.86	-0.03	0.75	0.75	-0.03
Age	40.71	40.02*	-1.78	37.70	37.62	-0.10
Physician Experience (Years)	11.08	10.83	-0.56	8.70	9.38	0.81
N-Physicians	321	369		399	87	
Joint Significance Test	0.85 (p-value = .5740)			0.80 (p-value = .6141)		

Note: The table presents descriptive statistics on patients served by non-deployer and deployer physicians in our sample (Panel A) as well as on the encounters performed by these two groups of physicians (Panel B). Columns 3 and 6 presents the t-statistic from a 2-sample test of means. Our measure of readmission is 90-day all-cause readmission.

Table 3:
Alternative Explanations

	(1) (Patient Volume)	(2) (Predicted Outcome)	(3) (Elixhauser)	(4) (Patient Age)
Panel A - Surgical	-0.068	0.003	-0.021	-0.297
	(0.099)	(0.002)	(0.044)	(0.907)
Sample Mean	2.38	0.135	0.51	34.12
Panel B - Obstetrics	0.285	-0.001	0.023	-0.215
	(0.231)	(0.003)	(0.060)	(0.523)
Sample Mean	3.27	0.255	0.25	26.94
Controls	No	Yes	Yes	Yes
Fixed Effects				
Match-Group - Physician	Yes	Yes	Yes	Yes
Match-Group - Relative-Time	No	Yes	Yes	Yes
MSDRG	No	Yes	Yes	Yes
Clinic-Year	Yes*	Yes	Yes	Yes
Month of Year	Yes	Yes	Yes	Yes

Note: This table presents outcomes for alternative explanations including a change in volume and patient risk. Results are +/- 6 months. Standard errors are clustered by surgeon. *Volume regression only has clinic fixed effect rather than clinic-year interacted.

Table 4:
Top Ten Surgical CPTs: Non-Deployed and Deployed

Panel A: Top Ten Non-Deployed Surgical CPTs and Frequency During Deployments

	Procedure	Annual Volume (All Deployments)
1	Laparoscopy, surgical; appendectomy	3
2	Total knee arthroplasty	0
3	Tonsillectomy and adenoidectomy; younger than age 12	0
4	Arthroscopically aided anterior cruciate ligament (ACL) reconstruction	0
5	Total hip arthroplasty	0
6	Laparoscopy, surgical; cholecystectomy	4
7	Thyroidectomy, total or complete	0
8	Septoplasty or submucous resection, with or without cartilage scoring, contouring or graft	0
9	Laminotomy, with decompression of nerve root(s); one interspace, lumbar	0
10	Tonsillectomy, primary or secondary; age 12 or over	0

Panel B: Top Ten Deployed Surgical CPTs (Facilities with ORs Only)

	Procedure	Annual Volume (All Deployments)
1	Laparoscopy, surgical; repair initial inguinal hernia	16
2	Repair initial inguinal hernia; reducible	11
3	Excision dermoid cyst, nose; simple, skin, subcutaneous	5
4	Laparoscopy, surgical; repair recurrent inguinal hernia	4
5	Adjacent tissue transfer or rearrangement, eyelids, nose, ears and/or lips	4
6	Laparoscopy, surgical; cholecystectomy	4
7	Laparoscopy, surgical; appendectomy	3
8	Suture of infrapatellar tendon; primary	3
9	Repair umbilical hernia, age 5 or older; incarcerated or strangulated	3
10	Repair umbilical hernia, age 5 or older; reducible	3

Note: The table presents the top ten complex surgeries while in traditional hospitals (panel A) and while deployed (panel B). Column 3 displays the total annual volume of each of these procedures across all deployed physicians each year.

Table 5:
Probability of Behavioral Health Diagnosis

	(1) (All)	(2) (Obstetrician)	(3) (Surgeon)
Probability of Mental Health Dx (All)	-0.001	0.006	-0.002
	(0.005)	(0.013)	(0.006)
Sample Mean	0.024	0.024	0.023
Probability of Anxiety/Stress Dx	0.002	0.008	0.005
	(0.004)	(0.010)	(0.005)
Sample Mean	0.013	0.014	0.011
Probability of Mood/Affect Dx	0.001	0.006	-0.004
	(0.003)	(0.014)	(0.004)
Sample Mean	0.012	0.011	0.013
Probability of Substance Abuse Dx	-0.001	0.000	-0.001
	(0.002)	(0.001)	(0.002)
Sample Mean	0.002	0.002	0.001
Probability of Elixhauser MH Diagnosis	0.004	-0.002	0.002
	(0.004)	(0.012)	(0.006)
Sample Mean	0.015	0.014	0.016
N-Provider-Months	55,095	31,565	23,521

Note: The outcomes are probabilities of receiving care for mental health related concerns after a deployment. Results are +/- 6 months. Standard errors are clustered by surgeon. ICD9 and ICD10 codes used to determine category of diagnose. In ICD10, values were determined using icd10data.com: F30-F39 (mood and affective disorders), F40-F48 (anxiety and stress-related disorders), and F10-F19 (substance abuse disorders). The (All) category is combining the above three categories if one was ever present. Other categories were considered, but no providers were ever diagnosed in the sample window: F50-59 (behavioral syndromes), F60-69 (adult personality), and F99 (unspecified).

Table 6:
Main Results

# of Months	(1) (1-2)	(2) (3-4)	(3) (5-6)
Surgical: Readmission Rate			
Cumulative Effect	0.062**	0.049*	0.039
	(0.031)	(0.029)	(0.028)
Marginal Effect	0.062**	0.035	0.020
	(0.031)	(0.029)	(0.029)
Obstetric: C-Section Rate			
Cumulative Effect	0.187***	0.196***	0.191***
	(0.070)	(0.065)	(0.064)
Marginal Effect	0.187***	0.204***	0.182***
	(0.070)	(0.064)	(0.066)
Controls	Yes	Yes	Yes
Fixed Effects			
Match-Group - Physician	Yes	Yes	Yes
Match-Group - Relative-Time	Yes	Yes	Yes
MSDRG*	Yes	Yes	Yes
Clinic-Year	Yes	Yes	Yes
Month of Year	Yes	Yes	Yes
Sample Mean (Readmission / C-Section)	0.13 / 0.25		
N (Readmission / C-Section)	49,915 / 70,903		

Note: The table presents coefficients on 1(post-deployment). The outcomes are 90-day probability of readmission at 90-day time intervals for Global90 procedures for surgeons and the probability of a C-section for obstetricians. Standard errors are clustered by physician. * indicates MSDRG fixed effects not included in C-Section outcome due to collinearity.

Table 7:
Main Results - ED Visits

	(1)	(2)	(3)
# of Months	(1-2)	(3-4)	(5-6)
Surgeons			
Cumulative Effect	0.072**	0.079***	0.084**
	(0.032)	(0.030)	(0.030)
Marginal Effect	0.072**	0.086***	0.094***
	(0.032)	(0.032)	(0.033)
Obstetricians			
Cumulative Effect	0.061	0.045	0.060
	(0.057)	(0.058)	(0.057)
Marginal Effect	0.061	0.030	0.089
	(0.057)	(0.063)	(0.058)
Controls	Yes	Yes	Yes
Fixed Effects			
Match-Group - Physician	Yes	Yes	Yes
Match-Group - Relative-Time	Yes	Yes	Yes
MSDRG	Yes	Yes	Yes
Clinic-Year	Yes	Yes	Yes
Month of Year	Yes	Yes	Yes
Sample Mean (Surgeons/Obstetricians)		0.20 / 0..15	
N		49,915 / 70,903	

Note: The table presents coefficients on 1(post-deployment). The outcomes are 90-day probability of being seen in the Emrgency Room following a Global90 inpatient stay. Standard errors are clustered by physician.

Table 8:
Heterogeneous Treatment Effects by Specialty

	DV: Readmissions				**DV: C-Sections**			
	Coef.	Std. Err.	t	P-Value	Coef.	Std. Err.	t	P-Value
multiple-procedures: (>2 procedures)								
β_1: Baseline DiD	0.039	0.029	1.33	0.184	0.202***	0.064	3.13	0.002
β_2: Triple difference	0.004	0.024	0.18	0.854	-0.016	0.033	-0.48	0.629
multiple-procedures: (>3 procedures)								
β_1: Baseline DiD	0.031	0.029	1.08	0.280	0.195***	0.064	3.04	0.003
β_2: Triple difference	0.057*	0.030	1.88	0.060	-0.026	0.043	-0.59	0.552
multiple-procedures: (>4 procedures)								
β_1: Baseline DiD	0.038	0.029	1.34	0.179	0.196***	0.064	3.07	0.002
β_2: Triple difference	0.014	0.040	0.34	0.734	-0.004	0.063	-0.07	0.948
Inexperienced in Years: (<10 years)								
β_1: Baseline DiD	0.021	0.031	0.67	0.502	0.194**	0.076	2.55	0.011
β_2: Triple difference	0.050**	0.022	2.26	0.024	-0.030	0.060	-0.51	0.613
Deployment Length: (>235 days)								
β_1: Baseline DiD	0.038	0.030	1.25	0.212	0.163**	0.069	2.36	0.019
β_2: Triple difference	-0.017	0.025	-0.69	0.491	-0.116	0.083	-1.40	0.161

Note: The table presents coefficients on the triple interaction of the number of procedures, experience, and deployment length for 90-day readmissions and C-section rates using equation 2 described in the text. Standard errors are clustered by physician. MSDRG fixed effects not included in C-Section outcome due to collinearity.

Table 9:
Learning Model

	(1) (Obstetrician)	(2) (Surgeon)
ln(rolling 90-day volume)	-0.044*	-0.022*
	(0.022)	(0.011)
ln(rolling prior year volume)	0.029	-0.017
	(0.044)	(0.013)
Predicted Outcome	0.313	0.936**
	(0.315)	(0.419)
Patient Age	-0.005	-0.000
	(0.015)	(0.004)
(Patient Age)^2	0.000	0.000
	(0.000)	(0.000)
Patient - White	-0.028	0.006
	(0.022)	(0.012)
Elixhuaser Score	0.042	0.013
	(0.047)	(0.015)
(Elixhauser Score)^2	-0.038	0.003
	(0.029)	(0.004)
# of Diagnoses	0.030***	0.007*
	(0.006)	(0.004)
Sample Mean (Volume)	14.17	9.14
N-Surgeries	1,712	3,472
N-Providers	45	128

Note: This table presents the results of the learning model as described in the text.