Baby bumps in the road: The impact of parenthood on job performance, human capital, and career advancement

Olivia Healy
Cornell University

Jennifer Heissel
Naval Postgraduate School

November 19, 2021

Abstract

We evaluate month-by-month impacts of parenthood on men and women's job performance, human capital accumulation, and career advancement using detailed data from the U.S. Marines. We estimate traditional event study models around the first birth followed by a matching strategy. Our matching strategy uses LASSO-selected predictors of a first birth to assign "placebo births" to non-parents. We study several standardized measures of job performance: job-relevant physical fitness test scores, supervisor-ratings of job proficiency, and scores on a firm-specific task (rifle and pistol marksmanship). We also examine months spent in firm-specific training, formal education degree attainment, and promotions.

We find negative impacts on parents' job performance and accumulation of firm-specific training in the two years after having a child, concentrated mainly among women. Consistent with these findings, women's promotion trajectories slow in response to childbirth. Men have slightly lower physical performance in the first year postbirth but are otherwise largely unaffected by parenthood. Slower promotion trajectories for mothers are driven by women missing key job performance tests due to pregnancy and postpartum waivers. In a complementary analysis, we exploit sudden policy changes to the length of paid maternity leave. Longer leaves exacerbate declines in women's job-related physical fitness but do not consistently change their promotion trajectories.

JEL Classification: J24, J16, J18, J45
Keywords: Parenthood, Promotion, Child Penalty, Parental Leave

1 Introduction

After having a first child, women experience large and persistent declines in earnings but men do not. This statistical fact, termed the "child penalty," is well established and holds across countries and contexts (Aguilar-Gomez et al., 2019; Andresen and Nix, 2020; Angelov et al., 2016; Barth et al., 2017; Bertrand et al., 2010; Kleven et al., 2019a,b). It also explains much of the overall gender earnings gap between men and women in higher-income countries (Kleven et al., 2020, 2019b). Prior research shows that child penalties result largely from differences in response to childbearing between mothers and fathers on three margins: (1) hours worked conditional on employment, (2) exit rates from the labor market, and (3) wages (Kleven et al., 2019b). Yet, the mechanisms that drive these gendered employment responses to childbearing are less understood. Few data sources track men and women's work outcomes over time and at the level of detail needed to investigate differential labor market responses following a birth. Better evidence on the mechanisms that drive these differences is critical for developing policies to address child penalties and gender gaps in earnings that systematically affect women.

In this paper, we use administrative records on servicemembers in the U.S. Marine Corps to overcome some of the limitations of prior research. Our primary advantage is access to detailed, consistently measured, longitudinal data that allow us to focus on outcomes beyond standard labor market measures (e.g., hours worked, employment, and wages). Key outcomes in this paper include high-frequency measures of individuals' on-the-job performance, job-specific and general human capital accumulation, and career advancement. Further, the Marine Corps is a setting where workers largely stay on-the-job working similar hours after becoming parents due to multiyear contract requirements. As such, our focus on this single employer helps hold two key drivers of the child penalty constant, limiting selection out of the workforce and minimizing reductions in hours worked in response to parenthood. In turn, our empirical analysis sheds light on the following question: if men and women were to stay on the job working similar hours after childbirth would a child penalty in earnings emerge, and why?

We put forth two possibilities in response to this question, both focused on the immediate postbirth experience of parents that return to work. First, mothers may struggle more than fathers to keep up their performance after having a child, whether due to greater child care responsibilities, unique biological impacts of birth, or other constraints. If mothers struggle on the job upon returning to work, they may ultimately exit the labor market or reduce hours worked a few years down the road. Changes in work performance in the immediate postbirth period among those who remain employed could then drive medium or long-term changes in labor market participation that give rise to lasting child penalties. Another possibility is that if women face unique challenges in job performance and skill advancement after having children, their their promotion rates may slow. Delays in career advancement due to childbearing could slow mothers' wage trajectories relative to fathers, with the potential to create child wage penalties for mothers in the long run.

We explore empirical evidence behind these two possibilities by analyzing the impacts of a first birth on job performance and human capital development, which jointly determine promotion decisions in the Marines. Our data allow a fine-grained assessment of the month-by-month consequences of parenthood in the immediate two years following a birth. Measures of job performance include scores on Marine-specific assessments of physical fitness performance and marksmanship skills, as well as supervisor ratings of overall job performance. Human capital outcomes include months spent in job-specific training and the accumulation of formal educational degrees (e.g., associate's or bachelor's degrees). We explore changes in these outcomes among a sample of parents who remain employed for at least two years after having a child. We then investigate whether having a child causes mothers or fathers to experience slowdowns in accumulated promotions during the same two-year period.

Our empirical strategy begins with a traditional event study approach, estimated separately for men and women. We leverage the precise month and year of childbirth as an exogenous shock to parents' work performance outcomes and use data on Marines who do not have a child during the study window to account for secular time trends in outcomes. We then move to a matching strategy that assigns nonparents to "placebo births" based on LASSO-selected observable characteristics that best predict the likelihood of a first birth. Using this strategy, we compare outcomes between parents and nonparents before and after a birth versus a placebo birth event, separately for men and women. The placebo birth matching strategy better corrects for observable differences between parents and nonparents that might have affected parents' postbirth outcome trajectories in the absence of a having a child.[1] The matching strategy also allows us to study cumulative changes in job outcomes (e.g., total months spent in training or number of promotions achieved), given that we can align parents and nonparents by event-time and then track successive changes in the outcome during pregnancy/placebo pregnancy and through the postbirth/placebo postbirth period. For this same reason, the matching strategy enables us to explore subgroup differences in the pattern of results based on prepregnancy characteristics.[2] Following Kleven et al. (2019b), the placebo birth matching strategy is our preferred approach.

We find small but persistent effects of parenthood concentrated among mothers on work performance, firm-specific human capital accumulation, and promotion trajectories. The largest negative impacts on mothers' physical performance and marksmanship scores (two firm-specific performance assessments) occur closest to the childbirth event, as soon as testing exemptions around pregnancy/birth for each assessment are lifted. Two years postbirth, mothers' physical performance remains 0.2 standard deviations lower than matched nonmothers'. Mothers' supervisor-rated job performance is consistently 0.1–0.2 standard deviations lower than nonmothers' postbirth. Mothers also accumulate fewer months of training relative to nonmothers, with gaps beginning during pregnancy and continuing through two years postbirth. Consistent with these patterns, mothers' promotion trajectories slow as a result of having a child. The only outcome unaffected for mothers is the number of formal educational degrees attained.

We observe minimal impacts of parenthood on fathers' outcomes. The birth of a child leads to small, short-lived declines in fathers' physical performance scores one month postbirth, but fathers' performance recovers. For other outcomes, results among fathers are sometimes inconsistent between the event study and placebo birth matching approach and, if anything, show positive albeit small effects of parenthood on marksmanship scores, training, and the number of educational degrees fathers attain in the two years postbirth.

[1] Put differently, if parents fundamentally differ from nonparents on observable characteristics, using all nonparents to estimate secular time trends in outcomes that would have occurred in the absence of parenthood (as we do under the event study approach) may provide a poor approximation of the counterfactual.

[2] In an event study approach, nonparents help to estimate secular time trends in the outcome (i.e., time fixed effects), but they are never assigned event time given that they do not experience the event. As such, it is not possible to pinpoint pre/postbirth periods for nonparents and use their outcomes as a counterfactual. When we assign placebo births, it creates a clear pre/post timeframe for control cases.

Taken together, our results suggest that the postbirth period may be uniquely challenging for women and that these challenges likely drive delays in mothers' career advancement. To investigate this, we explore potential mechanisms behind the negative impact of parenthood on mothers' promotions, using a back-of-the-envelope approach. We control for endogenous outcomes of childbirth (changes in job performance and human capital accumulation) when predicting the impact of birth on promotions. We find changes in physical performance partly account for mothers' promotion delays. However, the biggest factor driving mothers' promotion delays is the number of missing performance assessments mothers accumulated due to pregnancy- and birth-related test waivers. Specifically, physical fitness assessments taken every 6 months are a component of the promotion decision. Mothers are waived from taking these assessments during pregnancy and at least 6 months postbirth, but other individuals may miss these tests due to idiosyncratic mishaps like sprained ankles. Missing physical fitness assessments harms promotion rates for all Marines, and once we control for missing these tests, mothers' promotion trajectories are indistinguishable from matched nonmothers'.

Last, we explore whether variation in the length of maternity leave offered to women helps buffer against declines in performance, skill investment, and promotion accumulation for new mothers. Two major policy changes during our study window create variation in the length of paid leave available to women, increasing it from 6 to 18 weeks and then down to 12 weeks. We find longer paid leave exacerbates initial declines in physical performance, particularly during a brief policy period where mothers' additional weeks of leave could only be taken after they had already returned to work following initial postbirth leave (see Bacolod et al., 2021). However, differences in leave length do not consistently predict changes in training, education, or promotion among women after having a child. Results suggest that longer periods away from work may exacerbate declines in job-specific physical performance initially upon return to work, but these differences do not persist beyond one year nor do they hold for other outcomes.[3]

[3] Our findings do not rule out that longer leave might have positive effects in other domains. Balser et al. (2020) find maternity leave extensions from 6 to 12 weeks reduce postpartum depression diagnoses among mothers in the Army and Air Force. In U.S. and international contexts, access to paid parental leave improves mothers' physical and mental health (Bullinger, 2019; Butikofer et al., 2018), duration of breastfeeding (Pac et al., 2019), infant health (Bullinger, 2019), and time spent with children after return to work (Bailey et al., 2019; Trajkovski et al., 2019).

Our study contributes to a longstanding literature on the impact of fertility on parents' employment outcomes. Prior research has necessarily focused on new parents' labor force attachment, hours worked, or wages, as these are often the only measures available (Agüero and Marks, 2011; Angrist and Evans, 1998; Bronars and Grogger, 1994; Cáceres-Delpiano, 2006; Cools et al., 2017; Cruces and Galiani, 2007; Jacobsen et al., 1999). We add to this literature by exploring how children impact mothers' and fathers' job performance, training, and education—key precursors to worker output and productivity—as well as career advancement. We estimate these effects separately for women and men because we want to explore whether parenthood also affects men along the margins we study. If that is the case, fathers would not be a good comparison point for mothers. We find evidence of small impacts on fathers, particularly in terms of immediate postbirth physical performance, supporting this choice.[4]

We inform potential mechanisms driving child penalties because our data allow us to trace the dynamic responses of men and women month by month *within* the first and second years of work after becoming parents. Prior papers that estimate child penalties use annual earnings or income and therefore cannot detect initial, within-year impacts of childbirth (Andresen and Nix, 2020; Angelov et al., 2016; Barth et al., 2017; Bertrand et al., 2010; Kleven et al., 2019a,b). Our finding that parenthood begins to impact women's performance *within* the first year after birth points to one mechanism by which child penalties may arise. Immediate postbirth declines in mothers' ability to perform at work and invest in skill development—declines that do not recover—may lead women, more so than men, to exit the labor market, cut back hours worked, or receive lower wages by the time these outcomes have been measured one year postbirth in other papers. We also find evidence that having a child slows mothers' promotion trajectories but not fathers, suggesting that even absent changes in employment and hours worked, children might hinder mothers' career advancement, giving rise to child penalties over the long run.[5]

[4]Kleven et al. (2019b) also find some evidence that men are impacted by childbirth. They estimate difference-in-differences event study models for men and women separately across first birth and present results in online Appendix B. They find fathers' lifetime earnings *are* affected by parenthood but that impacts are small (a 3 percent decline 10 years after a birth). This finding coincides with our result that first births lead to a small decline in fathers' physical performance in the U.S. Marines upon the arrival of the child.

[5]Most similar in spirit to our paper is a set of studies focused on changes in mothers' and fathers' job productivity across the transition to parenthood. For example, Azmat and Ferrer (2017) find that female lawyers with young children are less productive compared to male lawyers with young children in terms of hours billed annually, a key productivity measure in the legal profession. Kim and Moser (2020) similarly show that female scientists in the 1950s patented less during their childbearing years. They posit lower patenting productivity drives lower rates and slower speed of promotion to tenure for

Finally, our results provide evidence for workers not currently represented in the literature, specifically those with low levels of education in a more physically demanding work setting. The Marine Corps as a work setting parallels civilian settings where maintaining a baseline level of physical health is important for work performance. About 45% of jobs in the civilian labor market require at least medium physical strength, defined as work that involves frequent lifting or carrying of objects weighing up to 25 pounds (Bureau of Labor Statistics, 2017). Given its size and role in the United States, studying the military is important in its own right, beyond any implications for the civilian setting.

2 Institutional Background

The DoD is the world's largest employer, with a total of 1.3 million active-duty servicemembers. We focus on the Marines Corps, where administrative records on job performance are readily available. The Marine Corps is an immediate response force, ready to deploy quickly to support combat missions on sea or land. The Marine Corps makes up nearly 15% of active-duty forces with roughly 185,000 active-duty Marines. The majority of Marines (92%) are male (Department of Defense, 2018), meaning our findings may be particularly relevant for other male-dominated industries or organizations.

Individuals begin in the Marine Corps either as a junior enlisted, akin to an entry-level civilian worker; or as an officer, akin to a civilian manager.[6] There are over 35 career fields in the Marines, and each has dozens of specializations, referred to as Military Occupational Specialties. Some career fields are specific to the military (e.g., infantry). Other career fields are also present in the civilian labor market such as food services, financial management, military police and corrections, and legal services.

A servicemember's occupational specialty and their assigned unit determines their day-to-day work

female scientists with children as compared to fathers and other women without children. Last, Gallen (2018) explores how firm output in Denmark varies by the gender and parenthood status of employees in private firms. She finds mothers are substantially less productive–according to a firm production function model–than other workers (nonmothers, fathers, and nonfathers), particularly during their childbearing years. By contrast, our paper homes in on a precursor to work output: work performance. Our results also provide evidence on a more diverse group of workers not currently represented in the literature.

[6]Enlisted servicemembers must have a high school degree and be between 18 and 29 years old when they begin. Officers must have at least a bachelor's degree upon entry. Roughly 89% of Marine Corps servicemembers are enlisted, and given age requirements, most are of prime childbearing age. Among Marines, 81% are under 30 years old, and more than one quarter are parents (Department of Defense, 2018).

environment. Our analytic sample predominantly consists of active-duty Marines who work full time, Monday through Friday. For these individuals, the workday typically begins with early morning physical training, followed by work assignments through the evening. Most servicemembers live and work on or near a military base and are stationed in the U.S. (83% of servicemembers; Department of Defense 2018), though some are stationed abroad.

Each military branch also has a Reserve component where individuals work part-time. Our sample also includes these reserve Marines. Part-time work as a Marine reserve requires (1) participating in training drills one weekend per month and (2) attending a two-week work program each year. Reservists typically work in civilian careers or are enrolled in higher education while they fulfill part-time Marine Corps reserve duties. Reservists can be called upon for active-duty deployment in times of war or national emergency, at which point they are active reserves and work for the Marines full-time. We group any active reserve Marines in our sample with active-duty servicemembers.

Marines sign a legally binding contract that outlines their required length of service. Initial enlisted contracts typically require 4 years of active-duty service, while the officer commitment is typically 3 years. Importantly for our purposes, these contracts limit the extent to which Marines can exit the labor force after they have a baby.

Effective job performance in the Marines requires both mental and physical acuity. The Marine Corps uses a standardized set of measures to evaluate performance among both active-duty and reserve Marines. Performance measures include physical performance, supervisor-scored job proficiency, and marksmanship tests. Scores on these measures are used to determine promotions, with different weights placed on each depending on the given promotion juncture.[7] Marines also go through long-term training exercises to develop job-specific skills (e.g., pilot training); training can last from a few weeks to several months. Many Marines also pursue educational degrees while on the job, either paid for as part of their job (e.g., being given orders to obtain a master's degree full-time) or paid for privately (e.g., an

[7] Promotions at lower ranks, E1 through E3, are relatively automatic after a given number of months in service and months in rank. A composite score based on performance metrics determines promotion from the rank of E4 to E5, conditional on meeting requirements for minimum time in service and minimum time in the current job level to be eligible for promotion. For promotion at ranks above E5, the same performance metrics, along with supervisor ratings, are reviewed by an evaluation board to determine promotion. As such, Marine Corps promotions are similar to civilian promotions. Both are based on work performance, but the Marine Corps promotion system is arguably more systematic.

enlisted Marine pursuing a bachelor's degree outside of work).[8] Training and education also contribute to promotion decisions. Based on these clearly identified measures, Marines can determine what they need to advance and, therefore, have especially strong incentives to perform well on measured assessments.

The DoD provides military parents with a number of family-friendly benefits, including fully paid parental leave. In a supplementary analysis, we focus on policy changes to the length of paid leave for primary caregivers (most often women) and refer to this as variation in maternity leave length.[9]

3 Data

We draw data from the Marine's Total Force Data Warehouse and obtain records on all active-duty and reserve Marines who served at any point during January 2010 through December 2019. Our data include basic descriptive information on servicemember characteristics (age, gender, race/ethnicity, education status, and AFQT/GCT scores—which measure aptitude and intelligence), dependent characteristics for spouses and children (exact date of birth, gender, and whether a spouse is in the military), and job characteristics of the servicemember (job type, rank, time in service, and unit location). Our first set of outcomes are the three primary measures of job performance used for promotion and retention decisions: physical ability assessments, supervisor ratings of job performance, and rifle marksmanship assessments.

Our first outcome, physical performance, measures job proficiency among Marines based on standard physical fitness tests. Marines take two tests per year: the physical fitness test (timed running, crunches, and upper-body strength) in the first half of the year and a combat fitness test (timed running, a combat-related obstacle course, and upper-body strength) in the second half of the year. Scores are awarded on a 300-point scale, which is adjusted for age and gender such that women do not need to do as many pull-ups as men, and older servicemembers do not need to run as fast as younger ones to achieve the same score. We standardize raw physical fitness scores by year, gender, and test type. We combine the Z-scores for the two tests into one measure, generally observed twice per year per Marine. Due to the

[8]Marines who obtain education as part of their job generally add time to their commitment as part of a "pay-back tour."

[9]Paternity leave (i.e., leave for "secondary caregivers") changed from 10 to 14 days during this time. We do not focus on paternity leave impacts in this paper, given the small magnitude of the change.

physically demanding nature of the assessment, women are not required to take the test when pregnant, and they are exempt from tests 6 months postbirth. We thus measure physical performance outcomes for mothers until 10 months before they give birth. We resume measurement at 8 months postbirth due to concerns that commanding officers may allow some women whose test dates should fall seven months after birth to skip the test during that assessment round (especially in December).

For our second outcome, supervisors evaluate Marines for performance purposes using one of two scales, depending on the Marine's rank. Junior enlisted receive proficiency and conduct marks ("Pro-Cons") at least twice per year, and senior enlisted and officers receive Fitness Report ("FitRep") scores at least annually. Both assessments require supervisors to rate a Marine's performance across a range of professional domains.[10] We standardize scores by year, gender, and measure. We then combine the Z-scores into one outcome we call job performance. We generally observe scores at least twice per year among junior enlisted, once in the first half and once in the second half. For senior enlisted and officers, we observe job performance ratings a minimum of once per year. If a Marine is transferred, discharged, or promoted, or if their supervisor changes, they receive additional performance ratings. Marines are relocated every few years, as are their designated supervisors. Decisions on relocation are made from a central location, which prevents Marines from manipulating their scores by selecting their supervisors (Cunha et al., 2018). Still, the subjective nature of these assessments means we cannot distinguish true changes in job performance from supervisors' *perceptions* of changes in performance using this measure.

Last, rifle and pistol marksmanship assessments evaluate Marines on their shooting skills, assessed with targets and associated points. Marksmanship is partially cognitive and requires concentration to do well. We standardize marksmanship scores by year, gender, and weapon. This exam is supposed to be taken once a year at junior levels and becomes optional for more senior Marines. Those who perform at the highest levels are exempt from re-testing the following year, making outcome data on this measure more sparse. Like physical performance tests, the requirement is waived for pregnant and postpartum

[10]Domains include mission accomplishment (job-specific aptitude, competence, technical knowledge, and practical skills), character (courage, effectiveness under stress, and initiative), leadership (setting the example, communication skills), and intellect and wisdom (professional military education, decision-making ability, and judgement), among others. We are missing ProCons for junior enlisted who left the Marines before October 2017. However, we observe the full history of performance ratings (including prior to October 2017) for any Marine who was active as of October 2017.

mothers on leave. Figure 1 displays the count of these three outcomes relative to birth for mothers; the figure shows a drop in observations around the birth event for physical performance and marksmanship scores. There is also ambiguity about whether the outcome occurred before or after the birth in the month of birth ($t = 0$). We exclude the shaded observations in our main parametric analyses.

Our second set of outcomes captures human capital development. For firm-specific human capital, we observe whether or not a Marine is in training on a month by month basis. We create a cumulative measure of months of training observed for each Marine since the time point before pregnancy. For general human capital, we observe the Marine's current level of formal education, which increases as the Marine gains a formal degree from an institution of higher education. We then count the number of degrees obtained since the time point before pregnancy. Formal education is part of the promotion process, so Marines are incentivized to keep this information up-to-date.

Finally, we track promotions over time. We count the number of promotions a Marine receives relative to 10 months before they have a child. A value of 1 on the variable indicates 1 promotion achieved since the time point before pregnancy.

Our preferred sample requires a semi-balanced panel to ensure our results are not driven by selective attrition. We require first-time parents be observed in the sample for 12 months prior to the birth and 24 months after (N=2,801 mothers). Under our placebo strategy, we also require mothers to have a measure of ability (AFQT or GCT), at least one observed pre-birth fitness score, and at least one same-gender non-parent of the same months of service, rank, and reserve status in the same $t = -10$ year; this reduces our sample to 2,568 mothers. For consistency, we use the same parent sample in the event study and placebo analysis. We also include a group of "nonparents," defined as Marines with at least 36 months total of service who do not experience a birth during the study window. Appendix Table **??** shows how the characteristics of the sample change based on a variety of possible sample restrictions.

We present characteristics in our preferred sample of first-time parents in Table 1, alongside characteristics of first-time civilian parents who are employed and have a child under the age of 1. First-time Marine parents are younger than their civilian counterparts and have much lower rates of college attendance and college completion. First-time Marine mothers and fathers also identify as Black or Hispanic at higher rates than first-time civilian parents. Marriage rates are generally similar: 87% of Marine vs. 83% of civilian fathers are married when they have their first child; and 68% of Marine vs. 78% of civilian mothers are married at first birth. We rely on the Standard Occupational Classification (SOC) system from O*NET, a federal standard used to classify workers into occupational categories, to explore the distribution of job types among Marines relative to civilians. We crosswalk Marine job codes to SOC codes and find that—outside of military-specific occupations—the largest share of first-time Marine fathers work in natural resources, construction, or maintenance (labeled "Construction/maint."), while the largest share of first-time Marine mothers work in sales or office roles. Most civilians who have a first child and stay in the workforce tend to be employed in management, business, science, or arts.

Only a small share of first-time Marine parents in our sample are officers (akin to civilian managers): 15% and 8% of Marine mothers and fathers, respectively. As such, the vast majority of first births occur to enlisted Marines. Finally, new Marine parents score just above average on required intelligence tests, including the AFQT and GCT.

Based on descriptive differences between Marine and civilian first-time parents, results from our analyses may generalize best to younger and less-educated workers as well as workers of color.

4 Empirical Approach

The ideal experiment to isolate the causal effect of fertility on men and women's work performance would randomly assign pregnancy and parenthood to workers. Random assignment would ensure that (on average) differences in outcomes were not driven by underlying characteristics of the types of people who chose to become parents but rather by the transition to parenthood itself. Of course, random assignment of childbirth is both unethical and unfeasible. Yet, a simple post hoc comparison of parents relative to nonparents is unlikely to recover a causal estimate of the effect of having a child. Those who opt into parenthood likely differ from nonparents in ways that might also correlate with work performance.

In the absence of a feasible experiment, we employ two alternative quasi-experimental approaches to estimate the impacts of parenthood: an event study approach and a placebo birth approach.

4.1 Event Study Strategy

We begin with a traditional event study strategy, which exploits variation in the precise timing of births to identify the effect of childbirth on first-time parents. If the transition to parenthood has an impact on health and performance, then the birth should generate a sharp change in these outcomes directly after it occurs. We can attribute any discontinuity in the outcomes at the time of the birth to the birth itself if we assume that other factors that shape job outcomes do not also undergo a sharp change in the same month as childbirth. In other words, while the choice to have a child may be endogenous, the exact timing of conception and subsequent childbirth serves as a shock to the outcomes of interest.

We build in a second source of variation to our event study approach by including Marines who do not give birth during the study window in the analysis. In this way, our event study is a form of a difference-in-differences model. Nonparents in the analysis help approximate counterfactual time trends in outcomes that all Marines would have experienced, assuming outcomes would have evolved similarly between first-time parents and nonparents absent childbirth. We also include nonparents in the event study analysis to limit the share of 2 x 2 comparisons that use early-treated units (i.e. parents with first births early in the study window) as controls in time trend estimation. In this way, we avoid some of the bias in estimates that can result from staggered treatment-timing event study approaches (see Baker et al., 2021 for an overview).

To implement our event study approach, we begin by estimating a fully flexible, dynamic specification separately for men and women, which allows us to test the assumption that outcomes change similarly for both parents and nonparents in that period. We implement this as follows:

$$Y_{it} = \alpha_i + \phi_t + \sum_{r=k_{min}}^{k_{max}} \mathbb{1}(t = t_i^* + r)\beta_r + \varepsilon_{it} \tag{1}$$

where β_r represents the effect of a birth in month t_i^* on outcomes r months later (or r months before, if $r < 0$). Effects are measured relative to month $r = -10$, which corresponds to 10 months prior to the birth and approximately 1 month before the start of the pregnancy. In other words, β_{12} would represent the average outcome 12 months after the birth, relative to $r = -10$ (the month before pregnancy). We censor r at $k_{min} = -18$, or $k_{min} = -24$, depending on the outcome.[11] We include α_i, an individual fixed effect to account for stable individual differences; ϕ_t, a month-by-year fixed effect to account for general changes over time in the outcome (e.g., changes in fitness test standards in a particular year); and error term ε_{it}. The estimation of month-by-year fixed effects is assisted by the inclusion of nonparents in the data, who provide an estimate of universal time-patterned changes to the outcome that affect all Marines similarly. We run all models separately by gender.

Beyond estimating individual month coefficients β_r, our goal is to identify any declines (or improvements) in performance during pregnancy, any drops immediately following birth, and any recovery following the immediate impact of birth. Similar to Lafortune et al. (2018), we create a more parsimonious model of performance changes over time relative to the prepregnancy period by defining:

$PregnancyDrop_{it} = 1$, if person i has a baby at time t_i^* and $t_i^* - 9 \leqslant t < t_i^*$, and 0 otherwise (for an intercept shift during pregnancy);

$PregnancyTrend_{it} = t - (t_i^* - 9)$, if person i has a baby at time t_i^* and $t_i^* - 9 \leqslant t < t_i^*$, and 0 otherwise (for monthly trends above and beyond the intercept shift during pregnancy);

$BirthDrop_{it} = 1$, if person i has a baby at time t_i^* and $t > t_i^*$ and 0 otherwise (for an intercept shift following birth);

$Recovery_{it} = t - (t_i^* + q)$, if person i has a baby at time t_i^* and $t > t_i^* + q$, where q is time point when the person is eligible to be observed for the given outcome after the birth, and 0 otherwise (for monthly trends above and beyond the intercept shift after birth); and

$\Delta Recovery_{it} = t - (t_i^* + 12)$, if person i has a baby at time t_i^* and $t > t_i^* + 12$, and 0 otherwise (for any change to the monthly recovery rate that begins at 13 months).

We exclude exclude $t = 0$ in this analysis because part of the month occurs before and part occurs after the birth. Using these different time frames, we estimate a semi-dynamic specification that fits

[11] We examine most outcomes 24 months prior to childbirth but measure supervisor ratings of job performance only 18 months prior to childbirth. Supervisor ratings cover roughly a 6-month retrospective period; therefore, any rating awarded 18 months before the birth implicitly covers an earlier time period of performance.

linear splines to portions of the data, as follows:

$$\begin{aligned}Y_{it} =& \alpha_i + \phi_t + \beta_0 PregnancyDrop_{it} + \beta_1 PregnancyTrend_{it} + \beta_2 BirthDrop_{it} \\ &+ \beta_3 Recovery_{it} + \beta_4 \Delta Recovery_{it} + \varepsilon_{it}\end{aligned} \tag{2}$$

Here, β_0 captures any immediate intercept shift and β_1 captures the monthly linear change in the outcome during the pregnancy period ($t = [-9, -1]$), relative to the prepregnancy period average ($t \leqslant -10$). The regression excludes $t = 0$ due to ambiguity about the timing of the outcome relative to birth for all outcomes; it also excludes fitness and marksmanship scores for excluded months. Coefficient β_2 represents the acute postnatal birth drop (if any) in the outcome in the first month parents are again assessed after childbirth. Then, β_3 captures the monthly linear recovery in the outcome following that initial drop, and β_4 captures any change in the monthly linear recovery rate in the child's second year of life ($t = [13, 24]$). All parameters are measured relative to the prepregnancy average ($t \leqslant -10$). We present a diagram of this model in Figure 2. We use this semi-dynamic spline specification to estimate postbirth effects for men and women at key time points (e.g., 12 and 24 months postbirth).

4.2 Matching Strategy to Assign Nonparents to Placebo Births

Our second approach follows Kleven et al. (2019b) by assigning a comparison group of nonparents to placebo births based on observable characteristics that predict selection into parenthood. Using this matched sample, we then compare promotion outcomes across the "birth event" for Marines with actual births and those with placebo births to trace whether gaps in outcomes arise.

Our placebo birth assignment process relies on an adaptive ridge LASSO (least absolute shrinkage and selection operator) model with 10-fold validation to determine the best predictors of a first birth. Possible predictors include age, race/ethnicity, military entrance exam scores (AFQT scores), marital status (including whether a spouse is also in the military), level of education, months of training, occupational field groups, reserve status, and most recent physical performance score, as well as interactions among all variables. We run this predictive model separately among women and men, measuring all characteristics of first-time parents at $t = -10$. We then use the predicted propensity from step one to match parents to up to five most observably similar nonparents. We require matches to have the same job rank, the same number of months in service, the same reserve status, and to have joined the Marines in the same year. We then assign placebo births to nonparents 10 months after the time of the match. The exact match on job rank, number of months in service, reserve status, and year of entry ensures that Marines with and without children have similar promotion histories and work contexts before the pregnancy begins.

Analyses then compare the changes in outcomes for first-time parents to the average change in outcomes for the 5 most observably similar nonparents to whom they match. Each parent receives a weight of 1 in the analysis, while each match receives a weight of 0.2 per match-month in the case where 5 distinct nonparents match to each parent.[12]

Appendix Table A.2 displays descriptive characteristics of the parents and their placebos separately by gender. We require the matches to be exact on months of service, rank (e.g, corporal or captain), reserve status, and year, so the groups exactly match on the first set of characteristics. The next set of characteristics are not exact matches but are included in the LASSO model (Along with interactions between all of these variables). The groups look almost identical, with a few small differences within gender. These differences are functionally small (e.g., mothers are 22.56 years old while their placebos are 22.77 years old). A joint test of statistical significance (predicting being a parent using the listed matching variables, months of service, reserve status, and fixed effects for year and rank) is statistically insignificant.[13] The slight differences arise because of small cell sizes from the exact match requirements; we retain the exact matches due to the importance of rank, time in service, and reserve status in the promotion process and year for accounting for any time-varying patterns (e.g., changes to a fitness test).

[12] We conduct nearest-neighbor matching with replacement, meaning the same nonparent can be matched to different first-time parents, or parents can match to less than 5 nonparents, who then get higher weights.

[13] For women, $F(41, 9065)=0.96$, p-value=0.5506. For men, $F(45, 81744)=1.27$, p-value=0.1034. Men have more variables because we never observe a mother at some senior grade levels.

Our primary parametric equation in this analysis is:

$$\begin{aligned} Y_{it} =& \alpha_i + \phi_t + \theta_0 PregnancyDrop_{it}^{all} + \theta_1 PregnancyTrend_{it}^{all} + \theta_2 BirthDrop_{it}^{all} + \\ & \theta_3 Recovery_{it}^{all} + \theta_4 \Delta Recovery_{it}{}^{parents} + \beta_0 PregnancyDrop_{it}^{parents} + \\ & \beta_1 PregnancyTrend_{it}^{parents} + \beta_2 BirthDrop_{it}^{parents} + \beta_3 Recovery_{it}^{parents} + \\ & \beta_4 \Delta Recovery_{it}{}^{parents} + \varepsilon_{it} \end{aligned} \tag{3}$$

Both parents and their matches have variables defined as above when labeled with superscript all, while variables with $parents$ are specific to parents. This creates a difference-in-difference-in-difference approach, where each β captures any change above and beyond that of nonparents with placebo births.

The variable difference we may be most worried about from Appendix Table A.2 is recent fitness score, which is slightly lower in the parents than the placebos. Our analysis assesses whether the $change$ postbirth differs for parents versus the placebos, so a small baseline gap would not artificially create our results. We can also estimate trend differences in outcomes between first-time parents and nonparents *before* $t = -10$. If nonparents provide a suitable comparison, we expect their outcomes will look similar to the parents' in the months before the pregnancy ($t < -10$), even if we do not $require$ them to move together. We find strong evidence that parents and nonparents are on parallel trends, lending confidence to this estimation strategy.

5 Results

5.1 Event Study

Figure 3 presents results from our flexible event study model estimated using Eq. 1. We first examine whether outcomes evolve smoothly leading into the pregnancy and birth, suggesting that nonparents provide a suitable counterfactual estimate of general time trends for the event study sample. The bottom left of each panel includes the p-value for whether the prepregnancy point estimates jointly equal zero. We begin with job-related physical performance scores (Panel A). Scores for women and men drop when initially assessed after having a child. We exclude outcomes for women during pregnancy and for seven months after birth, as policies allowed women to opt out of the physical assessments while pregnant and postbirth. Once women take the test, their performance declines are large and persistent. Even 24 months postbirth, women's physical performance scores continue to be lower than expected, after accounting for general time trends using the nonparent women. The prebirth periods are jointly statistically different from 0 (p-value of an F-test of joint significance=0.045), indicating that before giving birth mothers' trajectories differed from the nonparent women after accounting for time and individual fixed effects. For men, performance declines begin during the mother's pregnancy and reach their lowest point 1 month postbirth. But, the declines are short-lived.

Panel B shows evidence of lower supervisor ratings of job performance for women in the 2 years postbirth, though estimates are noisier than physical performance and do not appear to drop until late in the first year postbirth. This may occur because supervisor assessments encapsulate several months, and so the assessment is not the acute effect solely in the particular month. We also have fewer observations of job performance, which contributes to the noise. There appears to be limited impact of having a child on fathers' supervisor-rated job performance, though it may decline somewhat over time.

Finally, Panel C indicates that parenthood is unrelated to marksmanship for mothers, and if anything marksmanship improves postbirth for fathers. Given the rarity of this assessment, the estimates are noisy. Women are also exempt from these assessments during pregnancy and while on leave following birth.

5.2 Parametric Estimates

We next estimate a more parsimonious, semi-dynamic parametric model using the event study sample (using Eq. 2) and the placebo sample where our comparators are limited to those who are very similar to the parents on a variety of characteristics (using Eq. 3). The event study has the advantage of retaining more observations in the data, which may improve our statistical power for outcomes with less-frequent observations. The placebo analysis has the advantage that the comparison group is observably similar to parents at the time point before pregnancy, which may matter if subsequent outcome trajectories differ based on observable characteristics. For instance, if people in very physical job types tend to maintain their physical performance as they age, but those in less-physical jobs tend to decline, then it would be very important to ensure parents are being compared to a counterfactual with the same job type. Our matching approach allows us to account for this.

Rather than list point estimates (e.g., the intercept shift and slope of the postbirth period), we instead estimate the predicted effects at various time points.[14] This can be interpreted as how much more (or less) parents changed from the prebirth period to the given point in time, net of expected secular trends in the outcome, and whether this change statistically differs from zero.

5.2.1 Job Performance

Table 2 contains the estimates for job performance measures. Because women are not subject to physical fitness or marksmanship tests while pregnant or postbirth, there is a delay in when we make the prediction. Columns (1) and (2) in Panel A present results of the impact of parenthood on women's physical fitness performance for the event study and placebo samples, respectively. Beginning at 8 months postbirth, mothers perform 0.41 standard deviations below their expected average under the event study analysis and 0.52 standard deviations below their expected average under the placebo analysis. Mothers recover somewhat and by 12 months postbirth their predicted physical performance scores are 0.18–0.29 standard deviations below expectations. Mothers' physical fitness recovery slows in the child's second year of life. Two years after having a baby, mothers' predicted physical performance remains 0.10–0.17 standard deviations lower than before the pregnancy.[15] The placebo analysis consistently shows larger effects than the event study, and we prefer this estimate because it ensures that mothers are comparable on prebirth characteristics that likely matter for physical performance trajectories (e.g., job type and age).

The impact of childbirth on women's job performance accrues over time. Supervisor ratings do not change during pregnancy. By six months postbirth, mothers score 0.13 standard deviations lower than expected, relative to changes in the placebos. Marksmanship is much lower for mothers than their place-

[14] We include the slope parameters in Appendix Tables A.4 and A.5 for reference.

[15] Physical performance assesses a combination of cardiovascular health, endurance, and upper-body and core strength. Appendix Table A.5 presents raw scores for each item on the fitness assessments that measure overall physical performance. Declines in women's performance occur across almost all measured assessments. When women return to testing, the initial postbirth drop shows women run more slowly (i.e., run times increase) and complete fewer crunches and overhead lifts. Women could do pushups instead of pull-up before 2017, so we interpret the pullup results with caution.

bos postbirth, though by 24 months postbirth the differences from the placebos is no longer statistically significant.

The patterns for fathers' job-relevant physical performance is consistent with mothers' but smaller in magnitude (Panel B of Table 2). We see an immediate drop in performance of 0.08–0.12 standard deviations in the month after birth. By 12 months postbirth, the event study indicates that fathers' scores are slightly above expectations; the placebo analysis shows slight negative effects. Both point estimates are small in magnitude. We again prefer the placebo analysis in case trajectories differ by age, job type, or other important characteristics. For job performance, scores are slightly negative in the first month postbirth, but return to match the placebos by the child's first birthday. For marksmanship, if anything fathers may improve over time.

5.2.2 Training, Education, and Promotion

We next turn to human capital development and promotion. The expected trajectory of these outcomes depend on baseline characteristics, the values are based on growth over time, and the outcomes are observed every month.[16] Thus, we focus on the placebo birth sample. Figure 4 displays the weighted means of training, education, and promotion over time for parents and their placebos.

We begin with training, which we consider a measure of job-specific human capital. The Y-axis is measured as the running count of months observed in training status. Both the mothers and their placebos grow at the same rate before pregnancy, but starting in pregnancy the mothers flatten out while the placebos continue to grow. The mothers never make up the gap, and by 24 months postbirth mothers have about a month less training than the placebos (about a 13% difference).

We consider education a more general, transferable measure of human capital. Here, the Y-axis is the count of degrees (e.g., AA or BA) we observe the individuals gain over time, measured relative to the match time at $t = -10$. Again, a gap emerges between mothers and placebos; this gap does not close 24

[16]For instance, different grade levels have timeframes for promotion, so it is very important that we compare promotion trajectories from people starting at the same point in time. Moreover, we need a reference point in time for outcomes based on growth, and the placebo analysis provides a baseline point of comparison at $t = -10$. Finally, because we so consistently observe these outcomes, the estimates are fairly stable and we are less worried about statistical power, the main benefit of the event study approach.

months postbirth.

The final column displays promotions, which is a count relative to $t = -10$. Mothers and their placebos move together before pregnancy, but a gap emerges around the time of the birth. Mothers never catch up to their placebos; if anything the gap grows over time.

The placebo strategy relies on the placebos providing a good counterfactual to the parents. We require them to match at $t = -10$, so a natural test is whether the parents and placebos move together before the pregnancy. Visually, they appear to in Figure 4. Appendix Table A.3 tests whether the parents and placebos retroactively differ at $t = -12$ and $t = -24$ in the outcomes we observe every month (training, education, and promotion). They do not statistically differ, with the exception of training for fathers, which is 0.005 months higher for fathers-to-be than placebos at $t = -12$ (p-value=0.044).[17]

Table 3 displays the predicted postbirth gaps at various points based on Eq. 3.[18] We observe these outcomes at all times for parents, so we estimate the effect at $t = 1$ for mothers. Confirming the results from Figure 4, the training gap between mothers and placebos is about a half-month immediately after birth and grows to 0.9 months by 24 months postbirth. The educational gap is not statistically significant. The promotion gap is 0.06 at 6 months postbirth but grows to 0.10 promotions by 24 months postbirth. The placebos averaged 1.31 promotions in $t = [-10, 24]$, so this is a 7.4% reduction.

Fathers are indiscernible from their placebos in promotions in Figure 4, but fathers outgain their placebos in training and formal educational attainment postbirth. Table 3 confirms these patterns, and by 24 months the fathers have gained 0.09 months more training and 0.005 more degrees than their placebos. Though statistically significant, these point estimates are quite small. There is no gap in promotion between fathers and their placebos.

[17] The N is smaller for $t = 24$ because we only require the parents to be observed at $t = 12$ and later; parents who have their child early in their careers or at the beginning of our data are thus missing these outcomes for $t = -24$.

[18] The benefit of the parametric model over the raw means in Figure 4 is that it can smooth out any monthly noise and provide an estimate comparable to Table 2.

5.3 Heterogeneity across Subgroups

Parenthood may affect groups differently, and we explore heterogeneity in Figure 5. We conduct this analysis by interacting an indicator variable for a characteristic (e.g., being married) with the semi-dynamic event study variables from Eq. 3; this measures whether the change in the gap between parents and nonparents from prepregnancy to postpregnancy differs across group types.[19] The p-values listed in the text below test whether the gaps that emerge between parents and nonparents the same for these groups.

We begin with marital status. Unmarried parents may have less support at home and may struggle more at work. We generally confirm this hypothesis. Mothers who were unmarried at the time of birth have larger physical fitness declines than married mothers at 8 months postbirth (p=0.002), but the gap closes by 24 months postbirth (p=0.223). Single fathers consistently display lower performance on the physical fitness test than married fathers. By 24 months postbirth, the gap is negative for single fathers but married fathers actually outperform their placebos (p=0.000). The gap is also worse for single fathers at 24 months for training (p=0.000) and education (p=0.000). Overall, we have evidence that the negative impacts of childbirth on performance and human capital outcomes are larger for single mothers and fathers than for their married counterparts. However, this does not translate into worse promotion rates for the single parents.

Next, we explore variation in impacts by several job characteristics. We categorize Marines as working in a "high-physicality" ($> physical$) vs. "low-physicality" ($< physical$) job if their job responsibilities place them in the top half vs. bottom half of the distribution of O*NET's physicality index. We exclude individuals working in jobs that have no link to an O*NET classification. We then compare the magnitude and direction of the impacts between the two groups. We find mothers in high-physicality jobs have larger drops in physical performance at $t = 8$ than those in low-physicality jobs (p=0.010). By 24

[19]Definition of the subgroup is based on the matched parents characteristics for the placebos, rather than the characteristics of the placebos themselves. This ensures the placebos are in the same group as their matched parents, even if they don't exactly match on a characteristic. We use $t = -10$ to define the subgroup except for marital status, which we instead define at $t = 0$ because we are most interested in family characteristics when the baby arrives. This time gives couples surprised by a pregnancy some time to marry if they want to. Analysis excludes job performance and marksmanship due to small sample size; results are available in A.1.

months postbirth, the difference in the magnitudes of the effects among each group remains statistically significant (p=0.046). In the overall sample parenthood did not affect mothers' education, on average, but Figure 5 reveals that mothers in high-physicality jobs attain fewer educational degrees (relative to their placebos) after childbirth, while mothers in low-physicality jobs attain more educational degrees (relative to *their* placebos). This difference in impacts on degree attainment is statistically significant at $t = 1$ and $t = 24$ (p=0.006 and 0.001, respectively). The finding aligns with the notion that mothers with very demanding physical jobs may experience difficulty keeping up their performance after childbirth, perhaps leaving them with less time, ability, or incentive to focus on educational attainment. Fathers in high-physicality jobs also experience a larger negative impact on degree attainment than fathers in low-physicality jobs at $t = 1$ and $t = 24$ (p=0.000 and 0.000, respectively).[20] Overall, we have evidence that it is more difficult for parents, especially mothers, in physically demanding jobs to maintain their performance and invest in education. Yet, despite this, Marines in high- and low-physicality jobs realize similar and statistically indistinguishable promotion outcomes in response to parenthood.

Figure 5 also shows differences in impacts between junior enlisted, who may be less attached to their jobs, and senior enlisted or officers, who may be more attached. We find no consistent differences in physical performance or degree attainment when comparing impacts between junior enlisted (E1–4 as of $t = -10$) and more senior-ranking mothers. Instead, results show senior women have larger declines in time spent in training than junior mothers at $t = 1$ and $t = 24$ (with these differences testing as significant at p=0.001 and 0.010, respectively). Despite the larger impact of parenthood on senior women's time spent in training, it is the junior enlisted who show the largest declines in promotions. The differences in promotion outcomes between junior and senior mothers at $t = 1$ and $t = 24$ are statistically different (p=0.002 and 0.000, respectively).

For fathers, the negative impact of parenthood on physical performance is larger for junior enlisted men than for more senior-ranking men at $t = 1$ (p=0.000); however, differences in the size of impacts between junior- and senior-ranking fathers are only marginally statistically different from one another

[20] Among fathers, those with jobs that are less physically taxing participate in fewer months of job-specific training after childbirth. Childbirth has minimal to no impact on time spent in training for fathers with high-physicality jobs, and the difference in training impacts between Marines with high- vs. low-physicality jobs is statistically significant. We do not observe this same pattern among mothers, however, who across all job types accumulate less training after birth.

by $t = 24$ (p=0.076). More senior fathers actually outperform their placebos and receive more training, degrees, and promotions in response to childbirth, while junior enlisted fathers receive less (p=0.000 for all three outcomes). To some extent, then, fathers "rise to the challenge" of fatherhood by outperforming otherwise comparable non-fathers—but only the more senior-ranking fathers.

We also consider variation in impacts between reserve and active-duty Marines. Reservists provide a useful comparison to active-duty servicemembers, as most reservists have full-time civilian jobs but have the same physical, job performance, and marksmanship requirements as active-duty. If we see larger (or smaller) impacts of birth among reservists, it may give a sense of how civilians differ from those who make their primary living from being in the military. We do not have the power to conduct subgroup analyses for physical performance, job performance, and marksmanship scores for women due to only a small number of female reservists and sporadic nature of the assessments. Reductions in training after birth, however, are concentrated among active-duty not reservist mothers. There are no consistent differences in the size of impacts between reservist and active-duty mothers in terms of education or promotion. For fathers, the drop in physical performance immediately following birth is much larger for reservists than active-duty servicemembers (p=0.007). However, by 24 months postbirth, the size of the impacts does not differ for any outcome between active-duty and reserve fathers. Overall, then, it appears that for men who work only part-time for the military (and likely hold outside civilian jobs), physical fitness scores drop even more than for fathers who are active-duty, full-time Marines. We may expect that physical performance would drop even more in civilian settings in response to parenthood, if individuals' primary jobs are not explicitly evaluated on physical performance. Otherwise, impacts on active-duty and reserve Marines are very similar.

Finally, we assess whether the impact of parenthood is different for those who remain in the Marines for an additional year beyond our sample requirement, as compared to those who leave. A potential concern is that Marines who plan to leave the military soon after we observe them postpartum may not be as invested in their performance. To address this, our last subgroup splits the sample by those who stay in the Marines over 36 months postbirth (75% of parents) and those who leave in months 25–36. For mothers, the drop in physical performance is larger at $t = 8$ for those who get out in $t = [25, 36]$ than for those who remain until $t > 36$ (p=0.009). The differences remains at 24 months postbirth (p=0.025). By 24 months postbirth, there are no differences in the size of the impact of birth on training between mothers in each group. Somewhat surprisingly, there is a negative impact of parenthood on education for mothers who leave in $t = [25, 36]$ but not for stayers (p=0.000). If mothers *plan* to leave their job soon, we might expect them to be more incentivized to earn a degree. However, if mothers leave due to difficulty balancing parenting and work, perhaps they also struggle to invest in their education. Finally, the promotion gap is driven by the mothers about to leave (p=0.043). For fathers, those who leave in $t = [25, 36]$ consistently experience more negative impacts of parenthood than those who stay in; these are statistically significant at p=0.000 for all outcomes by t=24.

This pattern of worse outcomes for parents who leave could be because they knew they wanted to get out, put in less effort at work, and then did not get promoted – or because they struggled to perform after having a child, did not get promoted, and left despite wanting to stay in. Still, even among those who stay in, we generally find the same patterns as the main analysis: worse physical performance, less training, and lower promotion rates for mothers, and a short-term drop in physical performance but otherwise limited (or potentially even positive) effects for fathers.

5.4 Mechanisms for Changes in Promotion Rates

If mothers have promotion gaps relative to comparable women, an immediate question is to what extent changes in job performance and human capital development drive this outcome. We see four primary pathways by which motherhood could translate into promotion gaps. First, drops in performance on the job could explain the slower promotion. Second, changes in human capital development could translate into delayed promotion, especially if mothers need certain training to reach the next promotion level. Third, pregnant and postpartum mothers have the option of waiving physical fitness testing due to safety concerns during pregnancy and the physical toll of childbirth; this is why we excluded $t = [-9, 7]$ in the main analysis. However, promotion boards may view missed assessments negatively; female Marines we talked to believed missing fitness tests could be detrimental at promotion time. For these these pathways, many nonmothers have variation for nonmaternity-related reasons, while there is also variation in the changes among mothers.[21] Finally, a fourth pathway is that there is simply discrimination against mothers, and none of the observed characteristics fully explain the promotion gaps.[22]

We do a simplified mediation test of these pathways in Table 4. The outcome is the count of promotions accumulated from $t = [-10, 24]$. We include mothers, fathers, and their respective placebos in the same regression, with each person-match entered once. All models control for baseline characteristics and include an indicator for being female, an indicator being a parent, and an interaction between these two.[23] The coefficient on this interaction would be negative if mothers have low promotion levels, above and beyond being a parent in general, being a female in general, and the other variables included in the model. Column 1 provides a baseline. Fathers in general have 0.016 more promotions than would be predicted by their baseline characteristics; mothers' promotions are much lower than comparable females.

Next, we assess whether various interim outcomes explain this gap. Column 2 adds a variable for the mean fitness scores for $t > 0$. Beyond the baseline characteristics, every additional standard deviation higher fitness level is associated with more promotions, as expected. The coefficient on the mothers' interaction shrinks by 25% (see Sobel-Goodman mediation), which implies that changes in physical performance explain roughly a quarter of the motherhood promotion gap.[24] Job performance, marksmanship, training, and education (Columns 3–6) do not explain the promotion gap. Entering all of these variables simultaneously explains about a quarter of the motherhood promotion penalty (Column 7).

We next assess whether missing physical fitness assessments changes promotion rates. We operationalize this measure by counting how many of the 6-month fitness cycles that ended in $t > -10$ were missing fitness scores for each individual. We enter this as a total count per person-match, which averaged 2.6 waived tests for mothers, but had a range of waived tests for non-mother placebos, fathers, and

[21] For instance, some mothers still opt into taking the fitness test early, and some nonmothers get waivers for medical reasons such as spraining an ankle, so we have variation in the number of fitness tests waived.

[22] We note, however, that even if we do fully explain the gap, discrimination can come into play. Supervisor's job performance ratings may be biased, and assessments like physical fitness requirements may be unfair to women if they, say, test abdominal strength soon after childbirth, even if every Marine has to do crunches.

[23] Parents have a weight of 1; placebos have the weight from the main analysis. Controls are taken at $t = -10$ for months of service, an indicator for officer, an interaction between months of service and officer, and variables from the matching model.

[24] The Sobel-Goodman mediation test examines whether a mediator (e.g., fitness scores) carries the influence of some explanatory variable (e.g., motherhood) to the dependent variable (e.g., promotion). The test specifically computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls.

non-father placebos as well. In Table 4, having more waived fitness tests is associated with lower rates of promotion. Moreover, this explains 96% of the motherhood gap (Column 8). The patterns are very similar when all potential mechanisms are included (Column 9). Overall, missing evaluations seems to be particularly important for explaining why mothers are not getting promoted.[25]

5.5 Variation by Maternity Leave Length

Prior to 2015, all DoD branches provided active-duty women with 6 weeks of paid maternity leave. In July 2015, the Secretary of the Navy announced that primary caregivers in the Navy and Marine Corps would be entitled to 18 weeks of leave. Women who had given birth earlier in the year (as of January 2015) could retroactively take advantage of the 18-week leave policy before their child's first birthday. Women who had already returned to work following their initial 6 weeks of leave tended to use the additional 12 weeks of paid time off discontinuously. Women who were on leave at the time of the announcement of expanded leave, or gave birth after the announcement, generally took the additional leave consecutively (Bacolod et al., 2021). We analyze these "6 weeks + 12 flex" and "18 weeks" groups separately. In early 2016, the Secretary of Defense standardized maternity leave to 12 weeks for all services. The 12-week policy applied to pregnancies that began 31 days after the announcement (i.e., pregnancies that began on March 3, 2016 or later, per doctor estimation).

We disaggregate impacts according to the length of maternity leave in place when a woman gave birth. The key question of interest is whether longer or shorter leave predicts better or worse outcomes when women return to work. Variation in leave length is, at times, quasi-randomly assigned.[26] We

[25] Appendix Table A.7 explores several additional pathways: being deployed, months of being in a nondeployable status, and the average of whether a Marine achieved a "first-class" fitness score rather than the continuous measure to measure fitness. Actually deploying does not have a large mediation effect. Being in a limited duty status (which is strongly associated with waiving fitness tests) for medical or other reasons accounts for over half of the motherhood promotion penalty. Measuring fitness as attaining a first-class score does a slightly better job of mediating the motherhood penalty than do the standardized scores, possibly because promotion boards examine fitness in this dichotomized way.

[26] Some policy changes were unexpected and applied to women who were already pregnant, while other changes were prospective, allowing women to potentially select in. Selection presents the biggest concern for women who gave birth under the 12-week policy, given that the reduction in leave length was announced prospectively. Appendix Table A.8 presents descriptive characteristics of the women who gave birth under the four leave policy groups we analyze. We find minimal evidence of selection into leave length based on observable characteristics. Exceptions include significant differences in age, percent who identify as Hispanic, percent married, percent with a college degree, and months of training. When we estimate variation across policy groups, using the matched sample of parents and nonparents assigned to placebo births, our matching

conduct the analysis by defining indicator variables for the "6 weeks + 12 flex," "18 weeks," and "12 weeks" periods, with the "6 week" policy as the baseline group. We interact these indicators with the variables in Eq. 3 and make policy-specific predictions for the initial birth drop (i.e., 8 months for physical performance and 1 month for training, education, and promotion), 12 months postbirth, and 24 months postbirth.[27] We use F-tests to assess whether the parent-placebo gap for each timeframe statistically differs from the prebirth gap for each policy. We also assess whether the four policy gaps statistically differ from each other at each timeframe. Because babies born in November–December 2016 could have fallen under either the 18 or 12 week policy depending on date of conception, we exclude these mothers (and their matches) from the policy analysis. Table 5 replicates the primary analysis for the subsample excluding November–December 2016 births in the top section of each panel, and then shows the results from the regression with policy interactions.

Among mothers across all maternity leave policies, physical performance drops when initially observed 8 months after having a baby. However, women who had longer maternity leave had larger physical performance declines, particularly for the "6 weeks + 12 flex" policy. These mothers had returned to work following their initial 6 weeks of leave, then received an additional 12 weeks of leave they had to use by their child's first birthday. An F-test indicates that the magnitude of the drops across leave policies statistically differ from each other (p=0.039). By 12 and 24 months postbirth mothers under different policies are statistically the same. We find no policy differences for training or education.

At one year postbirth we find the largest promotion gaps for mothers who had very flexible leave. The estimated promotion gaps differ across the four policy periods (p=0.002), and this pattern is consistent with mothers who take leave flexibly throughout the year having worse promotion trajectories. By 24 months postbirth the gap is still the largest for the 6 weeks + 12 flex mothers, but the gaps no longer statistically differ across policies (p=0.239).

strategy implicitly controls for these underlying differences. We also test whether the density of first births is continuous across each policy threshold, providing additional evidence on possible selection. Figure A.2 shows month-by-month variation in the density of births, including a test for any discontinuity across policy thresholds, following Cattaneo et al. (2018). None of the differences across the policy thresholds reach statistical significance, suggesting the policies did not influence female fertility itself.

[27] We do not include supervisor-rated job performance or marksmanship here. Supervisor ratings for young enlisted Marines are only available for those who remained in service as of October 2017, which complicates analyses of policy changes that took place in 2015 and 2016. We also generally lack power to subdivide these outcomes given sparse observations.

6 Summary and Conclusions

We use repeated, direct measures of work performance, human capital accumulation, and career advancement to explore the link between the transition to parenthood and workers' outcomes. Our empirical strategy draws on an event study approach based on the precise month of birth and a matching design that assigns placebo births to observably similar nonparents. We find both men and women's health-related job performance responds negatively to the transition to parenthood. However, gender differences emerge in the magnitude and persistence of decline and recovery postbirth. Women experience large declines in job-related physical performance that remain for at least 2 years, while men experience short-lived declines in physical performance that fade by their child's second birthday. Women's supervisor-rated job performance and marksmanship scores also decline in the years after having a child, while men's do not. Mothers' accumulated training slows, while fathers' educational attainment actually increases slightly.

These patterns are consistent with our findings that women's promotion trajectories slow after having a child while men's do not. Among women, promotion delays accumulate over time after childbirth; the gap in number of promotions between mothers and nonmothers is largest 24 months postbirth. Promotion trajectories between fathers and nonfathers look almost identical for the 2 years after the birth or placebo birth event. Promotion gaps between mothers and fathers directly lead to a pay gap in this setting. By 24 months postbirth, the average mother would make $40,596 in basic pay according the Marine Corps pay schedule (excluding any bonuses or housing allowances).[28] The impact of birth on promotions means that mothers go from an average of $0 difference in pay compared to matched nonmothers 10 months before birth to $332 lower pay at 24 months postbirth. Fathers are generally more advanced in their careers when they have a child, so the mother-father wage differential grows from $5,789 at 10 months before birth to $5,890 at 24 months postbirth (a $101 increase in the gap). This highlights the importance of using similar "placebos" for the comparison: we would systematically undercount the mother's wage

[28] We use 2022 basic pay scales for this estimate. Basic pay is calculated by years of service and rank. Marines also have housings allowances that increase with number of dependents (thus counteracting the mothers' gap) and also increase with rank (thus increasing the gap if promotions are delayed), as well as other bonuses or incentives that may differ for mothers (e.g., combat pay). We focus on basic pay because it aligns most closely to civilian pay and is straightforward to calculate.

penalty using fathers as a comparison because mothers are earlier in their careers (when promotions *should* happen faster) than fathers. We find some evidence that the promotion gap is driven by mother's tendency to miss key evaluations used in determining promotions, as well as lower physical performance.

Last, we show that additional leave does not reduce the magnitude of the negative impact of birth on mothers' outcomes, but it largely does not increase them, either. The exception is when mothers are provided longer and particularly flexible leave, which is associated with a larger drop in their performance on job-relevant physical fitness assessments in the short-term (less than on year after the birth). These findings suggest longer periods away from work may erode job-specific skills in the short-term, and perhaps lower promotion rates in the medium-term. To the extent that the goal of maternity leave is providing bonding time with children and time for mothers to physically, medically, and mentally recover, this may be good news: the motherhood penalty was not exacerbated in this context by more generous parental leave policies, especially in the longer-term, two years after the birth.

Our findings provide a new angle on the longstanding literature that shows parenthood reduces mothers' employment, hours worked, and wages, while having no effect on fathers. We find having a child impacts mothers' job performance in the first 2 years of the child's life, highlighting the period after birth as a possible critical window that could give rise to long-term child penalties. Delays in promotion that accumulate for women, but not for men, in the years following birth also underscore the need for increased policy- and firm-level support for recent parents. Future research could explore whether alternate family support policies, such as increased access to affordable child care, further mitigate challenges parents face as they transition back into work after having a child.

References

Aguilar-Gomez, S., Arceo-Gomez, E., and De la Cruz Toledo, E. (2019). Inside the black box of child penalties. *Available at SSRN 3497089*.

Agüero, J. M. and Marks, M. S. (2011). Motherhood and Female Labor Supply in the Developing World: Evidence from Infertility Shocks. *The Journal of Human Resources*, 46(4):800–826.

Andresen, M. E. and Nix, E. (2020). What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms. page 69.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the Gender Gap in Pay. *Journal of Labor Economics*, 34(3):545–579.

Angrist, J. D. and Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, 88(3):450–477.

Azmat, G. and Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, 125(5):1306–1355.

Bacolod, M., Heissel, J. A., Laurita, L., Molloy, M., and Sullivan, R. (2021). Mothers in the military: The effect of maternity policy on leave up-take. In press.

Bailey, M., Byker, T., Patel, E., and Ramnath, S. (2019). The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data. Technical Report w26416, National Bureau of Economic Research, Cambridge, MA.

Baker, A., Larcker, D. F., and Wang, C. C. (2021). How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*.

Balser, C., Hall, C., and Bukowinski, A. (2020). The Effect of Expanding Paid Maternity Leave on Maternal Health: Evidence From the United States Air Force & Army. SSRN Scholarly Paper ID 3601018, Social Science Research Network, Rochester, NY.

Barth, E., Kerr, S. P., and Olivetti, C. (2017). The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data. Technical Report w23381, National Bureau of Economic Research, Cambridge, MA.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Bronars, S. G. and Grogger, J. (1994). The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. *The American Economic Review*, 84(5):1141–1156.

Bullinger, L. R. (2019). The effect of paid family leave on infant and parental health in the united states. *Journal of health economics*, 66:101–116.

Bureau of Labor Statistics (2017). Physical strength required for jobs in different occupations in 2016. Bureau of Labor Statistics (BLS) and United States Department of Labor.

Butikofer, A., Riise, J., and Skira, M. (2018). The impact of paid maternity leave on maternal health. *NHH Dept. of Economics Discussion Paper*, (04).

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Cools, S., Markussen, S., and Strøm, M. (2017). Children and Careers: How Family Size Affects Parents' Labor Market Outcomes in the Long Run. *Demography; Silver Spring*, 54(5):1773–1793.

Cruces, G. and Galiani, S. (2007). Fertility and female labor supply in Latin America: New causal evidence. *Labour Economics*, 14(3):565–573.

Cunha, J. M., Shen, Y.-C., and Burke, Z. R. (2018). Contrasting the impacts of combat and humanitarian assistance/disaster relief missions on the mental health of military service members. *Defence and Peace Economics*, 29(1):62–77.

Cáceres-Delpiano, J. (2006). The Impacts of Family Size on Investment in Child Quality. *The Journal of Human Resources*, 41(4):738–754.

Department of Defense (2018). 2018 demographics report: Profile of the military community". Technical report, Department of Defense (DoD), Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy (ODASD (MCFP)), under contract with ICF (https://www.icf.com/work/human-capital.

Gallen, Y. (2018). Motherhood and the gender productivity gap. *Becker Friedman Institute for Research in Economics Working Paper*, (2018-41).

Jacobsen, J. P., Pearce, J. W., and Rosenbloom, J. L. (1999). The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment. *The Journal of Human Resources*, 34(3):449–474.

Kim, S. and Moser, P. (2020). Women in science: Lessons from the baby boom. National Bureau of Economic Research Summer Institute (NBER SI) Session on Gender in the Economy.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2020). Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. Technical report, National Bureau of Economic Research.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child Penalties Across Countries: Evidence and Explanations. NBER Working Paper w25524, National Bureau of Economic Research, Cambridge, MA.

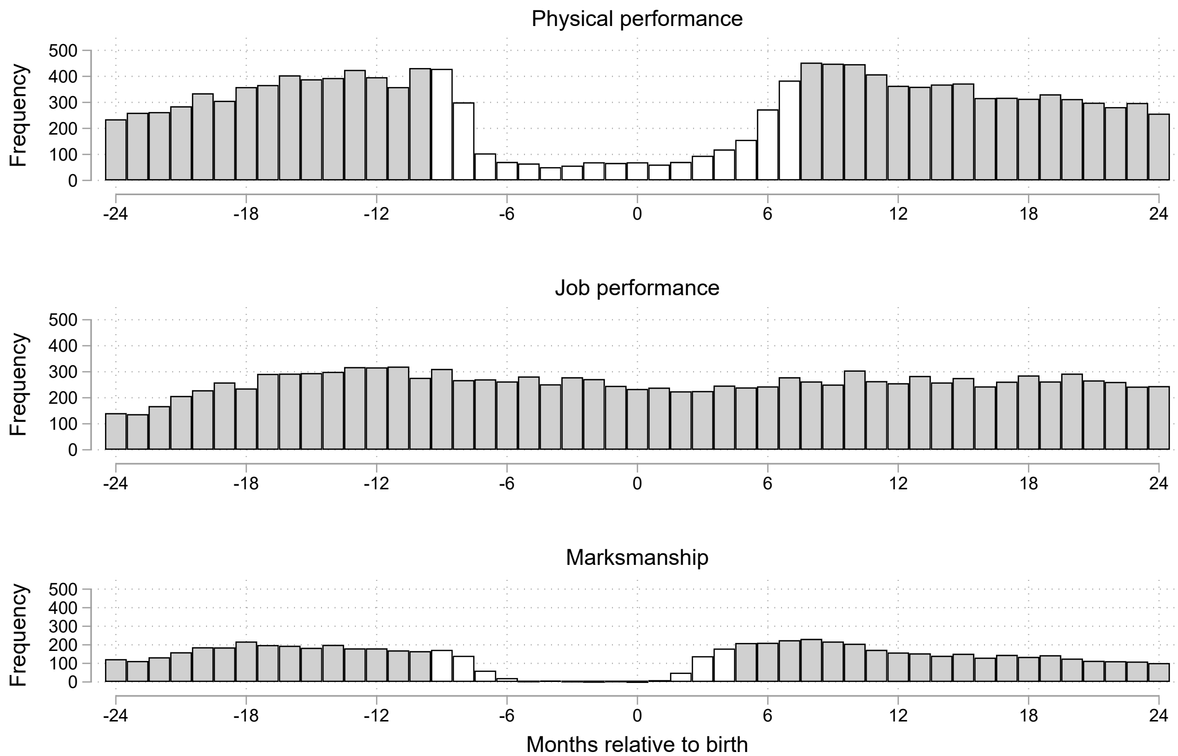
Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Pac, J. E., Bartel, A. P., Ruhm, C. J., and Waldfogel, J. (2019). Paid family leave and breastfeeding: Evidence from california. Technical report, National Bureau of Economic Research. NBER Working Paper.

Trajkovski, S. et al. (2019). California paid family leave and parental time use. Technical report, Center for Policy Research, Maxwell School, Syracuse University.

7 Figures and Tables

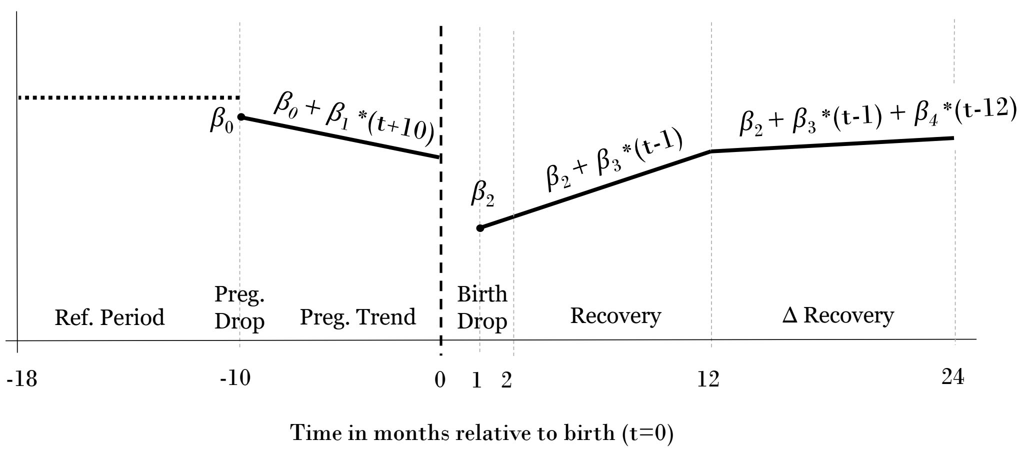
Figure 1: Distribution of Mothers' Outcomes Relative to Birth



Notes: Displays the count of physical performance, job performance, and marksmanship scores for mothers by month relative to birth. Physical performance tests are the most common test among these outcomes; all ranks are expected to take them twice a year (the Physical Fitness Test in January–June and the Combat Fitness Test in July–December). Main analysis excludes physical performance scores at $t = [-9, 7]$ (because mothers did not have to take the tests in pregnancy through 6 months postpartum, and commanders may give them some leeway in month 7) and marksmanship scores at $t = [-9, 4]$ (because mothers did not have to take the tests during pregnancy or while on leave). Semi-dynamic specifications always exclude $t = 0$ due to ambiguity about outcome timing relative to birth. Excluded outcomes are in gray.

Figure 2: Stylized Representation of the Semi-Dynamic Specification, Equation 2:

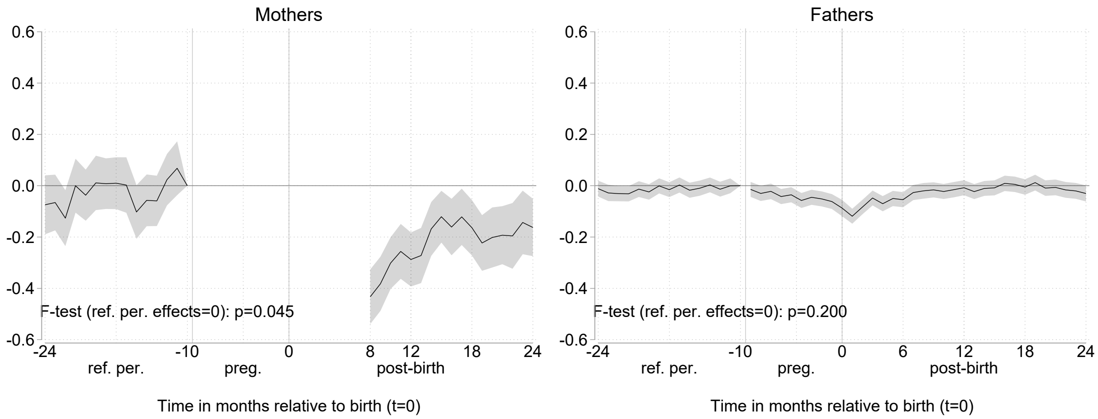
$$\begin{aligned} Y_{it} = &\alpha_i + \phi_t + \beta_0 PregnancyDrop_{it} + \beta_1 PregnancyTrend_{it} + \beta_2 BirthDrop_{it} \\ &+ \beta_3 Recovery_{it} + \beta_4 \Delta Recovery_{it} + \varepsilon_{it} \end{aligned}$$



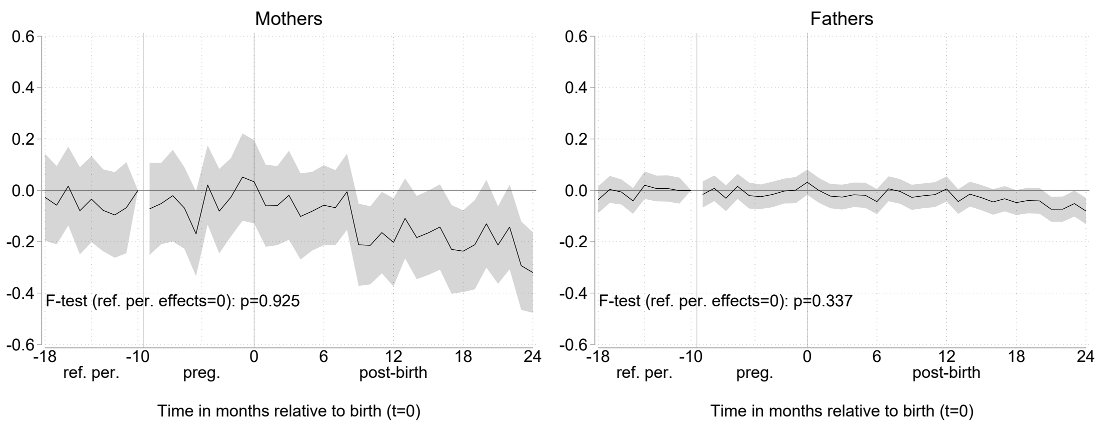
Notes: Figure displays a diagram of parameters defined in Equation 2 where the postbirth drop (β_2) is estimated in the first observed month following pregnancy. For women, we begin measuring the postbirth drop (β_2) for physical fitness performance at 8 months and marksmanship scores at 5 months after the birth. We cannot estimate β_0 or β_1, the pregnancy drop and trend, for women's physical fitness outcomes or marksmanship scores because women are not eligible to be assessed when pregnant.

Figure 3: Event Study Estimates of the Impact of Birth on Job Outcomes

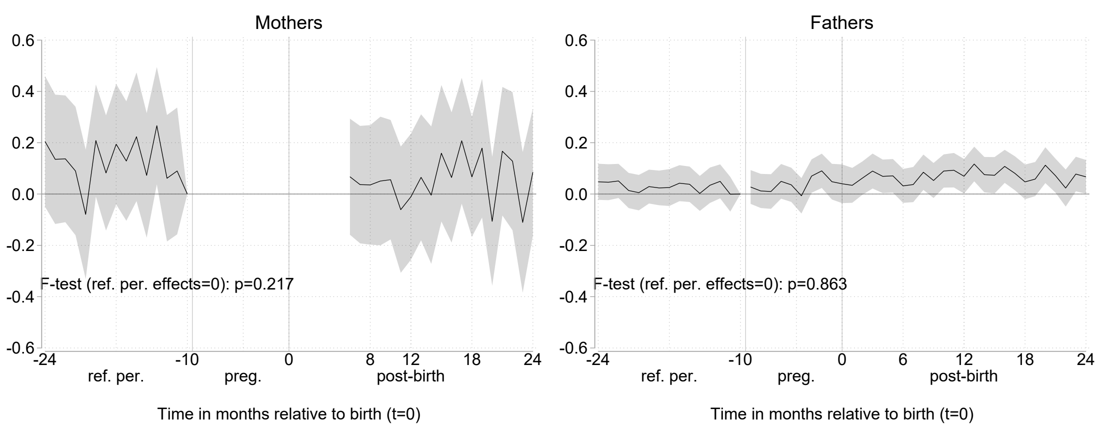
(a) Physical Performance (sd)



(b) Job Performance (sd)

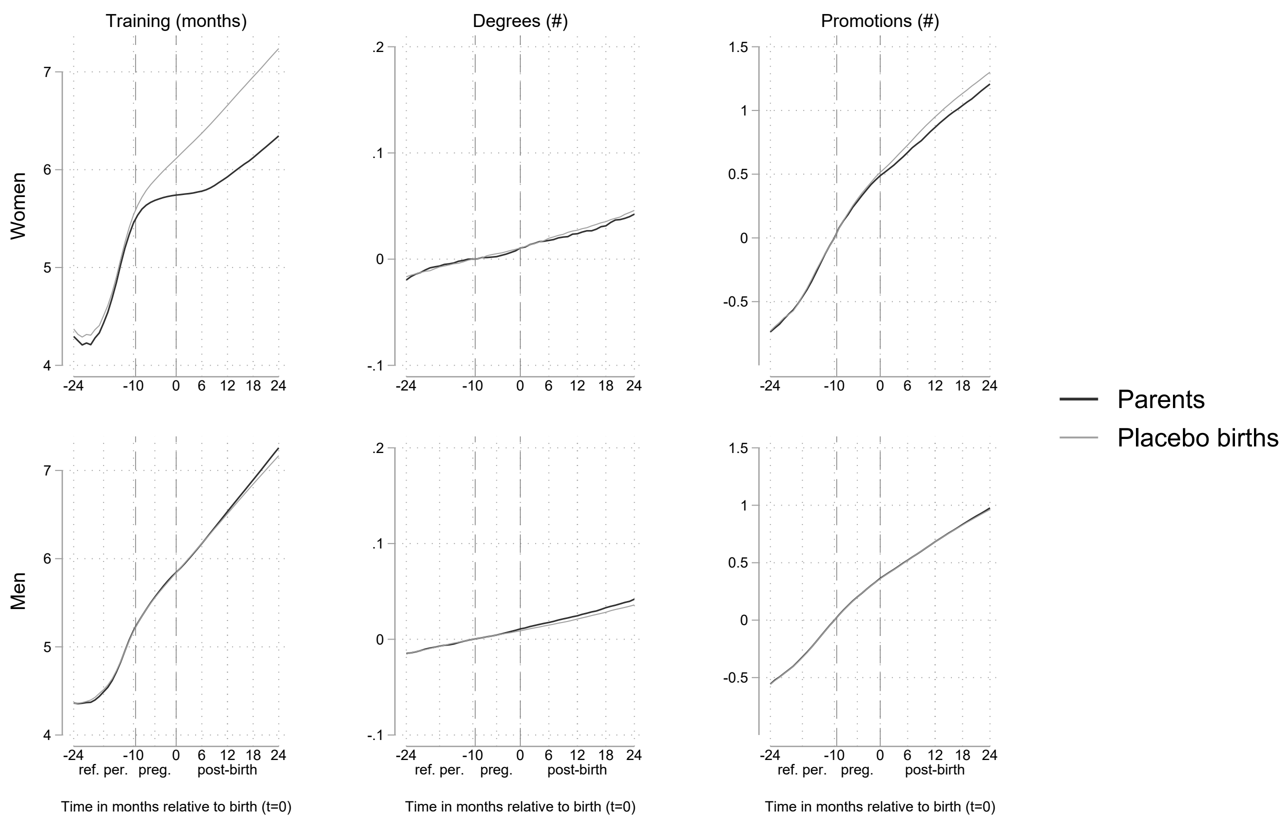


(c) Marksmanship (sd)



Notes: Displays coefficients from event study regressions. Outcomes include standardized scores from (1) physical/combat fitness tests, (2) job performance evaluations, and (3) rifle/pistol marksmanship evaluations, measured separately for males and females. The sample includes first-time parents observed at least 12 months before and 24 months after birth, as well as same-gender Marines with no birth during the study window and at least 36 months of service. Regressions include individual and month-year fixed effects. The reference month is $t = -10$, the month before the start of pregnancy. Vertical lines reflect the start of the pregnancy ($t = -9.5$) and the birth ($t = 0$). Standard errors are clustered by individual and included as shaded areas representing a 95% confidence interval.

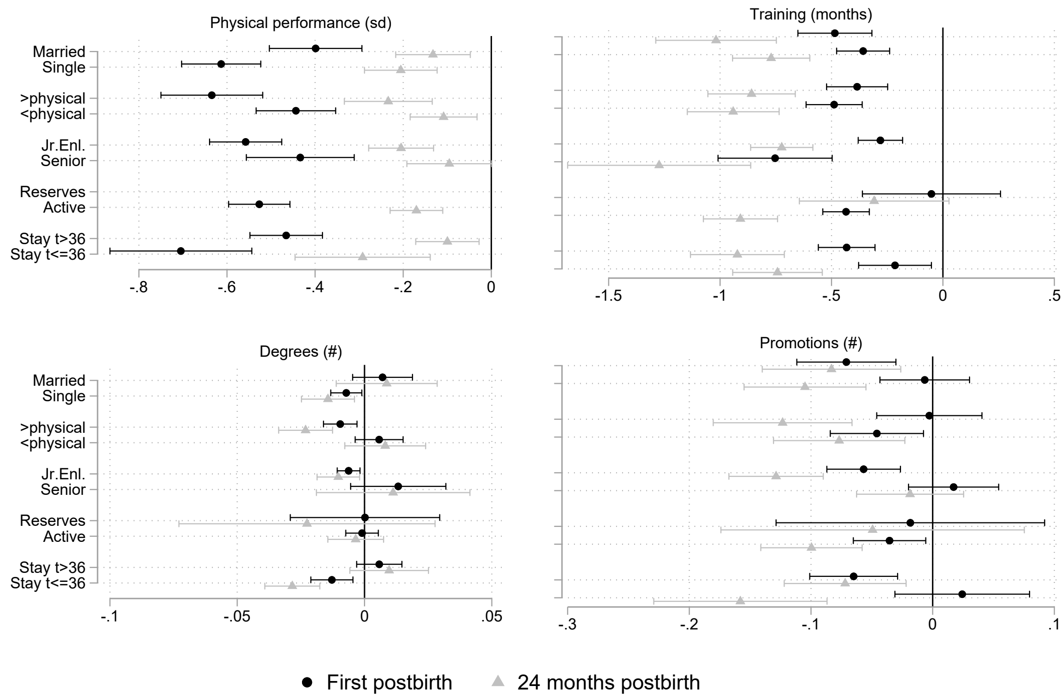
Figure 4: Placebo Birth Estimates of the Impact of Birth on Human Capital and Promotion



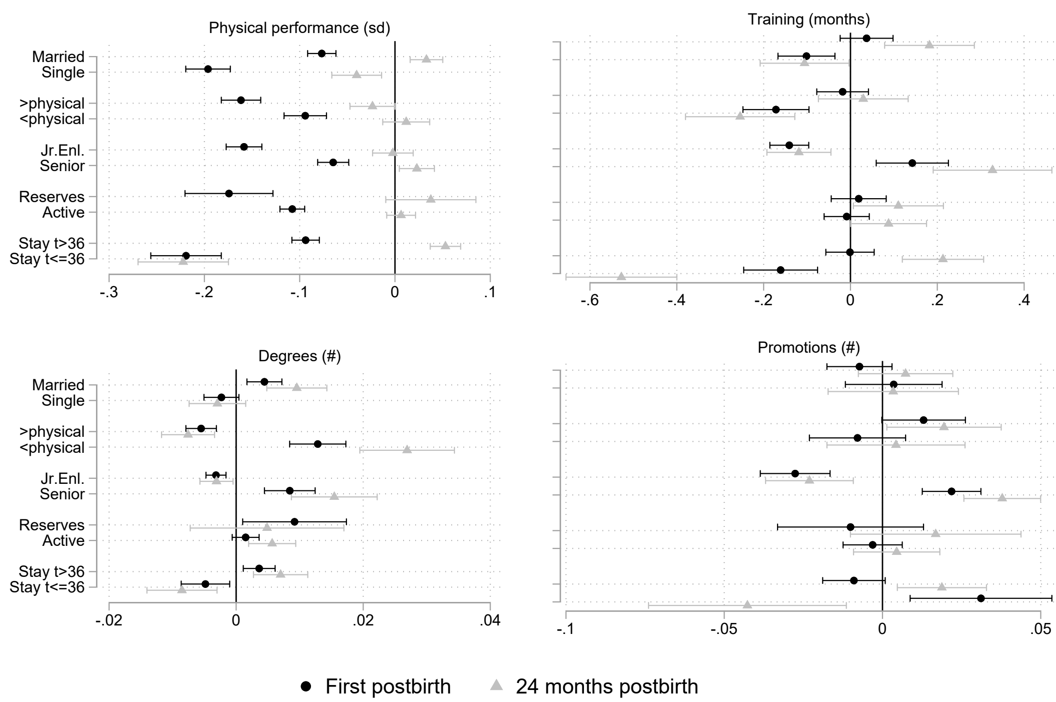
Notes: Displays weighted mean levels of cumulative months of training, cumulative count of educational degrees (relative to $t = -10$), and cumulative count of promotions (relative to $t = -10$) between first-time parents (solid line) and placebo parents (dashed line) over time. Nonparents assigned to placebo births are limited to those whose rank, number of months in service, reserve status, and year is an exact match with parents' 10 months before the birth. Among those with an exact match, each parent's outcomes are compared to the five nonparents most similar to parents in their propensity to have a child based on age, race/ethnicity, military entrance exam scores (AFQT scores), marital status (including whether a spouse is also in the military), level of education, months of training, occupational field, reserve status, and most recent physical performance scores.

Figure 5: Event Study Estimates of the Impact of Birth on Job Outcomes

(a) Mothers



(b) Fathers



Notes: Displays gaps in the physical performance, months of training, years of education, and number of promotions relative to prepregnancy between first-time parents and placebo parents across birth events for the first postbirth observation (black line) and 24 month postbirth (light gray line) by subgroups in the placebo sample. Each comparison (e.g., reserve vs. active) is based on one regression by interacting an indicator variable with a group indicator (e.g., reserve) with the paramaeters in Eq. 2. Classifications for parents are as follows: "Married" are married at $t = 0$; "Single" are not. "$>$physical" are those whose military job type above the median physicality level in our sampled based on O*NET classification; "$<$physical" are at or below the median, among those whose jobs are classified by O*NET. "Jr. Enl" are in enlisted grade E1–E4 at $t = -10$; "Senior" are E5 and up or officers. "Reserves" are not on active duty and likely working a civilian job; "Active" work their military job full-time. "Stay t$>$36" stay in the military at least 3 years after the birth event; "Stay t$<=$36" leave between $t = [24, 36]$. Vertical solid lines reflect a zero effect. Horizontal lies indicate 95% confidence intervals.

Table 1: Characteristics of First-Time Parents

	Mothers		Fathers	
	Marines	Civilians	Marines	Civilians
Age (mean)	23.4	29.9	25.5	31.6
Education				
Some College/Associates	5%	28%	5%	27%
Bachelor's Degree	10%	34%	16%	30%
Marital Status				
Married	68%	78%	87%	83%
Race/Ethnicity				
Black (Non-Hispanic)	15%	6%	10%	5%
Hispanic	23%	11%	14%	13%
Job Classification				
Mngmt./business/science/arts	14%	58%	10%	45%
Service	7%	15%	4%	11%
Sales/office	35%	24%	12%	15%
Construction/maint.	18%	0%	29%	15%
Production/moving/transpo.	19%	3%	14%	14%
Military Specific Chars.				
Officer	8%	–	14%	–
AFQT score (percentile)	58.5	–	63.3	–
GCT score (av=100; sd=20)	103.4	–	111.3	–
N of individuals	2,549	49,013	24,420	59,423

Notes: Displays characteristics of first-time parents in the Marine Corps in our sample alongside characteristics of first-time civilian parents in the labor market. Time-varying characteristics of Marines in our sample (e.g., age) are measured 10 months before the birth (t=-10). Data on civilians come from the American Community Survey 1-year estimates, 2010 to 2018. We limit the civilian sample to adults who are employed in the civilian labor market and have a first child under age 1. Job categories correspond to Standard Occupational Classification (SOC) system groups applied to U.S. Marine Corps job codes and available in the American Community Service. Military specific variables include whether a Marine is ranked as an officer (akin to manager) and AFQT and GCT scores, which are measures of intelligence. We do not observe these military-specific variables in the civilian sample.

Table 2: Impacts of Childbirth on Job Performance Among First-Time Parents

	Physical Performance (sd)		Job Performance (sd)		Marksmanship (sd)	
	Event Study (1)	Placebo (2)	Event Study (3)	Placebo (4)	Event Study (5)	Placebo (6)
A. Mothers						
6-month effect	–	–	0.000	-0.130***	-0.058	-0.138*
			[1.000]	[0.000]	[0.212]	[0.028]
8-month effect	-0.412***	-0.524***	-0.021	-0.145***	-0.053	-0.134**
	[0.000]	[0.000]	[0.476]	[0.000]	[0.120]	[0.002]
12-month effect	-0.182***	-0.289***	-0.062	-0.176***	-0.042	-0.127*
	[0.000]	[0.000]	[0.087]	[0.000]	[0.346]	[0.011]
24-month effect	-0.109***	-0.174***	-0.129**	-0.102*	-0.020	-0.074
	[0.000]	[0.000]	[0.002]	[0.034]	[0.704]	[0.201]
Parent mean ($t \leqslant -10$)	0.11		-0.01		-0.07	
Nonparent mean ($t = 24$)		0.14		-0.01		0.18
N of individuals	22,094	8,912	10,601	4,987	16,521	6,639
Observations	167,400	65,345	91,066	46,127	64,663	29,246
R^2	0.58	0.61	0.41	0.48	0.45	0.51
B. Fathers						
1-month effect	-0.077***	-0.115***	0.036**	-0.035**	0.040**	0.019
	[0.000]	[0.000]	[0.001]	[0.005]	[0.003]	[0.213]
12-month effect	0.018**	-0.026***	0.063***	0.001	0.083***	0.041**
	[0.001]	[0.000]	[0.000]	[0.958]	[0.000]	[0.002]
24-month effect	0.004	0.010	0.022	0.001	0.072***	0.045**
	[0.586]	[0.185]	[0.087]	[0.926]	[0.000]	[0.005]
Parent mean ($t \leqslant -10$)	0.25		0.01		0.02	
Nonparent mean ($t = 24$)		0.06		-0.02		0.20
N of individuals	321,760	81,667	136,108	45,774	215,302	60,443
Observations	2,515,047	1,018,968	1,198,963	451,124	817,386	374,101
R^2	0.59	0.61	0.41	0.50	0.48	0.49

Notes: Displays predicted values from Eq. 2, the semi-dynamic event study specification at listed time points postbirth in the odd columns using the full sample, and Eq. 3, the dynamic placebo specification in the even columns using the placebo sample. Outcomes include (1) standardized (mean=0, SD=1) scores from physical/combat fitness tests conducted 2x per year, (2) standardized scores (mean=0, SD=1) from supervisor-rated job performance evaluations conducted 1-2x per year, and (3) standardized scores (mean=0, SD=1) from rifle or pistol tests conducted 1 or fewer times per year. We exclude women's physical performance scores 9 months before through 7 months after birth because women are not required to take fitness tests during and after pregnancy. We exclude women's marksmanship scores 9 months before through 4 months after birth because women are not required to take marksmanship exams during pregnancy or while on leave. All outcomes for women and men exclude $t = 0$. Regressions include individual and month-by-year fixed effects. Predicted p-value of whether the value statistically differs from zero are shown in brackets, based on heteroscedasticity-robust F-test and standard errors clustered by individual. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table 3: Impacts of Childbirth on Human Capital & Career Advancement Among First-Time Parents

	Training (months)	Degrees (#)	Promotions (#)
	(1)	(2)	(3)
	A. Mothers		
1-month effect	-0.463***	-0.001	-0.035*
	[0.000]	[0.782]	[0.019]
6-month effect	-0.579***	-0.003	-0.058***
	[0.000]	[0.411]	[0.000]
12-month effect	-0.717***	-0.005	-0.086***
	[0.000]	[0.228]	[0.000]
24-month effect	-0.880***	-0.004	-0.097***
	[0.000]	[0.414]	[0.000]
Nonparent mean ($t = 24$)	7.22	0.05	1.31
N of individuals	9,066	9,066	9,066
Observations	615,797	706,319	691,232
R^2	0.85	0.38	0.68
	B. Fathers		
1-month effect	-0.006	0.002*	-0.004
	[0.820]	[0.029]	[0.383]
12-month effect	0.025	0.004*	-0.008
	[0.450]	[0.011]	[0.141]
24-month effect	0.091*	0.005**	0.006
	[0.025]	[0.002]	[0.355]
Nonparent mean ($t = 24$)	7.17	0.04	0.96
N of individuals	81,745	81,745	81,745
Observations	6,800,867	6,947,084	6,800,867
R^2	0.88	0.35	0.62

Notes: Displays predicted values from Eq. 3 using the placebo sample. Outcomes include monthly observations of (1) cumulative months of training, (2) cumulative degree counts relative to $t = -10$, and (3) cumulative promotion counts relative to $t = -10$. All outcomes exclude $t = 0$. Regressions include individual and month-by-year fixed effects. Predicted p-value of whether the value statistically differs from zero are shown in brackets, based on heteroscedasticity-robust F-test and standard errors clustered by individual. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

Table 4: Identifying Potential Mechanisms for Changes in Promotion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female X Parent	-0.103***	-0.077***	-0.098***	-0.104***	-0.101***	-0.102***	-0.077***	-0.004	-0.007
	(0.016)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)
Parent	0.019***	0.022***	0.020***	0.019***	0.019***	0.019***	0.022***	0.014**	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Female	0.112***	0.110***	0.101***	0.112***	0.112***	0.110***	0.102***	0.121***	0.110***
	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.008)
Avg post physical		0.081***					0.070***		0.063***
		(0.003)					(0.003)		(0.003)
Avg post job			0.125***				0.115***		0.113***
			(0.004)				(0.004)		(0.004)
Avg post mark				-0.010**			-0.020***		-0.021***
				(0.003)			(0.003)		(0.003)
Training (months)					0.002***		0.002***		0.002***
					(0.000)		(0.000)		(0.000)
Education (years)						0.030***	0.017***		0.017***
						(0.004)	(0.004)		(0.004)
# waived fit tests								-0.052***	-0.038***
								(0.002)	(0.002)
Prepregnancy controls	X	X	X	X	X	X	X	X	X
p(fem*par+par)=0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.509	0.443
Sobel-Goodman mediation		0.25	0.04	-0.01	0.02	0.01	0.25	0.96	0.94
Observations	161304	161304	161304	161304	161304	161304	161304	161304	161304
R-squared	0.2969	0.3040	0.3071	0.2970	0.2973	0.2975	0.3132	0.3009	0.3152
Adjusted R-squared	0.2969	0.3039	0.3070	0.2969	0.2972	0.2974	0.3130	0.3008	0.3151

Notes: Predicts count of promotions from $t = [-10, 24]$ in a combined women/men placebo sample, using one "wide" observation per match-ID. "Female X Parent" is a mother-specific indicator variable. "Parent" is an indicator for a parent in general. "Prepregnancy controls" includes controls for variables included in he LASSO model (see Appendix Table A.2) plus months of service, an indicator for officer, and an interaction between months of service and officer, and year at $t = -10$. The average post physical, job, and rifle scores are mean observed scores from $t > 0$; we do not drop any observed scores so observations for mothers who choose to take a fitness test soon after birth would be included. Training and education are the total months of training and the total years of education achieved, respectively, by $t = 24$. Number of missed tests is a count of the physical and combat fitness testing periods without an observed fitness test for $t > -10$; this is based on the USMC fitness test schedule rather than relative time to/from birth. We cluster standard errors by ID because some nonparent IDs are matched multiple times to more than one parent. Includes the p-value of an F-test of whether the parent and mother-specific coefficient sum to zero. Sobel-Goodman mediation test computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

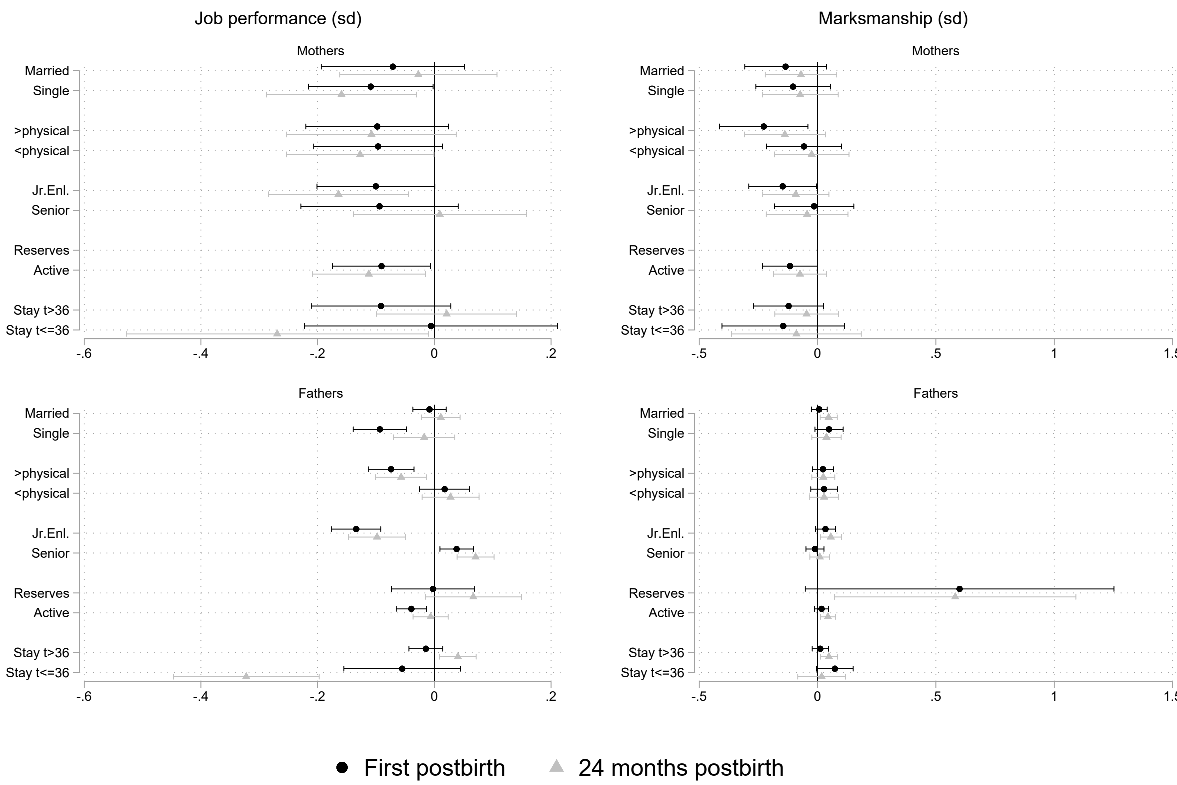
Table 5: Women's Outcomes by Leave Length

	Post-birth drop		12 months post		24 months post		
	Effect size	*p*	Effect size	*p*	Effect size	*p*	N
			A. Physical performance (sd)				
Main effect:	-0.517***	0.000	-0.283***	0.000	-0.171***	0.000	63,773
Effects by paid leave length:							
6 weeks	-0.466***	0.000	-0.283***	0.000	-0.131**	0.002	63,773
6 weeks + 12 flex	-0.808***	0.000	-0.136	0.196	-0.353**	0.009	
18 weeks	-0.629***	0.000	-0.309***	0.000	-0.156*	0.011	
12 weeks	-0.444***	0.000	-0.310***	0.000	-0.278***	0.000	
p(diff), all effects	0.039		0.491		0.214		
			B. Training (months)				
Main effect:	-0.412***	0.000	-0.720***	0.000	-0.870***	0.000	675,371
Effects by paid leave length:							
6 weeks	-0.370***	0.000	-0.683***	0.000	-0.809***	0.000	675,371
6 weeks + 12 flex	-0.239	0.099	-0.487*	0.025	-0.698*	0.018	
18 weeks	-0.474***	0.000	-0.850***	0.000	-1.130***	0.000	
12 weeks	-0.534***	0.000	-0.744***	0.000	-0.811***	0.000	
p(diff), all effects	0.090		0.384		0.141		
			C. Degrees (#)				
Main effect:	-0.000	0.934	-0.004	0.327	-0.004	0.513	675,371
Effects by paid leave length:							
6 weeks	-0.001	0.738	-0.005	0.320	0.000	0.969	675,371
6 weeks + 12 flex	0.003	0.760	-0.016	0.172	-0.024	0.177	
18 weeks	0.007	0.270	0.004	0.660	-0.001	0.948	
12 weeks	-0.008	0.181	-0.007	0.434	-0.012	0.323	
p(diff), all effects	0.376		0.573		0.667		
			D. Promotions (#)				
Main effect:	-0.033*	0.028	-0.086***	0.000	-0.098***	0.000	675,371
Effects by paid leave length:							
6 weeks	**	0.001	-0.115***	0.000	-0.112***	0.000	675,371
6 weeks + 12 flex	-0.001	0.989	-0.199**	0.003	-0.164*	0.027	
18 weeks	-0.025	0.320	-0.071*	0.035	-0.104**	0.009	
12 weeks	0.028	0.296	0.035	0.385	-0.031	0.492	
p(diff), all effects	0.018		0.002		0.239		

Notes: Regressions exclude November–December 2016 births in due to ambiguity about date of conception and policy assignment. Outcomes include physical performance, months of training, years of education, and count of promotion. Postbirth drop is measured at 8 months post-birth for physical performance and at 1 month postbirth for all other outcomes. Regressions include individual and month-by-year fixed effects. The first row replicates the main analysis for the smaller sample. The next rows display a separate regression from the policy interaction model. "6 weeks" is the predicted mother-placebo gap under the 6-week policy (for babies born December 2014 and prior). "6 weeks + 12 flex" is the predicted mother-placebo gap for mothers who gave birth under the 6-week policy but were retroactively given an additional 12 weeks of leave to use before their baby's first birthday after they had returned to work (for babies born January 2015–mid-May 2015). "18 weeks" value the predicted mother-placebo gap for mothers who gave birth knowing they would have 18 weeks of leave, 12 of which could be used flexibly before the baby's first birthday (for babies born mid-May 2015–October 2016). "12 weeks" is the predicted mother-placebo gap for mothers who gave birth knowing they would have 12 weeks of leave to use immediately following birth (for babies born January 2017 and later). The final row presents the *p*-value for an *F*-test of whether mother-placebo gaps are the same across all policy periods. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

A Supplemental Figures and Tables

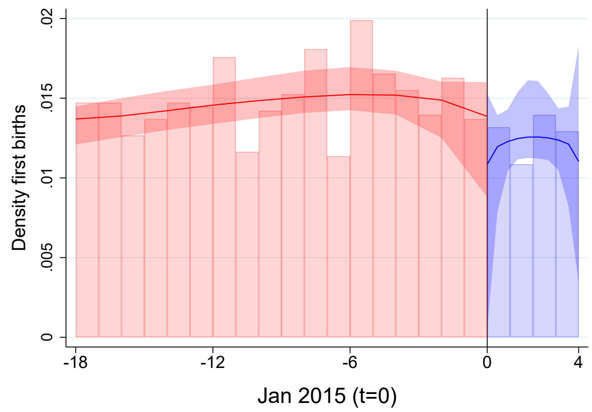
Figure A.1: Placebo Birth DiD Estimates of the Impact of Birth on Job Performance and Marksmanship by Subgroups



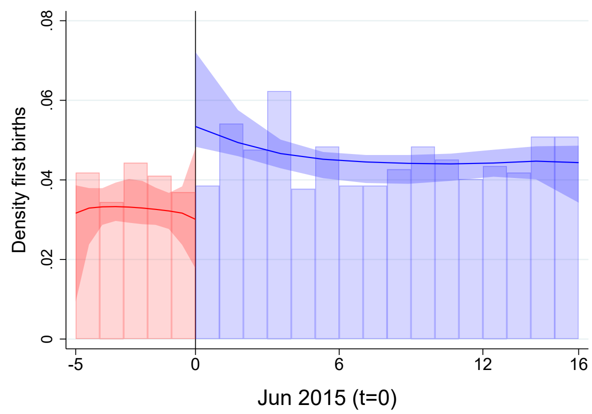
Notes: Displays gaps in job performance and rifle scores across birth/placebo birth events for the first postbirth observation (black line) and 24 month postbirth (light gray line) by subgroups. Each comparison (e.g., reserve vs. active) is based on one regression by interacting an indicator variable with a group indicator (e.g., reserve) with the paramaeters in Eq. 2. Classifications for parents are as follows: "Reserve" are those not on active duty and likely working a civilian job as well; "Active" are working their military job full-time. "> phys" are those whose military job type above the median physicality level in our sampled based on O*NET classification; "< phys" are those whose military job type at or below the median physicality level. "Married" are married at $t = -10$; "Single" are not. ">21" are above the median age (21) at $t = -10$; "<=21" are not. "Jr. Enl" are in enlisted grade E1–E4 at $t = -10$; "Other" are E5 and up or officers. Nonparents assigned to placebo births are limited to those whose rank, number of months in service, and year is an exact match with parents' 10 months before the birth. Among those with an exact match on rank and months in service, each parent's outcomes are compared to the five nonparents most similar to parents in their propensity to have a child based on age, race/ethnicity, military entrance exam scores (AFQT scores), marital status (including whether a spouse is also in the military), level of education, months of training, occupational field, reserve status, and most recent physical performance scores. Nonparent subgroups are defined based on their match's characteristics. The reference period is the time before pregnancy (($t = -10$) and earlier). Vertical solid lines reflect a zero effect. Horizontal lies indicate 95% confidence intervals.

Figure A.2: Density of First Births Across Policy Periods

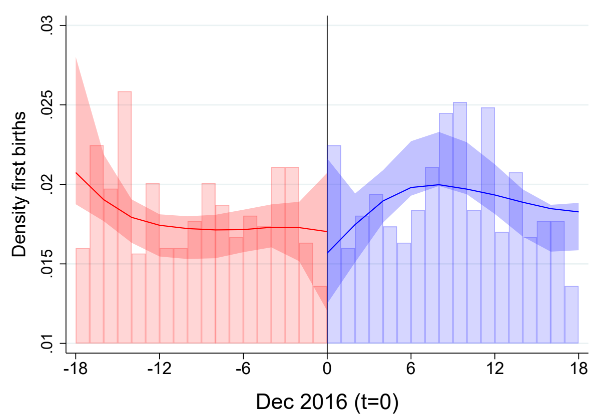
(a) 6 vs. 6 + 12 flex weeks



(b) 6 + 12 flex vs. 18 weeks



(c) 18 vs. 12 weeks



Notes: Histogram bars display the density of first births by month before and after t=0, which differentiates births subject to one leave-length policy period from another. Plotted curves and corresponding 95% confidence intervals come from a manipulation test using a local-polynomial density estimator developed by Cattaneo et al. (2018). The test for a discontinuity at t=0 is not statistically significant in Panel (a), (b), or (c). The sample includes all women in the Marines with a first birth during the time window.

Table A.1: Characteristics of First-Time Parents Across Samples

	Mothers			Fathers		
Obs. required around birth	–	-12/+12mo	-12/+24mo	–	-12/+12mo	-12/+24mo
Age	23.13	23.08	23.39	24.76	25.20	25.48
Education						
Some college	0.05	0.05	0.05	0.05	0.05	0.05
College	0.08	0.08	0.10	0.13	0.15	0.16
Marital Status						
Married	0.67	0.68	0.68	0.85	0.87	0.87
Race/Ethnicity						
Black	0.14	0.15	0.15	0.09	0.09	0.10
Hispanic	0.22	0.22	0.23	0.15	0.15	0.14
Job Classification						
Mngmt./Business/Science/Arts	0.12	0.12	0.14	0.09	0.10	0.10
Service	0.07	0.07	0.07	0.04	0.04	0.04
Sales/Office	0.36	0.36	0.35	0.12	0.12	0.12
Construction/Maint.	0.19	0.18	0.18	0.30	0.29	0.29
Production/Moving/Transpo.	0.20	0.20	0.19	0.13	0.14	0.14
Military	0.07	0.07	0.08	0.32	0.31	0.31
Military Specific Chars.						
Officer	0.06	0.06	0.08	0.11	0.13	0.14
AFQT score (percentile)	57.76	57.87	58.53	61.88	62.61	63.26
GCT score (av=100; sd=20)	102.74	102.86	103.40	110.30	110.84	111.32
Observations	6809	4502	2549	55296	36209	24420

Notes: Displays characteristics of first-time parents across samples. The columns labeled "–" include all first-time parents. The columns "-12/+12mo" restricts the sample of first-time parents to those observed at least 12 months before and 12 months after the birth, while the columns labeled "-12/+24mo" restrict to those observed at least 12 months before and 24 months after the birth. Time-varying characteristics (e.g., age) are measured 10 months before the birth ($t \leqslant 10$). Job categories correspond to Standard Occupational Classification (SOC) system groups applied to U.S. Marine Corps job codes. All averages reflect the share of the sample with the given characteristics, except average age and AFQT and GCT scores. AFQT and GCT scores are measures of intelligence and aptitude, with scoring scales described.

Table A.2: Descriptive characteristics for the Placebo Analysis

Variable	Women			Men		
	Mothers (1)	Placebos (2)	Difference (3)	Fathers (4)	Placebos (5)	Difference (6)
Exact match variables						
Months of service	39.113	39.113	0.000	58.571	58.571	0.000
	[42.620]	[42.613]	(1.035)	[49.633]	[49.632]	(0.392)
Officer	0.076	0.076	0.000	0.139	0.139	0.000
	[0.265]	[0.265]	(0.007)	[0.346]	[0.346]	(0.003)
Reservist	0.051	0.051	0.000	0.113	0.113	0.000
	[0.220]	[0.220]	(0.005)	[0.316]	[0.316]	(0.003)
Year of match	2013.235	2013.235	0.000	2013.082	2013.082	0.000
	[2.024]	[2.024]	(0.047)	[1.980]	[1.980]	(0.015)
Other matching variables						
Age	22.560	22.774	-0.214*	24.649	24.687	-0.038
	[4.184]	[4.372]	(0.105)	[4.632]	[4.848]	(0.038)
Black	0.154	0.149	0.004	0.096	0.091	0.005*
	[0.361]	[0.357]	(0.009)	[0.295]	[0.288]	(0.002)
Hispanic	0.222	0.216	0.006	0.143	0.133	0.010**
	[0.415]	[0.411]	(0.010)	[0.350]	[0.340]	(0.003)
Other	0.099	0.098	0.001	0.073	0.074	-0.001
	[0.299]	[0.298]	(0.008)	[0.260]	[0.262]	(0.002)
Cognitive test (Z-score)	-0.166	-0.155	-0.010	0.025	0.067	-0.042***
	[0.939]	[0.945]	(0.024)	[0.997]	[0.991]	(0.009)
Married	0.416	0.393	0.023	0.672	0.667	0.005
	[0.493]	[0.489]	(0.012)	[0.469]	[0.471]	(0.004)
Military spouse	0.266	0.251	0.015	0.040	0.040	-0.001
	[0.442]	[0.434]	(0.011)	[0.195]	[0.197]	(0.002)
Some college (0/1)	0.046	0.045	0.001	0.050	0.051	-0.001
	[0.210]	[0.208]	(0.005)	[0.217]	[0.219]	(0.002)
College (0/1)	0.093	0.092	0.001	0.158	0.158	0.000
	[0.291]	[0.290]	(0.008)	[0.365]	[0.365]	(0.004)
Months of training	5.455	5.537	-0.082	5.183	5.175	0.009
	[4.019]	[4.104]	(0.101)	[5.486]	[5.580]	(0.051)
Recent fitness score	0.069	0.110	-0.042	0.246	0.259	-0.012
	[0.903]	[0.876]	(0.021)	[0.839]	[0.831]	(0.007)
Combat job type	0.048	0.041	0.007	0.288	0.289	-0.001
	[0.213]	[0.198]	(0.005)	[0.453]	[0.454]	(0.004)
Combat support job type	0.627	0.637	-0.010	0.367	0.358	0.009*
	[0.484]	[0.481]	(0.012)	[0.482]	[0.479]	(0.004)
Aviation job type	0.192	0.190	0.002	0.242	0.248	-0.006
	[0.394]	[0.392]	(0.010)	[0.428]	[0.432]	(0.004)
Joint significance ($p(F)$)			0.5506			0.1034
Analytic weight	1.000	0.226		1.000	0.219	
Observations	2549	12100	14649	24420	116808	141228
Unique individuals	2549	6517	9066	24420	57325	81745

Notes: Displays means (SD in brackets) for parents (Columns 1 and 4) and their respective placebos (columns 2 and 5), and the difference in means (standard error in parentheses) between them (Columns 3 and 6) at the time of the match ($t = 10$), weighted by the analytic weight. Matching required exact matching on months of service, exact rank (e.g., corporal or captain), reservist, and year, with further matching based on predicted propensity score from the remaining variables and their interactions. Table includes the p-value of an F-test of a model predicting being a parent using the listed matching variables, months of service (linearly), reserve status, and fixed effects by year and rank. Table also displays the final analytic average weight number of unique person-month matches, and unique individuals. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

Table A.3: Falsification Test for the Placebo Analysis at $t = -12$

Variable	Women			Men		
	Mothers	Placebos	Difference	Fathers	Placebos	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
			Difference at =-12			
Training (0/1)	0.181	0.185	-0.004	0.136	0.131	0.005*
	[0.385]	[0.388]	(0.009)	[0.343]	[0.338]	(0.003)
Degrees (#)	-0.001	-0.003	0.001	-0.002	-0.002	-0.000
	[0.034]	[0.050]	(0.001)	[0.046]	[0.045]	(0.000)
Promotions (#)	-0.124	-0.121	-0.002	-0.091	-0.092	0.001
	[0.329]	[0.326]	(0.007)	[0.288]	[0.289]	(0.002)
Joint significance ($p(F)$)			0.5307			0.1891
Observations	2549	12100	14649	24420	116808	141228
Unique individuals	2549	6517	9066	24420	57325	81745
			Difference at =-24			
Training (0/1)	0.219	0.228	-0.008	0.159	0.163	-0.004
	[0.414]	[0.419]	(0.012)	[0.365]	[0.369]	(0.003)
Degrees (#)	-0.020	-0.017	-0.003	-0.015	-0.015	0.000
	[0.139]	[0.128]	(0.004)	[0.121]	[0.124]	(0.001)
Promotions (#)	-0.738	-0.730	-0.008	-0.556	-0.558	0.002
	[0.674]	[0.660]	(0.020)	[0.619]	[0.616]	(0.005)
Joint significance ($p(F)$)			0.6475			0.6381
Observations	1478	7016	8494	17621	84111	101732
Unique individuals	1478	3657	5135	17621	39098	56719

Notes: Displays means (SD in brackets) for parents (Columns 1 and 4) and their respective placebos (columns 2 and 5), and the difference in means (standard error in parentheses) between them (Columns 3 and 6) $before$ the match ($t = 12$), weighted by the analytic weight. Table includes the p-value of an F-test of a model predicting being a parent using the listed variables. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table A.4: Coefficients for Impacts of Childbirth on Job Performance Among First-Time Parents

	Physical Performance (sd)		Job Performance (sd)		Marksmanship (sd)	
	Event Study (1)	Placebo (2)	Event Study (3)	Placebo (4)	Event Study (5)	Placebo (6)
			A. Mothers			
Pregnancy drop	–	–	-0.052	-0.031	–	–
			(0.047)	(0.047)		
Pregnancy trend	–	–	0.012	-0.000	–	–
			(0.007)	(0.008)		
Postbirth drop $(birth - 24mos.)$	-0.412***	-0.524***	0.051	-0.091*	-0.061	-0.139
	(0.033)	(0.035)	(0.036)	(0.042)	(0.055)	(0.075)
Recovery $(birth - 24mos.)$	0.057***	0.059***	-0.010*	-0.008	0.003	0.002
	(0.010)	(0.011)	(0.004)	(0.005)	(0.011)	(0.014)
Δ recovery $(13 - 24mos.)$	-0.051***	-0.049***	0.005	0.014	-0.001	0.003
	(0.012)	(0.013)	(0.007)	(0.007)	(0.016)	(0.019)
Parent mean $(t \leqslant -10)$	0.11		-0.01		-0.07	
Nonparent mean $(t = 24)$		0.14		-0.01		0.18
N of individuals	22,094	8,912	10,601	4,987	16,521	6,639
Observations	167,400	65,345	91,066	46,127	64,663	29,246
R^2	0.58	0.61	0.41	0.48	0.45	0.51
			B. Fathers			
Pregnancy drop	-0.001	-0.018*	0.024	0.023	-0.029	-0.025
	(0.008)	(0.008)	(0.014)	(0.014)	(0.019)	(0.019)
Pregnancy trend	-0.005***	-0.011***	0.002	-0.002	0.008**	0.004
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Postbirth drop $(birth - 24mos.)$	-0.077***	-0.115***	0.035**	-0.035**	0.040**	0.019
	(0.006)	(0.006)	(0.011)	(0.012)	(0.014)	(0.015)
Recovery $(birth - 24mos.)$	0.009***	0.008***	0.003*	0.003*	0.004*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Δ recovery $(13 - 24mos.)$	-0.010***	-0.005***	-0.006**	-0.003	-0.005	-0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Parent mean $(t \leqslant -10)$	0.25		0.01		0.02	
Nonparent mean $(t = 24)$		0.06		-0.02		0.20
N of individuals	321,760	81,667	136,108	45,774	215,302	60,443
Observations	2,515,047	1,018,968	1,198,963	451,124	817,386	374,101
R^2	0.59	0.61	0.41	0.50	0.48	0.49

Notes: Displays coefficients from Eq. 2, the semi-dynamic event study specification at listed time points postbirth in the odd columns using the full sample, and Eq. 3, the dynamic placebo specification in the even columns using the placebo sample. Outcomes include (1) standardized (mean=0, SD=1) scores from physical/combat fitness tests conducted 2x per year, (2) standardized scores (mean=0, SD=1) from supervisor-rated job performance evaluations conducted 1-2x per year, and (3) standardized scores (mean=0, SD=1) from rifle or pistol tests conducted 1 or fewer times per year. We exclude women's physical performance scores 9 months before through 7 months after birth because women are not required to take fitness tests during and after pregnancy. We exclude women's marksmanship scores 9 months before through 4 months after birth because women are not required to take marksmanship exams during pregnancy or while on leave. All outcomes for women and men exclude $t = 0$. Regressions include individual and month-by-year fixed effects. The parameter "Pregnancy drop" captures any immediate shift from pre-birth to pregnancy, if observed. The parameter "Pregnancy trend" captures trends during pregnancy, if observed. "postbirth drop" is an indicator equal to 1 after the birth, starting in $t = 1$ for all men's outcomes; $t = 8$ for women's physical performance; and $t = 6$ for women's supervisor ratings. "Recovery trend" estimates monthly changes in the outcome for the entire postbirth period. "Δ Recovery trend" estimates any change in the slope in the second year postbirth. Robust standard errors are clustered by individual, shown in parentheses. $^{***} p < 0.001$, $^{**} p < 0.01$, $^{*} p < 0.05$.

Table A.5: Coefficients for Impacts of Childbirth on Human Capital & Career Advancement Among First-Time Parents

	Training (months) (1)	Degrees (#) (2)	Promotions (#) (3)
	A. Mothers		
Pregnancy drop	0.469***	0.001	0.388***
	(0.036)	(0.002)	(0.009)
Pregnancy trend	-0.030***	-0.000	-0.003
	(0.003)	(0.000)	(0.001)
Postbirth drop $(birth - 24mos.)$	-0.463***	-0.001	-0.035*
	(0.056)	(0.003)	(0.015)
Recovery $(birth - 24mos.)$	-0.023***	-0.000	-0.005***
	(0.004)	(0.000)	(0.001)
Δ recovery $(13 - 24mos.)$	0.009*	0.000	0.004*
	(0.004)	(0.000)	(0.002)
Nonparent mean $(t = 24)$	7.22	0.05	1.31
N of individuals	9,066	9,066	9,066
Observations	615,797	706,319	691,232
R^2	0.85	0.38	0.68
	B. Fathers		
Pregnancy drop	0.250***	0.005***	0.272***
	(0.014)	(0.001)	(0.003)
Pregnancy trend	0.002	0.000	-0.000
	(0.002)	(0.000)	(0.000)
Postbirth drop $(birth - 24mos.)$	-0.005	0.002*	-0.004
	(0.024)	(0.001)	(0.004)
Recovery $(birth - 24mos.)$	0.003	0.000	-0.000
	(0.001)	(0.000)	(0.000)
Δ recovery $(13 - 24mos.)$	0.003	0.000	0.002**
	(0.002)	(0.000)	(0.000)
Nonparent mean $(t = 24)$	7.17	0.04	0.96
N of individuals	81,745	81,745	81,745
Observations	6,800,867	6,947,084	6,800,867
R^2	0.88	0.35	0.62

Notes: Displays coefficients from Eq. 3 using the placebo sample. Outcomes include monthly observations of (1) cumulative months of training, (2) cumulative degree counts relative to $t = -10$, and (3) cumulative promotion counts relative to $t = -10$. All outcomes exclude $t = 0$. Regressions include individual and month-by-year fixed effects. The parameter "Pregnancy drop" captures any immediate shift from pre-birth to pregnancy, if observed. The parameter "Pregnancy trend" captures trends during pregnancy, if observed. "postbirth drop" is an indicator equal to 1 after the birth, starting in $t = 1$ for all men's outcomes; $t = 8$ for women's physical performance; and $t = 6$ for women's supervisor ratings. "Recovery trend" estimates monthly changes in the outcome for the entire postbirth period. "Δ Recovery trend" estimates any change in the slope in the second year postbirth. Robust standard errors are clustered by individual, shown in parentheses. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table A.6: Impact of Childbirth on Physical Performance Test Components

	Physical Fitness Test			Combat Fitness Test		
	(1) 3-Mile Run (time in secs)	(2) Crunches	(3) Pull-Ups	(4) 880-Yard-Run (time in secs)	(5) Lifts	(6) Shuttle Run (time in secs)
			A. Women			
Pregnancy trend	–	–	–	–	–	–
Postbirth drop $(birth - 24mos.)$	43.368***	-3.726***	0.285	8.393***	-1.930*	8.214***
	(5.712)	(0.655)	(1.368)	(1.052)	(0.772)	(1.223)
Recovery trend $(birth - 24mos.)$	-7.945***	0.414	0.381	-1.108***	0.475	-1.309***
	(1.871)	(0.218)	(0.408)	(0.333)	(0.266)	(0.393)
Δ Recovery trend $(13 - 24mos.)$	7.856***	-0.424	-0.582	0.903*	-0.361	1.100*
	(2.287)	(0.265)	(0.503)	(0.404)	(0.286)	(0.485)
12-month effect	11.587**	-2.069***	1.810*	3.962***	-0.029	2.977**
p-value	0.005	0.000	0.029	0.000	0.961	0.001
24-month effect	10.520*	-2.181***	-0.600	1.499	1.341	0.470
p-value	0.028	0.000	0.599	0.083	0.098	0.671
Prepregnancy mean	1520.168	94.233	20.483	215.152	60.638	196.777
Observations	82050	82428	24901	76368	76373	76365
N of individuals	19052	19129	8195	18446	18447	18445
R-squared	0.743	0.576	0.834	0.692	0.537	0.636
			B. Men			
Pregnancy trend	1.654***	0.025**	-0.013***	0.052*	0.002	0.066*
	(0.146)	(0.009)	(0.004)	(0.023)	(0.014)	(0.026)
Postbirth drop $(birth - 24mos.)$	19.571***	0.257***	-0.154***	1.272***	0.400	1.512***
	(1.155)	(0.069)	(0.032)	(0.172)	(0.205)	(0.198)
Recovery trend $(birth - 24mos.)$	-2.263***	0.042***	0.025***	-0.161***	0.038	-0.145***
	(0.136)	(0.009)	(0.004)	(0.020)	(0.021)	(0.023)
Δ Recovery trend $(13 - 24mos.)$	2.196***	-0.055***	-0.042***	0.160***	0.009	0.178***
	(0.230)	(0.015)	(0.006)	(0.034)	(0.032)	(0.040)

(Continued on next page)

Table A.6: Impact of Childbirth on Physical Performance Test Components *(Continued)*

	Physical Fitness Test			Combat Fitness Test		
	(1) 3-Mile Run (time in secs)	(2) Crunches	(3) Pull-Ups	(4) 800-Yard-Run (time in secs)	(5) Lifts	(6) Shuttle Run (time in secs)
12-month effect	-5.320***	0.720***	0.117***	-0.504**	0.820***	-0.088
p-value 0	0.000	0.000	0.000	0.001	0.000	0.609
24-month effect	-6.119***	0.567***	-0.086*	-0.526***	1.389***	0.306
p-value	0.000	0.000	0.015	0.005	0.000	0.152
Prepregnancy mean	1324.543	98.870	17.036	174.884	96.249	144.564
Observations	1099159	1103644	1099638	1015207	1015290	1015190
N of individuals	252102	252848	252691	242703	242709	242700
R-squared	0.721	0.634	0.745	0.654	0.578	0.642

Notes: Displays coefficients from the semi-dynamic specification in Eq. 2 and the average effect at 12 months and 24 months, with p-values below by item by fitness test type. Physical Fitness Test performance is assessed from January to June. Combat Fitness Test performance is assessed from July to December. Run times are measured in seconds, while crunches, pull-ups, and lifts are measured as raw counts. Prior to January 2017, women could opt for an alternative upper-body strength assessment to pull-ups. Limited pull-up outcome data exist for women prior to 2017. Column (4) captures scores on an 880-yard-run, the Movement to Contact drill, designed to mimic the stresses of running under pressure in battle. Column (5) measures the number of times a Marine can lift a 30-pound ammunition can overhead. Column (6) displays timed performance on a 300-yard shuttle run obstacle, Maneuver Under Fire, which includes crawls, ammunition resupply, grenade throwing, agility running, and the dragging and carrying of another Marine. Robust standard errors clustered by ID in parentheses.$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A.7: Identifying Additional Potential Mechanisms for Changes in Promotion

	(1)	(2)	(3)	(4)	(5)
Female X Parent	-0.103***	-0.091***	-0.035*	-0.073***	0.006
	(0.016)	(0.016)	(0.017)	(0.015)	(0.017)
Parent	0.019***	0.024***	0.018***	0.020***	0.021***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Female	0.112***	0.116***	0.121***	0.112***	0.124***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
Deployed post		0.067***			0.066***
		(0.005)			(0.005)
Months limited post			-0.009***		0.000
			(0.001)		(0.001)
Avg post 1st class fit				0.260***	0.213***
				(0.009)	(0.009)
# waived fit tests					-0.038***
					(0.003)
Prepregnancy controls	X	X	X	X	X
p(fem*par+par)=0	0.000	0.000	0.289	0.000	0.086
Sobel-Goodman mediation		0.12	0.66	0.29	1.06
Observations	161304	161304	161304	161304	161304
R-squared	0.2969	0.2987	0.2981	0.3046	0.3078
Adjusted R-squared	0.2969	0.2986	0.2980	0.3045	0.3077

Notes: Predicts whether an individual will remain until at least 36 months after the match birth event ($t = 36$) in a combined women/men placebo sample, using one "wide" observation per match-ID. "Female X Parent" is a mother-specific indicator variable. "Parent" is an indicator for a parent in general. "Prepregnancy controls" adds controls for variables included in Table 1 civilian job classifications and plus being married to a military spouse, months of service, an interaction between months of service and officer, year, an additional "other non-white" race/ethnicity category, and indicator variables for having a combat, combat support, or aviation-type job (relative to jobs not in these categories) at $t = -10$. Deployed post is an indicator equal to one if an individual was deployed at any point in $t > -10$. Months limited post is an count of the months on limited duty (nondeployable) status in $t > -10$. Average post first-class fitness is the percent of fitness test in the "first-class" category in in $t > 0$. Number of missed tests is a count of the physical and combat fitness testing periods without an observed fitness test for $t > -10$; this is based on the USMC fitness test schedule rather than relative time to/from birth. Count of promotions count in from $t = [-10, 24]$. We cluster standard errors clustered by ID because some nonparent IDs are matched multiple times to more than one parent. Includes the *p*-value of an F-test of whether the parent and mother-specific coefficient sum to zero. Sobel-Goodman mediation test computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table A.8: Descriptive characteristics for the Policy Analysis

	Length of Paid Maternity Leave				
	6 wks	6 wks + 12 flex	18 wks	12 wks	p(diff)
Age	22.41	22.62	23.01	22.47	0.05
Black	0.16	0.12	0.13	0.17	0.16
Hispanic	0.19	0.18	0.26	0.30	0.00
Other	0.11	0.11	0.10	0.07	0.23
Cognitive test (Z-score)	-0.16	-0.09	-0.16	-0.24	0.33
Married	0.40	0.40	0.48	0.39	0.01
Military Spouse	0.26	0.26	0.31	0.24	0.12
Some college (0/1)	0.04	0.06	0.06	0.04	0.24
College (0/1)	0.08	0.10	0.13	0.09	0.02
Months of training	4.26	6.04	7.04	7.31	0.00
Recent fitness score	0.07	-0.01	0.08	0.09	0.75
Combat job type	0.05	0.02	0.07	0.03	0.06
Combat support job type	0.63	0.63	0.60	0.66	0.28
Aviation job type	0.19	0.21	0.21	0.18	0.53
Reservists	0.06	0.07	0.04	0.03	0.08
Observations	1448	126	513	404	2491

Notes: Displays means for mothers by policy period (columns 1–4) and the p-value of an ANOVA test of whether the values differ across groups (column 5). Excludes mothers whose first birth was in November–December 2016 due to ambiguity about the policy for such mothers. Variables are those use in the matching procedure.