AEA Papers and Proceedings 2022, 112: 561–567
https://doi.org/10.1257/pandp.20221121

ECONOMICS OF THE FAMILY

Gender Disparities in Career Advancement across the Transition to Parenthood: Evidence from the Marine Corps†

By Olivia J. Healy and Jennifer A. Heissel*

Parenthood is a unique turning point in women's careers. Mothers, but not fathers, experience large and persistent child penalties to earnings after the birth of their first child.[1] Child penalties mainly result from three differences between mothers and fathers in response to childbearing: hours worked, labor market exit rates, and wages (Kleven, Landais, and Søgaard 2019). One possible explanation for these patterns is that the mental and physical strain of parenthood accrues more acutely to mothers, limiting their ability to perform on the job after having a child. Increased difficulty engaging in work may slow mothers' career advancement, driving longer-term gaps in earnings and labor force participation between mothers and fathers.

In this paper, we explore whether physical ability to perform work tasks changes differently for mothers than for fathers after a first birth. Researchers can rarely track on-the-job performance at a detailed, consistent level across the transition to parenthood. We use administrative data from the US Marine Corps (USMC) to investigate first-time parents' scores on twice-yearly standardized physical fitness tests. These tests measure performance in a key domain of Marines' jobs: physical ability.

The link between physical health and job performance in the military makes postpartum health especially relevant in our setting. The USMC setting also has the advantage of holding two drivers of the child penalty constant. Multiyear contracts limit Marines' ability to change hours worked or exit the labor market after a birth. Marine mothers could take 6–18 weeks of paid leave, but they continue to work similar hours after returning (Bacolod et al., forthcoming).

We estimate event study models around the first birth and include data on nonparents to estimate counterfactual trends. Following Kleven, Landais, and Søgaard (2019), we assign "placebo births" to nonparents and use least absolute shrinkage and selection operator (LASSO)–selected predictors of parenthood to identify observably similar parents and nonparents.

We find large and persistent effects of motherhood on physical performance. Two years postbirth, mothers' physical performance remains 0.2 standard deviations lower than nonmothers' relative to prepregnancy levels. For fathers, physical performance declines during pregnancy, reaching its lowest point one month postbirth. However, fathers' performance fully recovers by the child's second birthday. Negative effects are larger for junior- rather than senior-ranking

* Healy: Cornell University (email: ojh4@cornell.edu); Heissel: Naval Postgraduate School (email: jaheisse@nps.edu). The authors are grateful to Maj. Tamara Cordero, Capt. Amanda Henegar, Capt. Michael Larson, Maj. Laura Laurito, Capt. Julie Schumacher, Maj. Genevieve Studer, Capt. Ansley White, and all of the NPS Marines who provided insight into Marine processes and procedures. We also thank Jonathan Guryan, Matthew Notowidigdo, Diane Schanzenbach, and seminar participants at Stanford University and the ASSA, SOLE, MEA, WEA, and APPAM meetings. We are grateful for funding from the Robert Wood Johnson Foundation's Evidence for Action grant program (#77124), the Institute of Education Sciences' Multidisciplinary Program in Education Sciences (Award #R305B140042), and Northwestern University's Graduate Research Grant. Any errors are our own. The views expressed in this paper do not reflect the views of the US Department of Defense or the US Navy. Data from the Department of Defense are not available for sharing.

† Go to https://doi.org/10.1257/pandp.20221121 to visit the article page for additional materials and author disclosure statement(s).

[1] See, for instance, Aguilar-Gomez, Arceo-Gomez, and De la Cruz Toledo (2019); Andresen and Nix (forthcoming); Angelov, Johansson, and Lindahl (2016); Bertrand, Goldin, and Katz (2010); Kleven et al. (2019, 2020); and Kleven, Landais, and Søgaard (2019).

women, single rather than married fathers, fathers in more physically demanding jobs relative to those in less physically demanding jobs, junior- relative to more senior-ranking fathers, and fathers working part time (as reservists) relative to full time (active duty). Healy and Heissel (2022) demonstrate that mothers, but not fathers, have lower promotion rates in this same sample.

Understanding gender-related disparities in work outcomes in the military context is particularly important given that the US military is the world's largest employer. Military service members represent a large, diverse segment of the US population, with roughly 1.3 million full-time workers, 30 percent of whom identify as racial minorities (Department of Defense 2018). Our results are also relevant for civilian workers in physically demanding jobs. These groups are understudied in the child penalty literature, yet about 45 percent of civilian jobs in the United States require at least medium physical strength, work that involves frequent lifting or carrying of objects weighing up to 25 pounds (Bureau of Labor Statistics 2017).

Findings contribute to the literature on spillover of individual events (e.g., birth) within families (Fletcher and Marksteiner 2017; Heissel 2021). Because fathers are at least initially affected by parenthood, they may not be a good comparison for mothers when estimating the effect of parenthood. Performance declines among fathers also suggest that channels beyond the biological impact of birth shape parents' physical ability to perform on the job, consistent with findings from Andresen and Nix (forthcoming) and Kleven, Landais, and Søgaard (2021).

I. Institutional Background and Data

Our sample consists of active-duty Marines who work full time and reserve Marines who work part time. USMC jobs span more than 35 career fields, ranging from military-specific (e.g., infantry) to civilian-equivalent (e.g., food services or financial management) fields. Active-duty Marines typically work Monday through Friday, and their day usually begins with early morning physical training followed by work assignments through the evening. In contrast, reserve Marines participate in USMC training one weekend per month and two weeks during the year. The rest of the time, reservists lead civilian lives, usually working in the civilian sector or enrolled as students in higher education. The majority of Marines, including the majority in our sample, are active duty. In subgroup analyses, we focus on reservists separately from active duty to explore any differences in the effect of parenthood among a more civilian-oriented population.

Our data include descriptive information (age, gender, race/ethnicity, education level, and aptitude test scores), dependents' date of birth, and job characteristics (job type, rank, and time in service) for January 2010–December 2019. We observe data at the monthly level.

The USMC uses standardized measures to regularly evaluate Marines' job proficiency. This includes two required fitness tests per year: a physical fitness test in the first half of the year (including timed running, crunches, and pull-ups or push-ups) and a combat fitness test in the second half of the year (including a timed sprint, a combat-related obstacle course, and timed overhead ammunition can lifts to assess upper-body strength). Points are awarded based on raw fitness scores, and the points system adjusts for Marines' age and gender. We standardize points-based scores by calendar year, gender, and test type, then combine the Z-scores into one measure of physical performance, generally observed twice per year per Marine.

Fitness scores are one of three key inputs the USMC uses to determine promotions. As a result, changes in assessed physical ability can meaningfully impact career advancement.

II. Empirical Approach

If the transition to parenthood affects outcomes, a first birth should create a sharp change in outcomes when it occurs. We can attribute any discontinuity in the outcome to the pregnancy/birth if we assume that other factors that shape job outcomes do not also undergo a sharp change at the same time. In other words, while the choice to have a child may be endogenous, the exact timing serves as a shock.

We examine the event of a first birth and include nonparents to create a difference-in-difference model. One worry is that nonparents do not serve as an appropriate counterfactual for parents' outcomes absent birth. For instance, if nonparents are in more physical jobs, we may expect them to have different physical performance trajectories over time than those in less physical jobs, regardless of their parental status. To improve comparability, we follow Kleven, Landais, and Søgaard (2019) and assign a comparison group of nonparents to placebo births based on observable characteristics that predict selection into parenthood. We use a LASSO model to identify predictors of a first birth based on observable characteristics including age, race/ethnicity, military entrance exam scores, marital status, education, occupational field groups, recent physical performance scores, months of service, job rank (officer versus enlisted), reserve status, calendar year, and interactions among all variables. We run the LASSO prediction model separately for women and men. See Healy and Heissel (2022) for more details. We require nonparents to have the same job rank, number of months in service, active/reserve status, and to be in the same calendar year to match to parents. This ensures similar prepregnancy work contexts. Within these exact-match groups, we connect first-time parents to their five nearest neighbor nonparents (with replacement) based on parents' characteristics ten months before the birth. We assign placebo births to nonparents ten months after the match. Online Appendix Table A1 displays characteristics of the parents and placebo parents separately by gender. The groups look almost identical. Differences are functionally small (e.g., mothers are 22.6 years old, while placebos are 22.7) when they arise.

Our estimation strategy compares changes in outcomes for first-time parents (P_i) to the average change in outcomes for the five most observably similar nonparents to whom they match. We conduct analyses separately for men and women. Our fully flexible model is

$$(1)\quad Y_{igtr} = \sum_{r=k_{\min}}^{k_{\max}} \mathbf{1}\left(t = t_{ig}^{*} + r\right)\theta_r + \pi P_i + \sum_{r=k_{\min}}^{k_{\max}} \mathbf{1}\left[\left(t = t_{ig}^{*} + r\right)P_i\right]\beta_r + \alpha_g + \phi_t + \varepsilon_{igtr},$$

where t_{ig}^{*} is the month-year of the real or placebo birth for individual i in exact-match group g based on calendar time t. Coefficients θ_r represent the average change in outcomes r months after a placebo birth (or r months before, if $r < 0$) for nonparents, while β_r estimates whether this change is larger, smaller, or the same for parents r months before/after the actual birth. We measure effects relative to $r = -10$, corresponding to ten months prior to the birth and approximately one month before the pregnancy. Y_{igtr} is our outcome of standardized physical job performance. Women are exempt from physical fitness tests during pregnancy through six months postbirth. Some supervisors extend exemptions by a month, so we do not estimate effects on mothers' outcomes Y_{igtr} for $r = [-9, 7]$.

Our analysis focuses on effects for 2 years on either side of the birth ($r = [-24, 24]$). We require that parents and placebos be continuously observed in the data for $r = [-12, 24]$. We bin relative time endpoints, including a dummy variable for months below $r = -24$ and one for months above $r = 24$ in our model. Including binned endpoints allows us to estimate time fixed effects ϕ_t that account for month-by-year changes in the outcome (e.g., changes in fitness test standards in a particular year) separately from relative time fixed effects. We include α_g to create a within-match-group comparison and ε_{igtr} as the error term.

We include matched nonparents in our estimation to address bias from staggered treatment-timing event study designs (Baker, Larcker, and Wang 2021; Goodman-Bacon 2021; Roth et al. 2022). Nonparents approximate counterfactual time trends that all Marines would have experienced, assuming that outcomes would have evolved similarly absent childbirth. By relying on nonparents to help estimate the counterfactual, we limit the share of 2×2 "forbidden" comparisons (e.g., using parents with births early on to estimate counterfactual trends for those with births later). We also require that nonparents match to parents in the same year of the data, which more closely aligns birth/placebo birth timing with calendar time and further minimizes problematic 2×2 comparisons.

We also estimate a semiparametric model to smooth noise in monthly point estimates and improve statistical precision for estimates of subgroup differences among smaller samples. We model changes during pregnancy, the immediate postbirth period, and any postbirth recovery, relative to the prepregnancy period, as follows:

$$Y_{igtr} = (Preg_{igtr} + PregTrend_{igtr} + Post_{igtr} + Recovery_{igtr} + \Delta Recovery_{igtr})\,\theta_j + \pi P_i + [(Preg_{igtr} + PregTrend_{igtr} + Post_{igtr} + Recovery_{igtr} + \Delta Recovery_{igtr})\,P_i]\beta_j + \alpha_g + \phi_t + X_{igtr}\gamma_j + (X_{igtr} \times P_i)\,\delta_j + \varepsilon_{igtr}.$$

To capture level shifts in the outcome, we define $Preg_{igtr} = 1$ during the pregnancy ($r = [-9, -1]$) and $Post_{igtr} = 1$ during the postbirth period ($r > 0$) and 0 otherwise.[2] For monthly trends above and beyond any level shift during pregnancy, we define $PregTrend_{igtr} = [1, 9]$ corresponding to relative time $r = [-9, -1]$ or 0 otherwise. We estimate linear recovery trends during the postbirth period $Recovery_{igtr}$. We define $\Delta Recovery_{igtr} = [1, 12]$, which corresponds to postbirth months $r = [13, 24]$ and estimates changes to the monthly recovery rate that begin at 13 months postbirth. All recovery trend variables equal 0 outside of their relevant postbirth periods. The vector X_{igtr} includes 2 binary indicators for event time below $r = -24$ and above $r = 24$ to mirror our estimation strategy in equation (1).

Coefficients θ_j in equation (2) capture changes for nonparents, while coefficients β_j are specific to parents. We further interact equation (2) with subgroup variables of interest to estimate whether effects differ across subgroups of parents. We predict and report values for subgroups at 2 time points: the first postbirth observation ($r = 8$ for women and $r = 1$ for men) and 2 years postbirth ($r = 24$).

[2] We exclude $r = 0$ due to ambiguity about whether outcomes measured in the birth month reflect pre- or postbirth measures.

III. Results

Figure 1 presents results from equation (1). The bottom left of each panel displays the p-value for an F-test of whether prepregnancy point estimates, $r = [-24, -11]$, jointly equal 0. Parents' average outcomes do not differ from nonparents' before the pregnancy, bolstering confidence in our research design.

Once women resume testing at $r = 8$, we observe large and persistent physical performance declines. Mothers' performance on job-related fitness tests remains roughly 0.2 SD lower at the time of their child's second birthday ($r = 24$) than before the pregnancy ($r = -10$).

Fathers' physical performance also declines due to parenthood. Fathers' scores decline during the pregnancy and reach their lowest point one month postbirth. Unlike mothers, fathers recover to their prepregnancy performance levels shortly after the child's first birthday.

Figure 2 explores heterogeneity in effects by marital status, the physical intensity of one's job, job rank (junior versus senior), reserve/active-duty status, and one's likelihood to stay on the job (remain in the Marines) beyond 36 months postbirth.[3] The black marker indicates the predicted effect at $r = 8$ for mothers and $r = 1$ for fathers; the gray marker indicates the predicted effect at $r = 24$. Filled-in markers indicate that the difference between groups at those time points is statically significant at the 1 percent level. Online Appendix Table A2 displays p-values for tests of statistical significance for the cross-group comparisons.

There are no differences by marital status for mothers. Single fathers have lower performance than married fathers at $r = 1$ but not at $r = 24$.

The impact of a birth for mothers in high- versus low-physicality jobs does not vary. For fathers, those in high-physicality jobs have larger initial drops in performance at $r = 1$ than those in low-physicality jobs.

[3] Subgroup definitions are based on parents' characteristics at $r = -10$, except for marital status, which we measure at $r = 0$ to capture relationship circumstances when the baby arrives. Nonparents remain matched to parents regardless of whether their subgroup characteristics are the same. We define low/high-physicality jobs using median Marine O*NET physicality index scores, excluding jobs without O*NET classifications.

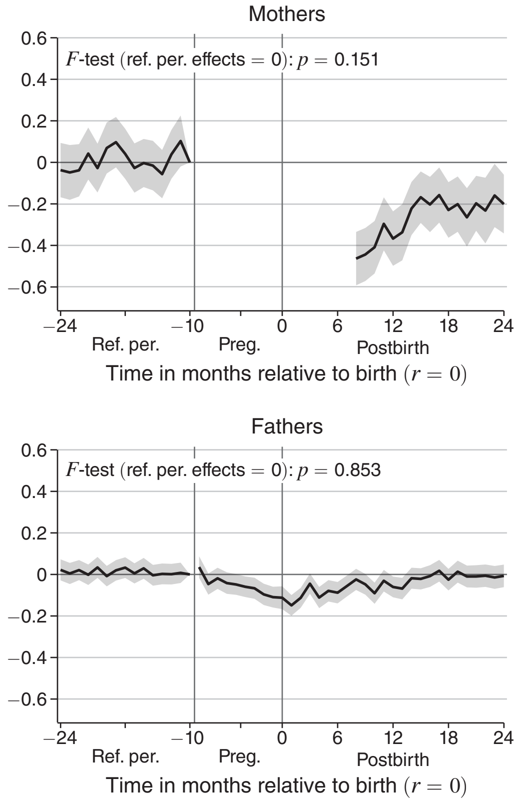

Figure 1. Effects of First Birth on Marines' Physical Fitness Performance Scores

Notes: Displays coefficients from event study regressions (equation (1)) comparing changes between parents and matched nonparents assigned to placebo births. The physical performance outcome is measured in standard deviation units based on scores from physical/combat fitness tests, standardized separately for men and women by year and test type. Nonparents are an exact match with parents at $r = -10$ on rank, number of months in service, reserve status, and observation year. A maximum of five exact-match nonparents are matched to each parent using nearest neighbor propensity score matching with replacement. LASSO-selected best predictors of a first birth generate nonparents' propensity scores. Predictors include age, race/ethnicity, military entrance exam scores, marital status, education, occupational field groups, recent physical performance scores, months of service, job rank (officer versus enlisted), reserve status, calendar year, and interactions among all variables as of $r = -10$. Regressions include exact-match group and month-year fixed effects. The reference month is $r = -10$. Vertical lines reflect the start of the pregnancy ($r = -9.5$) and the birth ($r = 0$). Standard errors are clustered by individual and exact-match group, included as shaded areas representing a 95 percent confidence interval.

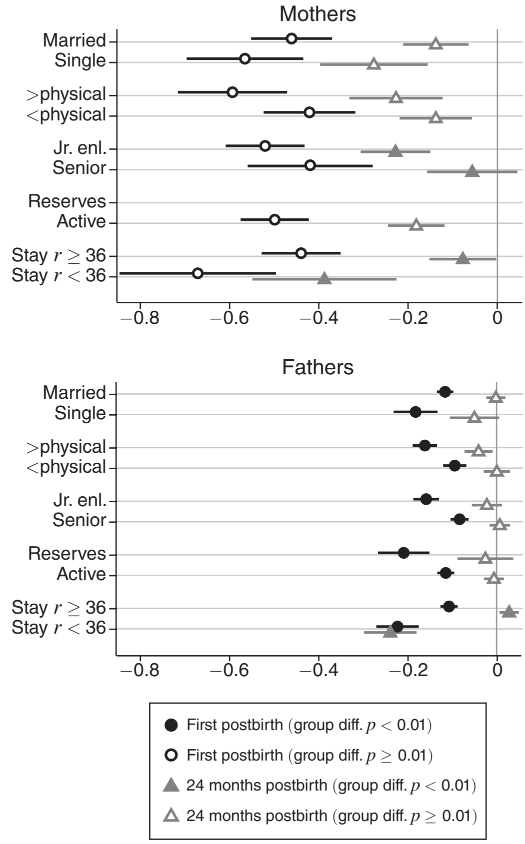

Figure 2. Predicted Effects by Subgroups

Notes: Displays predicted effects relative to prepregnancy at $r = 1$ months postbirth for men and $r = 8$ for women (each when first observed postbirth, shown in black) and 24 months postbirth (shown in gray) by subgroup. Each comparison (e.g., married versus single) is based on one regression that interacts model parameters from equation (2) with the subgroup indicator (e.g., married). Classifications are based on parents' characteristics. Models exclude observations $r = [-9, -7]$ for women due to test exceptions and $r = 0$ for fathers because tests may have occurred either before or after the birth. "Married" are married at $r = 0$; "Single" are not. ">physical" have jobs at $r = -10$ above the median physicality level in our sample based on O*NET classification; "<physical" are at or below the median. "Jr. Enl" are in grades E1–E4 at $r = -10$; "Senior" are E5 and higher or officers at $r = -10$. "Reserves" are not on full-time active duty at $r = -10$; "Active" are full-time Marines. "Stay r $\geq$ 36" stay in the Marines 3 years or longer after the birth event; "Stay r $\geq$ 36" leave before $r = 36$. Vertical solid lines reflect a zero effect. Horizontal lies indicate 95 percent confidence intervals.

Among mothers, negative longer-term effects at $r = 24$ are driven by junior enlisted women relative to more senior women. For fathers, the negative effect of birth is also larger for junior enlisted men than for more senior-ranking men initially at $r = 1$.

We do not have enough statistical power to conduct subgroup analyses for reservist women given that very few are in the reserves. For fathers, the negative effect of birth is larger for reservists than for active duty at $r = 1$. Reservists provide a useful comparison, as most have full-time civilian jobs. We may expect physical performance to drop more in civilian settings, suggesting that our main estimates among a sample of primarily active-duty Marines are a lower bound on the effects.

Marines who plan to leave the military soon after a birth may not be as invested in maintaining their physical job performance abilities. Our last subgroup splits the sample by those who stay in the Marines for 36 months or longer postbirth (75 percent of parents in our sample) and those who leave after 24 but before 36 months. For mothers and fathers, the drop in performance is generally larger for those who leave before 36 months relative to those who remain on the job past 36 months after a first birth. Worse outcomes for parents who leave could be because they already wanted to get out and thus put in less effort at work—or because they struggled to perform after having a child and left despite wanting to stay in.

IV. Discussion

We use repeated measures of job-relevant physical fitness tests to explore the effects of a first birth on workers' ability to maintain job performance. Both men's and women's physical performance drops after having a child. The effects for mothers are large and remain for at least two years, while for fathers, the declines are smaller and fade by their child's second birthday.

Our findings provide a new angle on the child penalty literature, highlighting that changes to job performance for women in the immediate 24 months following a first birth could lead to long-term child penalties. Results underscore the need for policy- and firm-level support for recent parents.

REFERENCES

Aguilar-Gomez, Sandra, Eva Arceo-Gomez, and Elia De la Cruz Toledo. 2019. "Inside the Black Box of Child Penalties." Unpublished.

Andresen, Martin, and Emily Nix. Forthcoming. "What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples." *Journal of Labor Economics*.

Angelov, Nikolay, Per Johansson, and Erica Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics* 34 (3): 545–79.

Bacolod, Marigee, Jennifer A. Heissel, Laura Laurita, Matthew Molloy, and Ryan Sullivan. Forthcoming. "Mothers in the Military: The Effect of Maternity Policy on Leave Up-Take." *Demography*.

Baker, Andrew, David F. Larcker, and Charles C.Y. Wang. 2021. "How Much Should We Trust Staggered Difference-In-Differences Estimates?" European Corporate Governance Institute Finance Working Paper 736/2021.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2 (3): 228–55.

Bureau of Labor Statistics. 2017. "Physical Strength Required for Jobs in Different Occupations in 2016." Bureau of Labor Statistics, April 10. https://www.bls.gov/opub/ted/2017/physical-strength-required-for-jobs-in-different-occupations-in-2016.htm.

Department of Defense. 2018. *2018 Demographics Report: Profile of the Military Community*. Department of Defense, Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy. https://download.militaryonesource.mil/12038/MOS/Reports/2018-demographics-report.pdf.

Fletcher, Jason M., and Ryne Marksteiner. 2017. "Causal Spousal Health Spillover Effects and implications for Program Evaluation." American Economic Journal: Economic Policy 9 (4): 144–66.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.

Healy, Olivia J., and Jennifer A. Heissel. 2022. "Baby Bumps in the Road: The Impact of Parenthood on Job Performance, Human Capital, and Career Advancement." Unpublished.

Heissel, Jennifer A. 2021. "Teen Fertility and Siblings' Outcomes Evidence of Family Spillovers Using Matched Samples." *Journal of Human Resources* 56 (1): 40–72.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11 (4): 181–209.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2021. "Does Biology Drive Child Penalties? Evidence from Biological and Adoptive Families." *American Economic Review: Insights* 3 (2): 183–98.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child Penalties across Countries: Evidence and Explanations." *AEA Papers and Proceedings* 109: 122–26.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2020. "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation." NBER Working Paper 28082.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." Unpublished.