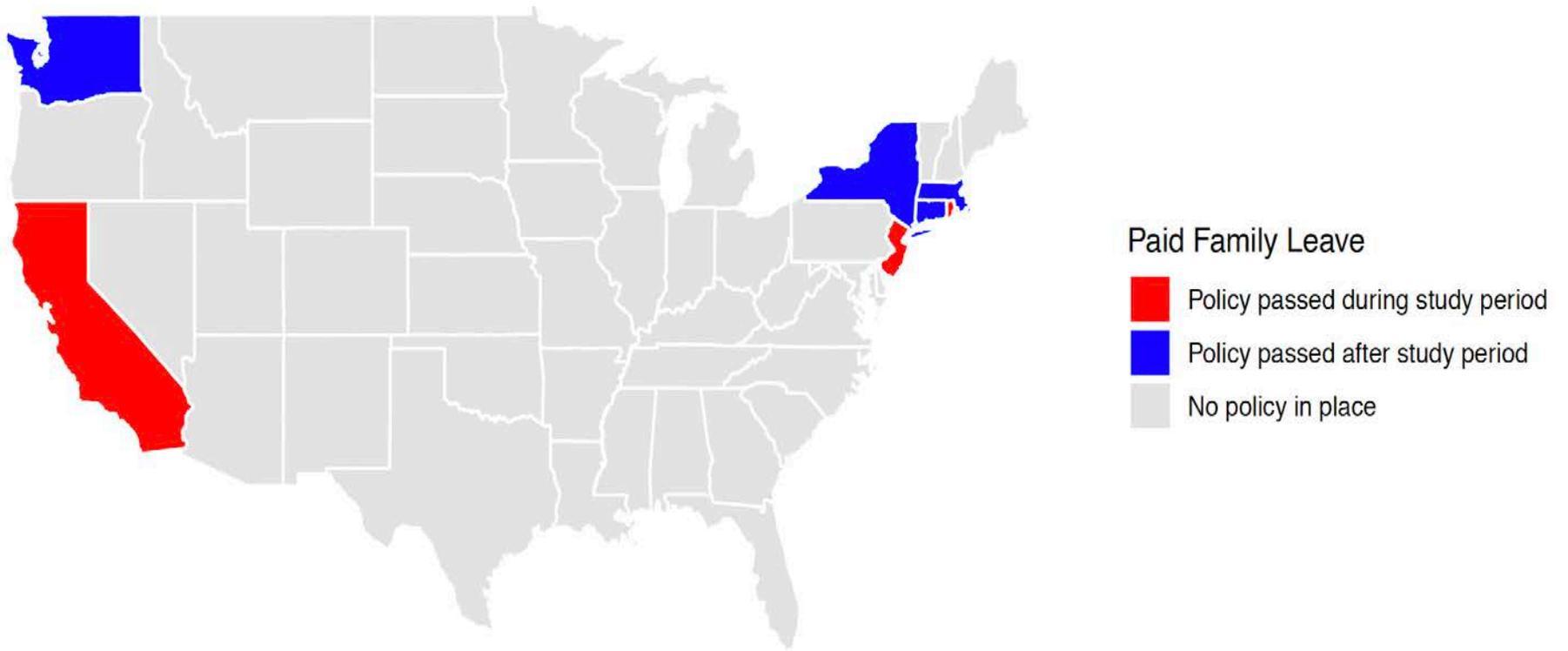**SUPPLEMENTAL METHODS**

In our primary analysis we cluster standard errors at the level of the state, as this is the level at which the "intervention" (i.e., paid family leave policy) occurs (1). However, in the presence of few treated clusters, standard errors may be biased (2). In a sensitivity analysis, we therefore implemented two variants of wild bootstrap inference—both restricted and unrestricted—using the *boottest* command in Stata (3). Confidence intervals roughly doubled using this procedure, such that only overall health and any alcohol consumption demonstrated a statistically significant or marginally significant effect (Supplemental Table 5). Results for other outcomes should therefore be interpreted with caution, although notably, these methods are thought to under-reject the null hypothesis in the case of the extreme case of only one treated cluster, so they may be overly conservative in this case.

Also, our initial analysis plan included the examination of body mass index, smoking, and infant birth weight. For these outcomes, however, there were violations of the pre-policy parallel trends assumption. We therefore omitted these results in our final analysis.

SUPPLEMENTAL REFERENCES

1. Abadie A, Athey S, Imbens GW, et al. When Should You Adjust Standard Errors for Clustering? National Bureau of Economic Research; 2017 (Accessed January 18, 2020).(http://www.nber.org/papers/w24003). (Accessed January 18, 2020)

2. MacKinnon JG, Webb MD. Wild Bootstrap Inference for Wildly Different Cluster Sizes. *J. Appl. Econom.* 2017;32(2):233–254.

3. Roodman D, MacKinnon J, Nielsen M, et al. Fast and wild: bootstrap inference in Stata using boottest. *Stata J.* 2019;

Supplemental Figure 1. States with paid family leave policies

Paid Family Leave

■ (red) Policy passed during study period
■ (blue) Policy passed after study period
□ (gray) No policy in place

Note: Map of the continental U.S. showing state-level paid family leave policies. While New Jersey and Rhode Island passed policies during the study period, these were excluded from our sample due to the small number of observations and subsequent unstable estimates. States with policies that were passed after the study period were still included in the control group. No policies have been passed in Alaska or Hawaii.

Supplemental Figure 2. Number of births and PFL claims paid in California, by year
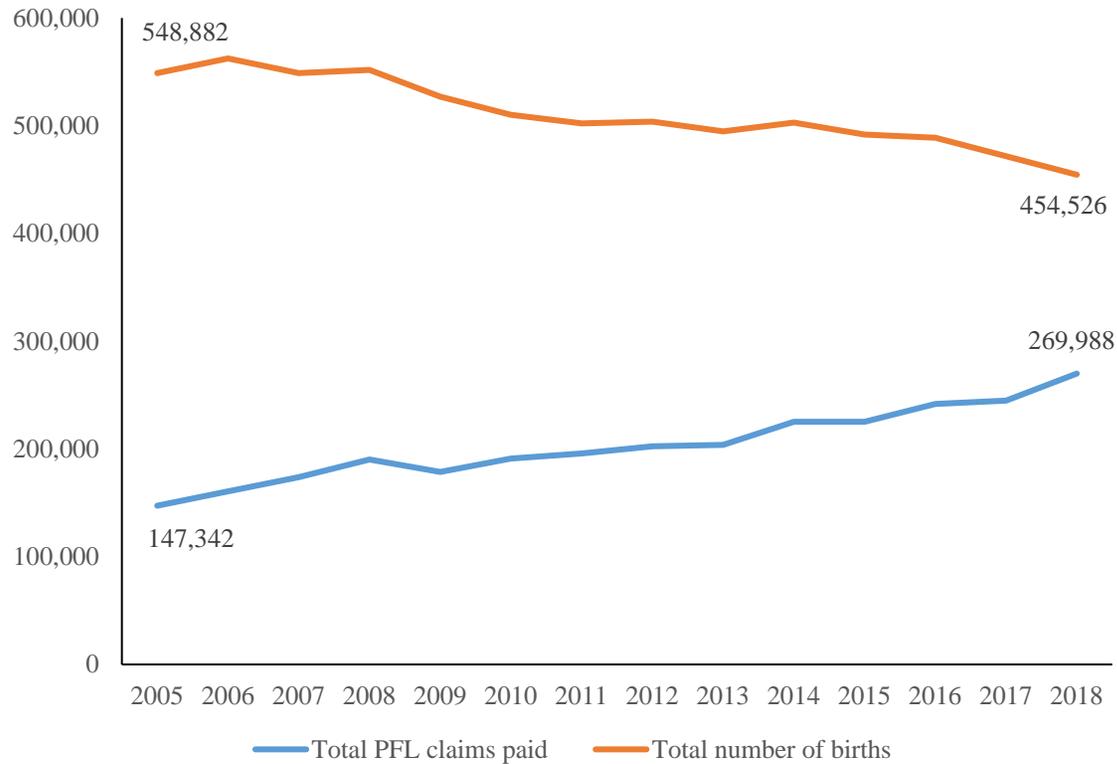
Note: Figure created based on authors' calculations using annual data from the National Vital Statistics Reports (https://www.cdc.gov/nchs/products/nvsr.htm) and the State of California Employment Development Department (https://data.edd.ca.gov/Disability-Insurance/Paid-Family-Leave-PFL-Monthly-Data/r95e-fvkm). We excluded 2004 because California's PFL policy was implemented in July, and therefore the number of claims paid that year was substantially lower. Natality data for 2019 are not yet available from the National Center for Health Statistics. Note that claims may be filed by one or both parents for a single birth. PFL: paid family leave.

Supplemental Figure 3. Graphical illustration of trends in outcomes
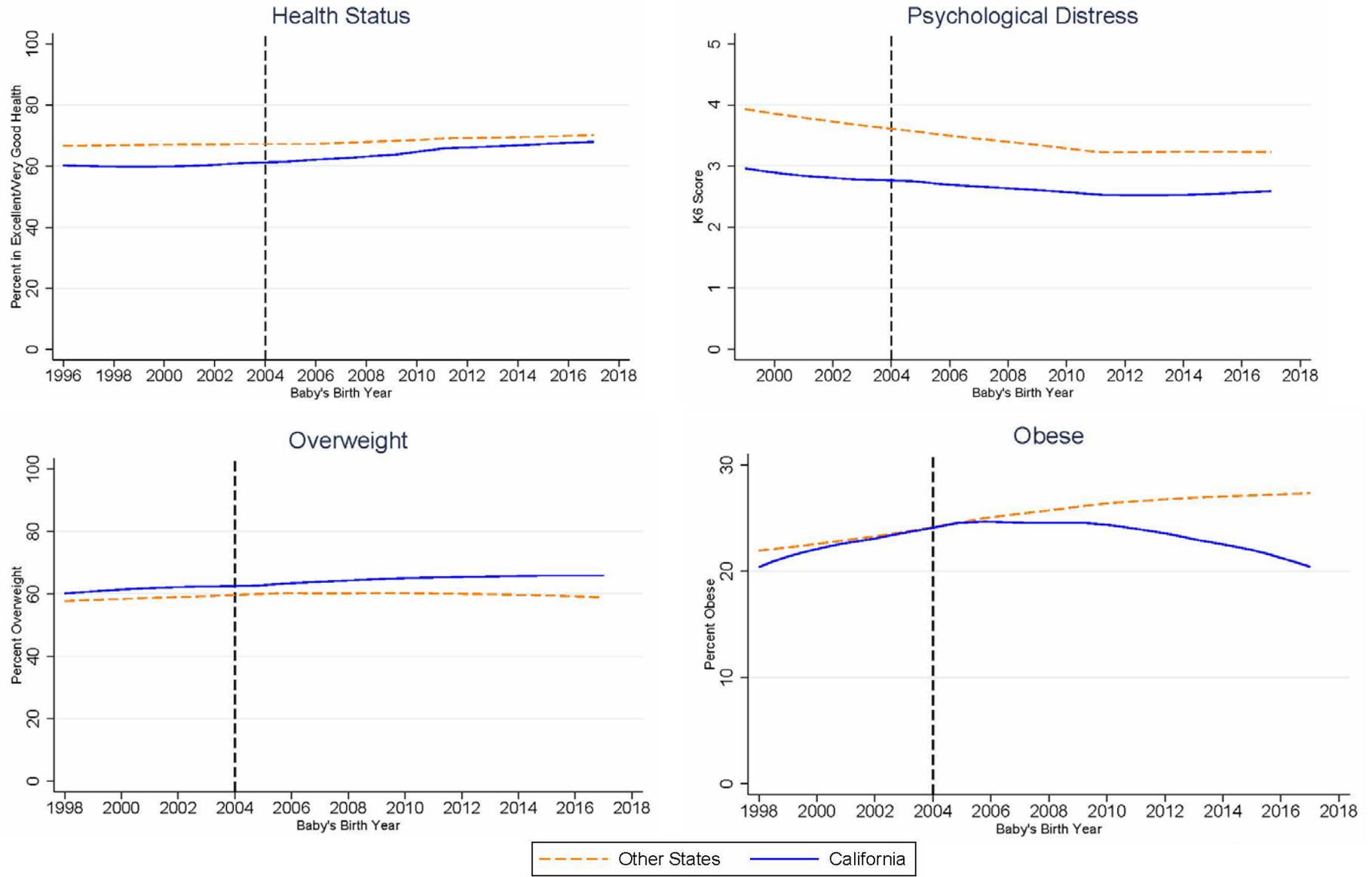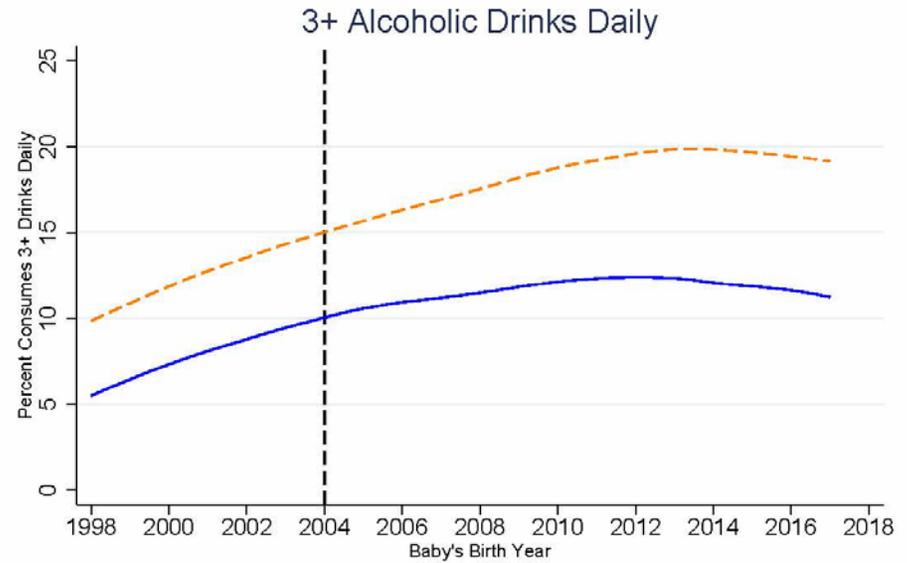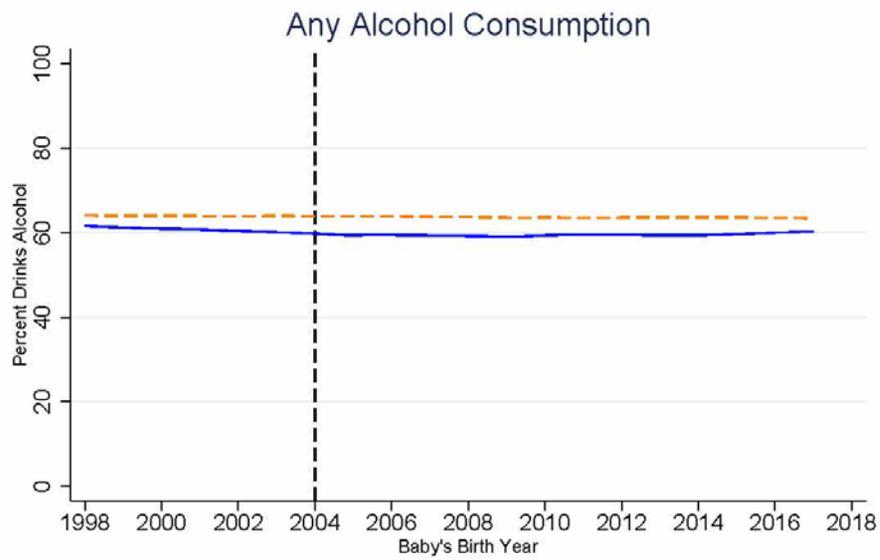
Supplemental Table 1. Test of parallel trends in outcomes during pre-policy period

	β [95% CI] No. person-years
	Test of trend in pre-policy period
Self-rated health	-0.0024
	[-0.017, 0.012]
	2,092
Psychological distress	-0.18
	[-0.74, 0.38]
	397
Overweight	-0.010
	[-0.035, 0.015]
	1,333
Obese	0.0071
	[-0.0098, 0.024]
	1,333
Drinks alcohol	0.010
	[-0.021, 0.042]
	1,358
3+ drinks daily	0.0019
	[-0.0081, 0.012]
	1,356

** $p < 0.01$, * $p < 0.05$

Sample was drawn from the 1993-2017 waves of the Panel Study of Income Dynamics. Analyses involved multivariable linear regressions adjusting for parental gender, age, and educational attainment, household size, inflation-adjusted household income in the year before the child's birth, the head of household's race, and several time-varying state characteristics. We also included fixed effects (i.e., indicator variables) for state and year of birth. Robust standard errors were clustered by state of birth. Coefficients represent a test of whether trends in outcomes were parallel for California and non-PFL states during the pre-policy period.

Supplemental Table 2. Test for differential compositional changes in California versus other states

	β [95% CI]
	Pre-post policy difference in composition
Female	-0.030
	[-0.25, 0.19]
Age	1.57*
	[0.33, 2.81]
Married	-0.062
	[-0.27, 0.14]
Race	
White	-0.11
	[-0.30, 0.075]
Black	0.037
	[-0.14, 0.22]
Hispanic	0.017
	[-0.057, 0.091]
Other	0.059*
	[0.0017, 0.12]
Education	
Less than high school	0.0061
	[-0.14, 0.16]
High school	-0.072
	[-0.29, 0.14]
Some college	-0.047
	[-0.25, 0.16]
College or more	0.11
	[-0.037, 0.26]
Number of children	0.091
	[-0.28, 0.46]
Household income	2,899
	[-13,652, 19,450]

** $p < 0.01$, * $p < 0.05$

Sample was drawn from the 1993-2017 waves of the Panel Study of Income Dynamics. Analyses involved linear regressions with fixed effects (i.e., indicator variables) for state and year of birth. Robust standard errors were clustered by state of birth. Coefficients represent a test of whether trends in outcomes were different in California versus non-PFL states across the pre-policy period.

Supplemental Table 3. Placebo test of the effect of paid family leave policy on parent health

	β [95% CI]	
	2001	*2007*
Self-rated health	-0.11	0.057
	[-0.42, 0.21]	[-0.085, 0.20]
Psychological distress	-1.87**	1.30*
	[-3.11, -0.63]	[0.059, 2.54]
Overweight	0.19	0.044
	[-0.13, 0.52]	[-0.070, 0.16]
Obese	-0.064	0.035
	[-0.27, 0.14]	[-0.062, 0.13]
Drinks alcohol	-0.18	0.075
	[-0.47, 0.11]	[-0.061, 0.21]
3+ drinks daily	-0.078	0.12*
	[-0.17, 0.014]	[0.015, 0.23]

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Sample was drawn from the 1993-2017 waves of the Panel Study of Income Dynamics. Analyses involved multivariable linear regressions adjusting for parental gender, age, and educational attainment, household size, inflation-adjusted household income in the year before the child's birth, the head of household's race, and several time-varying state characteristics. We also included fixed effects (i.e., indicator variables) for state and year of birth. Robust standard errors were clustered by state of birth.

Supplemental Table 4. Effect of paid family leave policy on parent health, with multiple imputation

	β [95% CI]
	Policy effect
Self-rated health	0.054
	[-0.047, 0.16]
Psychological distress	-0.12
	[-0.75, 0.50]
Overweight	-0.059*
	[-0.12, -0.003]
Obese	-0.038
	[-0.091, 0.015]
Drinks alcohol	-0.13***
	[-0.20, -0.049]
3+ drinks daily	-0.071**
	[-0.3, -0.015]

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

N = 6,690 parents. Sample was drawn from the 1993-2017 waves of the Panel Study of Income Dynamics, in which missing values were imputed using multiple imputation using chained equations. Analyses involved multivariable linear regressions adjusting for parental gender, age, and educational attainment, household size, inflation-adjusted household income in the year before the child's birth, the head of household's race, and several time-varying state characteristics. We also included fixed effects (i.e., indicator variables) for state and year of birth. Robust standard errors were clustered by state of birth.

Supplemental Table 5. Effect of paid family leave policy on parent health, with wild clustered bootstrap standard errors

	β [95% CI]	β [95% CI]
	Wild Clustered Restricted	*Wild Clustered Unrestricted*
Self-rated health	0.11+	0.11+
	[-0.014, 0.24]	[-0.15, 0.37]
Psychological distress	-0.79	-0.79
	[-1.87, 0.30]	[-2.93, 1.34]
Overweight	-0.082	-0.082
	[-0.24, 0.06]	[-0.38, 0.22]
Obese	-0.026	-0.026
	[-0.17, 0.12]	[-0.32, 0.27]
Drinks alcohol	-0.12*	-0.12*
	[-0.23, -0.01]	[-0.33, 0.10]
3+ drinks daily	-0.057	-0.057
	[-0.15, 0.03]	[-0.23, 0.12]

$* p < 0.05, + p < 0.1$

Sample was drawn from the 1993-2017 waves of the Panel Study of Income Dynamics. Analyses involved multivariable linear regressions adjusting for parental gender, age, and educational attainment, household size, inflation-adjusted household income in the year before the child's birth, and the head of household's race. We also included fixed effects (i.e., indicator variables) for state and year of birth. Standard errors were calculated using wild clustered bootstrap inference to account for the presence of a single treated cluster.